SUITABILITY OF GEOGRAPHICALLY WEIGHTED REGRESSION IN WATER DEMAND PREDICTION

Adrick Nyondo

A dissertation submitted to the University of Zambia in partial fulfillment of the requirements of the degree of Master of Science in Geo-Information Science and Earth Observation.

THE UNIVERSITY OF ZAMBIA

LUSAKA

2022

COPYRIGHT

DECLARATION

I, Adrick Nyondo, hereby declare that this dissertation is my own work. It has not previously been submitted for any other degree or examination at the University of Zambia or any other University. All published work or material from other sources incorporated in this dissertation have been duly acknowledged and referenced.

I therefore present the dissertation for examination for the Degree of Master of Science in Geo-Information Science and Earth Observation to the University of Zambia.

Name: ……………………………………

Signature: ……………………………………

Date: ……………………………………

APPROVAL

This dissertation of Adrick Nyondo (2017013100) has been approved as fulfilling the requirements for the award of Master of Science in Geo-Information Science and Earth Observation of the University of Zambia.

Name………………………………….	………………..…	………………
Internal Examiner 1	Signature	Date
Name………………………………….	………………..…	………………
Internal Examiner 2	Signature	Date
Name………………………………….	………………..…	………………
Internal Examiner 3	Signature	Date
Name…………………………………..	…………………..	………………
Chairperson – Board of Examiners	Signature	Date
Name…………………………………..	……………………	……..…………
Supervisor	Signature	Date

DEDICATION

To my children Sangwa and Tukuza so that they can be inspired to work hard, never give up and never procrastinate. And to my beloved wife for her support and continued encouragement as I pursued my Master of Science degree programme.

ABSTRACT

The aim of this study was to evaluate the use Geographically Weighted Regression (GWR) in predicting water demand in Lusaka City. The specific objective was to analyse the spatial distribution of water demand in the city, examine spatial relationships between water demand and its predictors and to assess the suitability of Geographically Weighted Regression for predicting water demand. Data of water demand (dependent variable) in cubic meters per day was obtained from a study done in 2011 by Lusaka Water and Sewerage Company. Independent variables used were population, income, tertiary education attainment, property values, level of spatial development, irrigated hectarage and temperature. Geographic Information System (GIS), Ordinary Least Squares (OLS) and GWR regression models were used to analyse the distribution and correlation between water demand its predictors and to predict it for the year 2035. The results showed that overall water demand was generally highest in high density neighbourhoods and lowest in low density neighbourhoods. Whereas percapita water demand was generally highest in low density and lowest in high density neighbourhoods. Further, the study found that water demand was not significantly related to temperature, irrigated hectarage and percentage of tertiary education attainment and property values but was significantly and positively correlated to population, income and size of spatial development. With a prediction power of $R^2 = 0.86$, it was concluded that GWR model is suitable for water demand prediction. Forward selection method was applied on the three significant variables and results indicated that population variable had the strongest influence on water demand followed by size of spatial development then income with 49, 37 and 14 percent contribution respectively. The study also concludes that GWR is an important and reliable tool for analysing and predicting water demand as it is able to account for spatial variations. It further concludes that since factors that influence water demand are spatially varying, institutions responsible for planning and water management can use GWR to localise demand management interventions based on each area's sensitivity to water demand predictors.

Keywords: Geographically Weighted Regression, Water demand, Relationship, Prediction

ACKNOWLEDGEMENTS

First and foremost, praises and thanks to the God the Almighty for His showers of blessings to enable me complete the research successfully.

I would like to express my deep and sincere gratitude to my research supervisor, Dr. Kawawa Banda, School of Mines, University of Zambia, for giving me the opportunity to do research and for providing invaluable guidance throughout this research study. I am grateful particularly for the times he made himself available to me even without proper appointment. It was a great privilege and honor to do this research under his guidance.

My sincere gratitude also goes to all lecturers and staff in the Department of Geography and Environmental Studies for their timely support. Special and heartfelt gratitude goes to Mr Garikai Membere and Dr Bridget Umar for their efforts in making sure that I did not miss out on any research stage. They ensured that all formalities were done for me to have a supervisor who is well vested in my research discipline.

To all my research participants from Lusaka City Council (LCC), Lusaka Water and Sewerage Company (LWSC) and Central Statistical Office (CSO) who allowed me to interview them and for their informative insights. I'm grateful for their time, comments, answers and data provided.

My sincere gratitude goes to my sponsors – Southern African Science Service Centre for Climate Change and Adaptive Land Management (SASSCAL) for paying my tuition fees to the University and Pilot Programme for Climate Resilience (PPCR) for funding this research study. Without their support it would have been a big strain on me.

I am also indebted to Dr Jackson Phiri. You literally are the reason I started pursuing this MSc programme. Thank you very much for your help.

My family members for their love, care, understanding and patience during the course of this work. I'm forever grateful to the love of my life, Mirriam for always being there for me. My children, Sangwa and Tukuza for always putting a smile on my face and being a constant reminder that I have to work harder.

TABLE OF CONTENTS

COPYRIGHT i
DECLARATION ii
APPROVAL iii
DEDICATION iv
ABSTRACT v
ACKNOWLEDGEMENTS vi
LIST OF TABLES iv
LIST OF FIGURES v
LIST OF APPENDICES vi
ACRONYMS vii

CHAPTER ONE: INTRODUCTION **1**
1.1 Background 1
1.2 Statement of the Problem 2
1.3 Aim 3
1.3.1 Specific Objectives 3
1.4 Hypotheses 3
1.6 Significance of the Study 3
1.7 Scope of the Study 4
1.8 Organisation of the Dissertation 4

CHAPTER TWO: LITERATURE REVIEW **6**
2.1 Introduction 6
2.2 The Case for Water Demand Estimation or Prediction 6
2.3 Approaches in Water Demand Estimation or Prediction 7
2.3.1 Time Series Models 8
2.3.2 Regression Models 9
2.3.3 Spatial Approach Models 10
2.4 The Case for Using a Spatial Model 12
2.5 Geographically Weighted Regression 13
2.6 Factors that Influence Water Demand 15
2.6.1 Demographic Variables 15
2.6.2 Socio-Demographic Variables 16
2.6.3 Climate Variables 17
2.6.4 Physical Variables 19
2.7 Case Studies 21
2.7.1 Mexico City 21
2.7.2 New York City 22
2.7.3 Phoenix City 23
2.7.4 Portland City 23
2.7.5 Kenya 24
2.7.6 Lusaka 25
2.8 Benefit of Literature Review to the Study 26

CHAPTER THREE: METHODOLOGY **27**
3.1 Description of the Study Area 27
3.1.1 Location 27
3.1.2 Population 27
3.1.3 Climate 27
3.1.4 Geology and Topography 27
3.1.5 Hydrogeology and Hydrology 28
3.1.6 Selection of the Study Area 29
3.2 Research Design 29
3.3 Data Collection, Processing Modelling and Analysis 29
3.3.1 Primary and Secondary Data 29
3.3.2 Dependent and Independent Variables 30
3.3.4 Data Analysis 29

CHAPTER FOUR: RESULTS AND DISCUSSIONS **32**
5.1 Results 32
5.1.1 Spatial Distribution of Water Demand 32
5.1.2 Relationships Between Water Demand and its Predictors 39
5.1.3 Suitability of GWR in Predicting Water Demand 50
5.2 Discussions 55
5.3 Study Limitations 57

CHAPTER SIX: CONCLUSIONS AND RECOMMENDATIONS **58**
6.1 Conclusions 58
6.2 Recommendations 59
6.3 Future Research 59

REFERENCES **60**

APPENDICES **68**

LIST OF TABLES

Table 1: Approaches in Water Demand Estimation and Prediction................................11
Table 2: GWR Output Parameters...14
Table 3: Variable Statistics and Description 32
Table 4: How Well Neighbourhoods Were grouped using each Variable 38
Table 5: Results of Exploratory Regression...40
Table 6: Direction of Relationship and Level of significance...41
Table 7: Summary of Ordinary Least Squares Results 42
Table 8: Summary of OLS Fitted Model Parameters..42
Table 9: OLS Diagnostics Output 43
Table 10: Results of Geographically Weighted Regression..46

LIST OF FIGURES

Figure 1: Location of the Study Area with Names of Its Neighbourhoods....................28
Figure 2: Work Flow of Methodology....................28
Figure 3: Hot Spot Analysis - Water Demand....................33
Figure 4: Scatter Plot- Water Demand (m^3/d)....................34
Figure 5: 2010 Overall Water Demand Distribution (m^3/d)....................35
Figure 6: 2010 Per Capita Water Demand Distribution (M^3/d)....................36
Figure 7: Groupings Based on Similarity of Variable Characteristics....................36
Figure 8: Parallel Box Plot Explaining Figure 7and Water Demand Variables....................37
Figure 9: Over and Under Prediction Map....................43
Figure 10: Spatial Autocorrelation (Moran's I) Results....................44
Figure 11: Estimated Water Demand (2010)....................47
Figure 12: Predicted Water Demand (2010)....................47
Figure 13: Local R - Squared (R^2)....................48
Figure 14: t-statistics Values....................48
Figure 15: Spatial Relationship Between Population and Water Demand....................49
Figure 16: Spatial Relationship Between Income and Water Demand....................49
Figure 17: Relationship Between Size of Built-up and Water Demand....................50
Figure 18: Projected Water Demand for 2035 (m^3/d)....................51
Figure 19: GWR Modelled Water Demand 2035 (m^3/d)....................51
Figure 20: Local R^2 – Strength of Prediction....................52
Figure 21: Projected vs GWR Modelled Water Demand for 2035 (M^3/d)....................53
Figure 22: Calibration of the GWR Model for 2010 (Logged Values)....................54
Figure 23: Validation of the GWR Model for 2035 (Logged Values)....................54

LIST OF APPENDICES

Appendix 1: Dependent and Independent Variables..68
Appendix 2: Distribution of Input Variables in the Study Area 70
Appendix 3: Output Results of Exploratory Regression..73
Appendix 4: Relationship Between Water Demand and its Predictors 75
Appendix 5: Comparison of Methods in Urban Water Demand Modelling...................76

ACRONYMS

CSO	Central Statistical Office
ER	Exploratory Regression
GIS	Geographical Information Systems
GRZ	Government of the Republic of Zambia
GWR	Geographically Weighted Regression
JICA	Japan International Cooperation Agency
LCC	Lusaka City Council
LWSC	Lusaka Water and Sewerage Company
MCA	Millennium Challenge Account
MCC	Millennium Challenge Corporation
NAS	National Academy of Sciences
NDMI	Normalized Difference Moisture Index
NDVI	Normalized Difference Vegetation Index
NIR	Near Infra-Red
OLS	Ordinary Least Squares
PPCR	Pilot Programme for Climate Resilience
SASSCAL	Southern African Science Service Centre for Climate Change and Adaptive Land Management
SDG	Sustainable Development Goals
UTM	Universal Transverse Mercator

CHAPTER ONE: INTRODUCTION

1.1 Background

Water is a very important natural resource not only for domestic use but also for industrial, agriculture, hydroelectric power production and a wide range of ecosystem services. Over two-thirds (97 percent) of earth's surface is covered by water (Biswas, 2003). This figure gives a perception that water is abundant yet only about 3 percent is fresh water of which only 1 percent is accessible fresh surface water, 29 percent is ground water and 79 is in glaciers and polar ice caps (Balasubramanian, 2015). The available fresh water is unevenly distributed across the globe such that majority of the world population lives in conditions of water scarcity (Mancosu *et al.*, 2015). As global population keep increasing water challenges are likely to worsen as availability of fresh water reduces with increasing population. This is because as population grows the water footprint of humanity, as identified by Ercin and Hoekstra (2014), will change under five drivers; population growth, economic growth, consumption patterns, global production and trade pattern and technology development.

There is a growing concern that about half the world's projected 9.7 billion people will live in water-scarcity regions by 2050 Guppy (2017) and that this growth combined with economic growth will lead to an increase in freshwater demand with potentially large-scale disruptions of economic activities. The demand-supply imbalance is expected especially in developing or emerging economies (Nechifor-Vostinaru, 2018) and Zambia is not an exception. This is why it is important to have water demand studies in order to understand consumption patterns and changes and incorporating them into management policies which take into account future water resource needs for a growing demand.

Urbanisation in Zambia will likely continue creating an imbalance of supply-demand for water. The situation is likely to be worse for Lusaka, one of the most populated and most urbanised cities in Zambia. Besides being the most urbanized and most populated the city of Lusaka is also the most densely populated district in the province with 4,853.2 persons per square kilometer which is about six hundred (600) times denser than Luangwa district which is the most sparsely populated in the province with 7 persons per square kilometer (CSO, 2010). While population keeps growing as a result of urbanization and natural growth, so does the built environment. The latter has, in fact, grown beyond the administrative boundary of the district, well into the perimeters of Chongwe, Chibombo and Chilanga districts. While the city has grown spatially and demographically, institutions charged with the mandate of service provisions are finding it hard to cope with the growing urbanization such that services like water supply has been outstripped by its demand (JICA, 2009; LWSC, 2011).

Lusaka Water and Sewerage Company (LWSC only produces approximately 230,000 cubic meters per day yet the demand for water in Lusaka stands at about 520,000 cubic meters per day (LWSC, 2011). Of the amount produced most of it is lost through Non Revenue Water such as leakages and only 44 percent (JICA, 2014) of supply can meet the existing demand. The water situation may be compounded by the fact that urbanization continues to increase and the impact of future climate scenarios on water resources (surface and groundwater) remain unpredictable, making sustainability of water supply and indeed any other water related development not guaranteed (LWSC, 2011).

1.2 Statement of the Problem

All the published studies involving water demand foresting done in Lusaka city used the water Requirements Approach for estimating water demand, whereby water demand is estimated/forecast strictly based on the current or future population without considering the level of spatial development or allowing for changes in consumer behaviour in response to external influences such as changes in weather or socio-economic fundamentals. This research study therefore, explored the use of the Economics Approach which takes into account people's decision processes to consume more or less water based on a variety of factors including physical, weather or socio-economic fundamentals (Baumann, Boland and Hanemann, 1997).

1.3 Aim

The aim of this research was to evaluate the suitability of Geographically Weighted Regression (GWR), which is one of the Economics Approaches, for predicting water demand at neighbourhood level in Lusaka City.

1.3.1 Specific Objectives

Based on the research need identified above this study sought:

i. To analyse the spatial distribution of water demand in the city
ii. To examine spatial relationships between water demand and its predictors
iii. To assess the suitability of Geographically Weighted Regression for predicting water demand

1.4 Hypotheses

i. There are no statistically significant relationships between water demand and predictor variables i.e. tertiary education attainment, income, property values, population, irrigated hectarage, temperature and spatial growth.
ii. There is no statistical significance in the suitability of Geographically Weighted Regression in predicting water demand.

1.6 Significance of the Study

Analysis, estimating or predicting urban water demand is essential because water is a key factor for sustaining development, and decision makers can plan well for future demands (House-Peters and Chang, 2011) if they have relaible information about water requiremnts. If responsible institutions are to meet current and future demands, the solution is not simply increasing supply of water. This is because such actions may result in unsustainable use such as over-investment or under-investments in areas where it is supplied. The issue of insufficient water supply and increasing demand must be considered in moderating the pattern of demand (House-Peters and Chang, 2011; Mareike, 2017) and this can be done better if there is more understanding of the influencing factors and current and future demand distribution. The results of this study are expected to provide new insights on water demand, its drivers and whether or not modelling using

GWR can improve an understanding of water demand in Lusaka. Results from the study also form a basis for policy formulation and planning for efficient water supply and management. Findings and literature reviewed by this study also contribute to the body of knowledge with regard to water demand assessment.

1.7 Scope of the Study

This study is limited to Lusaka city and only for the years 2010 and 2035. The basis for choosing 2010 as the base year and 2035 as the projected year is that water demand data at neighbourhood scale was only found in a study done in 2010 by Lusaka Water and Sewerage Company (published in 2011). The spatial scale is the neighbourhoods in Lusaka city. The study uses Grouping Analysis, Exploratory Regression, Ordinary Least Squares and Geographically Weighted Regression (GWR) models to assess the spatial distribution of water demand, examine the relationship between water demand and its predictors and predict water demand for 2035. The predictors this study is limited to are population, income, property values, temperature, size of spatial development, irrigated hectarage and percentage of tertiary education attainment. They were chosen based on their association with water demand and the availability or possibility to generate them. Several studies have noted the significant impact of their water consumption determinants (Mayer and DeOreo, 1999; Domene, Saurí and Parés, 2005; Troy, Holloway and Randolph, 2005; Gato, 2006; Guhathakurta and Gober, 2007; Schleich and Hillenbrand, 2009; House-Peters and Chang, 2011; Khudair, Sadeq and Mahmoud, 2018; Sanchez *et al.*, 2018).

1.8 Organisation of the Dissertation

This dissertation is divided into five chapters. Chapter One comprises the introduction and background to the study. It highlights the aim and key objectives that this study sought to address Chapter Two gives a review of the relevant literature on the subject matter. It outlines the case for water demand estimation, key concepts and models of water demand estimation and various cases studies are cited on water demand studies done in other regions. Chapter Three describes the location of study area, physical characteristics in terms of climate, hydrogeology, hydrology and topography. The chapter further provides details on how the study was conducted, i.e. data collection, processing, modelling and analysis. Chapter Four consists of research findings and discussions. This chapter links the first three chapters by presenting and discussing findings in line with objectives and the research hypotheses formulated in chapter one. Chapter Five concludes the study with some recommendations. This is the last chapter and it is followed by references and appendices.

CHAPTER TWO: LITERATURE REVIEW

2.1 Introduction

Understanding of water use in both developed and developing countries is important for efficient and effective management of the water resource. Since water is used for several purposes its demand has many dimensions. The various dimensions include influencing factors for different uses, economic aspects, variations in use, approaches in water demand estimation or prediction, etc. Several authors have examined the determinants of household water demand in developed countries, but little effort has been made to synthesize the growing body of literature evaluating household water demand in developing countries (Nauges and Whittington, 2010). This chapter reviews literature about water demand, factors that influence it and how water demand assessments are done. Further, the chapter seeks to use the insights and identify knowledge gaps from literature to attain the objectives of the study.

2.2 The Case for Water Demand Estimation or Prediction

The subject of urbanisation has been extensively researched by several scholars and the common agreement among them is that urbanisation and consequently population growth and urban spatial development increases water consumption. If water demand lags behind population growth, economic growth and most life functions get disrupted (Olmstead, Hanemann and Stavins, 2007; Mansur and Olmstead, 2012; Sanchez *et al.*, 2018). Globally therefore, there is an increasing need for its providers to keep pace with the growing demand for water as nearly every aspect of life requires it (Bigelow *et al.*, 2017). Further, there is a growing recognition that freshwater sources available in most countries are barely sufficient to maintain the quality of life and economy (NAS, 1999). In a study of challenges in managing urban water demand, (Araral, 2010) recognises that responsible institutions face challenges when populations are rising, water scarcity is growing and urbanization is on the rapid rise. He suggests that depending on each country's circumstances, both short term and long term solutions including tariffs, management, infrastructure, regulatory and public education solutions should be involved. However, the prerequisite to any water demand management solution is adequate information about water supply and demand requirements (Wentz and Gober, 2007; House-Peters, Pratt and Chang, 2010).

Additionally, scholars agree that water demand analysis is a fundamental principle in integrated resource planning (Howe and White, 1999; Kayaga and Smout, 2008; Shaban and Sattar, 2011). In other literature attention is drawn to a nexus among water, energy, food security and the underlying natural resources, and the need for integration of all these to ensure sustainable development as water is linked to the larger sustainable development agenda (Bizikova *et al.*, 2013; Rasul and Sharma, 2016).

2.3 Approaches in Water Demand Estimation or Prediction

Water demand estimation or prediction methods are the techniques and practices used to analyse the current and future water consumption. The basic principle in most approaches is that the forecasted values of one or more variables (e.g. population, income) are translated into estimates of future water requirements. Approaches to water demand estimation and prediction must employ strategies which provide accurate estimates so that there is cost-effective and reliable infrastructure expansion to meet demand requirements (House-Peters and Chang, 2011). Water demand estimating can be approached in two ways. (i) the time horizon on which the focus is based and (ii) the method employed to estimate it. On the time horizon used they are classified into short term, intermediate term and long term (Billings and Jones, 2008; Rinaudo, 2015). Short term is done to predict water demand over the coming hours to weeks in order to optimise the operation of water systems. Intermediate term is up to 10 years where estimating is done to increase customer base while taking into account consumption behaviour and weather cycles. Long-term estimating is over 10 years which focuses on building long-term supply infrastructures taking into consideration changes in customer base and economic growth (Rinaudo, 2015). One method employed by several studies focus on the degree of classification by customer type and how per capita water consumption estimating is obtained (Billings and Jones, 2008, in Rohrdrommel, 2017).

Two common approaches are the Water Requirements Approach and the Economics Approach. The water Requirements Approach is whereby water demand is estimated based on the variety of activities (residential, industrial, commercial) that use water. It ties future needs strictly to the number of users and demand coefficients are determined per inhabitant, per customer, per employee, or per unit of industrial output. The average (litres per capita per day) for each of the activities is then multiplied by projected population to predict future water demand. Total water demand can be estimated by adding the water demand of the customer categories together (Donkor *et al.*, 2014). Rinaudo (2015) contend that this method has a flaw since it does not account for changes in consumer behaviour in response to external influences such as changes in tariffs, weather, or socio-economic fundamentals. Donkor *et al.*, (2014) also question the reliability of this modelling method as it involves judgement instead of empirical analysis to estimate water demand for each customer category.

The eeconomics approach is whereby water demand is modelled as a behavioral phenomenon. It is based on the fact that people make decisions to consume more or less water based on a variety of factors including tariffs, weather, income, and so on (Baumann, Boland and Hanemann, 1997). The economic approach method relies mainly on regression analysis as the primary tool for estimating household water demand. An equation relating total consumption to multiple demand-related variables is estimated. Researchers in using this method consider water demand influencers that factor into the consumer's decision process such as weather and socio-economic factors which are estimated first. These estimates then become the basis for predictions or scenario analysis (e.g. what effect will higher temperatures or more educated population have on total water demand). These analyses or predictions inform what measures to take and the effectiveness of various demand management policies (Mareike, 2017). This approach uses several mathematical models which are discussed below.

2.3.1 Time Series Models

This model uses historical trends based on the understanding that future changes of water demand can be projected from past trends. Several mathematical models are used such as

Moving Average, Exponential Smoothing, or Bow–Jenkins models (Billings and Jones, 2008; Donkor *et al.*, 2014; Rinaudo, 2015). The main drawback for this model is that variables such as demographic, socio-economic changes as well as the implementation of water demand management options strategies during the observed period are not taken into consideration, which may compromise the accuracy of results (Billings & Jones, 2008). It is for this reason that Donkor *et al.*, (2014) suggests that the use of time-series models should be limited to short-term and medium-term predictions, because the variation of the influencing factors is expected to be negligible at these time periods.

2.3.2 Regression Models

These are statistical techniques for estimating the relationship between or among variables which have causal relationships. Regression models with one dependent variable and a single independent variable are called simple linear regression models while those with one dependent but more than one independent variables are called a multiple regression (Uyanık and Güler, 2013).

Regression models are based on theories about cause and effect whereby water demand is caused/influenced by factors such as water tariffs, income and population (Billings and Jones, 2008). These models can be created using cross-sectional data (data collected at one point without regard to differences in times), time series data (taken at successive equally spaced points in time), or panel data (observations of multiple phenomena obtained over multiple times) (Donkor *et al.*, 2014). Relationships between independent and dependent variables are analysed and the regression equation is determined (Liu and Xue, 2017). The relationship between dependent and independent variables is defined in the beginning of the analysis. For example an equation with n predictor variables X_1,X_2, X_3...., Xn and a dependent variable Y, takes the form (Arbués, García-Valiñas and Martínez-Espiñeira, 2003):

$$Y = B_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \text{ …….. } \beta_n Xn + \varepsilon_i$$

where Y is the water demand and X_1,X_2,X_3 …. Xn are independent variables such as income, population, etc. and B_0, β_1, β_2, β_3, ….. β_n are constants defining the nature of the relationship between the variables Y and X_1,X_2,X_3 …. Xn (independent variables). The

B_0 or Y-intercept is the value of Y when $X_n = 0$ and β_n is the slope of the line and is known as the regression coefficient and is the change in Y associated with a one-unit change in X_n.

After determining the necessary variables, the demand that would be obtained under a hypothetical evolution of the explanatory variables (assuming that the model coefficients hold true over the future time window considered) is calculated by taking projected explanatory variables into the regression equation (Liu and Xue, 2017). When the model is run, estimates of the values of the variables β_n and Xn, as well as other information valuable for diagnostic testing (i.e. goodness of fit, statistical significance) are generated. The goodness of fit of β_n is most commonly measured with the coefficient of determination (R^2) and is used for evaluating the regression (Billings and Jones, 2008). Each calculated coefficient shows the change in the dependent variable for a one-unit change in that specific explanatory variable, while keeping all other explanatory variables constant.

2.3.3 Spatial Approach Models

Most models use aggregated data for demand estimation. Aggregated data is at a large scale where groups of observations are replaced with summary statistics based on those observations. However, Wentz and Gober (2007) and House-Peters, Pratt and Chang (2010) note that the use of aggregated large-scale data in statistical models assumes a lack of a variation in spatial patterns and processes, which is not the case as these variations have since been recognized as important determinants of water consumption (House-Peters and Chang, 2011). The relevant classifications, their basic principles, data requirements and shortcomings of the various approaches in water demand estimation and prediction are listed in Table 1.

