TIME SERIES ANALYSIS OF WATER CONSUMPTION IN THE HOHOE MUNICIPALITY OF THE VOLTA REGION OF GHANA

By

COSMOS AGBE KWAME TODOKO (BSc. Statistics, HND Statistics)

A Thesis submitted to the Department of Mathematics, Kwame Nkrumah University of Science and Technology in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE (Industrial Mathematics)

Institute of Distance Learning

APRIL, 2013

CERTIFICATION

I hereby declare that this submission is my own work towards the MSc. degree and that, to the best of my knowledge, it contains no material previously published by another person nor material which has been accepted for the award of any other degree of the University, except where due acknowledgement has been made in the text.

Cosmos Agbe Kwame Todoko	…………………	…………………
(PG 6324811)	Signature	Date

Certified by:

Professor S. K. Amponsah	…………………	………................
(Supervisor)	Signature	Date
Mr. F. K. Darkwah	…………………	…………………
(HOD)	Signature	Date
Professor I. K. Dontwi	…………………	…………………
(Dean, IDL)	Signature	Date

ABSTRACT

Water is considered as lifeline of all living things especially humans, hence its availability is a critical component in the measurement of human wellbeing through the Human Development Index (HDI). Its production and distribution in Ghana particularly in the Hohoe Municipality of the Volta Region is a challenge. This study seeks to identify the best fit time series model to the water consumption data in the Hohoe Municipality and to forecast water consumption in the Municipality. This underpins the development of a time series model for forecasting water consumption levels of the residents, institutions and businesses in the municipality. Several time series models including AR, MA, ARMA, ARIMA and SARIMA were fitted to the data, and it emerged that the most adequate model for the data was ARIMA (2, 1, 2). There will be no astronomical increases in water consumption levels in the municipality over the next 4 years. It is recommended that the Ghana Water Company Limited should use the model and its forecasted figures in its operational and planning activities.

ACKNOWLEDGEMENTS

The successful completion of this work owes much to individuals who cannot go unnoticed. My heartfelt gratitude goes to Professor S. K. Amponsah, my supervisor, for painstakingly perusing through my work and most particularly his personal friendship with me. I sincerely appreciate the invaluable tuition and direction of all lecturers of the Department of Mathematics at the Kwame Nkrumah University of Science and Technology (KNUST), Kumasi especially, Dr. Richard K Avuglah, Nana Kena Frempong, Dr. F. T. Oduro, and Dr. Richard Koufie. I also appreciate the invaluable assistance and morale support of my dear wife, Mrs. Mabel Dosoo-Todoko, Selikem Moses Todoko, James Todoko and Akpene Tsogbe, Also, I thank my colleagues especially Francis Mawuli Abude, Japheth Arthur and Bernard Kudjo Nana Brany. My sincere thanks go to the officers particularly Mr. Avornu and Mr. Agboka at the Hohoe Branch of the Ghana Water Company Limited for freely given me the data needed for this work.

DEDICATION

To my wife, Mabel and children, James and Moses.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AR – Autoregressive

ARMA – Autoregressive Moving Average

ARIMA – Autoregressive Integrated Moving Average

GSS – Ghana Statistical Service

KPSS – Kwiatkowski-Phillips-Schmidt-Shin test

MA – Moving Average

SARIMA – Seasonal Autoregressive Integrated Moving Average

UN – United Nations

WHO – World Health Organisation

CHAPTER 1

INTRODUCTION

1.0 Overview

This chapter presents the background of the study, the area profile, the problem statement and objectives of the study. It is also include the methodology adopted, the justification of the study, the scope and limitation and the theories organisation.

1.1 Background of Study

Water is required by all living creatures for survival. It is also required for economic growth and development (Hagan, 2007). According to the UN World Water Development Report (2006), "water is an essential life sustaining element. It pervades our lives and is deeply embedded in our cultural backgrounds." The achievement of the millennium development goals depend largely on improved water supply and sanitation in the developing countries. The World Health Organisation (1997) recommended that 75 litres of water a day is necessary to protect against household diseases and 50 litres a day necessary for basic family sanitation. The international consumption figures released by the 4th World Water Forum (2006), indicate that a person living in an urban area, uses an average of 250 litres/day; but individual consumption varies widely around the globe (THD, 2007). WHO/UNICEF Joint Monitoring Programme (2012) on water supply and sanitation estimated that 783 million people all over the world do not use an improved drinking-water. The number of people who rely on the earth's limited freshwater reserves is increasing everyday with Arms (2008) stating that a scarcity of clean, fresh water is one of the world's most pressing environmental problems.

Also, at the 2002 World Summit on Sustainable Development in Johannesburg, South Africa, great concern was expressed about the 1.1 billion people in the world who do not have access to safe drinking water and the 2.4 billion who live without proper sanitation (Cech, 2005). The resulting human toll is roughly 3.3 billion cases of illness and 2 million deaths per year. Moreover, even as the world's population grows, the limited easily accessible freshwater resources in rivers, lakes and shallow groundwater aquifers are dwindling as a result of over-exploitation and water quality degradation (IAEA, 2004). The UN predicts that by 2025, two-thirds of the world population will experience water scarcities, with severe lack of water blighting the lives and livelihoods of 1.8 billion. According to the UN World Water Assessment Programme, by 2050, 7 billion people in 60 countries may have to cope with water scarcity (Chenoweth, 2008). People in many parts of the world today are faced with the problem of water paucity and insecurity (Udoh & Etim, 2007). The most obvious concern about an unsafe water supply is the health risk to family or guests. Wastewater contamination serves as a source of bacteria, viruses, and parasites that can cause gastrointestinal problems or transmit contagious diseases (Arms, 2008; UMES, 2008).

In a study conducted by WaterAid (2013) in five African countries including Ghana, Rwanda, Niger, Sierra Leone and Uganda, it was revealed the political commitment of governments towards providing portable and accessible water and sanitation was low. WaterAid, therefore, concluded that unless investment is increased, the challenges of urbanisation, inequality of access, climate change and population growth risk turning back the clock even further, and urged the various African leaders to deliver on their past water and sanitation commitments.

The above finding on Ghana is corroborated by the recent water crisis in some parts of the country. WaterAid-Ghana (2013) further estimates that 7 million Ghanaians lack access to clean water. The value of water is determined by two elements, *supply* – the cost of providing the resource in a certain quality, quantity and location which varies in different parts of the country and *demand* – the utility to humans and their willingness to pay for that utility (Cech, 2005).

Water treatment cost is determined by factors such as: electricity cost, chemical cost, availability of raw water, capability of raw water pumps and treated water pumps, number of filters and their efficiency, number of clarifiers and their efficiency, availability and quality of pipelines and labour. Some of the aforementioned factors are fixed whilst others vary with time.

Water supply can be increased either by making more effective use of existing supply capacity or by adding additional supply capacity. To the consumer, the additional capacity supplied will either displace and/or add to already existing water sources. Every person uses water for drinking, cooking, bathing, washing of clothes, for sanitation purposes, etc. Sources of water include piped water supply systems, dug wells, hand pumps, canals, ponds, rivers, bottled water, water from vendors and rainwater among others. If the additional supply of water is used to displace already existing sources, it is called non-incremental demand. For example, a household which obtains a new connection to the piped water supply system may no longer make use of the existing dug well. If the additional supply of water generates an increase in existing consumption, it is called incremental demand. For example, a household obtaining its water from a well at a distance of 300 meters may increase its water consumption from 450 liters to 650 liters per day after a public tap is installed in closer proximity to the house.

1.1.1 Study area profile

The Hohoe Municipality is one of the eighteen districts in the Volta Region of Ghana with Hohoe as its capital and administrative centre. According to the GSS (2010), it has a population of 262,046 made up of 126,239 and 135,807 males and females respectively. The population size has grown by almost 81% from year 2000. In terms of water supply in the municipality, the Ghana Water Company Limited and DANIDA have been effective in resolving many problems identified in 1992 where more than 60% of the population lack good drinking water and sanitation facilities. Very old machines, broken down hand pumps and other equipment have either been replaced or repaired for efficient water production and distribution in the municipality.

Majority of the people of the Municipality (about 65%) are engaged in agricultural production. The technology employed in agricultural production in the municipality is largely the traditional cutlass and hoe. Mechanised farming is very limited and the rate of adoption of other agricultural-related technologies is equally low. Farming is entirely rain-fed as there are no irrigation facilities, and this culminates in low productivity. Access roads to farming centres are also poor thus hampering the marketing of the products. These together with the absence of storage facilities give rise to high post harvest losses.

1.2 Problem statement

Water is considered as lifeline of all living things especially humans, hence its availability or otherwise is a critical component in the measurement of human wellbeing through the Human Development Index (HDI). Although successive governments of Ghana have tried to provide water to her citizenries by investing heavily in the water sector, not much have been achieved due to improper planning, dysfunction organisation, corruption, inadequate capital injection, absolute equipment, and intermittent power supply among others.

Although water demand in the Hohoe Municipality is high, due to the above challenges, the municipality is always in water crisis. It is estimated that the optimal water demand of the municipality is 3,000,000m^3 per day but the available data shows that only between 11,000 and 15,000 m^3 of water is been produced fortnightly. For the challenges of insufficient and erratic water situation in the municipality to be surmounted, all aforementioned challenges must be resolved including the development of an adequate time series model for forecasting water consumption of the residents and businesses in the municipality. This is what this study seeks to determine for the Hohoe Branch of the Ghana Water Company Limited.

1.3 Objectives of the study

The objectives of the study include:

1. to identify the best fit times series model to the water consumption data in the Hohoe Municipality,
2. to forecast water consumption in the municipality for the next 4 years on a fortnight basis.

1.4 Methodology

In modeling and forecasting water consumption in the Hohoe Municipality, a time series data is required. Therefore, secondary data will be obtained from the Ghana Water Company Limited, Hohoe from 2009 to 2012 on fortnight bases. The statistical tools include tables, graphs, Moving Average (MA), Autoregressive (AR), ARIMA and seasonal Autoregressive Integrated Moving Average (SARIMA) model. A time series is a set of statistics, usually collected at regular intervals. Time series data occur naturally in many application areas like the following:

1. Economics - e.g. monthly data for unemployment, hospital admissions, etc.
2. Finance - e.g. daily exchange rate, a share price, etc.
3. Environmental - e.g. daily rainfall, air quality readings.
4. Medicine - e.g. ECG brain wave activity every 2-8 second.

The methods of time series analysis pre-date those for general stochastic processes and Markov Chains. The aims of time series analysis are to describe and summarise time series data, fit low-dimensional models, and make forecasts. We write our real-valued series of observations as $\ldots X_{-2}, X_{-1}, X_0, X_1, X_2, \ldots$ a doubly infinite sequence of real-valued random variables indexed by Z. In gathering literature on the subject, the researcher consults the University of Cape Coast's Library, GWCL Annual Reports (2009 – 2012), and the internet. The R, SPSS (version 17.0), and MINITAB (version 14.0) were used in the analysis.

1.5 Justification

Water is a limited resource, the demands for which are fast increasing. Populations in some towns in the Hohoe Municiaplity, like Hohoe, Likpe Kukurantumi, Ve Kolloenu, Akpafu Odomi, Logba Alapati, Wli Afegame, Lolobi Kumasi, Agate, Have Etoe, etc., for instance, are expanding at a fast rate. The result is that water managers must struggle to keep taps flowing without compromising water supplies for future generations. The research findings would therefore be crucial to the Ghana Water Company Limited (GWCL), Hohoe, the Hohoe Municipal Assembly and its residence, and the academia. The GWCL would depend on the model and forecasted values to guide its operations. Also, researchers in the academia would use the study as literature in other related areas.