Table 1: Approaches in Water Demand Estimation and Prediction

Method	**Principle**	**Applications**	**Data requirements**	**Shortcomings**
Time - Series	Projection of past observed tendencies	Development of a "business as usual scenario" assuming a continuation of prevailing socio-economic conditions	Time series of water consumption	Limited predicted capability – does not account for changes in socioeconomic context
Unit water demand analysis (water requirements approach)	Estimation based on "unit water demand coefficient" multiplied by the number of users in each category	Development of sectoral demand forecast accounting for expected future population growth, change in economic activity per branch. Demand can easily be represented spatially	Unit water consumption coefficients (per type of users). Estimated future number of users per category	Does not account for possible future changes in unit water consumption due to evolving water tariffs, household income, etc.
Multivariate statistical models(Regression models)	Estimates per capita consumption as a function of explanatory variables such as water rates, household income, level of economic activity (employment/turnover), housing characteristics), weather conditions, etc.	Allows forecasting future demand considering changes in (i) population and economic activity and (ii) changes in socioeconomic variables (water rates, households' characteristics and income, etc.)	Time series for water consumption and all explanatory variables. Estimated future number of users per category	Does not account for changes in plumbing code or campaigns to promote water conservation
Land use based models	Demand assessed on the scale of uniform spatial entities using unit ratio	Spatially accurate water demand forecast, integrated with urban planning	Long range urban planning scheme. Unit consumption ratio per category of urban development	Does not account for changes in economic conditions (prices, income) nor evolution of technologies/ plumbing code

Source: (Rinaudo, 2015:3)

2.4 The Case for Using a Spatial Model

In water demand modelling most models take a global approach whereby the relationships being modelled are homogeneous i.e. they are the same everywhere within the study area (Martínez-Fernández, Chuvieco and Koutsias, 2013). The assumption is that there is no spatial variability and therefore an equation describing water demand as a function of dependent (water demand) and independent variables (its influencing factors) is the same regardless of location. However, water demand cannot be the same everywhere because factors which influence vary across space which are better described by spatial dependence and spatial heterogeneity. Spatial dependence is a situation where an observation of a given variable at place *i* depends on the observations in other location *j*. Spatial heterogeneity refers to a situation where the relationship among variables varies in space (Wentz and Gober, 2007).

Therefore, besides population, income, irrigated hectarage, temperature, property values and percentage of tertiary education attainment as variables used for water demand estimation and prediction, this study incorporates spatial development pattern in trying to analyse the influence of location characteristics on water consumption. The use of spatial variable has grown in recognition because it helps in achieving a better understanding of the complex spatial and temporal patterns of water usage (Lee and Wentz, 2008). Studies that use GIS techniques have shown that different variables have different influences at different spatial locations and scales. There are several examples of proponents of spatial approach. A study done by Franczyk and Chang (2009) explored biophysical and socio-economic factors that explain spatial patterns using Moran's I, local index of spatial autocorrelation (LISA), and spatial regression models. The results were that water use patterns are not distributed evenly over space and time. Spatially explicit methodologies in urban water demand modelling have improved the ability of water managers to model the influence of significant variables at multiple spatial scales and such studies make it possible to determine the scales at which certain processes are most influential which helps in understanding patterns of demand (Lee and Wentz, 2008; Franczyk and Chang, 2009; Polebitski and Palmer, 2009; Chang, Parandvash and Shandas, 2010; House-Peters and Chang, 2011). In order to assess and predict water demand that takes into account neighbourhood characteristics Geographically Weighted Regression (GWR), a GIS based regression model, is therefore, used in this study (Wentz and Gober, 2007; Balling, Gober and Jones, 2008).

2.5 Geographically Weighted Regression

Geographically Weighted Regression (GWR) is a GIS based spatial regression analysis technique that takes non-stationary (spatially varying) variables into consideration (e.g., climate; demographic factors; physical environment characteristics) and models the local relationships between these predictors and an outcome of interest - water demand in this case (House-Peters, Pratt and Chang, 2010; Bivand, 2020). It is a statistical method that accounts for spatial autocorrelation and it is an improvement over Ordinary Least Squares (OLS) method. It can be run in ArcGIS, R or GWR 4.0 software packages. GWR estimates a regression equation for each location from a subset of nearby observations, allowing the relationships being modelled to vary from neighbourhood to neighbourhood (Ahmadi and Sedghamiz, 2007). Coefficients and intercept estimates are subjected to weighted values of neighbouring observations, commonly defined by a distance-decay kernel function. In other words, it is assumed that points located further from a given location are more likely to present differing coefficients than points that are located closer (Fotheringham, Charlton and Brunsdon, 1998). Examples of studies that used the GWR model are discussed by House-Peters and Chang (2011) (see Appendix 5) who identified varying degrees of the GWR coefficients for the household size variable, suggesting different sensitivity of water consumption with an increase in household size across different census tracks in the city of Phoenix. Similar studies and outcomes are available (Fotheringham, Charlton and Brunsdon, 1998) and suggest that water resource planning and management should incorporate spatial and neighborhood effects to effectively manage limited natural resources.

In order to successfully perform GWR in ArcGIS software two GIS based regression tools are performed one after the other. These are Exploratory Regression (ER) and Ordinary Least Squares (OLS) in that order. ER is used for selecting the appropriate predictors based on their strength of correlation with water demand – the dependent variable. OLS is used for diagnostic or specifying (fitting) the best model (best combination of variables) to be used in GWR using the identified variables (Caron and Ngui, 2012; Wafula and Ngigi, 2015). OLS works on the assumption that (Frost, 2018):

i. The linear regression model is linear in parameters
ii. There is a random sampling of observations
iii. The conditional mean should be zero.
iv. There is no multi-collinearity
v. There is homoscedasticity and no autocorrelation

Based on their significance in predicting water demand, some or all the variables are used in GWR to analyse relationships between dependent and independent variables and or to predict future water demand. GWR assumes that the strength and direction of the relationship between a dependent attribute and the independent attributes may be modified by contextual factors. The main output from GWR for each observation point is a set of parameter estimates and associated diagnostics (Table 2) that can be visualized within a GIS environment (Charlton and Fotheringham, 2009). Analysis of the outputs maps are used to explore variations of the independent variables factors and local R^2 values show the performance of the GWR model in different areas (Martínez-Fernández, Chuvieco and Koutsias, 2013).

Table 2: GWR Output Parameters

Parameter	Description
Observed	The observed (estimated) value of the dependent (y) variable
Condition number	The condition number of the data matrix – local collinearity produces unreliable coefficient estimates – the results should be treated with caution. Values around 5 to 10 suggest weak dependencies in the data, whereas values greater than 30 suggest moderate or stronger dependencies in the data.
Local R^2	The locally weighed r^2 between the observed (estimated) and fitted values. The statistic is a measure of how well the model replicates the local *y* values around each observation. A simple dimensionless measure that is readily calculated as the proportion of variance explained by model terms.
Predicted	The local prediction of the y variable (fitted value)
Intercept	The local intercept

Residual	The residual – the difference between the observed (estimated) and fitted value
StdError	The standard error of the residual
tdErr_Int	The locally weighed standard error of the Intercept
StdErrC*n*_P	The locally weighed value of the coefficient for the *n*th variable in the model
StdResid	The standardised residual – these have a mean of zero and a standard deviation of unity.
Source_ID	The FID of the corresponding feature in the Input feature class attribute table.

Source: (Charlton and Fotheringham, 2009)

2.6 Factors that Influence Water Demand

There are many factors that influence water demand and understanding which factors are significant is critical both in theory terms and also in technical and policy related matters. In order to have an accurate estimation or prediction and effective demand management policies it is important to have a clear understanding of the drivers of water demand. These factors are categorized into demographic (population, household size, age group, etc.), socio-demographic (income, education, etc.), climatic (temperature, precipitation, wind speed, etc.) and physical (location, spatial, etc.) characteristics (Wentz and Gober, 2007; House-Peters, Pratt and Chang, 2010). Along with explanations about their relationships with water demand, literature about how population size, income level, irrigated hectarage, temperature, property values, size of spatial growth and tertiary level attainment (the explanatory variables used in this study) influence water demand distribution is presented. The explanatory variables were used because according to literature and theory, they are common influencing factors of water demand (Ahmadi and Sedghamiz, 2007; House-Peters and Chang, 2011; Caron and Ngui, 2012; Wafula and Ngigi, 2015; Mareike, 2017).

2.6.1 Demographic Variables

Depending on which one is available population characteristics include the population size, household size, age and gender profiles. The more detailed the information about demographic characteristics the better it allows exploration of the influence of different population compositions on water demand (Makki *et al.*, 2015). Since detailed data about population characteristics is usually not available especially at lower scales, general population size is included for the development of reliable water demand prediction models (Makki *et al.*, 2013). In several studies involving water demand estimation, consumption increases with increase in population (Mayer and DeOreo, 1999; Gato, 2006). The increase though is not one-to-one due to economies of scale - i.e., doubling household members does not double water demand, as some major water uses (notably outdoor irrigation) is not family size dependent (Höglund, 1999). Generally, however, water consumption is positively correlated with population size. Water demand increases due to more water being used for bathing, laundry, toilet flushing and dishwashing, but per capita use is negatively related to occupancy size (Wentz and Gober, 2007; Schleich and Hillenbrand, 2009).

2.6.2 Socio-Demographic Variables

Socio-demographic variables include employment status, educational level and annual income level of household members. Water consumption in households with working residents is significantly higher than that in households with unemployed residents, and this is mainly due to bathing and laundry related consumption behavior of the employed and unemployed (Makki *et al.*, 2015). Similarly, Makki *et al.* (2013) postulate that consumption related to bathing takes a large share of water use and it is positively correlated with employment status, education level and income level. People with higher income levels are more likely to have received a higher education and more likely to be employed therefore a positive correlation with water demand. Income is also incorporated in water demand studies because of the assumption that higher income is likely to lead to more possession of water-using devices. Hence research on the link between income and demand has shown that high volume water users tend to be wealthier (Guhathakurta and Gober, 2007; Klein *et al.*, 2007; Schleich and Hillenbrand, 2009). Since income is not always available especially at fine resolutions such as household or neighbourhood level, Agthe and Billings (1997) suggest that income groups could be identified among neighborhoods and be used for demand estimation. They used this technique during surveys of water users in Tucson (United States of America) from 1979 and 1989 and categorised three income groups - low, middle and high. Results of their survey was that water demand in high income households exceed low income groups by 56 percent and middle income users by 37 percent.

2.6.3 Climate Variables

Climatic variables which impact on water consumption mainly include precipitation, temperature and drought. Researchers generally agree that urban water demand is directly related to temperature and inversely related to precipitation (Agthe and Billings, 1997; Gutzler and Nims, 2005; Hoffmann, Worthington and Higgs, 2006). Balling, Gober and Jones (2008) conducted a study in Phoenix where multivariate analyses were made using monthly climatic data (temperature, rainfall and drought respectively) whose model coefficients +0.55, -0.69, and -0.52 respectively, indicating that temperature, precipitation and drought have a statistically significant relationship with water use. For temperature, the principle is that hotter days lead to increase in activities and behaviour that consume more water such as irrigation of lawns and gardens, the need to replace water in pools and other water features, as water is lost to evaporation and and personal hygiene (Hoffmann, Worthington and Higgs, 2006; House-Peters and Chang, 2011). Morgan and Smolen (1976) and Hansen and Narayanan (1981) used temperature and rainfall as weather features for regression models to estimate water demand. Bakker *et al.* (2014) tested three models with and without the inclusion of weather features. They concluded that models which take into account climatic features performed better than models without climatic features. Gutzler and Nims (2005) conducted a research in New Mexico using multiple regression where temperature and precipitation were used as independent variables and found that over 60% of variance in water demand is explained by climate variables. These studies point to the fact that climate variables should be taken into consideration when estimating or predicting water demand.

Several studies have been conducted suggesting varying thresholds beyond which increase in temperature leads to increase in water consumption. Polebitski and Palmer (2009) found that a 10% increase in mean maximum temperature led to a 10% increase in water use in Seattle while Balling, Gober and Jones (2008) estimated a 6% increase in Phoenix for a 10% increase in mean maximum temperature. With regards to the link between affluence and temperature water use was found to be most drought-sensitive in townships with higher incomes and more swimming pools per capita. Balling, Gober and Jones (2008) and Miaou (1990) examined weekly municipal water use in Texas and compared it to a heat function, a temperature function and rainfall events. Using the heat function (maximum air temperature with rainy days removed), the authors found that in Texas cities, water use rises when temperature reaches about 21^{o}c and then rises sharper at about 32^{o}c, indicating that the response of households to temperature changes is not a linear function. Guhathakurta and Gober (2007) in Arizona using multiple regression and minimum daily temperature range, household income and size, plot size, house age, swimming pool evaporative coolers, NDVI, percent owner-occupied homes, water source and property value found that 0.6°c increase in temperature results in 1,098 liters increase in water use per household. Ruth *et al.* (2007) in New Zealand, using multiple regression and temperature, precipitation and wind speed as independent variables found that projected climate change and population growth scenarios result in 30-40% probability of water shortages, implying that precipitation, temperature and other climate related variables have a positive relationship with water demand. Klein *et al.* (2007) in Colorado used Fixed Effects to model water demand using irrigation and temperature as independent variables and the findings were that water use increases by 2% for every 0.6°c rise in temperature and decreases by 4% for every inch of rain.

House-Peters, Pratt and Chang (2010), in their study on residential water consumption in Hillsboro, Oregon used spatial analysis techniques and found that although water demand in that area was not sensitive to dry conditions at all, specific areas presented higher water consumption under such conditions. Also, the authors showed that in areas where water demand is more sensitive to weather, these areas presented higher concentration of new and big residences, with high property values and with residents with higher education level. They confirmed that the response to independent variables varies from location to location and depends on the affluence of residences and property values of each location.

2.6.4 Physical Variables

Physical variables that influence water demand are spatial development patterns, irrigated hectarage (such as lawns, gardens, etc.) and property values.

2.6.4.1 Built Up Area

There is a relationship between spatial patterns of developed land use and water consumption and this is confirmed by several studies (Wentz and Gober, 2007; Khudair, Sadeq and Mahmoud, 2018; Sanchez *et al.*, 2018). In fact, Bouziotas, Rozos and Makropoulos (2015) examined aspects of the interplay between the dynamics of urban spatial growth and the urban water cycle and found that the most important driver of urban water demand is urban spatial growth. Although these studies have recognised the importance of spatial patterns of development, models for demand estimation or forecasting rarely incorporate it as a variable (Stoker and Rothfeder, 2014). Studies that have used spatial variable have examined this relationship at fine scale, incorporating variables like number of bedrooms and bathrooms, outdoor area, presence of a swimming pool (Chang, Parandvash and Shandas, 2010), whose data is rarely available especially at a large scale. A study using fine scale data was done in the city of Portland about the spatial development pattern and water demand. The conclusion was that water consumption was directly related to residence size and inversely related to building density and age of buildings (House-Peters, Pratt and Chang, 2010). A few studies have been conducted, however, which incorporated spatial growth as a variable at a coarse scale. In south-eastern United States for example the relationship between built up areas and water demand was examined by quantifying the spatial pattern of developed land within township boundaries and results were that metrics describing the spatial patterns of development land explained significantly more variation (53%) in water use than socio-economic and environmental variables (Sanchez *et al.*, 2018).

2.6.4.2 Irrigated Hectarage

Accurate mapping of the distribution of irrigated land can facilitate an improved understanding of patterns of water use. The assumption is that wealthier households are able to and more interested in maintaining lawns and gardens which tend to be more extensive the wealthier the households are (Domene, Saurí and Parés, 2005). Maintenance of gardens as Robbins (2007) observes tends to denote status or they are simply maintained to respond to neighbourhood expectations. Robbins concludes therefore, that these contextual aspects jointly shape water consumption behaviours. Apart from the socio-economic and behavioural reasons for maintaining lawns and gardens climatic variables such hot and dry weather increases water use for irrigation of lawn and gardens (Zhou *et al.*, 2000; House-Peters and Chang, 2011). In order to investigate how irrigation of lawns and gardens influence water demand, methods such as land use classification combining Normalized Difference Vegetation Index (NDVI) and Normalized Difference Moisture Index (NDMI) indicators are performed to map irrigated areas (Wentz and Gober, 2007; Wu and De Pauw, 2011). NDVI indicates the presence of green, healthy vegetation and if this is done during the dry season it can most likely be attributed to irrigated vegetation. It describes the vigour level of the vegetation and is calculated as the ratio between the difference and the sum of the refracted radiations in the near infrared and in the red (NIR-RED)/(NIR+RED). The interpretation of the absolute value of the NDVI allows recognition of the areas of the vegetation that are stressed which is a sign of not being irrigated**.** The values of NDVI vary between -1 and 1, with -1 being most stressed and 1 being the healthiest. NDMI describes the vegetation's water stress level and is calculated as the ratio between the difference and the sum of the refracted radiations in the near infrared and SWIR (NIR-SWIR)/(NIR+SWIR). The interpretation of the absolute value of the NDMI allows recognition of the areas of vegetation with or without water stress problems**.** The values of NDMI vary between -1 and 1, with -1 being the most water stressed and 1 being the most watered (Pervez and Brown, 2010; Wu and De Pauw, 2011).

Values of NDVI greater than 0.25 indicate healthy vegetation (high vigour) while values of NDMI greater than 0.4 indicate irrigated vegetation. Therefore, in order to identify irrigated regions of an area classification of a satellite image is done. After classification a logical operation (NDVI > 0.25) AND (NDMI > 0.4) is performed to differentiate the irrigated areas with heathy vegetation from the non-irrigated ones (Senturk, Bagis and Berk, 2014).

2.6.4.3 Property Values

There is a relationship between property values and household water consumption. High property values are proxy measures of wealth and in theory wealthier households are more likely than their poor counterparts to reside in neighbourhoods with higher property values. Because of their affluence status they are more likely to have extensive lawns or gardens and/or significant water consuming facilities such as swimming pools all of which are associated with high water consumption (Troy, Holloway and Randolph, 2005). Therefore, water demand increases with increase in property values.

2.7 Case Studies

2.7.1 Mexico City

Mexico City is the capital city of Mexico and has a population of 8.8 million. A large majority of the city's water supply comes from an underground aquifer that is being drained at a rate faster than it can refill. Water demand studies. Several water demand measurement methods were embarked on by the local government since 1994, but this target had not been fully accomplished because water demand was mainly considered an issue of population growth and no other determinant or diagnosis was given (Ramos-Bueno, Perevochtchikova and Chang, 2021).

A study was conducted in an effort to explore water demand determinants other than population growth to understand residential water demand comprehensively (Ramos-Bueno, Perevochtchikova and Chang, 2021). The Ordinary Least Squares (OLS) and Geographically Weighted Regression (GWR) models were used to explore relationships between residential water demand and determinants in Mexico City. The unity of analysis was the neighbourhood scale in a study area comprising of 1,346 neighborhoods and the data used was for the year 2010 due to census data availability for that year. The dependent variable was annual residential water demand and the independent variables were residential water users with metering device (%), total dwellings with internal water connection, intermittent water supply (0 or 1), social development index (used as a proxy for income), average household size and dwelling density (dwellings per ha). The results of the study were that high levels of residential water demand were associated with high levels of the social development index, small household size (low density neighbourhoods) and low incidence of intermittent water supply.

2.7.2 New York City

Kontokosta and Jain (2015) undertook a study in New York City whereby determinants of large-scale building water use were modeled. The background of this study was that growth in urban population in the United States in general and New York City in particular has led to a corresponding increase in the population living in multi-family housing. This is the city where multi-family buildings account for over two-thirds of the total housing units and represent a large majority of potable water consumption (Kontokosta and Jain, 2015).

Using water use, land use and demographic data the study applied Weighted Robust Regression (Bhar, 2009) and Geographically Weighted Regression models to analyze the determinants and spatial patterns of water consumption in over 2,307 multi-family buildings located in New York City (Kontokosta and Jain, 2015). The aim of the study was to understand the drivers of multi-family housing water consumption, to analyze patterns of use both spatially and by building type and to examine differences in consumption patterns and intensity by socioeconomic status of households and by neighborhood income and demographics. The variables used were occupancy, building size, building age, ownership structure, neighborhood demographic and socioe-conomic characteristics and the energy use intensity of a building. The results were that water consumption in multi-family housing in New York City is statistically significantly related to the independent variables used. Further, the results of the GWR spatial analysis demonstrated significant spatial variability across building characteristics, demographic variables and house-hold income.

For individual variables resulted indicated that occupancy is positively correlated with water consumption such that for every unit increase in house-hold occupancy, water use increases by 18%. As for build size the results indicated that as building size increases efficiency in water consumption increases such that for every 10% of additional floor area water use decreases by approximately 0.8%. With regard to energy intensity every 10% increase in energy consumption is asso-ciated with a 2.8% increase in water consumption. Generally, GWR results confirmed the existence of significant spatial heterogeneity across our study area.

2.7.3 Phoenix City

Study was done the city of Phoenix to understand the demand side of water consumption. This was because with Rapid population growth in the face of an uncertain climate future challenges there was need to find ways of consuming water more prudently. Determinants of water consumption ware identified for detached single-family residential units using ordinary least squares regression (OLS), whose results were compared with the results of a Geographically Weighted Regression (GWR) model to determine whether there are spatial effects beyond the effects of the OLS variables. The determinants used in this study were household size, the presence of swimming pools, plot size and the prevalence of landscaping that requires a moist environment. The findings were that household size, the presence of a pool, landscaping practices and plot size showed statistically significant correlation with water demand. The other finding was that there was an improvement of the GWR over the OLS, with the former showing spatial besides the other determinants of water demand. It was, therefore, concluded that model parameters can be used to investigate the effects of policies designed to regulate plot size, pool construction, and landscaping practices on water consumption and to forecast water demand in areas of new construction (Wentz and Gober, 2007).

2.7.4 Portland City

A study in Portland City was done to assess the role of urban development patterns on water demand. Billing and plot tax records for single-family residential buildings were used to test the role of structural independent variables (plot size, building size, building density, and building age) on water consumption. To accomplish this, individual household billing and tax data was aggregated into census block groups after which structural attributes, socio-demographics, and water consumption data were integrated. GIS and statistical models were used to analyze single-family residential water consumption in the city. The findings were that residential water consumption at household level is best explained by average building size, followed by building density and building age. The study also found that spatial variables explain up to 87% of variations in water consumption. The study, as a result, helped to develop a water demand framework that incorporated existing factors with urban development policies to more effectively manage limited water and land resources (Chang, Parandvash and Shandas, 2010).

2.7.5 Kenya

A study was undertaken by Manetu, Mutua and Kenduiywo (2019) on water demand in in Kenya the Athi River using two GIS-based regression models (OLS and GWR models). It is located between 35 ° and 60° N and S and it has been experiencing rapid population growth and industrial development leading to increased water demand. As of 2019 the region was experiencing water shortages such that only 40 percent of the required amount was supplied by the responsible utility company. The study to project water demand in the region was necessitated by the need to developing alternative water supply sources, integrating water demand management programs and planning a cost effective and reliable infrastructure. The study used household characteristics as explanatory variables and water consumption for 2017 as the dependent variable. The explanatory variables used were household room sizes, garden presence, meter connection, education level, household size and household income. Ordinary Least Squares (OLS) and Geographically Weighted Regression (GWR) models were applied to understand domestic water demand in Athi river region. GWR model was then adopted to examine the spatial relationship between the significant variables and water demand and to predict water demand in the year 2022.

The results of the study were that the variables household size, household income, meter connections and household rooms were statistically significant independent variables and that there was a spatial relationship between these variables and water demand across the study region. Results of water demand prediction using GWR model indicated that water demand would increase from 721,899 m^3 in 2017 to an estimated 880,769 m^3 in 2022, an increase of about 22%. The study concluded that household size, household income, meter connection and household room size are important variables when prediction future domestic water demand. Further, results of the GWR confirmed the importance of spatial effects in influencing water demand for the variables used in the study.

2.7.6 Lusaka

Several studies have been done in Lusaka on both surface and groundwater and they produced key outputs or closed several knowledge gaps with regard to groundwater and geology (JICA, 2009, 2014; LWSC, 2011; Kang'omba and Baumle, 2013). With regard to groundwater JICA (2009) found that that the annual abstraction of groundwater surpasses the recharge capacity during a typical rainy season. This is a concern since the rate of water supply by LWSC lags behind the rate of urban growth making people provide their own water through boreholes, implying future groundwater resource will be under stress. JICA in the same study did a water demand projection study for Greater Lusaka and their methodology for current and 2030 total maximum daily demand for water was estimated, based on the 2030 population projection**,** to be 757,300 m^3/day. LWSC formulated an Investment Master Plan for water and sanitation of 2011 in which they focused on improving infrastructure capacity for water supply. Water demand for 2010 was estimated at 249,791 m^3/day and with a population projection of 2,818,173 people for 2035 water demand was estimated at 577,327 m^3/day (LWSC, 2011). Projection for water demand by LWSC was also formulated based on demographic and income projections. Studies by Federal Institute for Geosciences and Natural Resources (BGR) also modeled current and future water budget under different water management options which revealed lack of capacity by Lusaka city Council, and that although the current abstraction scheme in Lusaka is sustainable new commercial boreholes are needed to meet the future water demand of Lusaka (Kang'omba and Baumle, 2013).

All the published studies on Lusaka are related to this research but there are gaps that the research study seeks to address. Beside population projections, land use and income JICA and LWSC did not use other factors which influence water demand and they did not explore the statistical relationships between water demand and its predictors. Simon &

Roland (2013) in their development of a Groundwater Information & Management Program for the Lusaka groundwater systems modelled current and future water budget using temperature, rainfall and evapotranspiration, but these variables were used to estimate ground water budget as opposed to how they influence water demand. Their study also did not explore local variability of water demand. This study includes other variables besides population and income and it explores the statistical relationships between water demand and variables that influence it, then uses spatial regression to predict water demand.