1.7 Scope and Limitation

The thesis is restricted to the objectives of the research. Research work is characterised by some constraints. Some of these setbacks include time constraints and the difficulties in obtaining relevant materials on the topic.

1.7 Thesis organisation

Chapter 1 is made up of introduction, which comprises the background of the study, study areas, problem statement and objective of the study. It also presents the justification and limitations of the study. Chapter 2 highlights related literature on the topic with ideas of different authors whose findings have been defined in relation to the topic under study.

Chapter 3 focuses on methodological review in the light of mathematical and statistical tools that are relevant to the analyses of the data. Basically, the study seeks to use time series model for the analyses. Chapter 4 deals with the data collection and analysis, and the findings from the application of the various time series models. In the same way, chapter 5 consists of summary, conclusion and recommendations.

1.8 Chapter summary

The chapter gave an introduction to the thesis report highlighting on issues relating to background of the study, problem statement and objectives guiding the study, methodology and justification of the study. In addition to these are limitations as well as thesis organisation. The chapter concludes with this summary.

CHAPTER 2

LITERATURE REVIEW

2.0 Introduction

In this section, there is a review of the work of several authors concerning concept definitions and various researches done. Research works, empirical work and authors' opinion are looked at. Below are the focuses of the review.

2.1 Urban Water Management

2.1.1 Historical background

Many cities throughout the world are growing rapidly and are facing water shortage. Urbanisation has result to the increasing of water demand whereas not every country able to provide the water supply for their municipalities. WHO data on water supply coverage 2006 shows that only 86% of urban population has an access to improved drinking water (WHO/UNICEF, 2006).

The development of urban area resulted in the change of physical properties of land surface, due to the increasing paved surface area. And it has brought a change to hydrological cycle, such as pattern of surface water, lost functions of rainwater storage and infiltration, and accelerated transport of pollutant from urban areas. As a consequences, an essential approach with respect to centralized waste water system also take place, whereas the detention, retention and recharge approach during 1970 has shifted to the source of pollution control during 1980s and 1990s (Niemczynowicz, 1999). At that time, water authorities in many countries found a positive correlation between poor sanitation and high mortality, and prompt to manage water supply, storm water drainage and wastewater disposal separately.

Since then, a variety of storm water handling and treatment methods have been developed, whereas it is understood that storm water should be handled locally (Anderson, 1996;

Mitchell, et al., 2003; Niemczynowicz, 1999; Zaizen, et al., 2000). Evaluation to conventional water supply and disposal methods, such as the reuse of storm water and waste water has been taken. These alternatives appear to offer many benefits. For example, less water is imported into towns and cities, and less storm water and wastewater are discharged. Thus, storm water and wastewater are being re-evaluated as resources to be utilised, rather than as waste products for disposal. The wisdom of importing large volumes of high quality water into urban areas, and exporting even larger quantities of storm water and wastewater out of them, is now being reviewed. Consequently, approach to urban water management should be done in a more holistic.

2.1.2 Integrated Urban Water Management

As the increasing demand of water sources in urban areas, the integrated urban water management promotes the consideration of water supply, storm water and waste water as components of total urban water cycle. By quantifying the water balance of the component urban water cycle, various alternatives of both water inputs into and water outputs from urban catchments will be taking into account in order to making more use of the total water resources available in urban area (Mitchell, et al., 2001). Within this new paradigm, decentralized approach on sewage treatment was introduced, where wastewater collected from individual or/and clusters homes are treated at or near the point of gener2ation, and maximise its potential reuse opportunities (Maher & Lustig, 2003).

The principle of these components is to minimise the consumption of water supply by recycling the storm water and waste water, and consequently will reduce the discharged water to be treated. Some ways that can be implemented are through retention of roof rainwater, onsite treatment of grey water and black water (toilet), storm water detention.

Harvesting rainwater and reuse waste water would not only protect the valuable natural water resources, take the pressure off some ground and river water extraction, but also reduce the energy used for treating and conveying water. Water recycling measures trigger a chain reaction of savings, as when using water efficiently means less amount water discharges and less energy and chemicals are required for water and sewage treatment and distribution (Cheng, 2002).

The ultimate objective of integrated water management is to achieve the sustainable, coordinated management of water resources within a region, with the objectives of controlling and conserving water, minimising adverse affects, and achieving specified water management and social objectives (Marsalek, et al., 2001).

The adoption of Integrated Urban Water Management is introduced as an approach to the urban water cycle. This system includes the integration of storm water, groundwater and surface water use, re-use of treated wastewater, and recycling. It emphasizes the developments of technologies that reduces the use of treated water for sanitation and take advantage of rainwater as a resource that will lead to a fuller recycling and reuse of urban water (UNESCO, 2001). Since urban water services includes the drinking water supply, treatment, reuse and disposal of water and storm water management, the interaction of urban water-cycle system and their boundaries (physical, temporal and spatial) need to be well defined (Hardy et al., 2005).

2.2 Revamping Ghana's Water Sector: Lessons from South Africa

According to Lamiley (2012), even with scarce water resources and an uneven rainfall pattern, South Africa provides 91% of its 50 million citizens with improved drinking water and 77% with access to improved sanitation. The Department of Water Affairs (DWA), the national body that formulates and implements policies regarding water in South Africa, described the country as water scarce, as fresh water resources availability are almost equivalent to the demand for them. The country, therefore, prioritises water resources management and insists on reconciling demand with availability at all times.

Ghana on the other hand has abundant water resources but with only 63 per cent of the population having access to improved drinking water and 13 per cent with access to improved sanitation. A team of professionals drawn from various sectors but with a direct impact on water issues embarked on a one-week study tour of South Africa to, among others, learn and share best practices with their counterparts in the water sector. More specifically, the team was interested in learning about urban and rural water supply management systems, and environmental sanitation management approaches, integrated water resources management and programme-based budgeting.