2.8 Benefit of Literature Review to the Study

The literature review has provided a firm argument for Water Demand estimation or prediction and has provided knowledge about the approaches to do so. An assessment of the range of existing materials dealing with knowledge and understanding in water demand field has been provided by this review. It has also provided insights into previous work in the context of what has already been done and the extent to which it has been researched. Review of literature on studies done in Zambia has helped to identify a gap and thus providing a framework and direction taken by the researcher. The review has contributed to the development of the researcher's knowledge on the different methodologies and data collection methods used by previous researchers in similar studies. The literature also revealed the experts in water demand field with regard to estimation and prediction, which experts have been a key resource for reference in the current research. The reviewed literature also determined the methodology used in this study and method of analysis of results.

CHAPTER THREE: METHODOLOGY

3.1 Description of the Study Area

3.1.1 Location

Lusaka City is in Lusaka Province of Zambia, located at latitude 15^{o} 30' north and longitude 28^{o} 10' east (Figure 1). The City has a surface area of 360 km^2, and shares boundaries with Chongwe, Chilanga, Chibombo and Kafue Districts (JICA, 2009).

3.1.2 Population

The district population is 1,747,152 after increasing from 421,000 and 1,084,703 million people in 1974 and 2000, respectively with a current growth rate of 5% per annum. It is the most densely populated district in the province with 4,853.2 persons per square kilometre, while Luangwa district which is the most sparsely populated has 7.0 persons per square kilometre (CSO, 2010).

3.1.3 Climate

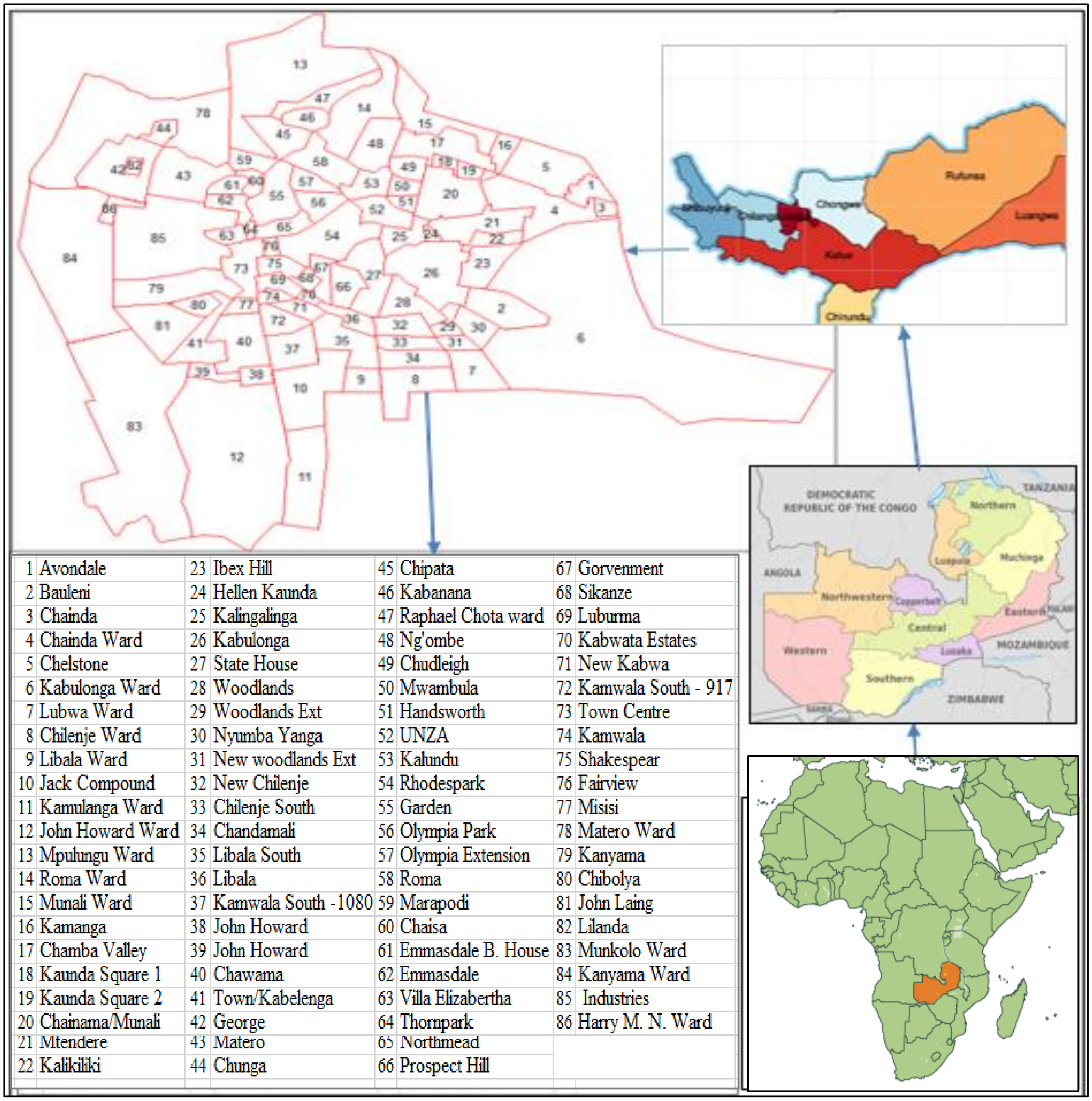
The city's climate is humid subtropical and it is characterized by three distinct seasons: the cool dry season, which extends from May to August; the hot dry season (August to November); and the rainy season (November to April) (Muvombo, 2017). The average rainfall of Lusaka is 837 mm per annum received between November and March with the mean annual temperature being between 20˚C in September and 30.6˚C in October and the lowest is about 10℃ in June and July (Bäumle and Museteka, 2009). The 30-year average annual rainfall for Lusaka is 857mm with 77 rainy days per year (Kang'omba and Baumle, 2013; Muvombo, 2017).

3.1.4 Geology and Topography

Its topography is mostly flat with an elevation ranging from 1,200 m to 1,300 m above sea level. Escarpments lie to the east and north of the City, which ends in the Luangwa Valley. The geology of the City comprises a PreCambrian basement complex consisting of granites, gneises and quartzites which is overlaid by limestones and dolomites (JICA, 2009).

3.1.5 Hydrogeology and Hydrology

The City is divided into three drainage basins, i.e. Chongwe, Chunga-Mwembeshi and Kafue Basins. There are only small-scale rivers in the City, namely Ngwerere and Chunga Streams, flowing to the north-eastern and north-western directions respectively.



1	Avondale	23	Ibex Hill	45	Chipata	67	Gorvenment
2	Bauleni	24	Hellen Kaunda	46	Kabanana	68	Sikanze
3	Chainda	25	Kalingalinga	47	Raphael Chota ward	69	Luburma
4	Chainda Ward	26	Kabulonga	48	Ng'ombe	70	Kabwata Estates
5	Chelstone	27	State House	49	Chudleigh	71	New Kabwa
6	Kabulonga Ward	28	Woodlands	50	Mwambula	72	Kamwala South - 917
7	Lubwa Ward	29	Woodlands Ext	51	Handsworth	73	Town Centre
8	Chilenje Ward	30	Nyumba Yanga	52	UNZA	74	Kamwala
9	Libala Ward	31	New woodlands Ext	53	Kalundu	75	Shakespear
10	Jack Compound	32	New Chilenje	54	Rhodespark	76	Fairview
11	Kamulanga Ward	33	Chilenje South	55	Garden	77	Misisi
12	John Howard Ward	34	Chandamali	56	Olympia Park	78	Matero Ward
13	Mpulungu Ward	35	Libala South	57	Olympia Extension	79	Kanyama
14	Roma Ward	36	Libala	58	Roma	80	Chibolya
15	Munali Ward	37	Kamwala South -1080	59	Marapodi	81	John Laing
16	Kamanga	38	John Howard	60	Chaisa	82	Lilanda
17	Chamba Valley	39	John Howard	61	Emmasdale B. House	83	Munkolo Ward
18	Kaunda Square 1	40	Chawama	62	Emmasdale	84	Kanyama Ward
19	Kaunda Square 2	41	Town/Kabelenga	63	Villa Elizabertha	85	Industries
20	Chainama/Munali	42	George	64	Thornpark	86	Harry M. N. Ward
21	Mtendere	43	Matero	65	Northmead		
22	Kalikiliki	44	Chunga	66	Prospect Hill		

Figure 1: Location of the Study Area with Names of Its Neighbourhoods

The major rock type that contributes to the ground water of the area is dolomite. The dolomitic marbles underlying most of the city constitute a karstic aquifer, providing almost half of the total amount of water to Lusaka for agricultural, industrial and domestic use (Kang'omba and Baumle, 2013).

3.1.6 Selection of the Study Area

Lusaka district was chosen as a study area because the high population growth rate and the construction sector is growing faster than the rate at which water is supplied (JICA, 2009; LWSC, 2011; Simwanda and Murayama, 2018)). Lusaka City was chosen also because there is available data on water demand from studies done before this, which data enabled water demand prediction possible. Based on these studies this research study explored whether water demand analysis and prediction methods that include other variables than just population, and income can achieve similar results as the previous studies.

3.2 Research Design

The study design is both (i) qualitative - where water demand distribution is analysed and relationships between dependent and independent variables is explored and (ii) quantitative - where modelling techniques are used to estimate future water requirements for the 86 townships in Lusaka. The design outlines the sources and types of data input requirements and describes the methods of analysis which was applied to the data.

3.3 Data Collection, Processing Modelling and Analysis

This study involves three main methods: (i) collecting or generating then processing both primary and secondary data about dependent and independent variables (input data), (ii) modelling using Exploratory Regression, Ordinary Least Squares and Geographically Weighted Regression (iii) analyzing water demand distribution, relationships between water demand and its determinants and predicting water demand.

3.3.1 Primary and Secondary Data

For primary data, interviews were conducted with relevant key informants from LWSC, LCC, CSO. Secondary data was obtained from official records and reports from LWSC and CSO. This was supplemented by perusal of books, publications such as newspapers and journals on the study discipline. Landsat 2010 satellite image was obtained from U.S.

Geological Survey (2010).Shape files were collected from institutions with such data, such as LCC, LWSC and CSO. The other input data was the Lusaka City boundary shape file which was obtained from Lusaka City Council. The input data is either dependent or independent variables.

3.3.2 Dependent and Independent Variables

The Dependent Variable for this study is Water Demand and the Independent Variables are Population, Income, Irrigated Hectarage, Property Values, Tertiary Level Education Attainment, Temperature and Size of Built up Area. All the input data are for the year 2010 and the water demand prediction is for the year 2035. In the same 2010 study the prediction year was 2035, which study also used for model prediction in order to compare the projection done by Lusaka in the study in 2010 and this study since different methods were employed. Additionally, key data such as population, income and water demand data was disaggregated into neighbourhoods. Since this study's objective was exploring spatial variations of relationships between water demand and its predictors (which is the methodology adopted in this study), it used the 2010 and 2035 data.

Water demand and population data set for the neighbourhoods in the study area were obtained from a study done by (LWSC, 2011). Based on water demand for 2010 and projected population for 2035 the said study projected water demand for each neighbourhood for 2035. Income data was adapted from the same study and was ranked into 1, 2, 3, and 4 for informal, low, medium and high respectively as dummy variables. Data for property values for each neighbourhood was obtained from Lusaka City Council (LCC). Education attainment data was obtained from CSO and was based on 2010 data. It was ward level data and was disaggregated to each township. The data on education had limitations because after disaggregation (from ward level data) to townships found in each ward, each township would have the same figures on education.

Temperature data was obtained from gridded global climate and weather data documentation (Fick and Hijmans, 2017). This is a global 1km resolution gridded dataset for temperature and it was used because Lusaka city only has one recording station (City

Airport) so the resolution from the measuring stations is coarser than the gridded data. Irrigated hectarage dataset was generated by combining Normalized Difference Moisture Index (NDMI) which is used to determine vegetation water content (Senturk, Bagis and Berk, 2014).

The idea of combining the two was to identify health vegetation and also identify which of the health vegetation has high water content. This indicated vegetation that is being irrigated, especially that images were taken in the dry season. Data on the size of the built up area for 2010 was obtained by classification of the Landsat 2010 image of Lusaka. This was done in order to extract the built and non-built land cover information for Lusaka City. All variables were based on 2010 because data for water demand was for 2010. All variables were aggregated to each township using zonal statistic tool in ArcGIS 10.4.

3.3.3 Workflow

With reference to Figure 2 below, Water Demand, Tertiary Education Attainment, Population, Income, Size of Built up Area, Temperature, Irrigated Hectarage and Property Values were collected as input data variables and were processed in the next stage. Water Demand and Population data underwent data transformation by logging their data values. Tertiary Education was disaggregated to townships from ward level the rest of variable were processed and aggregated to townships. The third stage indicates the processed dependent and independent (explanatory) variables which were input data modelling (the fourth stage). Modelling was done in GIS environment using ArcMap Software by first running Exploratory Regression (ER) trials after which Ordinary Least Squares (OLS) was performed in order to specify the best global model and Geographically Weighted Regression (GWR) was run after the best model was specified. The last stage is the results stage research questions or hypotheses were tested. The first question regarding the spatial distribution of water demand was done by exploring and analysing Water demand spatial patterns. Existence or non-existence of relationships between water demand and independent variables was confirmed when OLS was run and existence or non-existence of local relationships was confirmed when GWR was run. Suitability of

GWR was tested by using GWR model to predict water demand for 2035 and then performing model calibration and validation.

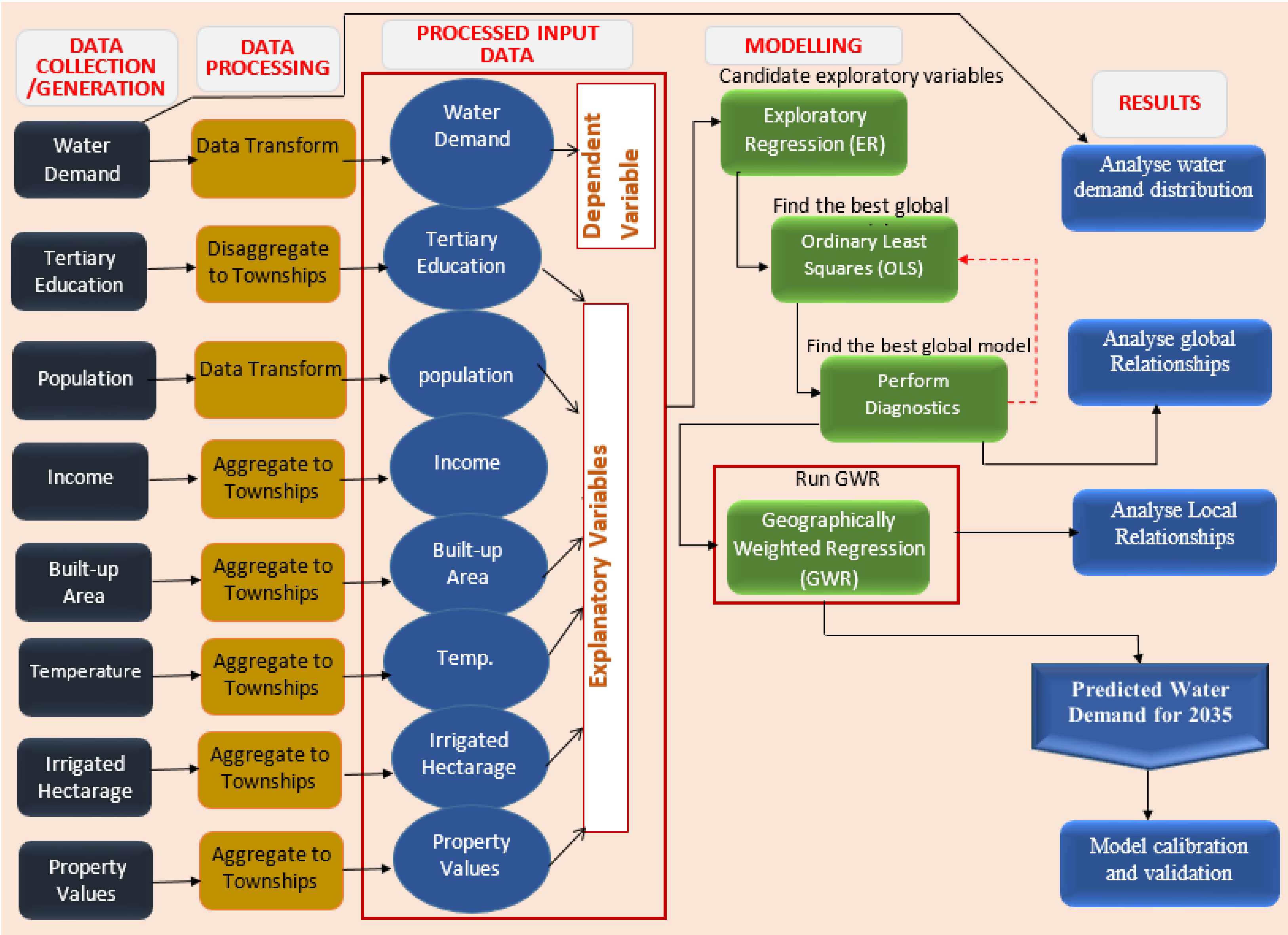


Figure 2: Work Flow of Methodology

3.3.4 Data Analysis

This study uses qualitative and quantitative regression methods to identify patterns, make analyses as well as identify correlations. The level of analysis is the neighbourhood level because the focus of this study was to explore spatially varying relationships between dependent and independent variables.

3.3.4.1 Spatial Distribution of Water Demand in Lusaka

Maps were generated showing demand patterns for overall and per capita water demand. This was done in order to visualize the spatial distribution and it is from this that spatial and non-spatial analysis were made. Further analysis was done using Grouping Analysis – a GIS analysis tool.

3.3.4.2 Relationships Between Water Demand and its Predictors

In order to test the spatial relationship between water demand and its predictors, Exploratory Regression (ER), Ordinary Least Squares (OLS) and Geographically Weighted Regression (GWR) was run one after the other in that order (Pavelescu, 2004). These are all geo-processing tools within ArcGIS 10.4 software. ER was used for selecting the appropriate predictors based on their strength of correlation with water demand – the dependent variable. This was done to identify which of the explanatory variables had significant relationship with water demand, so that they could be used for modelling in the next stage (OLS). ER tool works by iterating through different combinations of the explanatory variables and identifying the most suitable combination for each iteration. Seven iterations were done and the best 1-7 combination were made. OLS was used for diagnostic in order to specifying the best model for GWR and for testing the null hypothesis that the relationship between water demand and it predictors is insignificant (Fotheringham, Charlton and Brunsdon, 1998; Charlton and Fotheringham, 2009; Caron and Ngui, 2012; Wafula and Ngigi, 2015; Bivand, 2020). The first trial run for OLS resulted in an unspecified model and Data Transformation was of some variables was done in order to address this.

Spatial Linear Regression analysis assumes normality (random and not clustered distribution) of the model residuals. If they are not normally distributed a transformation is applied on the original data (Parker and Wilby, 2013). Therefore, after several unsuccessful trials without a properly specified model, property values with wrong sign and the three insignificant variables were dropped. If observation data deviate significantly from the normal distribution, Osborne (2002) suggests that data transformations techniques such square root, log, inverse and box-Cox should be used to improve the normality of variables. Therefore Water Demand and Population variables were transformed by logging the former and square rooting the latter figures. After this transformation OLS tool was re-run with the remaining three variables (population, income and spatial growth). GWR was done test the null hypothesis that relationships between water demand and its predictors and spatially stationary. It was also used to explore and visualise (on maps) the spatial variations of the relationships between dependent and independent variables. These regression tools were performed one after the other, starting with ER, OLS then GWR in that order. Modelling began with eight variables – Water demand, income, property values, tertiary education, population, irrigated hectarage, spatial growth and temperature. Water demand was the dependant variable while the rest were independent (predictors) variables.

3.3.4.3 Geographically Weighted Regression

Geographically Weighted Regression tool was run after the OLS model passed all the 6 required checks for a well fitted model (Howe and White, 1999; Charlton and Fotheringham, 2009) . However, since OLS is a global model, it does not take into account local variability (relationships are fixed) in townships in the study area, therefore it was only used to come up with a properly specified model, after which Geographically Weighted Regression, which takes into account local variability (relationships can vary across the study area) was run. GWR was used to test the null hypothesis that Suitability of GWR in Water Demand prediction is insignificant. WGR calibrates a regression equation for each township, such that each neighbourhood has its own R^2. Like in the OLS model water demand was used as a dependant variable while population, income and spatial growth were used as independent variables. For tool settings the kernel type was set to ADAPTIVE and the bandwidth method to Akaike's Information Criterion (AICc). AICc is a measure of model fit/performance just like adjusted R^2.

3.3.4.4 Suitability of Geographically Weighted Regression for Water Demand Prediction

Geographically Weighted Regression was used to predict water demand for 2035 for locations within the study area. Water demand for 2010 was used as the dependant variable and 2010 population and income as the explanatory variables. Spatial growth was not used for predicting because it would require simulating urban spatial growth for each township for 2035, which was beyond the scope of this study. Since the objective of this section is only to assess whether GWR is suitable or not for predicting water demand, the two variables used were sufficient.

To assess the power of prediction of GWR, model calibration and validation was performed using scatter plots. The model was calibrated using water demand, population and income variables for 2010. However, the explanatory variables for predicting were the projected population and income for 2035. Parameter settings were the same as what were used in OLS and GWR model fitting such that the Kernel type was set to ADAPTIVE and the Bandwidth method was set to AICc.

CHAPTER FOUR: RESULTS AND DISCUSSIONS

5.1 Results

5.1.1 Spatial Distribution of Water Demand

Refer to Table 3 for the summary statistics and Appendix 1 for values of the dependent and independent variables for 86 townships in Lusaka and Appendix 2 for the spatial distribution of independent variables in the city.

Table 3: Variable Statistics and Description

Variable	Description and Presentation	Mean	Standard Deviation	Minimum	Maximum
Water demand	Presented as WATER DEMAND – Water demand in cubic meters per day (m^3/d) for each township (depend variable). Data from LWSC report	2928	7,926	51	70,204
Tertiary Education	Presented as TTIARYEDU - Percentage of township population that has attained tertiary education. Generated by author based on CSO 2010 data	35	17.8	5.3	72.6
Income	Presented as INCOME_1 - Represents High, Medium, Low and informal income groups for each township ranked into 1, 2, 3, and 4. Based on LWSC report 2011	2.7	1.1	1	4
Property Values	Presented as PROPERTYVALUES – Values of properties in each township in Zambian Kwacha. Data from Lusaka City Council	1,536,046	911,032	900,000	4,000,000
Population	Presented as POP_2010_1 - 2010 population for each township. From LWSC report	17563	21,472	1,148	110,718
Irrigated Hectarage	Presented as GREENERY_1 – Size of land in hectares in each township that was being irrigated in 2010. Data generated by author using NDVI and NDMI indices	8	25	0	153
Temperature	Presented as TEMPERATURE – Average temperatures (°C) in each township in 2010. Generated by author based on Fick and Hijmans (2017) climate data	24.3	0.4	21.7	24.6
Spatial Growth	Presented as GROWTH84_2010 – Percentage of built-up area in each township in 2010. Data generated by author based on land use classification of Landsat image	81	17	28	100

In order to plot water demand pattern Hot Spot Analysis (Getis-Ord Gi*) was run in ArcGIS 10.4, using Zone of Indifference as a parameter for conceptualization of spatial relationships and 1000 meter as Distance Band parameter. The result of this geo-processing was that only Industries (one of the neighbourhoods in the study area (Figure 3) were identified as a significant Hot Spot implying that it was significantly higher than other neighbourhoods at 99% confidence level. None of the areas in Lusaka are significant cold spots (significantly lower than other neighbourhoods).

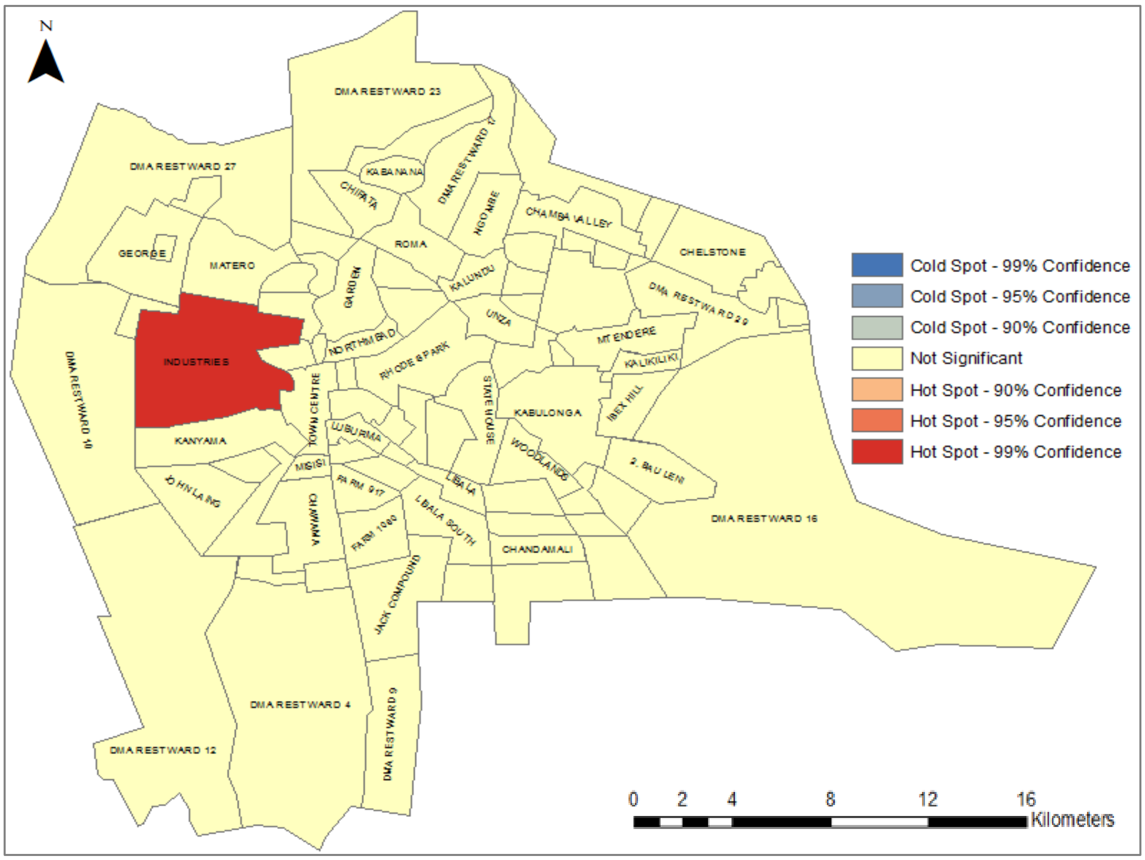


Figure 3: Hot Spot Analysis - Water Demand

This is confirmed by the results of the scatter plot (Figure 4) which showed industries with an outlying water demand figure. Even though the results of the Hot Spot Analysis does not show many patterns, visual analysis of the overall water demand (Figure 5) and per capita water demand (Figure 6) distribution maps shows some clusters. For example, on the map showing overall water demand two cluster are observed. Water demand is generally high in western and northern part of the city which is dominated by high density residential areas and industries while in centre of the city, which is dominated by low density neighbourhoods, demand is comparatively lower.

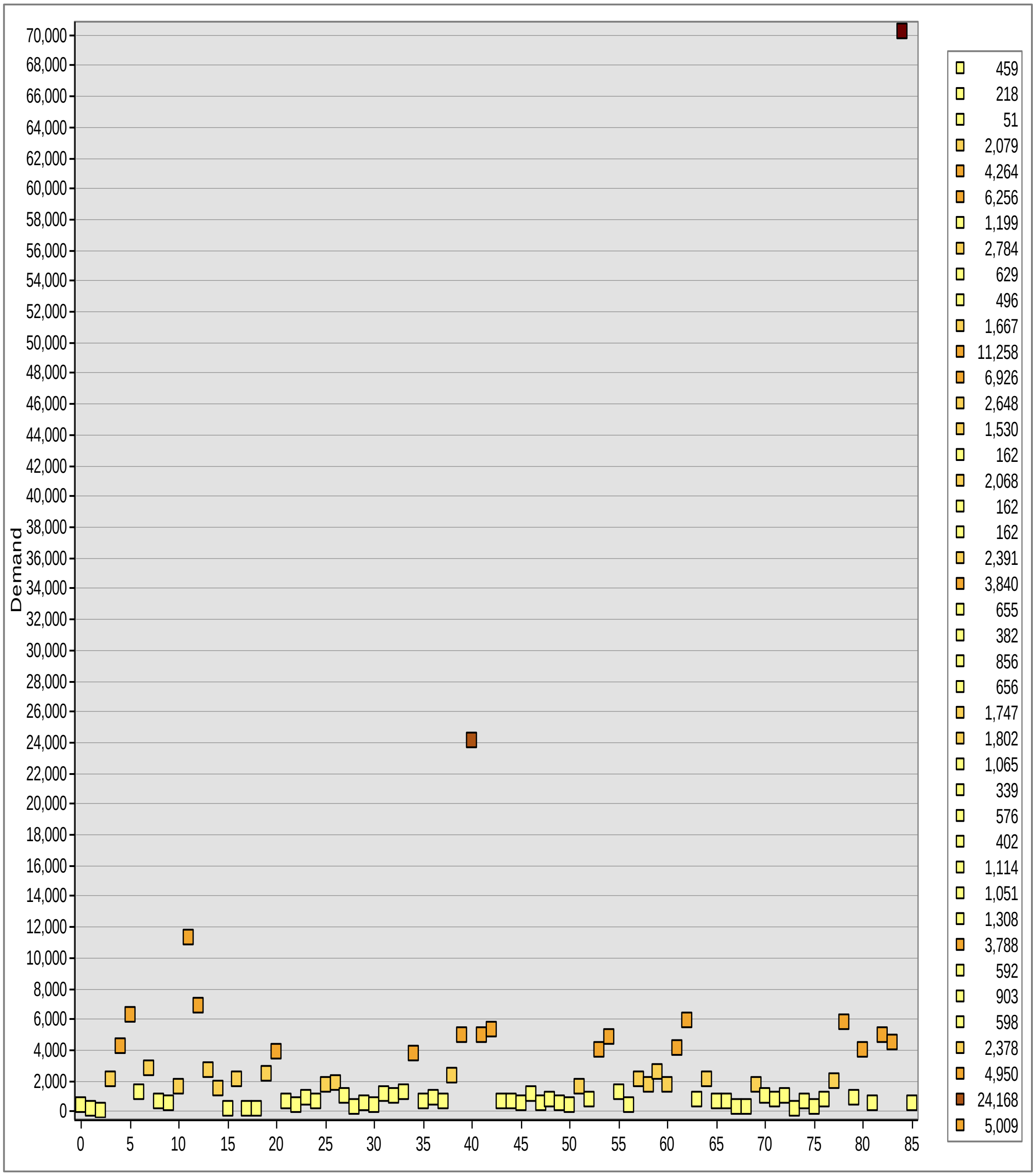


Figure 4: Scatter Plot- Water Demand (m^3/d)

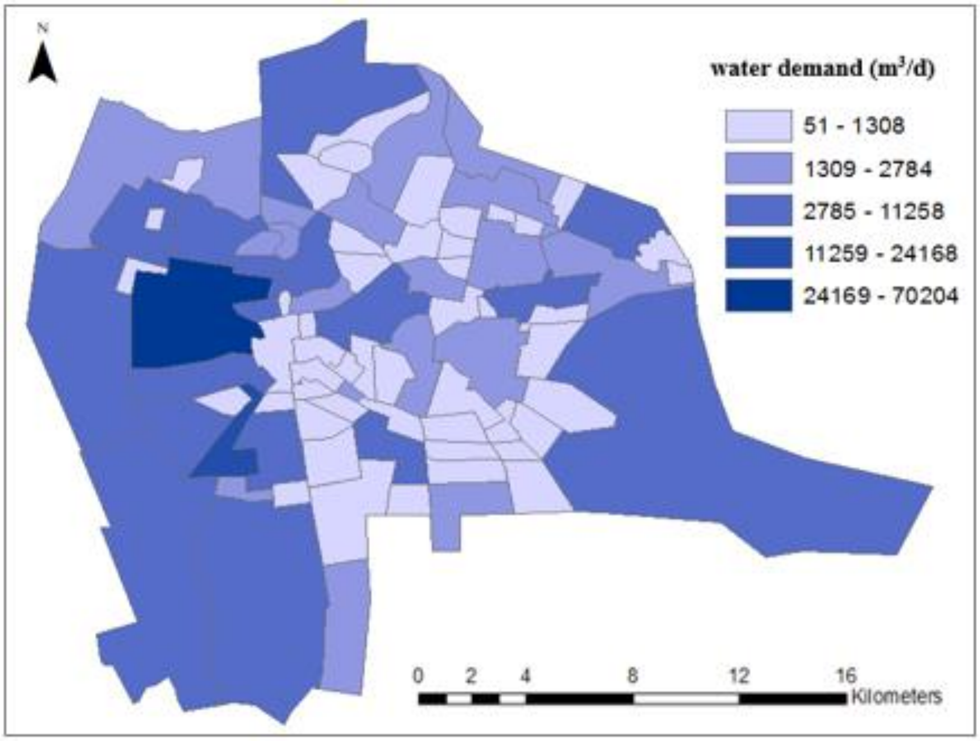


Figure 5: 2010 Overall Water Demand Distribution (m^3/d)

On the map showing per capita water demand (Figure 6) the reverse is observed i.e. per capita demand is higher in the low density areas (centre) than in high density neighbourhoods (north and west).

The existence of patterns on the 2 maps (Figure 5 and Figure 6) however, were not apparent hence no conclusions could be drawn as doing so could obscure or over-emphasize patterns. Therefore, further analysis was made using Grouping Analysis (a geo-reprocessing tool in ArcGIS 10.4) which creates neighbourhood boundaries based on their collective similarity or difference (Figure 5) (Jain, 2010). Here, neighbourhoods were analysed with regard to seven variables, population, income, property values, spatial growth size, irrigated hectarage, per capita water demand and percentage of tertiary education attainment (Figure 7). The variables were split into four groups so that within groups features are as similar as possible and between groups features are as different as possible.

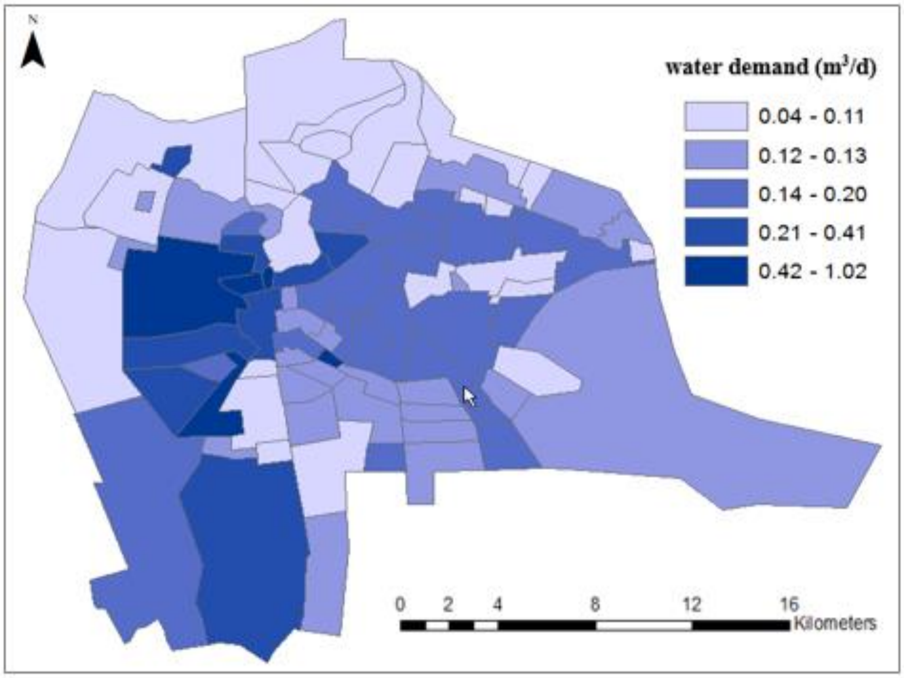


Figure 6: 2010 Per Capita Water Demand Distribution (M^3/d)

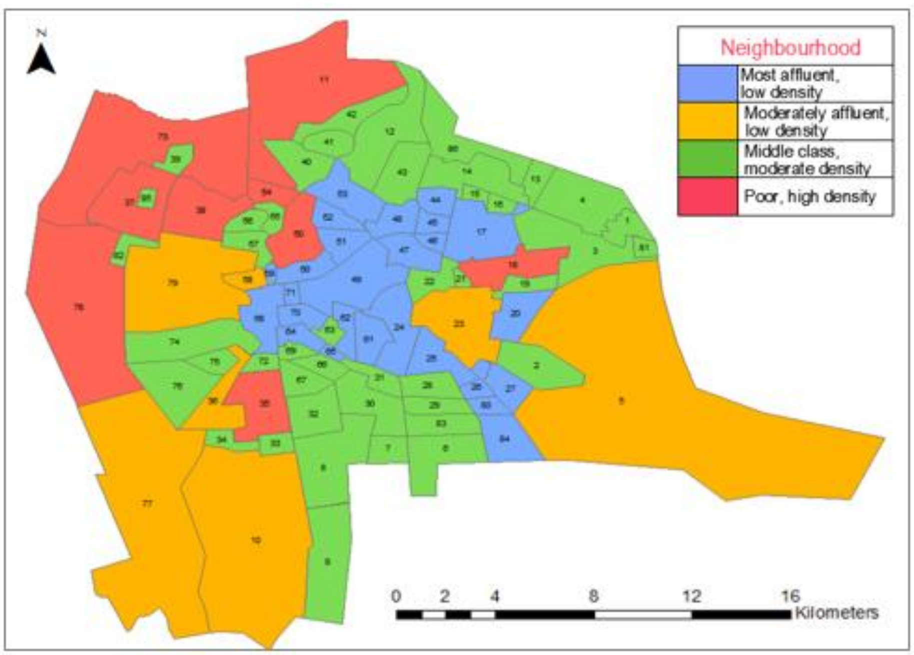


Figure 7: Groupings Based on Similarity of Variable Characteristics

As can be seen in (Figure 7) and (Figure 8) neighbourhood averages were split into four groups. The blue group (generally the most affluent and low density townships) has the highest income, property values and level of tertiary education attainment but has the highest income, property values and level of tertiary education attainment but has the lowest population.

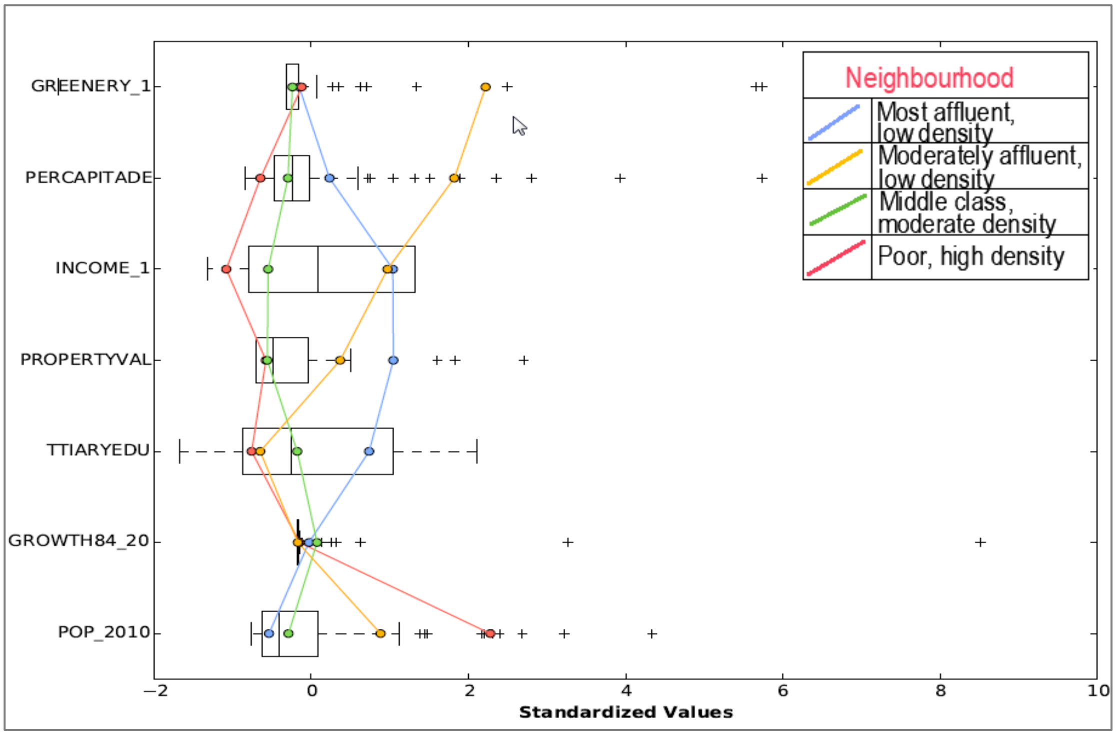


Figure 8: Parallel Box Plot Explaining Figure 7and Water Demand Variables

This group, however, has the second highest per capita water demand, spatial growth size and irrigated hectarage (greenery-1). The red group are generally the high density and lowest affluence neighbourhoods. Although this group has the highest population it has the lowest per capita water demand, property values, level of tertiary education and fall in the lowest income class. This group also has the second lowest irrigated hectarage and spatial growth size. The orange group, just like the blue group, is generally low density affluent neighbourhoods. They have the second highest population, property values and income class and the highest irrigated hectarage and per capita water demand, the second lowest tertiary level attainment and the lowest spatial growth size. The green group is generally a middle class and had the third highest population, per capita water demand, tertiary education attainment, property values and income class. It has the highest spatial growth rate and lowest irrigated hectarage. Table 4 shows the R^2 of how the Grouping Analysis tool grouped each variable based on neighbourhoods. The larger the R^2 value for a particular variable the better the grouping analysis tool categorised it.

Table 4: How Well Neighbourhoods Were Grouped Using Each Variable

Overall Variable Statistics: Count = 86, Std. Distance = 911389.1698, SSD = 339.9892

Variable	Mean	Std. Dev.	Min	Max	R2
POP_2010	17563.1395	21471.9557	1148.0000	110718.0000	0.7371
INCOME_1	8.4767	5.6870	1.0000	16.0000	0.6785
PROPERTYVALU	1536046.5116	911032.2271	900000.0000	4000000.0000	0.5358
GREENERY_1	8.0698	25.2682	0.0000	153.0000	0.4384
PERCAPITADEM	0.1626	0.1490	0.0370	1.0175	0.3744
TTIARYEDU	34.9913	17.7979	5.2523	72.5673	0.2747
GROWTH84_201	2436.8586	13764.2334	0.4746	119650.0000	0.0078

Focusing on variable associations with per capita water demand for each group in Figure 8 highlights can be made. The red, green, orange and blue groups, in that order, are generally lowest to highest income groups and it can be observed that neighbourhoods with high incomes have generally the highest per capita water. This finding is consistent with theory and findings of similar studies (Klein *et al.*, 2007; Wentz and Gober, 2007; Schleich and Hillenbrand, 2009) where high volume water users tend to have higher incomes. For population variable the red, orange green and blue clusters, in that order, have highest to lowest population but have lowest to highest per capita water demand. This is because, according to theory, per capita water demand is negatively related to household size (Wentz and Gober, 2007; Schleich and Hillenbrand, 2009). Neighbourhoods at the centre of the city have lowest population sizes and likely to have smaller household sizes hence higher per capita water demand while the converse is true for neighbourhoods in the west of the city. In addition, despite having lower populations for neighbourhood at the centre they have higher incomes and therefore are more likely to have more facilities like lawns and swimming pools per capita or lifestyles which are more water consumptive (Miaou, 1990; Balling, Gober and Jones, 2008). For spatial growth the green, blue, orange and red groups, in that order have the highest to lowest sizes of the build areas. Generally, the results show that spatial size of the built up area has a direct relationship with water demand - the bigger the size of the built area the higher water demand. As can be observed the orange group which has the second lowest built up area is the one with the highest per capita demand. This could be because the size of the built up area was calculated as a percentage of the neighborhood size. The scale of development in the orange group was not comparatively as high as in the other groups. The slow spatial growth in the orange group community could be attributed to these areas being formerly farm areas which have been slowly undergoing subdivision. Therefore, the low percentage of spatial growth but high per capita water demand is expected since these predominantly affluent communities.

5.1.2 Relationships Between Water Demand and its Predictors

5.1.2.1 Exploratory Regression

The results of Exploratory Regression are summarised in the parameters in Table 5 and in Appendix 3. Table 6 shows that there are correlations between water demand and its predictors. The level of significance and direction (positive or negative) of these relationships are summarized in this table.

Table 5: Results of Exploratory Regression

```
******************************************************************************
Choose 1 of 7 Summary
     Highest Adjusted R-Squared Results
AdjR2   AICc   JB K(BP)  VIF   SA   Model
 0.61 196.46 0.43  0.04 1.00 0.00  +POP_2010_1***
 0.09 270.54 0.00  0.25 1.00 0.00  +GREENERY_1***
 0.06 272.92 0.41  0.28 1.00 0.07  -TTIARYEDU**
       Passing Models
AdjR2 AICc JB K(BP) VIF SA   Model
******************************************************************************
Choose 2 of 7 Summary
               Highest Adjusted R-Squared Results
AdjR2   AICc   JB K(BP)  VIF   SA   Model
 0.72 169.32 0.40  0.16 1.09 0.00  +POP_2010_1***  +INCOME_1***
 0.63 194.99 0.35  0.03 1.06 0.00  +PROPERTYVALUES***  +POP_2010_1***
 0.62 196.68 0.15  0.05 1.09 0.00  +POP_2010_1***  +GREENERY_1**
       Passing Models
AdjR2 AICc JB K(BP) VIF SA   Model
******************************************************************************
Choose 3 of 7 Summary
                     Highest Adjusted R-Squared Results
AdjR2   AICc   JB K(BP)  VIF   SA   Model
 0.73 166.66 0.49  0.17 1.92 0.00  -PROPERTYVALUES**  +POP_2010_1***  +INCOME_1***
 0.72 170.93 0.55  0.13 1.39 0.00  -TTIARYEDU  +POP_2010_1***  +INCOME_1***
 0.72 171.52 0.37  0.23 1.09 0.00  +POP_2010_1***  +INCOME_1***  +GROWTH84_2010
       Passing Models
AdjR2 AICc JB K(BP) VIF SA   Model
******************************************************************************
Choose 4 of 7 Summary
                            Highest Adjusted R-Squared Results
AdjR2   AICc   JB K(BP)  VIF   SA   Model
 0.73 168.26 0.68  0.20 2.08 0.00  -PROPERTYVALUES**  -TTIARYEDU  +POP_2010_1***  +INCOME_1***
 0.73 168.62 0.41  0.22 1.97 0.00  -PROPERTYVALUES**  +POP_2010_1***  +GREENERY_1  +INCOME_1***
 0.73 168.90 0.52  0.23 1.93 0.00  -PROPERTYVALUES**  +TEMPERATURE  +POP_2010_1***  +INCOME_1***
       Passing Models
AdjR2 AICc JB K(BP) VIF SA   Model
******************************************************************************
Choose 5 of 7 Summary
                                   Highest Adjusted R-Squared Results
AdjR2   AICc   JB K(BP)  VIF   SA   Model
 0.73 170.32 0.60  0.23 2.09 0.00  -PROPERTYVALUES**  -TTIARYEDU  +POP_2010_1***  +GREENERY_1  +INCOME_1***
 0.73 170.42 0.75  0.22 2.08 0.00  -PROPERTYVALUES**  -TTIARYEDU  +TEMPERATURE  +POP_2010_1***  +INCOME_1***
 0.73 170.63 0.68  0.26 2.11 0.00  -PROPERTYVALUES**  -TTIARYEDU  +POP_2010_1***  +INCOME_1***  -GROWTH84_2010
       Passing Models
AdjR2 AICc JB K(BP) VIF SA   Model
******************************************************************************
Choose 6 of 7 Summary
                                          Highest Adjusted R-Squared Results
AdjR2   AICc   JB K(BP)  VIF   SA   Model
 0.73 172.54 0.67  0.25 2.09 0.00  -PROPERTYVALUES**  -TTIARYEDU  +TEMPERATURE  +POP_2010_1***  +GREENERY_1  +INCOME_1***
 0.73 172.76 0.60  0.28 2.12 0.00  -PROPERTYVALUES**  -TTIARYEDU  +POP_2010_1***  +GREENERY_1  +INCOME_1***  -GROWTH84_2010
 0.73 172.85 0.75  0.28 2.11 0.00  -PROPERTYVALUES**  -TTIARYEDU  +TEMPERATURE  +POP_2010_1***  +INCOME_1***  +GROWTH84_2010
       Passing Models
AdjR2 AICc JB K(BP) VIF SA   Model
******************************************************************************
Choose 7 of 7 Summary
                                                 Highest Adjusted R-Squared Results
AdjR2   AICc   JB K(BP)  VIF   SA   Model
 0.72 175.04 0.67  0.29 2.12 0.00  -PROPERTYVALUES**  -TTIARYEDU  +TEMPERATURE  +POP_2010_1***  +GREENERY_1  +INCOME_1***  +GROWTH84_2010
       Passing Models
AdjR2 AICc JB K(BP) VIF SA   Model
******************************************************************************
```

Asterisk (*) indicates that a coefficient is statistically significant (p < 0.01)

Table 6: Direction of Relationship and Level of Significance

Summary of Variable Significance

Variable	% Significant	% Negative	% Positive
POP_2010_1	100.00	0.00	100.00
INCOME_1	68.75	0.00	100.00
GREENERY_1	56.25	0.00	100.00
PROPERTYVALUES	53.12	68.75	31.25
TTIARYEDU	50.00	75.00	25.00
TEMPERATURE	18.75	0.00	100.00
GROWTH84_2010	10.94	21.88	78.12

Results of the Exploratory Regression analysis (Table 5.4) indicate variables with 50% significance or more are POP_2010_1 (100%), INCOME_1 (68.75%), GREENERY_1 (56.25%), PROPERTYVALUES (53.12%) and TTIARYEDU (50.0%). TEMPERATURE and GROWTH84_2010 had 18.75% and 10.94% respectively. Some variables were more significantly related to water demand than others. However, all of the seven variables were made to participate in next stage (OLS) because even though the level of significance were low for some of them, they are still useful for analysing variations and could still be classified as significant at OLS stage.