The tour, funded by the European Union through the Improvement of Water Sector Performance Management Framework (IWSPMF) project, according to the Director of Water, came at a time when Ghana was at the threshold of revamping the water sector to improve water resource management, production and delivery, and consequently accessibility to the larger population that are not yet served.

2.2.1 Evolution of Ghana's Water Management System

The first water supply system in Ghana, then Gold Coast, was established in Accra just before World War I. Extensions were made exclusively to other urban areas among them the colonial capital of Cape Coast, Winneba and Kumasi in the 1920s. However, there have been challenges hampering Ghana's effective water resource management, production and delivery since independence.

2.2.2 Institutional capacity

Lamiley (2012) posited that like the DWA in South Africa, Ghana has the Water Directorate that is responsible for the formulation and implementation of policies and strategies for the water sector. The difference, however, is that the DWA directly supervises implementation of programmes, whereas Ghana's Water Directorate rather coordinates activities of several agencies that do the implementation. Unlike the DWA of South Africa, the Water Directorate is currently hugely under-resourced to play any effective role. The DWA with a workforce of 8,000 is one of the largest public institutions in South Africa, directly coordinating all water management, supply and research activities and appears to be in full control of policy and implementation of both water resources management and water supply issues.

Ghana's Water Directorate compared with the DWA is not strong enough in terms of staffing (only 15 workers over 90 per cent of who are short-term project staff and those on secondment from other agencies) and structure to take full control although Ghana on the other hand has a very strong institutional arrangement with agencies such as the Community Water and Sanitation Agency (CWSA), the Water Resources Commission (WRC) and the Ghana Water Company Limited (GWCL) playing specialised roles and with the PURC playing a regulatory role.

2.2.3 Taking Financial Control of the Water Sector

According to Lamiley (2012), as delegates listened keenly to the various presentations, one issue that clearly came up was the fact that the Central Government funds the entire water and sanitation sector in South Africa upon advice from the DWA with little or no financial support from international donors. According to officials, about 60 per cent of the Municipal Infrastructure Grant is meant for water and sanitation interventions. Most of their water supply agencies (the Water Boards), though government owned, are self-reliant and they hardly make losses. For this reason, the departments are able to draw clear plans and to implement the plans. Sanitation for instance is supply-driven and, therefore, little time is spent discussing models and concepts.

The opposite is, however, the case for Ghana, which currently depends heavily on foreign donors in pushing the agenda of the water sector.

Again, whereas South Africa has one major Water Act for both water and sanitation, which provides direction to all institutions in the WASH sector, Ghana has fragmented sub-sector specific acts, making coordination relatively more difficult.

She reported that in South Africa, all municipalities and the various Water Boards report directly to the DWA. In Ghana, though the agencies report to the MWRWH, local government institutions report to the Ministry of Local Government and Rural Development under the Local Government Act. Coordination would be easier if all WASH agencies and MMDAs were to report to one central institution (e.g. the Water Directorate).

Comparatively, South Africa has stronger private sector participation and stronger monitoring and enforcement of water use regulations. There is a monitoring task force (the Blue Scorpions) in place and are made up of staff from the DWA. They also have the Water Resources Tribunals to trial water use-related offences. These enforcement bodies are currently absent in Ghana. South Africa again has stronger data management systems in place as there is a functional Water Research Commission providing data collection and data management services. Waste water is usually treated and pumped back into catchments.

2.2.4 Water Supply

Lamiley (2012) indicated that in South Africa, responsibility for the provision of water and sanitation services lies with the districts and municipalities though they all report directly to the DWA. There are Water Boards that supply bulk water to municipalities, who in turn distribute to households and collect the bills. This means that for both rural and urban water supply, municipalities play a direct role by buying water in bulk from the water supply companies (called Water Boards) and resell to households. There are no private water supply companies; all water boards belong to the public. Water tariffs vary between municipalities and even between localities depending on ability to pay and a programme of cross-subsidisation, decided upon at the municipal level. As a policy, every household receives the first six cubic metres of water per month for free (even though 75 per cent of the population, according to a research, can afford to pay). The policy was introduced gradually since 2000 within the means of each municipality. In 2010 the programme reached 86% of all households. South Africa also works a lot with the Private Sector in an effort to reduce Non-Revenue Water. A number of experts from the private sector are, therefore, engaged on contract basis to manage water leakage and illegal connections. Some of the major strategies applied are pressure control and routine physical checks in the field. To make this programme effective, metering is considered imperative.

Ghana can significantly improve its water delivery if some of these key lessons in South Africa could be considered. The 3rd Ghana Water Forum for instance recommended the breaking of the monopoly of the Ghana Water Company and promotion of either bulk supply to MMDAs or different companies providing specialised services in urban water supply. Ghana may also consider beefing up her pro-poor policies through cross-subsidisation or increased amount of free water to all or to the poor, as well as consider seeking private sector support in reducing the Non-Revenue Water, which was more 49% as of 2010.

2.2.5 Research and Knowledge Management in the Water Sector

According to Lamiley (2012), South Africa has a strong research and training infrastructure in the water sector. The Water Research Commission (WRC) supports water research and development, as well as the building of a sustainable water research capacity in South Africa. It serves as the country's water-centred knowledge ‘hub’ leading the creation, dissemination and application of water-centred knowledge, focusing on water resources management, water-linked ecosystems, water use and waste management and water utilisation in agriculture.

The WRC is funded by the government through levies on water bills and the Commission in turn provides grants to researchers based on the relevance of their research proposals to the country. The WRC then assembles all research reports and generates necessary information for the government and all stakeholders. The WRC has a unit that engages in dissemination of available research findings. By way of capacity development and provision of more researchers for the water sector, researchers are usually mandated to work with students through attachments among others so that they develop an interest and gain research experience in the sector.