5.1.2.2 Ordinary Least Squares

After several trial runs resulted an improperly specified model (some variables were not significant as there is no (*) against them (Table 7) in the probability (b) column and the direction (-) of relationship for some variable in the coefficient (a) column is against theory. Charlton and Fotheringham (2009) states that if one or more variables are not significant or have an unexpected sign at OLS stage they must not be used in the next stage (GWR). Data for the remaining variables were transformed. After data transformation OLS was re-run and the results was that all variables were significant with correct type of relationships with water demand (Figure 8). The results also indicates the Adjusted R-Squared (R^2) of 0.72 (Figure 9), implying that the properly specified OLS model explains 72% of water demand variations. Visual analysis of the output map in Figure 9 shows a random distribution of residuals which confirms that the model was properly specified.

Table 7: Summary of Ordinary Least Squares Results

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]	VIF [c]
Intercept	-2.153610	5.058801	-0.425716	0.671493	6.521619	-0.330226	0.742120	--------
TTIARYEDU	-0.008141	0.005370	-1.515973	0.133575	0.005171	-1.574319	0.119468	1.414268
TEMP	0.328108	0.210999	1.555019	0.123996	0.270355	1.213616	0.228557	1.129168
PROPERTYVALU	-0.204573	0.114652	-1.784289	0.078269	0.093343	-2.191621	0.031387*	1.689095
GREENERY_1	0.001528	0.003205	0.476729	0.634893	0.001881	0.812290	0.419092	1.015278
POP_2010_1	0.000033	0.000005	6.201473	0.000000*	0.000007	4.546344	0.000021*	2.003728
INCOME_1	0.423363	0.102608	4.126005	0.000095*	0.116194	3.643572	0.000487*	2.056895
SPATIALGROWT	0.000560	0.000292	1.916694	0.058941	0.000275	2.040985	0.044634*	1.842983

Asterisk (*) indicates that a coefficient is statistically significant ($p < 0.01$)

Table 8: Summary of OLS Fitted Model Parameters

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]	VIF [c]
Intercept	5.466595	0.249550	21.905765	0.000000*	0.297632	18.366932	0.000000*	--------
POP_2010_1	0.000036	0.000005	6.803028	0.000000*	0.000008	4.708467	0.000011*	1.899899
INCOME_1	0.305097	0.077251	3.949426	0.000169*	0.088575	3.444520	0.000909*	1.099223
SPATIALGROWT	0.000607	0.000295	2.056283	0.042934*	0.000313	1.934891	0.056452	1.769384

Asterisk (*) indicates a coefficient is statistically significant ($p < 0.01$)

Table 9: OLS Diagnostics Output

Input Features:	agregate_220719	Dependent Variable:	DEMAND_LOG
Number of Observations:	86	Akaike's Information Criterion (AICc) [d]:	170.572630
Multiple R-Squared [d]:	0.731923	Adjusted R-Squared [d]:	0.722115
Joint F-Statistic [e]:	74.627273	Prob(>F), (3,82) degrees of freedom:	0.000000*
Joint Wald Statistic [e]:	135.567336	Prob(>chi-squared), (3) degrees of freedom:	0.000000*
Koenker (BP) Statistic [f]:	4.771327	Prob(>chi-squared), (3) degrees of freedom:	0.189328
Jarque-Bera Statistic [g]:	3.985136	Prob(>chi-squared), (2) degrees of freedom:	0.136345

Asterisk (*) indicates a coefficient is statistically significant (p < 0.01)

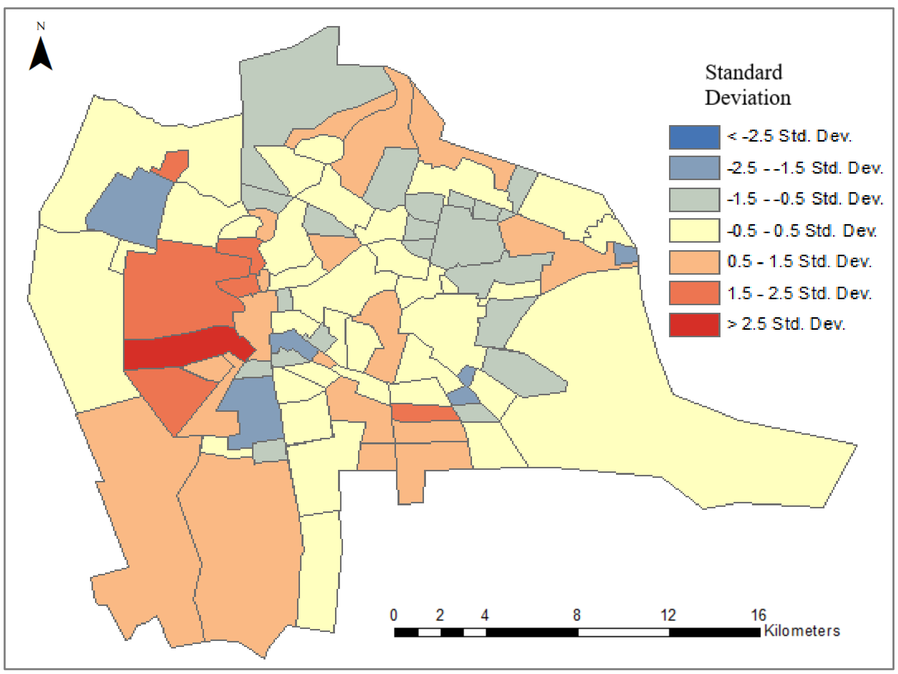


Figure 9: Over and Under Prediction Map

In order to confirm results of Figure 9 Spatial Autocorrelation (Moran's I) tool was used to test the randomness of the residuals whose results (Figure 10) indicate that residuals were spatially random and, given the z-score of 1.5, the pattern does not appear to be significantly different from random.

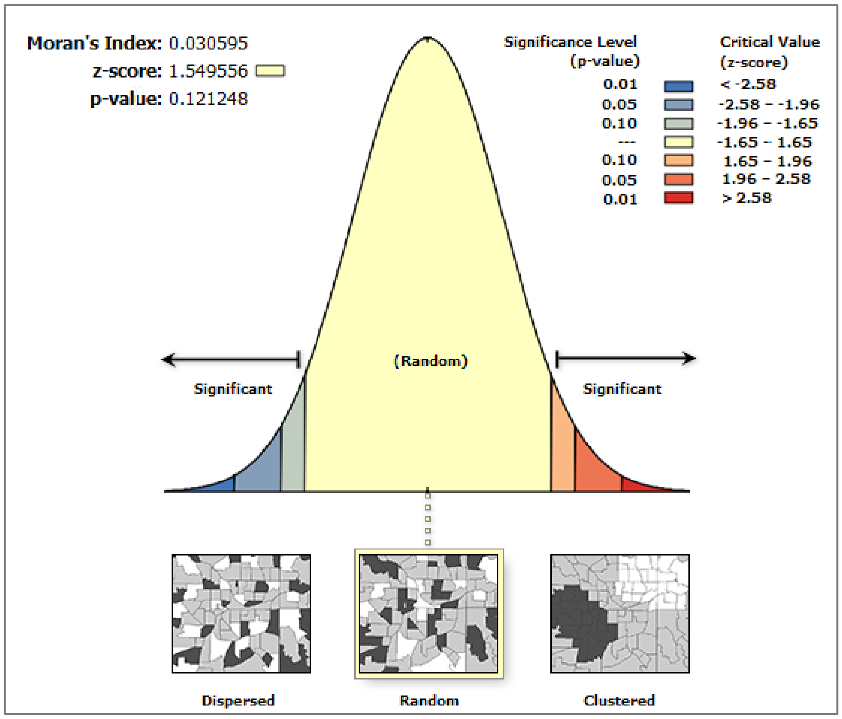


Figure 10: Spatial Autocorrelation (Moran's I) Results

Therefore, (Figure 8) and (Figure 9) above are results of the fitted model which was successfully specified (after several trials) in accordance with the procedure for performing OLS using ArcGIS software (Howe and White, 1999; Charlton and Fotheringham, 2009). According to Charlton and Fotheringham (2009) and (Howe and White, 1999) a properly specified or valid model has to pass six checks (all of which this model passed) as follows:

5.1.2.2.1 Expected Sign of Coefficients

This means that coefficients should support theory or hypothesis. All variable coefficients for population, income and spatial growth are positively related to water demand (Figure 8). This means that as the number of people, income and size of spatial urban development goes up the water demand also goes up. According to theory this is the relationship expected of water demand as a dependant variable and population, income and size of urban development as its predictors.

5.1.2.2.2 Significance of Predictors

All predictors must have statistically significant coefficients (they should be telling an important part of the story). In (Figure 8) the Probability [b] measures coefficient's statistical significance. An asterisk (*) next to the probability indicates the coefficient is significant. This is why the other four explanatory variables (irrigated vegetation, property values, temperature and percentage of tertiary education attainment) which were not significant were removed from the model as they were not helping in the model specification.

5.1.2.2.3 Pattern and Value of Residuals

To test the distribution of residuals Spatial Autocorrelation (Moran's I) tool was used and in this case (Figure 9 and Figure 10) they are spatially random (free from spatial autocorrelation - spatial clustering of over and under predictions).

5.1.2.2.4 Distribution of Residuals

This test is done using the Jarque-Bera test (Figure 9) which must not be statistically significant. For a properly specified model OLS model, the model residuals (the over- and under-predictions) are normally distributed. The Jarque-Bera test measures whether or not the residuals from a regression model are normally distributed. When it is statistically significant it means the model is biased - one or more key explanatory variables are missing. The results of OLS (Figure 9) show that the Jarque-Bera test was not statistically significant (did not have an asterisk).

5.1.2.2.5 Redundancy Among Explanatory Variables

This test is done by Variance Inflation Factor (VIF) (Figure 8) values of each variable and it is done to ensure that there is no redundancy or multicolinearity among variables - each variable should not only tell an important but a different part of the story. In order to pass this test VIF values should be lower than 10 (Olusegun et tal., 2015). If the VIF value for any of the variables is higher than 10, it means one or more variables are telling the same story. This leads to an over-count type of bias. This model passed this test since all the three variables (Figure 9) had VIF values lower than 10.

5.1.2.2.6 Model Performance

This is done by checking the strength of the adjusted R^2 (Figure 9) from the OLS output. The adjusted R^2 value ranges from 0 to 1.0 and indicate how much of the variation in the dependent variable has been explained by the model. Generally, R^2 values of 0.5 or higher indicate that the model performed well, but how good R^2 value should be depends on what is being modelled. This model's adjusted R^2 is 0.72 meaning the model has explained 72% of variations in the dependant variable (water demand). At this stage the model was properly specified and objective 2 of this research study was partially answered.

5.1.2.3 Geographically Weighted Regression (GWR)

The third and last stage for objective 2 was to use GWR to explore spatial relationships between dependent and independent variables. The results showed that adjusted R^2 improved from 0.72 in OLS (Table 9) to 0.75 (Table 10) in GWR and AICc improved (the smaller the value of AICc the better) from 171 in OLS (Table 9) to 165 in GWR (Table 10).

Table 10: Results of Geographically Weighted Regression

OBJECTID *	VARNAME	VARIABLE	DEFINITION
1	Neighbors	80	
2	ResidualSquares	25.799272	
3	EffectiveNumber	12.188729	
4	Sigma	0.591211	
5	AICc	164.86132	
6	R2	0.785834	
7	R2Adjusted	0.75337	
8	Dependent Field	0	Demand_log
9	Explanatory Field	1	Pop_2010_sqrt
10	Explanatory Field	2	Income_1
11	Explanatory Field	3	SpatialGrowth

Running GWR with 2010 data results in (Figure 11) and (Figure 12) show how well the model was fitted. These maps show the estimated (values from LWSC) and the predicted (by GWR) water demand for 2010. The maps indicate highest water demand in regions with the deepest blue and regions with the lowest water demand shown in the lightest blue colour.

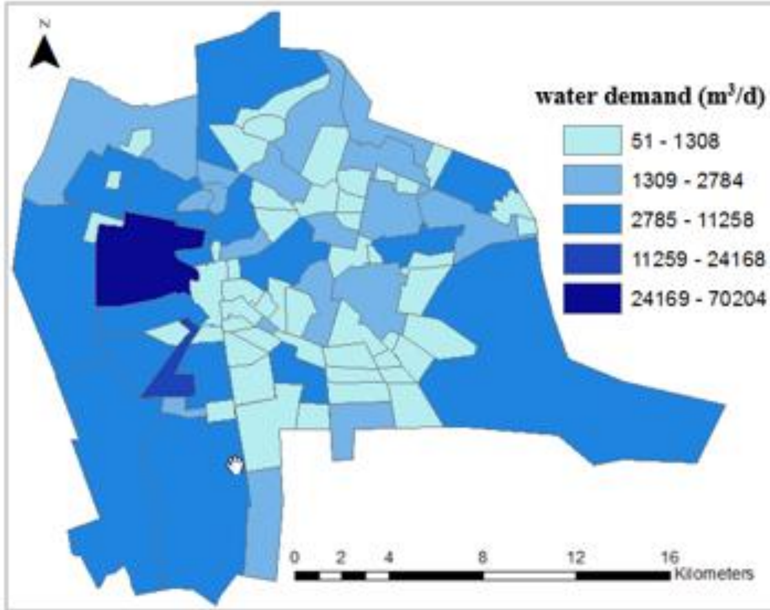


Figure 11: Estimated Water Demand (2010)

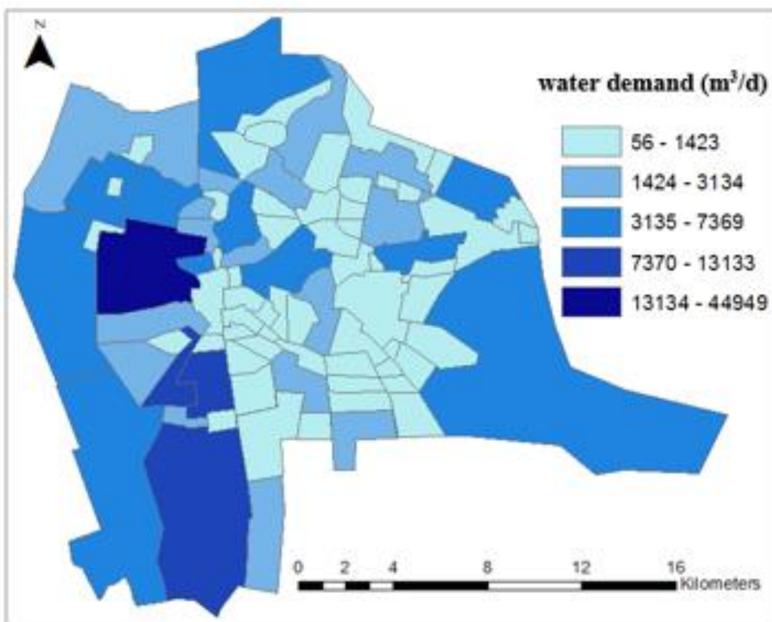


Figure 12: Predicted Water Demand (2010)

Figure 13 show variations in the R^2 (measure of how well the model was fitted) across different townships of the study region. The lowest R^2 is 0.67 and the highest is 0.82, implying that the model is explaining 67 to 82 percent of the variations in the water demand in each neighbourhood. The t-test values of the coefficients are presented in (Figure 14) in order to visualize the level of significance of coefficients.

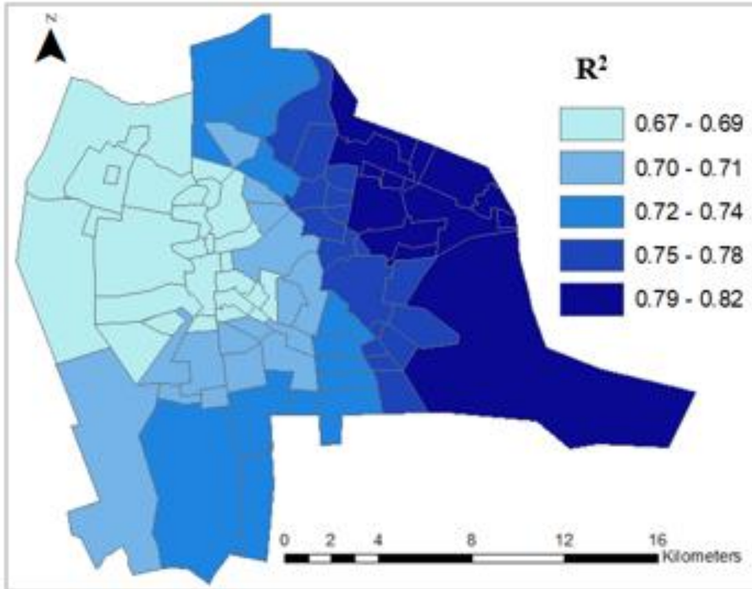


Figure 13: Local R - Squared (R^2)

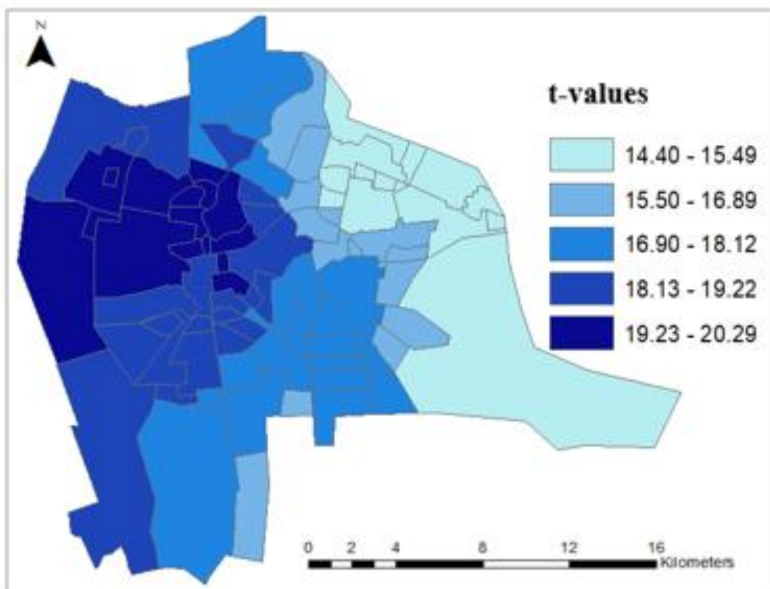


Figure 14: t-statistics Values

Out of the seven predictors only three were used in fitting the model hence only these explanatory variables are analysed. Figure 15 - Figure 17 show that spatial relationships of population, income and spatial growth variables with water demand exist. Relationships are represented by coefficients of each variable whereby, the deepest blue indicate regions where the relationship is strongest and the lightest blue the weakest for the respective variables in a particular neighbourhood of the study area.

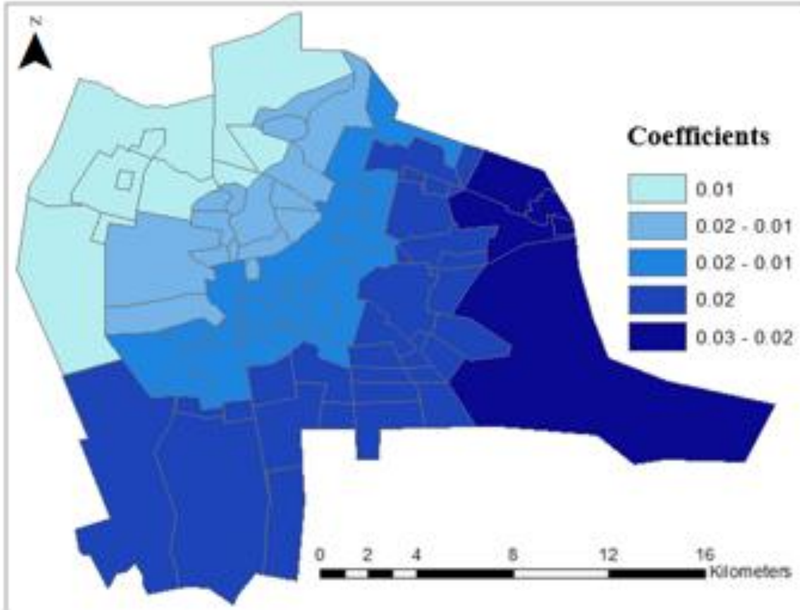


Figure 15: Spatial Relationship Between Population and Water Demand

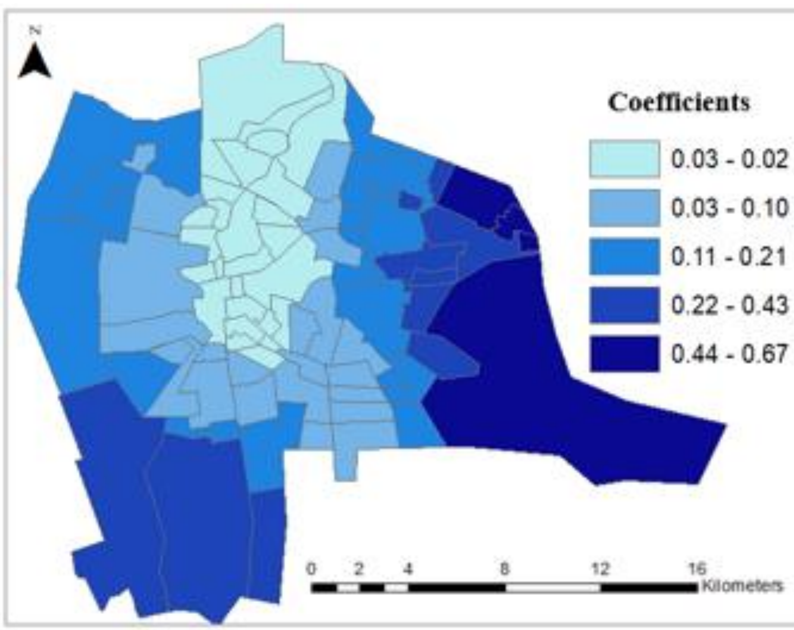


Figure 16: Spatial Relationship Between Income and Water Demand

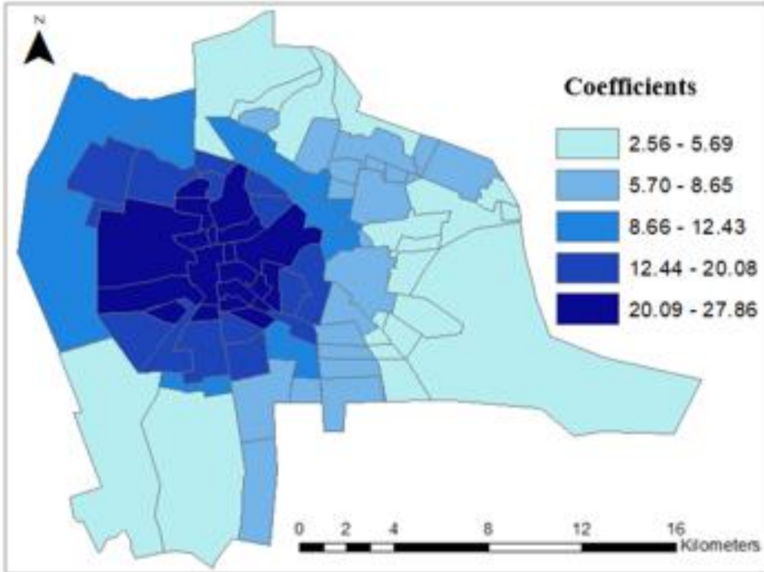


Figure 17: Relationship Between Size of Built-up and Water Demand

5.1.2.4 Test of Hypothesis

In relation to the null hypothesis that there are no significant relationships between water demand and its predictor variables, the following conclusions are made:

i. At 99% confidence level, Temperature, Property Values, Tertiary Education Attainment and Irrigated Hectarage predictor variables have no significant relationship with Water Demand. Therefore null hypothesis is not rejected.
ii. At 99% confidence level, Population, Income and Size of Built up Area predictor variables have a significant relationship with Water Demand and these relationships are location specific. Therefore, the null hypothesis is rejected.

5.1.3 Suitability of GWR in Predicting Water Demand

The total projected (total from LWSC) water demand for 2035 is 583,932 cubic meters per day (m^3/d) while the GWR modelled total is 667, 897 cubic meters per day (m^3/d), making the result from the model reliable as the difference is not significant (See totals for 2035 projected and modelled Appendix 1). Figure 18 is the map of the projected (by LWSC) water demand and Figure 19 is a map of the output of the GWR modelled water demand for 2035.

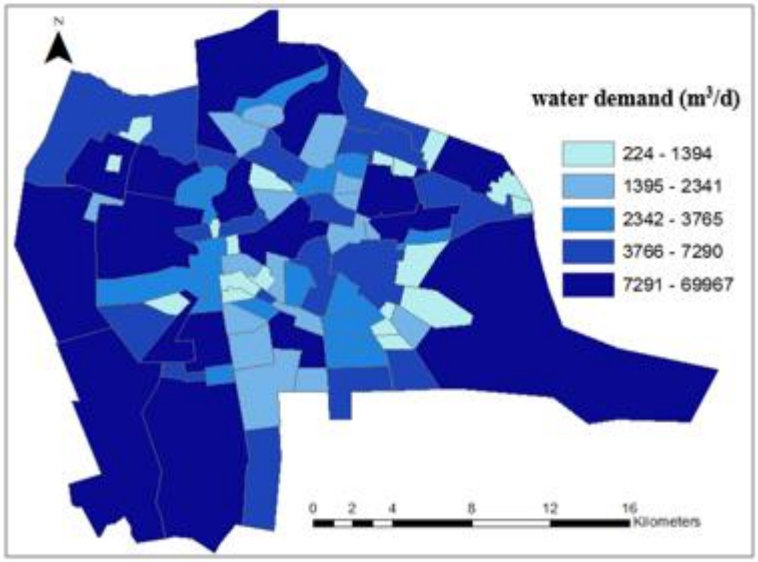


Figure 18: Projected Water Demand for 2035 (m^3/d)

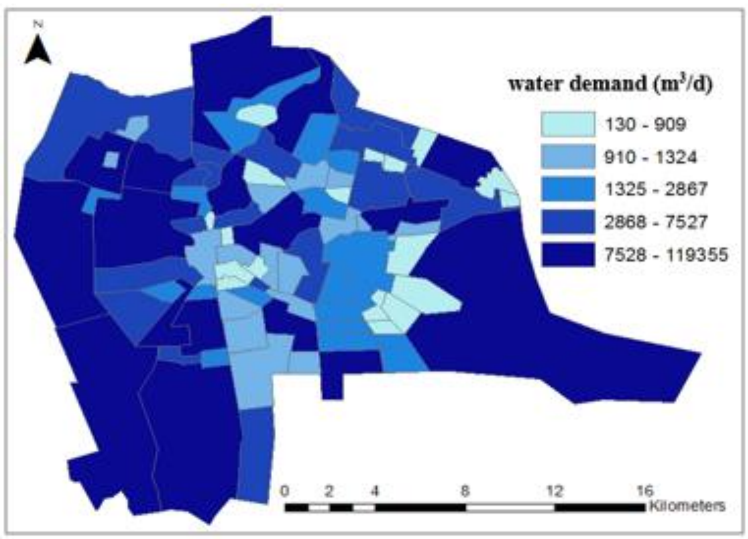


Figure 19: GWR Modelled Water Demand 2035 (m^3/d)

The variation in the strength of prediction of the model across different neighbourhoods is represented by Local R-squared (R2) in Figure 20 showing variation in model performance ranging from 59% - 91%. The reliability of the GWR model was further revealed in Figure 21, a graph of some townships, which shows that the projected and modelled graphs have minimal deviations from each other.

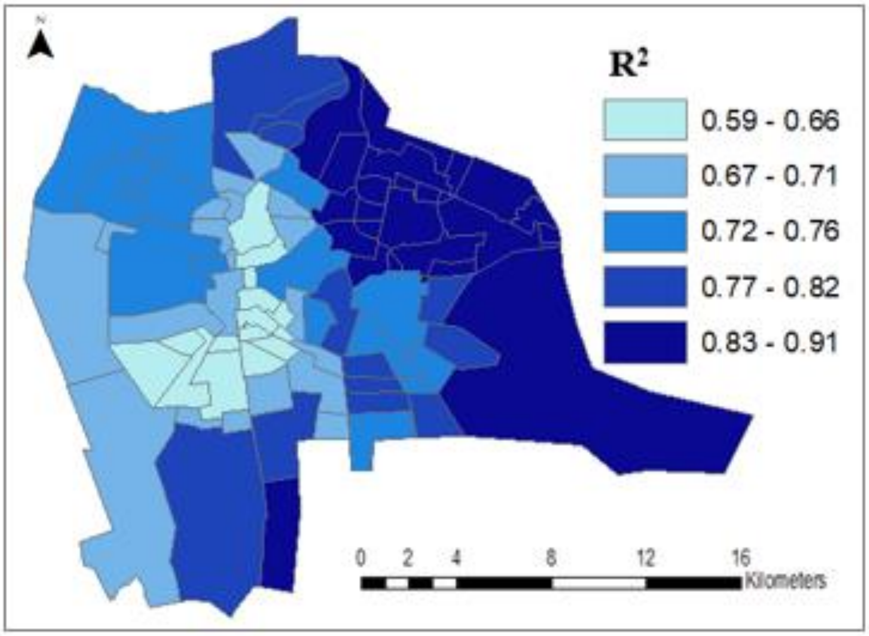


Figure 20: Local R^2 – Strength of Prediction

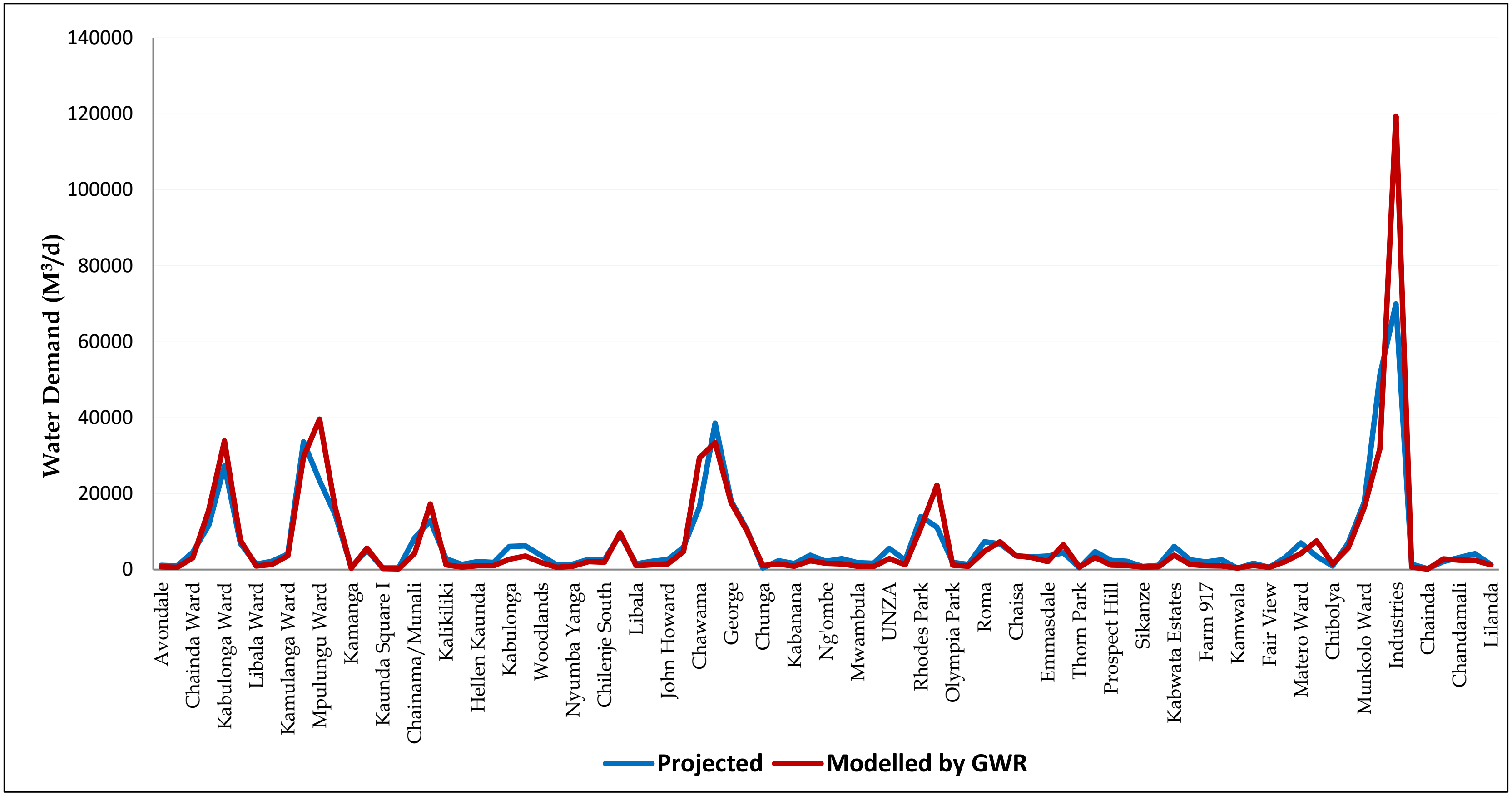


Figure 21**:** Projected vs GWR Modelled Water Demand for 2035 (M^3/d)

5.1.3.1 Strength of Prediction of the GWR Model

The estimated water demand, population and income values for each location done by LWSC (2011) for 2010 were used as calibration data sets and the corresponding 2035 estimates were used as the validation data sets (Appendix 1). The results of the model calibration (Figure 22) and validation (Figure 23) produced R^2 values of 0.85 and 0.86 respectively indicating that the projected and the water demand values are very similar. Therefore, GWR is suitable for water demand prediction since water demand estimates done by the GWR model were not significantly different from the projected figures for 2035. A similar GWR model validation method was used by Koutsias, Martínez-Fernández and Allgower (2010) when they modelled wildfire occurrence from socio-economic and demographic indicators together with land cover and agricultural statistics.

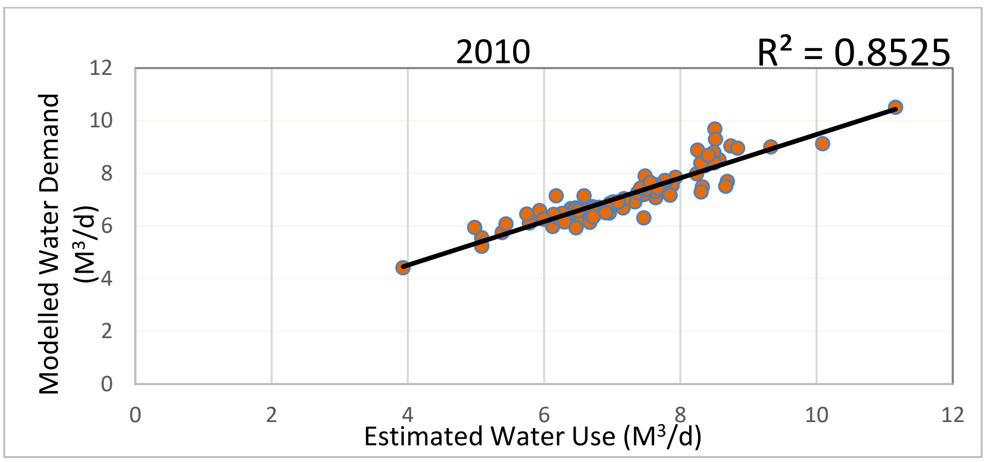


Figure 22: Calibration of the GWR Model for 2010 (Logged Values)

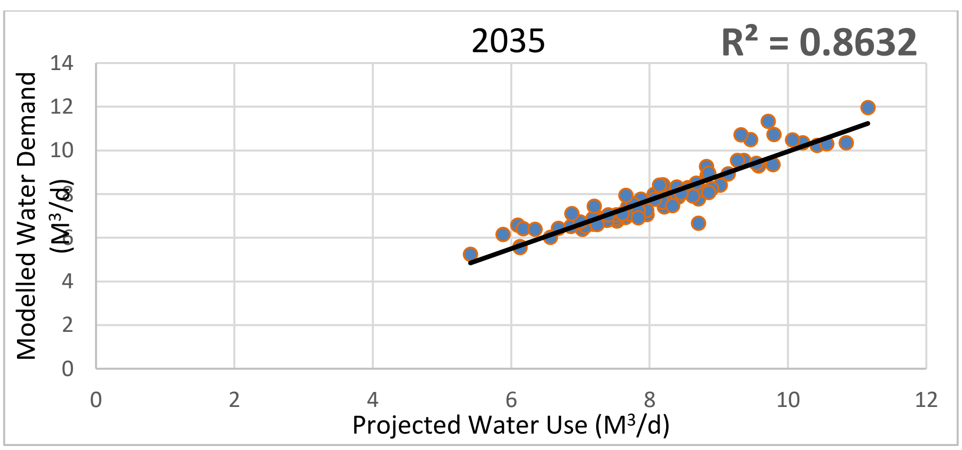


Figure 23: Validation of the GWR Model for 2035 (Logged Values)

5.1.3.2 Test of Hypothesis

In relation to the second null hypothesis that there is little or no significance in the suitability of Geographically Weighted Regression in predicting Water Demand, the null hypothesis is rejected at 99% confidence level and the study concludes, therefore, that GWR model is suitable for water demand prediction.

5.2 Discussions

For the irrigated hectarage variable there are several studies that have showed that is a significant variable when estimating water demand (Zhou et al. 2000; Guhathakurta and Gober, 2007; House-Peters and Chang, 2011). This variable, however, was insignificant in majority of neighbourhoods of the study area. This could be because the input data (Landsat images) used were of low spatial resolution. The correlation between population and water demand was positive with coefficient values higher on the south-east than on the north-east of the city (Figure 15) indicating that the relationship is stronger in the former than in the later. In all townships it implies a unit increase in population would lead to increase in water demand by a factor of the coefficient indicated in each neighborhood, assuming other variables are constant. This type of relationship is expected as it is consistent with theory and is similar to studies done by Mayer (1999), Gato (2006) and Wentz & Gober (2007) who found that water consumption increases with increase in population.

For income the type of relationship is positive and stronger in the, generally, more affluent residential areas as indicated by coefficients in Figure 16. Similar to population variable the interpretation is that household incomes and demand for water rise and fall together. This is expected because according to theory the increase in former would lead to the increase in the latter. The possible explanation to the strength of relationship being predominant in the relatively low density neighborhoods is that residents in these areas use water for extra activities such as watering of lawns and other outdoor activities likely not found in high density neighborhoods. Since water demand in high density areas is relatively not sensitive to income growth, increase in income will lead to increase in water consumption but at a smaller scale than in the low density neighborhoods. This consumption behavior was confirmed by Balling et al. (2008) who studied the link between affluence and water use and found that townships with higher incomes had more swimming pools per capita and other water consuming facilities than those with lower incomes.

For Built-Up Area the relationship is positive in the entire study area and it implies that the size of spatial development is directly related to water demand. This relationship is stronger in the central and western side of the city. This means that if there is 1unit increase in level of spatial development water demand will increase by factors ranging from 2.56 to 27. 86 depending on each neighborhood (Figure 17), assuming other variables are constant. Similar studies confirm this kind of relationship (House-Peters et al. 2010; Sanchez et al, 2017) which studies concluded that metrics describing the spatial patterns of developed land significantly explain variations in water use. Visual analysis of the distribution of coefficients shows that the association between water demand and spatial development is generally stronger in high density neighbourhoods. This could be because higher overall water consumption is associated with areas where development density is high (Sanchez et al. 2017). Population in these areas is dense and they are either unplanned or plot sizes are relatively small hence dense pattern of development. The GWR results demonstrate significant spatial variation in the relationship between Water Demand and different influencing factors. Even the insignificant variables which were not used during model fitting at OLS stage contribute locally (Appendix 4).

For tertiary education attainment variable, which had an unexpected direction of relationship with water demand, no firm conclusion can be made from this study regarding its significance. The data that was used was coarse (at ward level) which may have contributed to the result unexpected result. Property value variable was significant but its direction of relationship was not expected based on several studies reviewed (Troy et al. 2005; House-Peters and Chang, 2011). Although the relationship was positive in some neighbourhoods, it was negative in most of them. Literature reviewed does not provide an explanation to this relationship and it is recommended that further studies be carried out in this area. Significantly, the study did not find the correlation between temperature and water demand. This finding is not in agreement with studies done in similar climates in humid arid and semi-arid regions (Ruth et al. 2007; Kenney et al. 2008; Balling et al. 2008). The size of the study area could have contributed because the temperature differences were very minimal hence could not, possibly, influence variations in water demand. This variable may only be useful for demand studies whose scale is larger than neighbourhoods or time series studies whose periods span decades where temperatures changes are likely to have high impacts.

5.3 Study Limitations

The study in general had limitations of input data. For example, the resolution of raw data used to generate input data on Tertiary Education Attainment and Irrigated Hectarage variables was coarse. For Irrigated Hectarage variable Landsat images were used whose spatial resolution is 30 meters hence it is possible that irrigated lawns or gardens that are smaller than $900m^2$ could not be captured during image classification. Tertiary Education Attainment the data that was used was at ward level. An attempt was made to disaggregate the data to the respective neighbourhoods. However, the challenge was that all neighbourhoods that fell in a particular ward carried the same percentage of tertiary education attainment, which may have affected the aspect of spatial variation since this was the emphasis of the study. In some wards there is a mix of different classes of neighbourhoods. For example, Kabulonga ward has Bauleni compound and Ibex Hill among others and the percentage of residents that attained tertiary education by 2010 was 79%. This may not be the figure for all the neighbourhoods in that ward. These limitations could have contributed to data imperfections about these two variables which might have affect the model Specification. However, these and other limitations do not invalidate the findings of this study, because its main objective was to demonstrate whether relationships between water demand and its predictors vary with location and to assess the suitability of GWR in forecasting water demand and these objectives were achieved.

CHAPTER SIX: CONCLUSIONS AND RECOMMENDATIONS

6.1 Conclusions

This study revealed that overall water demand is generally highest in high density neighbourhoods whereas per capita water demand is generally highest in low density neighbourhoods. Generally, neighbourhoods where per capita Water Demand, Income levels, Property Values and level of Tertiary Education Attainment are high, have relatively low population than areas where these variables are lower. Additionally, areas that have high overall Water Demand have relatively high Population, low Income levels, low Property Values and low level of Tertiary Education Attainment.

The study revealed that relationships between Water Demand and Temperature, Education, Irrigated Hectarage are insignificant while Population, Income, Property Values and Size of the Built Area are significantly correlated with Water Demand. Exploration of relationships between water demand and its predictors is that there is a strong spatial dependence in water demand. It can also be concluded from the findings that since there is a relationship between Water Demand and Size of the Built Area, the finding is a significant contribution to the literature and, water managers would need to include it when conducting water demand studies. Inclusion of this variable is easy because raw data to generate it is available at any space and time since it just requires image classification.

It can be concluded that more or less demand for water depends on the type neighbourhood and how its inhabitants respond to factors which influence water demand. This finding is important to Lusaka Water and Sewerage Company as they are able to localise interventions based on sensitivity to these factors.

With the prediction power of 86% the projected and the modeled water demand were not significantly different from each other. Hence the study concludes that GWR is a reliable tool for water demand prediction. Since the study revealed that all variables showed non-stationarity in relation to water demand, future estimates of water demand will benefit from the use of GWR since it reveals spatially varying relationships and provides estimates at a local (neighbourhood) level as opposed to city level. The findings are not only important to water managers (Lusaka Water and Sewerage Company) but also Local Authorities responsible for spatial planning and physical development because they will be able to ensure that future urban growth planning is directed towards urban settlement types that make more efficient use of water. This is important, more so because future climate variations and their impact on water resources is uncertain.

6.2 Recommendations

i. The use of Geographically Weighted Regression in water demand prediction is recommended to Water supply managers, as it reveals spatially varying relationships between dependent and independent variables and provides reliable estimates at neighbourhood level.
ii. It is recommended that whenever there is need to conduct demand studies involving Irrigated Hectarage variable, acquisition of higher resolution images should be made. There was a challenge to obtain such images because the National Remote Sensing Centre, which archive Spot 5 which are high resolution images, did not have them for the year 2010 – the focus year for this study.

6.3 Future Research

This study used cross-sectional regression model (did not examine demand patterns over space and time) hence it is recommended that future studies should focus on spatio-temporal variations of water demand. However, this requires that data about water consumption or demand and factors that influence it must be regularly collected to make time space analysis possible.

REFERENCES

Agthe, D.E. and Billings, R.B. (1997) 'Equity and conservation pricing policy for a government-run water utility', *Aqua - Journal of Water Supply Research and Technology*, 46(5), pp. 252–260.

Ahmadi, S.H. and Sedghamiz, A. (2007) 'Geostatistical Analysis of Spatial and Temporal Variations of Groundwater Level', *Environmental Monitoring and Assessment*, 129(1–3), pp. 277–294. doi:10.1007/s10661-006-9361-z.

Araral, E. (2010) 'Urban Water Demand Management in ASEAN Countries: Challenges and Solutions', *SSRN Electronic Journal* [Preprint]. doi:10.2139/ssrn.1569630.

Arbués, F., García-Valiñas, M.Á. and Martínez-Espiñeira, R. (2003) 'Estimation of residential water demand: a state-of-the-art review', *The Journal of Socio-Economics*, 32(1), pp. 81–102. doi:10.1016/S1053-5357(03)00005-2.

Bakker, M. *et al.* (2014) 'Improving the Performance of Water Demand Forecasting Models by Using Weather Input', *Procedia Engineering*, 70, pp. 93–102. doi:10.1016/J.PROENG.2014.02.012.

Balasubramanian, A. (2015) *The World's Water. Technical report Number 2, researchgate*. Available at: https://www.researchgate.net/publication/315123891_The_World's_Water (Accessed: 26 March 2022).

Balling, R.C., Gober, P. and Jones, N. (2008) 'Sensitivity of residential water consumption to variations in climate: An intraurban analysis of Phoenix, Arizona', *Water Resources Research*, 44(10). doi:10.1029/2007WR006722.

Baumann, D.D., Boland, J.J. and Hanemann, W.M. (1997) 'Urban Water Demand Management and Planning', in *Urban Water Demand Management and Planning*. 1st edn. McGraw-Hill Professional, pp. 95–134. Available at: https://www.researchgate.net/publication/276083077_Urban_Water_Demand_Management_and_Planning (Accessed: 26 March 2022).

Bäumle, R. and Museteka, L. (2009) *Development of a Groundwater Information & Management Program for the Lusaka Groundwater Systems. Groundwater Chemistry of Springs and Water Supply Wells in Lusaka: Results of the sampling campaigns conducted in 2008, Federal Ministry for Economic Cooperation and Development (BMZ: Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung)*. Lusaka. Availableat: https://www.bgr.bund.de/EN/Themen/Wasser/Projekte/abgeschlossen/TZ/Zambia/lu_report_1.pdf?__blob=publicationFile&v=4.

Bhar, L. (2009) *Robust Regression, Indian Agricultural Statistics Research Institute Indian Agricultural Statistics Research Institute*. Available at: https://www.researchgate.net/publication/308792821_Robust_Regression (Accessed: 31 March 2022).

Bigelow, D.P. *et al.* (2017) ‘How Does Urbanization Affect Water Withdrawals? Insights from an Econometric-Based Landscape Simulation’, *Land Economics*, 93(3), pp. 413–436. doi:10.3368/le.93.3.413.

Billings, B.R. and Jones, C. V. (2008) *Forecasting urban water demand*. 2nd ed. Edited by C. Valentine, M.; Howard, R. S.; Armstrong. American Water Works Association, cop. Available at: https://www.worldcat.org/title/forecasting-urban-water-demand/oclc/860508265?referer=di&ht=edition.

Biswas, A.K. (2003) ‘Water Availability and Use’, in *Water Resources of North America*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 163–174. doi:10.1007/978-3-662-10868-0_19.

Bivand, R. (2020) *Geographically Weighted Regression*, *Encyclopedia of Geography*. doi:10.4135/9781412939591.n478.

Bizikova, L. *et al.* (2013) ‘The Water-energy-food Security Nexus: Towards a Practical Planning and Decision-support Framework for Landscape Investment and Risk Management’, *International Institute for Sustainable Development*, (February), p. 28. Available at: http://www.iisd.org/sites/default/files/pdf/2013/wef_nexus_2013.pdf.

Bouziotas, D., Rozos, E. and Makropoulos, C. (2015) ‘Water and the city: exploring links between urban growth and water demand management’, *Journal of Hydroinformatics*, 17(2), pp. 176–192. doi:10.2166/hydro.2014.053.

Caron, J. and Ngui, A. (2012) ‘Using Geographically Weighted Regression to Assess Spatial Distribution of Mental Health in an Urban Environment’, *Advances in psychology research*, 95(9), pp. 167–186. Available at: https://www.researchgate.net/publication/235940745

Chang, H., Parandvash, G.H. and Shandas, V. (2010) ‘Spatial Variations of Single-Family Residential Water Consumption in Portland, Oregon’, *Urban Geography*, 31(7), pp. 953–972. doi:10.2747/0272-3638.31.7.953.

Charlton, M. and Fotheringham, A.S. (2009) *Geographically Weighted Regression, ARC GIS Tutorial*, *National Centre for Geocomputation. National University of Ireland Maynooth*. Available at: https://www.geos.ed.ac.uk/~gisteac/fcl/gwr/gwr_arcgis/GWR_Tutorial.pdf.

CSO (2010) *2010 Census of Population and Housing. Lusaka Province Analytical Report*, *Central Statistical Office*. Available at: https://www.scirp.org/%28S%28vtj3fa45qm1ean45vvffcz55%29%29/reference/referencespapers.aspx?referenceid=2975998 (Accessed: 27 March 2022).

Domene, E., Saurí, D. and Parés, M. (2005) ‘Urbanization and Sustainable Resource Use: The Case of Garden Watering in the Metropolitan Region of Barcelona’, *Urban Geography*, 26(6), pp. 520–535. doi:10.2747/0272-3638.26.6.520.

Donkor, E.A. *et al.* (2014) ‘Urban Water Demand Forecasting: Review of Methods and

Models', *Journal of Water Resources Planning and Management*, 140(2), pp. 146–159. doi:10.1061/(ASCE)WR.1943-5452.0000314.

Ercin, A.E. and Hoekstra, A.Y. (2014) 'Water footprint scenarios for 2050: A global analysis', *Environment International*, 64, pp. 71–82. doi:10.1016/J.ENVINT.2013.11.019.

Fick, S.E. and Hijmans, R.J. (2017) *WorldClim Version2 | WorldClim - Global Climate Data*. Available at: http://www.worldclim.com/version2 (Accessed: 28 March 2022).

Fotheringham, A.S., Charlton, M.E. and Brunsdon, C. (1998) 'Geographically Weighted Regression: A Natural Evolution of the Expansion Method for Spatial Data Analysis', *Environment and Planning A: Economy and Space*, 30(11), pp. 1905–1927. doi:10.1068/a301905.

Franczyk, J. and Chang, H. (2009) 'Spatial Analysis of Water Use in Oregon, USA, 1985–2005', *Water Resources Management*, 23(4), pp. 755–774. doi:10.1007/s11269-008-9298-9.

Frost, J. (2018) *7 Classical Assumptions of Ordinary Least Squares (OLS) Linear Regression - Statistics By Jim, Statistics by Jim*. Available at: https://statisticsbyjim.com/regression/ols-linear-regression-assumptions/ (Accessed: 31 March 2022).