Research is not harmonised in Ghana’s WASH sector as it is usually conducted by those who need information. Though there is a Water Research Institute (WRI) in Ghana there is currently too little interface between it and the WASH Sector. This is partly because the WRI is under the Ministry of Environment, Science and Technology and does not report to the MWRWH or the MLGRD. It would be very appropriate if a similar unit is created at the Water Directorate to harmonise research in the WASH sector.

CHAPTER 3

METHODOLOGY

3.0 Introduction

This chapter examines thoroughly the basic definitions and concepts of time series analysis, assumptions, conditions, principles and processes involved in the application of moving averages (MA), autoregressive (AR), autoregressive moving averages (ARMA), and autoregressive integrated moving average (ARIMA).

3.1 Basic Concepts and Definitions of Time Series

3.1.1 Time Series Analysis

Time series analysis comprises methods or processes that breakdown a series into components and explainable portions that allows trends to be identified, estimates and forecasts to be made. Basically time series analysis attempts to understand the underlying context of the data points through the use of a model to forecast future values based on known past values. Such time series models include MA, AR, ARIMA, GARCH, TARCH, EGARCH, FIGARCH, CGARCH and ARIMA among others but the main focus of this study is based on MA, AR, ARMA, and ARIMA models.

3.1.2 Lag

Lag is the time periods between two observations. For example, lag 1 is between Y_t and Y_{t-1}. Lag 2 is between Y_t and Y_{t-2}. Time series can also be lagged forward, Y_t and Y_{t+1}. The observation at the current time, Y_t , depends on the value of the previous observation, Y_{t-1}.

3.1.3 Differencing

Differencing simply means subtracting the value of an earlier observation from the value of a later observation. Calculating differences among pairs of observations at some lag to make a non stationary series stationary. There are possible shifts in both the mean and the dispersion over time for this series. The mean may be edging upwards, and the variability may be increasing. If the mean is changing, the trend is removed by differencing once or twice. If the variability is changing, the process may be made stationary by logarithmic transformation. Differencing the scores is the easiest way to make a non stationary mean stationary (flat). The number of times you have to difference the scores to make the process stationary determines the value of d. If dC0, the model is already stationary and has no trend. When the series is differenced once, dC1 and linear trend is removed. When the difference is then differenced, dC2 and both linear and quadratic trend are removed. For non-stationary series, d values of 1 or 2 are usually adequate to make the mean stationary.

3.1.4 Stationary and Non-stationary Series

Stationary series vary around a constant mean level, neither decreasing nor increasing systematically over time, with constant variance. Non-stationary series have systematic trends, such as linear, quadratic, and so on. A non stationary series that can be made stationary by differencing is called “non stationary in the homogenous sense.”

Stationarity is used as a tool in time series analysis, where the raw data are often transformed to become stationary. For example, economic data are often seasonal or dependant on a non-stationary price level. Using non-stationary time series produces unreliable and spurious results and leads to poor understanding and forecasting. The solution to the problem is to transform the time series data so that it becomes stationary. If the non-stationary process is a random walk with or without a drift, it is transformed to stationary process by differencing.

Differencing the scores is the easiest way to make a nonstationary mean stationary (flat). The number of times you have to difference the scores to make the process stationary determines the value of *d*. If *d*=0, the model is already stationary and has no trend. When the series is differenced once, *d*=1 and linear trend is removed. When the difference is then differenced, *d*=2 and both linear and quadratic trend are removed. For nonstationary series, *d* values of 1 or 2 are usually adequate to make the mean stationary. If the time series data analysed exhibits a deterministic trend, the spurious results can be avoided by detrending. Sometimes the non-stationary series may combine a stochastic and deterministic trend at the same time and to avoid obtaining misleading results both differencing and detrending should be applied, as differencing will remove the trend in the variance and detrending will remove the deterministic trend.

A non-stationary process with a deterministic trend becomes stationary after removing the trend, or detrending. For example, $Y_t = \alpha + \beta_t + \varepsilon_t$ is transformed into a stationary process by subtracting the trend β_t: $Yt - \beta_t = \alpha + \varepsilon_t$. No observation is lost when detrending is used to transform a non-stationary process to a stationary one. Non-stationary data, as a rule, are unpredictable and cannot be modeled or forecasted. The results obtained by using non-stationary time series may be spurious in that they may indicate a relationship between two variables where one does not exist. In order to receive consistent, reliable results, the non-stationary data needs to be transformed into stationary data. In contrast to the non-stationary process that has a variable variance and a mean that does not remain near, or returns to a long-run mean over time, the stationary process reverts around a constant long-term mean and has a constant variance independent of time.

3.2 Components of Time Series

A vital step in choosing appropriate modeling and forecasting procedure is to consider the type of data patterns exhibited from the time series graphs of the time plots. The sources of variation in terms of patterns in time series data are mostly classified into four main components:

(i) Horizontal – when data values fluctuate around constant value

(ii) Trend – when there is long term increase or decrease in the data

(iii) Seasonal – when a series is influenced by seasonal factors and recurs on a regular periodic basis.

(iv) Cyclic – when the data exhibit rises and falls that are not of a fixed period.

3.2.1 The Trend (d)

The trend is simply the underlying long term behavior or pattern of the data or series. The Australian Bureau of Statistics (ABS, 2008) defined trend as the 'long term' movement in a time series without calendar related and irregular effects, and is a reflection of the underlying level. It is the result of influences such as population growth, price inflation and general economic changes. A model with two trend terms (dC2) has to be differenced twice to make it stationary. The first difference removes linear trend, the second difference removes quadratic trend, and so on.

3.2.2 Seasonal Variation (S)

A seasonal effect is a systematic and calendar related effect. Some examples include the sharp escalation in most Retail series which occurs around December in response to the Christmas period, or an increase in water consumption in summer due to warmer weather. Other seasonal effects include trading day effects (the number of working or trading days in a given month differs from year to year which will impact upon the level of activity in that month) and moving holidays (the timing of holidays such as Easter varies, so the effects of the holiday will be experienced in different periods each year). Seasonal adjustment is the process of estimating and then removing from a time series influences that are systematic and calendar related. Observed data needs to be seasonally adjusted as seasonal effects can conceal both the true underlying movement in the series, as well as certain non-seasonal characteristics which may be of interest to analysts. Seasonality in a time series can be identified by regularly spaced peaks and troughs which have a consistent direction. Other techniques that can be used in time series analysis to detect seasonality include:

1. A seasonal subseries plot is a specialised technique for showing seasonality.
2. Multiple box plots can be used as an alternative to the seasonal subseries plot to detect seasonality.
3. The autocorrelation plot can help identify seasonality.

3.2.3 Cyclical Variations (C)

Cyclical variations are the short term fluctuations (rises and falls) that exist in the data that are not of a fixed period. They are usually due to unexpected or unpredictable events such as those associated with the business cycle sharp rise in inflation or stock price, etc. The main difference between the seasonal and cyclical variation is the fact that the former is of a constant length and recurs at regular intervals, while the latter varies in length. More so, the length of a cycle is averagely longer than that of seasonality with the magnitude of a cycle usually being more variable than that of seasonal variation.

3.2.4 Irregular Variations (I)

The irregular component (sometimes also known as the residual) is what remains after the seasonal and trend components of a time series have been estimated and removed. It results from short term fluctuations in the series which are neither systematic nor predictable. In a highly irregular series, these fluctuations can dominate movements, which will mask the trend and seasonality.

3.3 Common Assumptions in Time Series Techniques

A common assumption in many time series techniques is that the data are stationary. A stationary process has the property that the mean, variance and autocorrelation structure do not change over time. Stationarity can be defined in precise mathematical terms as

1. The mean $\mu(t)= E(\gamma(t))$
2. The variance $\sigma^2(t) = Var(y(t)) = \gamma(0)$
3. The autocovariance $\gamma(t_1, t_2) = cov(y(t_1), y(t_2))$

Hence a time series is said to be strictly stationary if the joint distribution of any set of n observations $(t_1, t_2)= cov(y(t_1),y(t_2))$ is the same as the joint distribution of $y(t_1), y(t_2), y(t_3) \dots y(t_n)$ for all n and k. If the time series is not stationary, we can often transform it to stationary with one of the following techniques:

1. We can difference the data. That is, given the series Z_t, the differenced data will contain one less point than the original data. Although you can difference the data more than once, one difference is usually sufficient.
2. If the data contain a trend, we can fit some type of curve to the data and then model the residuals from that fit. Since the purpose of the fit is to simply remove long term trend, a simple fit, such as a straight line, is typically used.

3. For non-constant variance, taking the logarithm or square root of the series may stabilise the variance. For negative data, you can add a suitable constant to make the entire data positive before applying the transformation. This constant can then be subtracted from the model to obtain predicted (i.e., the fitted) values and forecasts for future points.

3.4 Autocorrelation Function (ACF)

Autocorrelation refers to the correlation of a time series with its own past and future values. Autocorrelation is also sometimes called "lagged correlation" or "serial correlation", which refers to the correlation between members of a series of numbers arranged in time. The pattern of autocorrelations in a time series at numerous lags; the correlation at lag 1, then the correlation at lag 2, and so on. Correlations among sequential scores at different lags. The lag 1 autocorrelation coefficient is similar to correlation between the pairs of scores at adjacent points in time, rY_t, Y_{t-1} (e.g. the pair at time 1 and time 2, the pair at time 2 and time 3, and so on). The lag 2 autocorrelation coefficient is similar to correlation between the pairs of scores two time periods apart, rY_t, Y_{t-2} (e.g., the pair at time 1 and time 3, the pair at time 2 and time 4, and so on). Positive autocorrelation might be considered a specific form of "persistence", a tendency for a system to remain in the same state from one observation to the next. Three tools for assessing the autocorrelation of a time series are:

1. The time series plot,
2. The lagged scatter plot, and
3. The autocorrelation function.

An important guide to the persistence in a time series is given by the series of quantities called the sample autocorrelation coefficients, which measure the correlation between observations at different times. The set of autocorrelation coefficients arranged as a function of separation in time is the sample autocorrelation function or the acf.

The first- order autocorrelation coefficient is the simple coefficient of the first N–1 observations, t=1, 2,…, N-1 and the next N - 1 observations, X_t, t=2,3,…, N. The correlation between X_t and X_{t+1} is given by,

$$r_1=\frac{\sum_{t=1}^{N-1}(x_t-\bar{X}_1)(x_{t+1}-\bar{X}_2)}{[\sum_{t=1}^{N-1}(x_t-\bar{X}_1)^2][\sum_{t=1}^{N-1}(x_t-\bar{X}_1)^2]} \qquad \text{...(3.1)}$$

Where $\bar{X}_1$ is the mean of the first N – 1 observations and is the mean of the last N-1 observation. As the correlation coefficient given above measure correlation between successive observations it is called the autocorrelation coefficient or serial correlation coefficient. For N reasonably large, the difference between the sub-period means $\bar{X}_1$ and $\bar{X}_2$ can be ignored and r_1 can be approximated as by

$$r_1=\frac{\sum_{t=1}^{N-1}(x_t-\bar{X})(x_{t+1}-\bar{X})}{\sum_{t=1}^{N}(x_t-\bar{X})^2} \qquad \text{.....................................(3.2)}$$

Equation (3.2) can be generalised to give the correlation between observations separated by k years:

$$r_k=\frac{\sum_{t=1}^{N-k}(X_t-\bar{X})(X_{t+k}-\bar{X})}{\sum_{t=1}^{N}(X_t-\bar{X})^2} \qquad \text{... ... ... (3.3)}$$

The quantity is called the autocorrelation coefficient at lag k. The plot of the autocorrelation function as a function of lag is also called the correlogram. The autocorrelation function can be used for the following two purposes:

- To detect non-randomness in data.
- To identify an appropriate time series model if the data are not random.