Gato, S. (2006) *Forecasting Urban Residential Water Demand*. RMIT University. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.733.7569&rep=rep1&type=pdf.

Guhathakurta, S. and Gober, P. (2007) 'The Impact of the Phoenix Urban Heat Island on Residential Water Use', *Journal of the American Planning Association*, 73(3), pp. 317–329. doi:10.1080/01944360708977980.

Guppy, L. and Anderson, K. (2017) *Water Crisis Report. United Nations University Institute for Water, Environment and Health, UNU-INWEH Reports*. Available at: https://inweh.unu.edu/global-water-crisis-the-facts/ (Accessed: 28 March 2022).

Gutzler, D.S. and Nims, J.S. (2005) 'Interannual Variability of Water Demand and Summer Climate in Albuquerque, New Mexico', *Journal of Applied Meteorology*, 44(12), pp. 1777–1787. doi:10.1175/JAM2298.1.

Hansen, R.D. and Narayanan, R. (1981) 'A Monthly Time Series Model of Municipal Water Demand', *Journal of the American Water Resources Association*, 17(4), pp. 578–585. doi:10.1111/j.1752-1688.1981.tb01263.x.

Hoffmann, M., Worthington, A. and Higgs, H. (2006) 'Urban water demand with fixed volumetric charging in a large municipality: the case of Brisbane, Australia', *The Australian Journal of Agricultural and Resource Economics*, 50(3), pp. 347–359. doi:10.1111/j.1467-8489.2006.00339.x.

Höglund, L. (1999) 'Household demand for water in sweden with implications of a potential tax on water use', *Water Resources Research*, 35(12), pp. 3853–3863. doi:10.1029/1999WR900219.

House-Peters, L., Pratt, B. and Chang, H. (2010) 'Effects of Urban Spatial Structure, Sociodemographics, and Climate on Residential Water Consumption in Hillsboro, Oregon 1', *JAWRA Journal of the American Water Resources Association*, 46(3), pp. 461–472. doi:10.1111/j.1752-1688.2009.00415.x.

House-Peters, L.A. and Chang, H. (2011) 'Urban water demand modeling: Review of concepts, methods, and organizing principles', *Water Resources Research*, 47(5). doi:10.1029/2010WR009624.

Howe, C. and White, S. (1999) 'Integrated Resource Planning for Water and Wastewater', *Water International*, 24(4), pp. 356–362. doi:10.1080/02508069908692188.

Jain, A.K. (2010) 'Data clustering: 50 years beyond K-means', *Pattern Recognition Letters*, 31(8), pp. 651–666. doi:10.1016/j.patrec.2009.09.011.

JICA (2009) *The Study on the Comprehensive Urban Development Plan for the City of Lusaka. Masterplan of Sub-Programmes, Water Supply and Sewerage/Drainage. Vol 2,*. Lusaka. Available at: http://open_jicareport.jica.go.jp/pdf/11932878.pdf.

JICA (2014) *Data Collection Survey on the Urban Water Supply Sector in the Republic Of Zambia*. Lusaka. Available at: http://open_jicareport.jica.go.jp/pdf/12153516.pdf.

Kang'omba, S. and Baumle, R. (2013) *Development of a Groundwater Information & Management Program for the Lusaka Groundwater Systems - Key Recommendations and Findings. Final Report*. Lusaka. Available at: https://www.whymap.org/EN/Themen/Wasser/Projekte/abgeschlossen/TZ/Zambia/lu_report_final.html.

Kayaga, S. and Smout, I. (2008) 'Water demand management: a key building block for sustainable urban water management', in *Annual Conference of Allied Social Sciences Association: 2008 Annual Meeting*. New Orleans: Transportation and Public Utilities Research Group (TPUG). Available at: http://coba.georgiasouthern.edu//tpug/conferences.htm.

Khudair, B.H., Sadeq, N.S. and Mahmoud, R.S. (2018) 'Determining and Predicting the Water Demand Dynamic System Model Mapping Urban Crawling and Monitoring Using Remote Sensing Techniques and GIS', *Journal of Engineering*, 24(6), pp. 103–113. doi:10.31026/j.eng.2018.06.08.

Klein, B. *et al.* (2007) 'Factors Influencing Residential Water Demand: a Review of the Literature', *Water*, pp. 1–44. Available at: http://sciencepolicy.colorado.edu/admin/publication_files/2006.28.pdf.

Kontokosta, C.E. and Jain, R.K. (2015) 'Modeling the determinants of large-scale building water use: Implications for data-driven urban sustainability policy', *Sustainable*

Cities and Society, 18, pp. 44–55. doi:10.1016/j.scs.2015.05.007.

Lee, S.-J. and Wentz, E.A. (2008) ‘Applying Bayesian Maximum Entropy to extrapolating local-scale water consumption in Maricopa County, Arizona’, *Water Resources Research*, 44(1). doi:10.1029/2007WR006101.

Liu, Z. and Xue, L. (2017) ‘Forecast of water demand in Beijing in 2030’, in *AIP Conference Proceedings*, p. 020125. doi:10.1063/1.4992942.

LWSC (2011) *Water Supply Investment Master Plan Lusaka, Zambia*. Lusaka. Available at: www.scirp.org/(S(czeh2tfqyw2orz553k1w0r45))/reference/referencespapers.aspx?referenceid=1945876.

Makki, A.A. *et al.* (2013) ‘Revealing the determinants of shower water end use consumption: enabling better targeted urban water conservation strategies’, *Journal of Cleaner Production*, 60, pp. 129–146. doi:10.1016/j.jclepro.2011.08.007.

Makki, A.A. *et al.* (2015) ‘Novel bottom-up urban water demand forecasting model: Revealing the determinants, drivers and predictors of residential indoor end-use consumption’, *Resources, Conservation and Recycling*, 95, pp. 15–37. doi:10.1016/j.resconrec.2014.11.009.

Mancosu, N. *et al.* (2015) ‘Water Scarcity and Future Challenges for Food Production’, *Water*, 7(12), pp. 975–992. doi:10.3390/w7030975.

Manetu, W.M., Mutua, F. and Kenduiywo, B.K. (2019) ‘Spatial Modelling of Current and Future Piped Domestic Water Demand in Athi River Town, Kenya’, *Journal of Geographic Information System*, 11(02), pp. 196–211. doi:10.4236/jgis.2019.112014.

Mansur, E.T. and Olmstead, S.M. (2012) ‘The value of scarce water: Measuring the inefficiency of municipal regulations’, *Journal of Urban Economics*, 71(3), pp. 332–346. doi:10.1016/J.JUE.2011.11.003.

Mareike, R. (2017) *Drivers of residential water demand - Analysis and outlook of water utility consumption data*. Aalto University. Available at: https://aaltodoc.aalto.fi/bitstream/handle/123456789/27994/master_Rohrdrommel_Mareike_2017.pdf?sequence=1&isAllowed=y.

Martínez-Fernández, J., Chuvieco, E. and Koutsias, N. (2013) ‘Modelling long-term fire occurrence factors in Spain by accounting for local variations with geographically weighted regression’, *Natural Hazards and Earth System Sciences*, 13(2), pp. 311–327. doi:10.5194/nhess-13-311-2013.

Mayer, P.W. and DeOreo, W.B. (1999) *Residential End Uses of Water*. Available at: https://blackwells.co.uk/bookshop/product/Residential-End-Uses-of-Water-by-Peter-W-Mayer-William-B-DeOreo-AWWA-Research-Foundation/9781583210161 (Accessed: 29 March 2022).

Miaou, S.-P. (1990) ‘A class of time series urban water demand models with nonlinear climatic effects', *Water Resources Research*, 26(2), pp. 169–178. doi:10.1029/WR026i002p00169.

Morgan, W.D. and Smolen, J.C. (1976) 'Climatic Indicators in the Estimation of Municipal Water Demand', *Journal of the American Water Resources Association*, 12(3), pp. 511–518. doi:10.1111/j.1752-1688.1976.tb02714.x.

Muvombo, M.M. (2017) *Road users perception of outdoor advertisements in Lusaka*. University of Zaamabia.

NAS (1999) *Water for the Future: The West Bank and Gaza Strip, Israel, and Jordan*. Washington, D.C.: National Academies Press. doi:10.17226/6031.

Nauges, C. and Whittington, D. (2010) 'Estimation of Water Demand in Developing Countries: An Overview', *The World Bank Research Observer*, 25(2), pp. 263–294. doi:10.1093/wbro/lkp016.

Nechifor-Vostinaru, V. (2018) *Modelling freshwater resources use and the economic impacts of demand-driven water scarcity*. University College London. Available at: https://discovery.ucl.ac.uk/10049597/1/Victor Nechifor - thesis final print.pdf.

Olmstead, S.M., Hanemann, M.W. and Stavins, R.N. (2007) 'Water demand under alternative price structures', *Journal of Environmental Economics and Management*, 54(2), pp. 181–198. doi:10.1016/j.jeem.2007.03.002.

Osborne, J. (2002) 'Notes on the use of data transformations', *Practical Assessment, Research & Evaluation*, 8(6). doi:10.7275/4vng-5608.

Parker, J.M. and Wilby, R.L. (2013) 'Quantifying Household Water Demand: A Review of Theory and Practice in the UK', *Water Resources Management*, 27(4), pp. 981–1011. doi:10.1007/s11269-012-0190-2.

Pavelescu, F.M. (2004) 'Features Of The Ordinary Least Square (Ols) Method. Implications For The Estimation Methodology', *Journal for Economic Forecasting*, 1(2), pp. 85–101. Available at: https://ideas.repec.org/a/rjr/romjef/v1y2004i2p85-101.html.

Pervez, M.S. and Brown, J.F. (2010) 'Mapping Irrigated Lands at 250-m Scale by Merging MODIS Data and National Agricultural Statistics', *Remote Sensing*, 2(10), pp. 2388–2412. doi:10.3390/rs2102388.

Polebitski, A.S. and Palmer, R.N. (2009) 'Seasonal Residential Water Demand Forecasting for Census Tracts', *Journal of Water Resources Planning and Management*, 136(1), pp. 27–36. doi:10.1061/(ASCE)WR.1943-5452.0000003.

Ramos-Bueno, A., Perevochtchikova, M. and Chang, H. (2021) 'Socio-spatial analysis of residential water demand in Mexico City', *Tecnología y ciencias del agua*, 12(2), pp. 59–110. doi:10.24850/j-tyca-2021-02-02.

Rasul, G. and Sharma, B. (2016) 'The nexus approach to water–energy–food security: an option for adaptation to climate change', *Climate Policy*, 16(6), pp. 682–702.

doi:10.1080/14693062.2015.1029865.

Rinaudo, J.-D. (2015) ‘Long-Term Water Demand Forecasting’, in Wai, G.Q.. D.K.A.. N.C.. R.J.-D.& and N., W.C. (eds) *Understanding and managing urban water in transition*. pringer, pp. 239–268. doi:10.1007/978-94-017-9801-3_11.

Robbins, P. (2007) *Lawn People: How Grasses, Weeds, and Chemicals Make Us Who We Are*. Philadelphia: Temple University Press.

Ruth, M. *et al.* (2007) ‘Adaptation of urban water supply infrastructure to impacts from climate and socioeconomic changes: The case of Hamilton, New Zealand’, *Water Resources Management*, 21(6), pp. 1031–1045. doi:10.1007/s11269-006-9071-x.

Sanchez, G.M. *et al.* (2018) ‘Spatial Patterns of Development Drive Water Use’, *Water Resources Research*, 54(3), pp. 1633–1649. doi:10.1002/2017WR021730.

Schleich, J. and Hillenbrand, T. (2009) ‘Determinants of residential water demand in Germany’, *Ecological Economics*, 68(6), pp. 1756–1769. doi:10.1016/j.ecolecon.2008.11.012.

Senturk, S., Bagis, S. and Berk, U.B. (2014) ‘Application of remote sensing techniques in locating dry and irrigated farmland parcels’, in *2014 The Third International Conference on Agro-Geoinformatics*. IEEE, pp. 1–4. doi:10.1109/Agro-Geoinformatics.2014.6910630.

Shaban, A. and Sattar, S. (2011) ‘Water security and sustainability in urban India’, *International Journal of Global Environmental Issues*, 11(3/4), p. 231. doi:10.1504/IJGENVI.2011.044552.

Simwanda, M. and Murayama, Y. (2018) ‘Spatiotemporal patterns of urban land use change in the rapidly growing city of Lusaka, Zambia: Implications for sustainable urban development’, *Sustainable Cities and Society*, 39, pp. 262–274. doi:10.1016/j.scs.2018.01.039.

Stoker, P. and Rothfeder, R. (2014) ‘Drivers of urban water use’, *Sustainable Cities and Society*, 12, pp. 1–8. doi:10.1016/j.scs.2014.03.002.

Troy, P., Holloway, D. and Randolph, B. (2005) *Water use and the built environment: patterns of water consumption in Sydney, of New South Wales, Sydney*. Sydney. Available at: http://www.be.unsw.edu.au/sites/default/files/upload/researchpaper1.pdf.

Uyanık, G.K. and Güler, N. (2013) ‘A Study on Multiple Linear Regression Analysis’, *Procedia - Social and Behavioral Sciences*, 106, pp. 234–240. doi:10.1016/J.SBSPRO.2013.12.027.

Wafula, P.N. and Ngigi, T.G. (2015) ‘Gis Based Analysis of Supply and Forecasting Piped Water Demand in Nairobi’, *International Journal of Engineering and Science Invention (IJESI)*, 4(2), pp. 1–11. Available at: http://figshare.com/articles/Gis_Based_Analysis_of_Supply_and_Forecasting_Piped_Water_Demand_in_Nairobi/1393168.

Wentz, E.A. and Gober, P. (2007) ‘Determinants of Small-Area Water Consumption for the City of Phoenix, Arizona’, *Water Resources Management*, 21(11), pp. 1849–1863. doi:10.1007/s11269-006-9133-0.

Wu, W. and De Pauw, E. (2011) ‘A simple algorithm to identify irrigated croplands by remote sensing’, *34th International Symposium on Remote Sensing of Environment - The GEOSS Era: Towards Operational Environmental Monitoring*, (January), pp. 2–5. Available at: https://www.researchgate.net/publication/259780693_A_Simple_Algorithm_to_Identify_Irrigated_Croplands_by_Remote_Sensing.

Zhou, S.L. *et al.* (2000) ‘Forecasting daily urban water demand: a case study of Melbourne’, *Journal of Hydrology*, 236(3–4), pp. 153–164. doi:10.1016/S0022-1694(00)00287-0.

.

APPENDICES

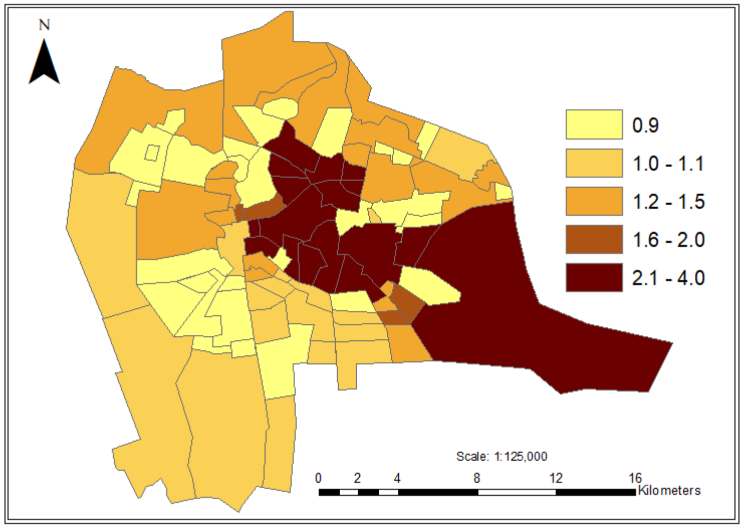
Appendix 1: Dependent and Independent Variables

Neighbourhood	Population 2010	Population 2035	Income 2010	Income 2035	Water Demand 2010 (M3/d)	Projected Water Demand -2035 (M3/d)	Property Values (Million Kwacha)	Tertiary Education Attainment (%)	Temperature (Degrees Celcious)	Irrigated Hectarage (Greenery)	2010 Built-up (Spatial Growth) (Ha)	Modelled Water by GWR-2035 (M3/d)	Local R^2
Avondale	3612	6744	3	3	459	1127	1.2	43.4	24.25	0	134	731	0.91
Bauleni	5873	10966	1	2	218	955	0.9	27.5	24.13	3	289	553	0.79
Chainda Ward	14482	27043	3	3	2079	4518	1.2	43.4	24.22	0	483	3089	0.90
Chelstone	33560	62669	3	3	4264	11641	1.1	43.4	24.26	24	348	15669	0.91
Kabulonga Ward	49241	91952	4	4	6256	27314	4	27.5	24.12	151	1694	33902	0.87
Chilenje Ward	21915	40922	3	3	2784	6836	1.1	49.6	24.20	0	382	7707	0.76
Libala Ward	4594	8577	3	3	629	1433	1.1	55.1	24.35	7	153	988	0.71
Jack Compound	13375	24977	1	2	496	2174	0.9	19.5	24.31	1	553	1324	0.77
Kamulanga Ward	13119	24497	3	3	1667	4092	1	19.5	24.30	3	234	3638	0.84
Lilayi Ward	47215	88166	3	3	11258	33646	1	22.5	24.52	153	1178	29647	0.77
Mpulungu Ward	75237	140496	2	3	6926	23470	1.2	11.2	24.30	2	1525	39640	0.78
Roma Ward	25945	48450	2	4	2648	14392	1.2	16.9	24.31	15	422	16291	0.84
Kamanga	4367	8155	2	2	162	710	0.9	42.2	24.25	0	211	337	0.88
Chamba Valley	16275	30393	3	3	2068	5077	1.2	38.4	24.27	3	96	5662	0.90
Kaunda Square I	2098	3917	2	2	162	458	0.9	38.4	24.30	0	351	236	0.88
Kaunda Square Ii	2098	3917	2	2	162	458	0.9	43.4	24.30	0	32	220	0.88
Chainama/Munali	14936	27891	4	4	2391	8285	1.2	43.4	24.24	10	61	4248	0.88
Mtendere	67311	125694	2	2	3840	12827	0.9	19.9	24.23	5	403	17242	0.87
Kalikiliki	17680	33015	1	1	655	2874	0.9	27.5	24.20	1	314	1282	0.88
Ibex Hill	2388	4459	4	4	382	1325	3.2	27.5	24.27	1	107	646	0.87
Hellen Kaunda	6740	12587	3	3	856	2103	1.1	27.5	24.30	3	240	1007	0.82
Kalingalinga	8516	15904	2	2	656	1862	0.9	26.6	24.30	8	16	1060	0.86
Kabulonga	10916	20383	4	4	1747	6055	3.2	27.5	24.21	71	126	2683	0.84
State House	11258	21023	4	4	1802	6245	4	58.1	24.24	42	304	3536	0.76
Woodlands	6655	12429	4	4	1065	3692	3	58.1	24.18	4	234	1875	0.78
Woodlands Ext	2120	3959	4	4	339	1176	1.2	58.1	24.23	0	204	585	0.75
Nyumba Yanga	4536	8471	3	3	576	1415	2	58.1	24.10	0	104	891	0.74
New Chilenje	8766	16370	3	3	1114	2735	0.9	49.6	24.30	0	181	2139	0.76
Chilenje South	8274	15451	1	3	1051	2581	1.1	49.6	24.30	0	194	1915	0.78
Libala South	29815	55676	3	3	3788	9301	1.1	55.1	24.32	0	152	9696	0.79
Libala	4663	8708	3	3	592	1455	1.1	55.1	24.30	0	382	1053	0.67
Farm 1080	7108	13274	3	3	903	2217	1.1	40.0	24.37	0	144	1247	0.68
John Howard	16139	30138	1	2	598	2624	0.9	16.2	24.40	5	302	1511	0.70

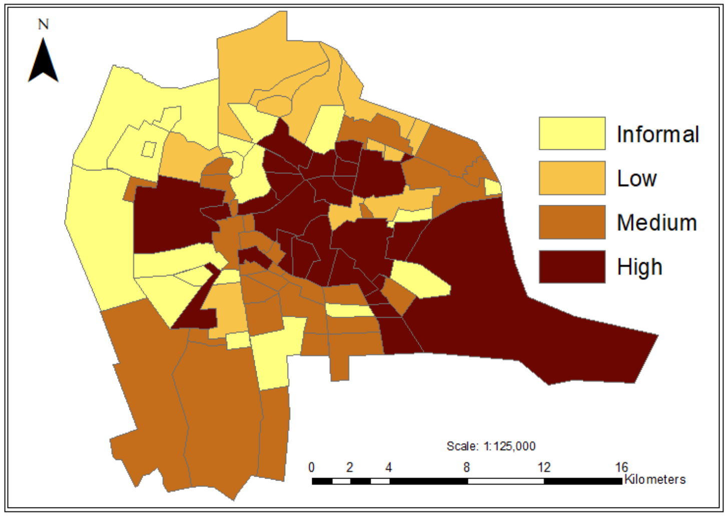
John Howard Ward	18718	34955	3	3	2378	5839	0.9	16.2	24.50	0	103	4733	0.70
Chawama	86760	162012	2	2	4950	16533	0.9	17.0	24.47	5	75	29398	0.66
Town/Kabelenga	41659	77793	4	4	24168	38528	0.9	29.7	24.47	3	445	33436	0.64
George	110718	206753	1	1	5009	17998	0.9	7.5	24.28	0	243	17590	0.64
Matero	48470	90512	2	2	5259	10594	0.9	26.6	24.38	1	606	10194	0.73
Chunga	2026	3784	1	2	645	443	0.9	31.1	24.30	0	513	1004	0.72
Chipata	13050	24370	1	2	640	2341	0.9	12.9	24.35	5	105	1517	0.73
Kabanana	7149	13349	2	2	551	1563	0.9	12.9	24.40	0	225	908	0.71
Raphael Chota Ward	12070	22539	2	3	1111	3765	1.2	12.9	24.30	2	134	2312	0.77
Ng'ombe	13311	24856	1	1	493	2164	0.9	16.9	24.30	2	231	1682	0.82
Chudleigh	5161	9638	4	4	826	2863	1.2	38.4	24.20	1	280	1463	0.86
Mwambula	3289	6143	4	4	526	1825	3	19.5	24.30	3	93	910	0.87
Handsworth	2909	5432	4	4	466	1614	3	19.5	24.24	0	54	834	0.87
UNZA	10010	18692	4	4	1602	5552	3	26.6	24.35	0	41	2867	0.87
Kalundu	4444	8298	4	4	711	2465	3	26.6	24.28	4	187	1225	0.86
Rhodes Park	25170	47001	4	4	4028	13962	3	26.6	24.29	17	134	11051	0.87
Garden	64207	119899	1	2	4898	11157	0.9	72.6	24.33	6	472	22250	0.73
Olympia Park	3321	6200	4	4	1288	1842	3	58.3	24.35	1	348	1171	0.63
Olympia Extension	2367	4421	4	4	379	1313	3	58.3	24.40	8	135	909	0.69
Roma	13143	24542	4	4	2104	7290	3	58.3	24.32	1	96	4806	0.67
Marrapodi	41649	77773	1	2	1785	6770	0.9	5.3	24.26	0	222	7335	0.74
Chaisa	22262	41570	1	2	2574	3619	0.9	21.0	24.30	0	132	3667	0.67
Emmasdale Bank House	10709	19999	3	3	1669	3341	1.2	26.6	24.35	0	76	3203	0.66
Emmasdale	11523	21519	3	2	4134	3595	1.5	26.6	24.20	0	101	2102	0.69
Villa Elizabetha	7942	14830	4	4	5941	4405	1.5	13.7	24.40	0	143	6535	0.71
Thorn Park	1535	2866	3	3	789	479	2	60.5	24.20	0	101	670	0.74
Northmead	8501	15873	4	4	2144	4715	2	60.5	24.20	4	30	3187	0.70
Prospect Hill	4408	8231	4	4	706	2445	4	58.1	24.20	8	113	1181	0.64
Government	3992	7455	4	4	639	2215	4	60.5	24.30	5	127	1081	0.73
Sikanze	2559	4780	3	3	325	798	0.9	58.4	24.30	0	125	612	0.67
Luburma	1964	3668	4	4	314	1090	1.5	40.0	24.40	0	69	709	0.62
Kabwata Estates	3923	7326	3	3	1747	6055	1.1	58.4	24.30	0	119	3820	0.59
New Kabwata	8140	15200	3	3	1034	2539	1.1	58.4	24.40	0	36	1355	0.63
Farm 917	6442	12029	3	3	818	2009	1.1	40.0	24.40	0	114	1017	0.65
Town Centre	3693	6845	3	3	1000	2529	1.1	60.5	24.40	0	171	937	0.65
Kamwala	1148	2143	3	3	146	358	1.5	40.0	24.40	0	293	390	0.67
Shakespear	5249	9802	3	3	667	1637	3	40.0	24.35	1	66	1072	0.61
Fair View	1820	3399	3	3	231	568	3	60.5	24.30	0	138	551	0.61
Misisi	19583	36569	1	1	726	3183	0.9	7.2	24.41	0	39	2126	0.65
Matero Ward	39129	73069	1	2	1919	7018	1.5	21.6	24.40	26	64	4251	0.63
Kanyama	21195	39579	1	1	5816	3445	0.9	11.9	24.40	0	1322	7527	0.74