Autocorrelation plots are formed by:

- Vertical axes: autocorrelation coefficient.
- Horizontal axes: Time lag k= 1,2,3………….
- Confidence band

The confidence band uses the following formula if the autocorrelation plot is used to check for randomness in the data.

$$\frac{\pm Z_{1-\frac{\alpha}{2}}}{\sqrt{N}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots..(3.4)$$

Where N is the sample size, z is the percent point function of the standard normal distribution and α is the significance level. If autocorrelation plots are also used in the model identification stage for fitting ARIMA models, the confidence band uses the following formula:

$$\pm Z_{1-\frac{\alpha}{2}}\sqrt{\left(\frac{1}{N}(1+2\sum_{i=1}^{k} y_1^{2})\right)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(3.5)$$

Where k is the lag, N is the sample size; z is the percent point function of the standard normal distribution and α is the significance level.

3.5 Partial Autocorrelation Function

Partial autocorrelation function measures the degree of association between Y_t and Y_{t+k} when the effect of other time lags on Y are held constant. The partial autocorrelation function

PACF denoted by the set of partial autocorrelations at various lags *k* are defined by (r_{kk}, k=1, 2, 3…). The set of partial autocorrelations at various lags *k* are defined by

$$r_{kk} = \frac{r_k - \sum_{j=1}^{k-1} r_{k-1,j}\, r_{k-1}}{1 - \sum_{j=1}^{k-1} r_{k-1,j}\, r_j} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (3.6)$$

where, $r_{k,j} = r_{k-1,j} - r_{kk}\, r_{k-1,k-1}$ $\qquad j=1,2\ldots\ldots\ldots\ldots\ldots\ldots k-1$

Specifically, partial autocorrelations are useful in identifying the order of an autoregressive model. The partial autocorrelation of an AR (*p*) process is zero at lag *p*+1 and greater. The approximate 95% confidence interval for the partial autocorrelations is at $+2/_{N}$. Partial autocorrelation plots are formed by:

- Vertical axes: partial autocorrelation at coefficient at lag, *k*,
- Horizontal axes: time lag *k* (*k* = 0, 1, 2...).

In addition, 95% confidence interval bands are typically included on the plot.

3.6 Common Approaches to Univariate Time Series

There are a number of approaches to modeling time series. Few of the most common approaches are below:

3.6.1 Decomposition

One approach is to decompose the time series into a trend, seasonal, and residual component. In other words decomposition refers to separating a time series into trend, cyclical, and irregular effects. Decomposition may be linked to de-trending and de-seasonalising data so as to leave only irregular effects, which are the main focus of time series analysis. Triple exponential smoothing is an example of this approach. Another example, called seasonal loess, is based on locally weighted least squares and is discussed by Cleveland (1993).

3.6.2 Autoregressive (AR) Models

An autoregressive model is simply a linear regression of the current value of the series against one or more prior values of the series. The value of *p* is called the order of the AR model. AR models can be analyzed with one of various methods, including standard linear least squares techniques. They also have a straightforward interpretation. A common approach for modeling univariate time series is the autoregressive (AR) model:

$$X_t=\delta+\emptyset_1X_{t-1}+\emptyset_2X_{t-2}+\ldots\ldots\emptyset_pX_{t-p}+A_t \ldots\ldots (3.7)$$

where X_t is the time series, A_t is white noise, and

with μ denoting the process mean. An autoregressive model of order *p*, denoted by AR(p) with mean zero is generally given by the equation:

$$X_t=\phi_1X_{(t-1)}+\phi_2X_{(t-2)}+\phi_3X_{(t-3)}+\ldots\ldots+\phi_pX_{(t-p)}+\varepsilon_t \ldots\ldots (3.8)$$

Or

$$X_t=(\phi_1L+\phi_2L^2+\phi_3L^3+\ldots\ldots+\phi^PL^P)X_t+\varepsilon_t \ldots\ldots (3.9)$$

Where $\phi(L)=\varepsilon_t$

$$\phi(L)=(1-\phi_1L+\phi_2L^2+\phi_3L^3+\ldots\ldots+\phi^PL^P)\ldots\ldots (3.10)$$

Where L, is the lag operator

$\phi_1, \phi_2, \phi_3 \ldots\ldots \phi_p(\phi_{p\neq}0)$ are the autoregressive model parameters and ε_t is the random shock or white noise process, with mean zero and variance σ_ε^2. The mean of X_t is zero. If the mean, μ of X_t is not zero, replace X_t by $X_t-\mu$. That is

$$X_t-\mu=\phi_1(X_{t-1}+\mu)+\phi_2(X_{t-2}+\mu)+\ldots+\phi_p(X_{t-p}+\mu)+\varepsilon_t \ldots\ldots (3.11)$$

Or write

$$X_t=\alpha+\phi_1X_{t-1}+\phi_2X_{t-2}+\ldots\ldots\phi_PX_{t-p}+\varepsilon_t \ldots\ldots (3.12)$$

Where $\alpha=\mu(1-\phi_1-\phi_3 \ldots\ldots, \phi_p)$

In this general case, the ACF damps down and the PACF cuts off after p lags. An AR (p) model is stationary if the roots of $\phi(L) = 0$ all lie outside the unit circle. A necessary condition for stationary is that $r_k = 0$ as $k \longrightarrow \infty$.