Chibolya	5970	11148	1	2	831	970	0.9	13.7	24.40	0	553	1489	0.69
John Laing	18211	34005	1	1	4055	6992	0.9	13.7	24.49	0	121	5670	0.66
Munkolo Ward	29731	55520	3	3	4915	17709	1	29.7	24.59	71	391	16349	0.65
Kanyama Ward	64990	121360	1	2	4520	51138	1	11.9	24.33	0	1541	31915	0.68
Industries	68995	128839	4	4	70204	69967	1.5	13.7	24.38	0	1709	119355	0.67
New Woodlands Ext	2514	4694	4	4	402	1394	2	58.1	24.20	0	1338	676	0.72
Chainda	1376	2569	1	2	51	224	0.9	43.4	24.20	0	85	130	0.76
H M Nkumbula Ward	3807	7110	3	4	484	2112	0.9	13.7	21.70	0	50	2771	0.91
Chandamali	10293	19221	3	3	1308	3211	1.1	49.6	24.30	0	81	2487	0.70
Lubwa Ward	7494	13995	4	4	1199	4157	1.2	58.1	24.30	0	216	2419	0.78
Lilanda	4288	8007	1	3	545	1338	0.9	7.5	21.90	0	271	1276	0.79
Munali Ward	16616	31029	2	3	1530	5183	1.2	38.4	24.32	8	43	7486	0.72

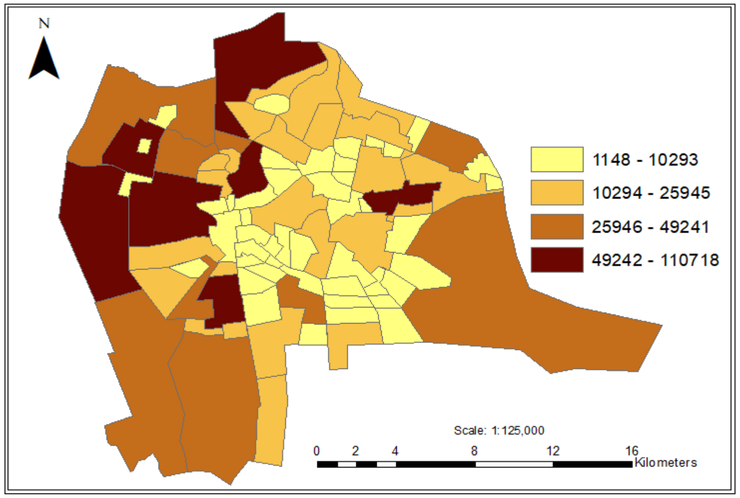
Appendix 2: Distribution of Input Variables in the Study Area



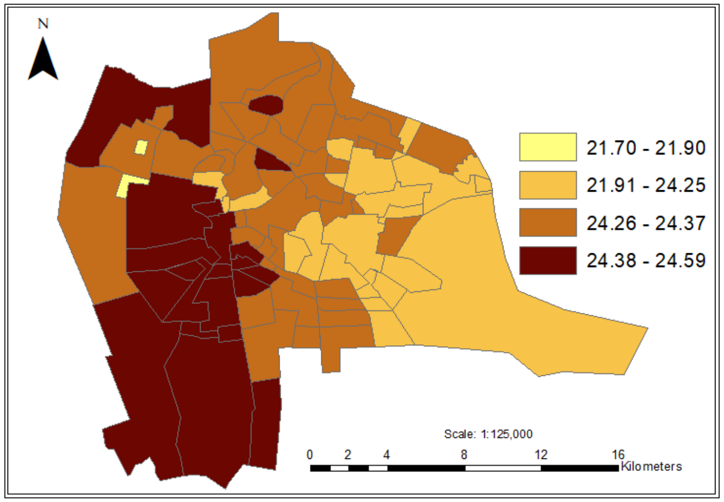
Property Values in Lusaka in (million kwacha)



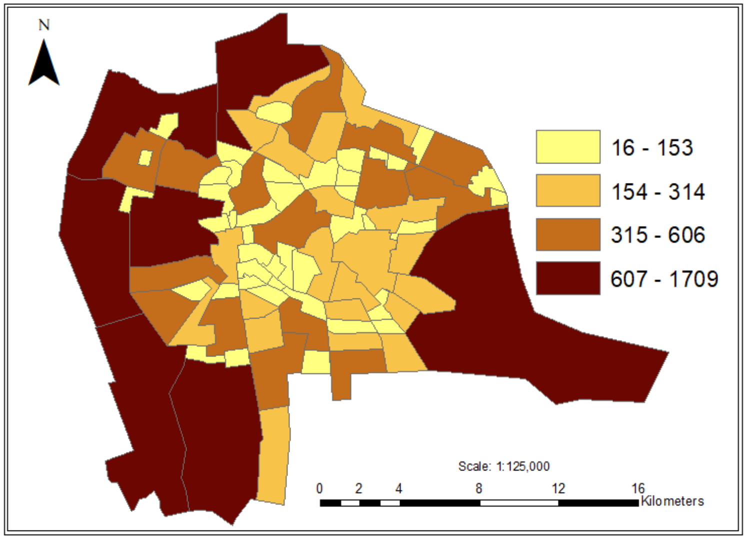
Income Distribution in Lusaka City (2010)



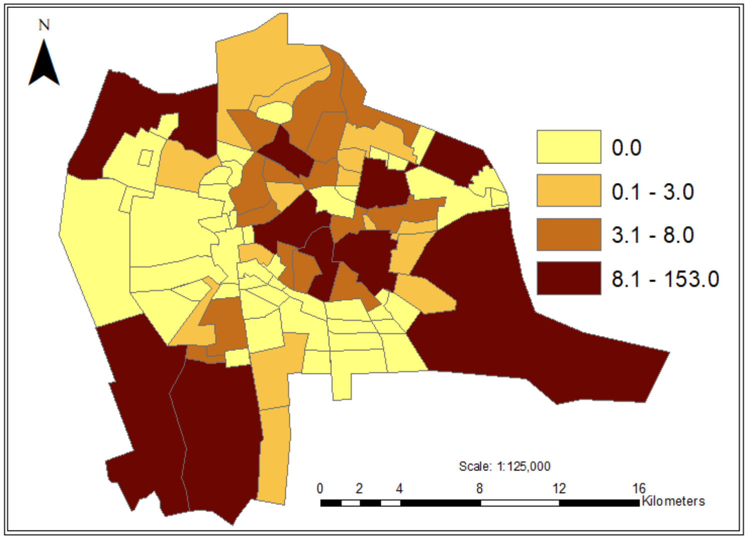
Population Distribution in Lusaka City (2010)



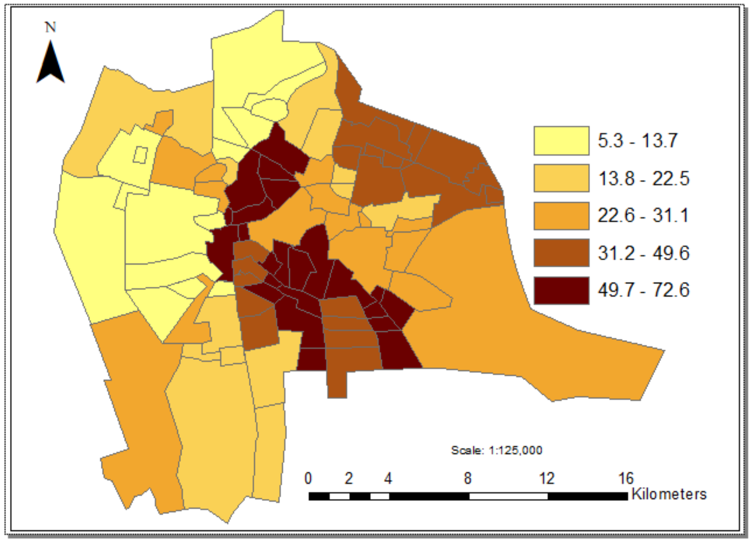
Temperature Values (°c) in 2010



Built-up Area in Lusaka City (in hectares) in 2010



Irrigated Area in Lusaka City (in hectares) in 2010



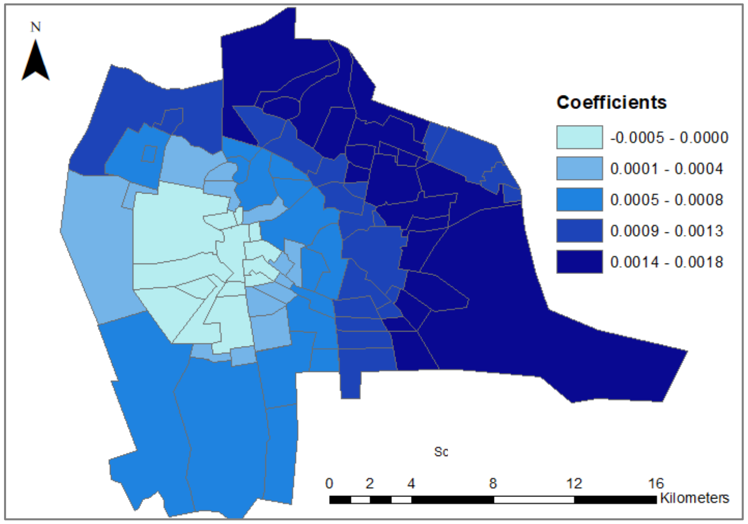
Tertiary Education Attainment in Lusaka City (%) in 2010

Appendix 3: Output Results of Exploratory Regression

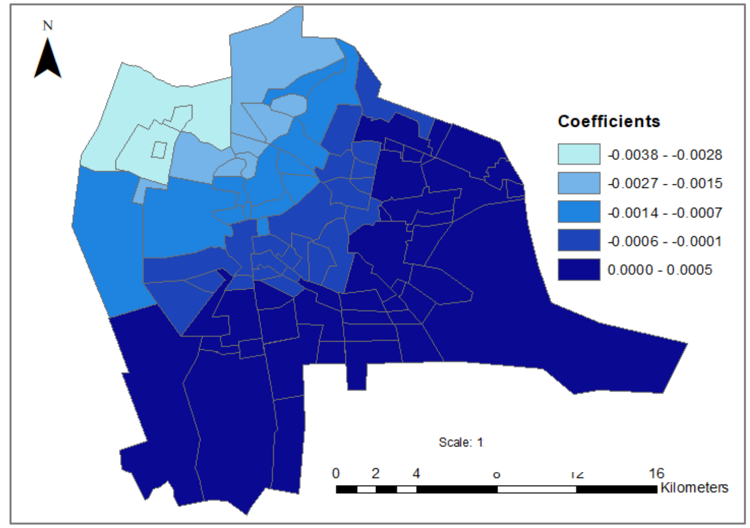
```
********************************************************************************
Choose 1 of 7 Summary
     Highest Adjusted R-Squared Results
AdjR2   AICc   JB K(BP)  VIF   SA   Model
 0.61 196.46 0.43  0.04 1.00 0.00  +POP_2010_1***
 0.09 270.54 0.00  0.25 1.00 0.00  +GREENERY_1***
 0.06 272.92 0.41  0.28 1.00 0.07  -TTIARYEDU**
       Passing Models
AdjR2 AICc JB K(BP) VIF SA   Model
********************************************************************************
Choose 2 of 7 Summary
               Highest Adjusted R-Squared Results
AdjR2   AICc   JB K(BP)  VIF   SA   Model
 0.72 169.32 0.40  0.16 1.09 0.00  +POP_2010_1***  +INCOME_1***
 0.63 194.99 0.35  0.03 1.06 0.00  +PROPERTYVALUES***  +POP_2010_1***
 0.62 196.68 0.15  0.05 1.09 0.00  +POP_2010_1***  +GREENERY_1**
       Passing Models
AdjR2 AICc JB K(BP) VIF SA   Model
********************************************************************************
Choose 3 of 7 Summary
                     Highest Adjusted R-Squared Results
AdjR2   AICc   JB K(BP)  VIF   SA   Model
 0.73 166.66 0.49  0.17 1.92 0.00  -PROPERTYVALUES**  +POP_2010_1***  +INCOME_1***
 0.72 170.93 0.55  0.13 1.39 0.00  -TTIARYEDU  +POP_2010_1***  +INCOME_1***
 0.72 171.52 0.37  0.23 1.09 0.00  +POP_2010_1***  +INCOME_1***  +GROWTH84_2010
       Passing Models
AdjR2 AICc JB K(BP) VIF SA   Model
********************************************************************************
Choose 4 of 7 Summary
                            Highest Adjusted R-Squared Results
AdjR2   AICc   JB K(BP)  VIF   SA   Model
 0.73 168.26 0.68  0.20 2.08 0.00  -PROPERTYVALUES**  -TTIARYEDU  +POP_2010_1***  +INCOME_1***
 0.73 168.62 0.41  0.22 1.97 0.00  -PROPERTYVALUES**  +POP_2010_1***  +GREENERY_1  +INCOME_1***
 0.73 168.90 0.52  0.23 1.93 0.00  -PROPERTYVALUES**  +TEMPERATURE  +POP_2010_1***  +INCOME_1***
       Passing Models
AdjR2 AICc JB K(BP) VIF SA   Model
********************************************************************************
Choose 5 of 7 Summary
                                   Highest Adjusted R-Squared Results
AdjR2   AICc   JB K(BP)  VIF   SA   Model
 0.73 170.32 0.60  0.23 2.09 0.00  -PROPERTYVALUES**  -TTIARYEDU  +POP_2010_1***  +GREENERY_1  +INCOME_1***
 0.73 170.42 0.75  0.22 2.08 0.00  -PROPERTYVALUES**  -TTIARYEDU  +TEMPERATURE  +POP_2010_1***  +INCOME_1***
 0.73 170.63 0.68  0.26 2.11 0.00  -PROPERTYVALUES**  -TTIARYEDU  +POP_2010_1***  +INCOME_1***  -GROWTH84_2010
       Passing Models
AdjR2 AICc JB K(BP) VIF SA   Model
********************************************************************************
Choose 6 of 7 Summary
                                          Highest Adjusted R-Squared Results
AdjR2   AICc   JB K(BP)  VIF   SA   Model
 0.73 172.54 0.67  0.25 2.09 0.00  -PROPERTYVALUES**  -TTIARYEDU  +TEMPERATURE  +POP_2010_1***  +GREENERY_1  +INCOME_1***
 0.73 172.76 0.60  0.28 2.12 0.00  -PROPERTYVALUES**  -TTIARYEDU  +POP_2010_1***  +GREENERY_1  +INCOME_1***  -GROWTH84_2010
 0.73 172.85 0.75  0.28 2.11 0.00  -PROPERTYVALUES**  -TTIARYEDU  +TEMPERATURE  +POP_2010_1***  +INCOME_1***  +GROWTH84_2010
       Passing Models
AdjR2 AICc JB K(BP) VIF SA   Model
********************************************************************************
Choose 7 of 7 Summary
                                                 Highest Adjusted R-Squared Results
AdjR2   AICc   JB K(BP)  VIF   SA   Model
 0.72 175.04 0.67  0.29 2.12 0.00  -PROPERTYVALUES**  -TTIARYEDU  +TEMPERATURE  +POP_2010_1***  +GREENERY_1  +INCOME_1***  +GROWTH84_2010
       Passing Models
AdjR2 AICc JB K(BP) VIF SA   Model
********************************************************************************
************ Exploratory Regression Global Summary (WATERDEMAND) *************
```

```
*********************************************************************************
************ Exploratory Regression Global Summary (WATERDEMAND) *************

             Percentage of Search Criteria Passed
                  Search Criterion Cutoff Trials # Passed % Passed
            Min Adjusted R-Squared > 0.50    127       64    50.39
           Max Coefficient p-value < 0.05    127       16    12.60
                    Max VIF Value < 7.50     127      127   100.00
           Min Jarque-Bera p-value > 0.10    127       97    76.38
Min Spatial Autocorrelation p-value > 0.10    22        0     0.00

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

        Summary of Variable Significance
Variable       % Significant % Negative % Positive
POP_2010_1           100.00       0.00     100.00
INCOME_1              68.75       0.00     100.00
GREENERY_1            56.25       0.00     100.00
PROPERTYVALUES        53.12      68.75      31.25
TTIARYEDU             50.00      75.00      25.00
TEMPERATURE           18.75       0.00     100.00
GROWTH84_2010         10.94      21.88      78.12

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

     Summary of Multicollinearity
Variable        VIF Violations Covariates
PROPERTYVALUES 2.02      0      --------
TTIARYEDU      1.49      0      --------
TEMPERATURE    1.09      0      --------
POP_2010_1     1.48      0      --------
GREENERY_1     1.24      0      --------
INCOME_1       2.12      0      --------
GROWTH84_2010  1.05      0      --------

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

                                               Summary of Residual Normality (JB)
      JB      AdjR2      AICc     K(BP)      VIF       SA   Model
0.899398 0.159474 267.017202 0.744768 2.072980 0.000771  -PROPERTYVALUES  -TTIARYEDU***  +TEMPERATURE**  +INCOME_1***
0.892441 0.149410 269.345114 0.669079 2.098621 0.000795  -PROPERTYVALUES  -TTIARYEDU***  +TEMPERATURE**  +INCOME_1***  +GROWTH84_2010
0.843694 0.141372 267.591631 0.679298 1.242975 0.000664  -TTIARYEDU***  +TEMPERATURE**  +INCOME_1**

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

                  Summary of Residual Spatial Autocorrelation (SA)
      SA     AdjR2      AICc       JB     K(BP)      VIF   Model
0.065634 0.060764 272.923625 0.410820 0.281197 1.000000  -TTIARYEDU**
0.003167 0.625650 194.985426 0.347112 0.025644 1.056947  +PROPERTYVALUES***  +POP_2010_1***
0.002644 0.613942 196.462690 0.434764 0.036616 1.000000  +POP_2010_1***

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

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)

------------------------------------------------------------------------------
```

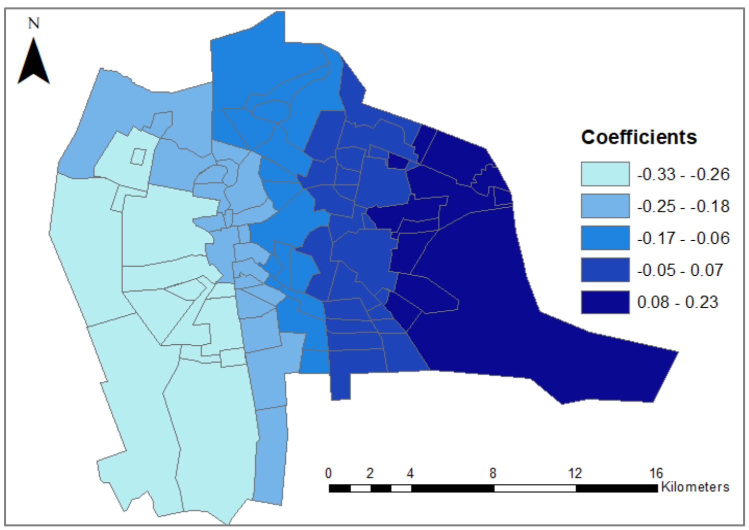
Appendix 4: Relationship Between Water Demand and its Predictors



Relation Between Water Demand and Tertiary Education



Relation Between Water Demand and Irrigated Hectarage



Relation Between Water Demand and Property Values

Appendix 5: Comparison of Methods in Urban Water Demand Modelling

Method	Data	Time Scale	Model Scale	Sources	Characteristics	Limitations
Multiple regression	time series data	daily or monthly	municipal	Maidment et al. [1985], Maidment and Miaou [1986], Gato et al. [2007], Adamowski [2008], Ghiassi et al. [2008], Caiado [2010], Agthe and Billings [1980], Al-Qunaibet and Johnston [1985], Maidment et al. [1985], Miaou [1990]	(1) short-term forecasting, (2) identify peak and off-peak daily demand to ensure necessary treatment and distribution capacity to meet demand, (3) estimate price and income elasticities, (4) assess the differential effect of demand determinants during peak versus off-peak periods	(1) lacking or highly aggregated spatial data, (2) aggregated data used instead of microlevel data, (3) OLS assumption of independence in the error term is easily violated, (4) endogeneity bias due to price based on consumed quantity under multipart tariff pricing, (5) serial correlation bias
Multiple regression	time series data	seasonal/sinusoidal	municipal	Maidment et al. [1985], Maidment and Miaou [1986], Miaou [1990], Zhou et al. [2000], Syme et al. [2004], Gutzler and Nims [2005], Gato et al. [2007], Praskievicz and Chang [2009], Wong et al. [2010]	(1) short-term forecasting, (2) separate water use into two components: weather-insensitive base use (winter or indoor) and weather-sensitive seasonal use (summer or outdoor), (3) examine climatic effects on demand	(1) lacking or highly aggregated spatial data, (2) difficult to determine the correct functional form of the model, (3) conventional models may underestimate water use response to climate variables because of the influence of stochastic events on seasonal use
Multiple regression	dynamic panel data	monthly/bimonthly	household level or census tract scale	Nauges and Thomas [2003], Arbués et al. [2004], Arbués et al. [2010],	(1) incorporates both temporal and subject-based variability into coefficient estimates, (2) more	(1) analysis of both disaggregated temporal and spatial data is more time and data intensive than traditional methods,

				Polebitski and Palmer [2010]	efficient and consistent parameter estimates than OLS, (3) fine spatial scale data, (4) integrate lagged independent variables (e.g., price)	(2) determining the regression method (e.g., fixed effects versus random effects) to account for spatial variability remains uncertain
Piecewise linear regression	time series data;	cross-sectional data	daily or monthly municipal or census block group scale	Maidment et al. [1985], Maidment and Miaou [1986], Chang et al. [2010]	(1) determine structural or temporal regime shift in a regression model, (2) create discrete linear segments connected at a point of change, (3) capture nonlinear responses in slope when thresholds are passed, (4) simple to interpret	(1) difficult to determine a priori the knot (point of change), (2) statistical testing required to ensure that the slopes are statistically significantly different before and after the knot
Spatially explicit ordinary least squares (OLS) regression	cross-sectional data; geotagged data	monthly/bimonthly	census block group scale or census tract scale	Chang et al. [2010], House-Peters et al. [2010], Shandas and Parandvash [2010], Guhathakurta and Gober [2007], Wentz and Gober [2007], Balling et al. [2008], Lee and Wentz [2008], Lee et al. [2010]	(1) visualize and quantify water use patterns at fine spatial scales, (2) elucidate spatial patterns of clustering and dispersion of high and low water users, (3) model individual household level consumption data, (4) correct for heterogeneity due to spatial autocorrelation, which otherwise causes biased parameter estimation	(1) water provider service areas do not match administrative boundaries (e.g., census block), (2) data usually must be aggregated to protect customer privacy, (3) no consistency between water providers regarding collection of water use data

weighted regression (GWR)	cross-sectional data; geotagged data	monthly individual water consumption	data aggregated to census tract scale	Wentz and Gober [2007]	(1) forecast small-area water consumption, (2) accounts for spatial autocorrelation, (3) calculates a set a unique regression equation for each observation defined by geographic coordinates, (4) improvement over OLS (R2 = 0.64) versus GWR (mean R2 = 0.85)	(1) computationally and data intensive, (2) each sample has it own unique regression, difficult to interpret results for a large sample, (3) availability of geotagged data continues is lacking, (4) model demonstrates spatial variation but does not have explanatory power
Simultaneous equation demand model	panel data	monthly	municipal	Agthe et al. [1986], Espey et al. [1997], Torregrosa et al. [2010]	(1) simultaneously and jointly determines the endogenous dependent variables on the basis of exogenous variables, (2) corrects for multicollinearity and serial autocorrelation	(1) data aggregation at large spatial scales, (2) violation of the economic assumption that households are perfectly informed of water price
Autoregressive integrated moving average (ARIMA)	time series data	daily	municipal	Bougadis et al. [2005], Adamowski [2008], Praskievicz and Chang [2009]	(1) accounts for the autocorrelation in the water demand time series, (2) uses the previous day's water use as an independent variable	(1) data aggregation at large spatial scales, (2) difficult to determine a priori correct model form and parameters
State-space forecasting model	time series data	monthly	municipal	Billings and Agthe [1998]	(1) Computes forecasts based on the dependence of a variable upon its own lags and the cross lags of the independent variable, (2) simpler than ARIMA	(1) The values of the independent variables must be forecasted in order to compute a forecast of the dependent variable

Bayesian maximum entropy (BME)	cross-sectional data; soft data	monthly	census tract scale	Lee and Wentz [2008], Lee et al. [2010]	(1) assimilates data uncertainty into the data extrapolation and mapping process, (2) obtain downscaled estimates from spatially aggregated data, (3) project future water use, (4) inclusion of data uncertainty as soft data	(1) data extrapolation can lead to dubious results because the values are estimated beyond the scope of the known data, (2) computationally intensive, (3) variances can be overestimated if interaction terms are neglected in the models used to build the probabilistic soft data
Artificial neural networks (ANNs)	time series data	daily or hourly	municipal	Adya and Collopy [1998], Bougadis et al. [2005], Adamowski [2008], Ghiassi et al. [2008], Bárdossy et al. [2009], Firat et al. [2009], Adamowski and Karapataki [2010], Herrera et al. [2010]	(1) highly effective for forecasting short-term demand (99% accuracy), (2) alternative to traditional linear modeling approach, (3) explicitly analyze nonlinear time series events, (4) minimize relative error, (5) maximize robustness	(1) complex, data and computationally intensive training and testing requirements, (2) loss of parsimony, (3) lack of explanatory power of the results, (4) sensitive to misspecification error
System dynamics models (SDMs)	time series data	monthly	regional	Rosenberg et al. [2007], Winz et al. [2009], Ahmad and Prasha [2010]	(1) incorporate diverse variables and submodels, (2) visualize the effects of intervention strategies, (3) continuously test assumptions and system sensitivity under scenarios	(1) unlike ABMs, the behavior of neighbors and the influence of this behavior cannot be simulated, (2) data, software, and computationally intensive

Source: House-Peters (2011)