3.6.3 Moving Average (MA) Models

Moving average model is conceptually a linear regression of the current value of the series against the white noise or random shocks of one or more prior values of the series. The random shocks at each point are assumed to come from the same distribution, typically a normal distribution, with location at zero and constant scale. The distinction in this model is that these random shocks are propagated to future values of the time series. Fitting the MA estimates is more complicated than with AR models because the error terms are not observable. This means that iterative non-linear fitting procedures need to be used in place of linear least squares. MA models also have a less obvious interpretation than AR models. Moving Average (MA) is another common approach for modeling univariate time series models is the moving average (MA) model:

$$X_t=\mu+A_t-\theta_t A_{t-1}-\theta_2 A_{t-2}-\ldots\ldots -\theta_q A_{t-q} \ldots\ldots \quad (3.13)$$

Where, X_t is the time series, μ is the mean of the series, A_{t-i} are white noise, and $\theta_1, \ldots, \theta_q$ are the parameters of the model. The value of q is called the order of the MA model.

A moving average model of order *q*, with mean zero, denoted by MA (*q*) is generally given by:

$$X_t = \theta_1\varepsilon_{(t-1)}+\theta_2\varepsilon_{(t-2)}+\theta_3\varepsilon_{(t-3)}+\ldots\ldots +\theta_q X_{(t-q)}+\varepsilon_t \ldots\ldots \quad (3.14)$$

Or

$$\theta(L)=(\theta_1 L+\theta_2 L^2+\theta_3 L^3+\ldots\ldots+ \theta_q L^q)X_t+\varepsilon_t \ldots\ldots \quad (3.15)$$

Where $X_t = \theta(L)\varepsilon_t$

$$\theta(L)=(\theta_1L+\theta_2L^2+\theta_3L^3+\ldots\ldots\ldots+\theta_qL^q)\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (3.16)$$

Where L, is the lag operator and ε_t is the random shock or white noise process and $\theta_1, \theta_2, \theta_3 \ldots \theta_q$ are the moving average model parameters. An MA (q) is said to be invertible if $\phi(L)$ can be inverted, in other words if it can be expressed as an AR. An MA (q) is invertible if the roots of $\phi(L)=0$ all lie outside the unit circle. A finite AR is always invertible. The random shocks at each point are assumed to come from the same distribution, typically a normal distribution, with location at zero and constant scale. The distinction in this model is that these random shocks are propagated to future values of the time series. Sometimes the ACF and PACF will suggest that a MA model would be a better model choice and sometimes both AR and MA terms should be used in the same model. It is also important to note, however, that the error terms after the model is fit should be independent and follow the standard assumptions for a univariate process. Box and Jenkins popularised an approach that combines the moving average and the autoregressive approaches (Box, Jenkins, and Reinsel, 1994). This resulted in autoregressive moving average model (ARMA). The Box-Jenkins model assumes that the time series is stationary. Box and Jenkins recommend differencing non-stationary series one or more times to achieve stationarity. Doing so produces an ARIMA model, with the "I" standing for "Integrated." This is described in detail below since it is the main method used in the analysis of data in this research.

3.6.4 Autoregressive Moving Average (ARMA) Models

Autoregressive and Moving Average processes can be combined to obtain a very flexible class of univariate processes (proposed by Box and Jenkins), known as ARMA processes. The time series X_t is an ARMA (p, q) process, if it is stationary and

$$X_t=\phi_1 X_{(t-1)}+\phi_2 X_{(t-2)}+\phi_3 X_{(t-3)}+\ldots+\phi_p X_{(t-p)}+\varepsilon_t+\theta_1\varepsilon_{(t-1)}+\theta_2\varepsilon_{(t-2)}+\ldots+\theta_q X_{(t-q)} \quad (3.14)$$

$$X_t=\phi_1 X_{t-1}+\theta(L)\varepsilon_t \quad (3.15)$$

$$\phi(L)X_t=\theta(L)\varepsilon_t \quad (3.16)$$

Where θ, ϕ, ε and L are as defined above with $\phi_p \neq 0$ and $\theta_q \neq 0$.

An ARMA process is stationary if the roots of $\phi(L)$ all lie outside the unit circle and invertible if the roots of $\theta(L)$ all lie outside the unit circle.

The acronym for an auto-regressive integrated moving average model. The three terms to be estimated in the model are auto-regressive (*p*), integrated (trend-*d*), and moving average (*q*).The ARIMA (auto-regressive, integrated, moving average) model of a time series is defined by three terms (*p*, *d*, *q*). Identification of a time series is the process of finding integer, usually very small (e.g., 0, 1, or 2), values of *p*, *d*, and *q* that model the patterns in the data. When the value is 0, the element is not needed in the model. The middle element, *d*, is investigated before *p* and *q*. The goal is to determine if the process is stationary and, if not, to make it stationary before determining the values of *p* and *q*. Recall that a stationary process has a constant mean and variance over the time period of the study.

3.7 Box-Jenkins ARIMA Process

In statistics and econometrics, and in particular in time series analysis, an autoregressive integrated moving average (ARIMA) model is a generalisation of an autoregressive moving average (ARMA) model. The Box-Jenkins methodology, named after the statisticians George Box and Gwilym Jenkins, applies ARIMA models to find the best fit of a time series to past values of this time series, in order to make forecasts. They are applied in some cases where data show evidence of non-stationarity, where an initial differencing step (corresponding to the “integrated” part of the model) can be applied to remove the non-stationarity. The model is generally referred to as an ARIMA (*p, d, q*) model where *p, d* and *q* are non-negative integers that refer to the order of the autoregressive, integrated, and moving average parts of the model respectively.

3.7.1 Box-Jenkins Modeling Approach

The Box-Jenkins model uses iterative three-stage modeling approach which is:

1. Model identification and model selection: making sure that the variables are stationary, identifying seasonality in the dependent series (seasonally differencing it if necessary), and using plots of the autocorrelation and partial autocorrelation functions of the dependent time series to decide which (if any) autoregressive or moving average component should be used in the model.
2. Parameter estimation using computation algorithms to arrive at coefficients which best fit the selected ARIMA model. The most common methods use maximum likelihood estimation or non-linear least-squares estimation.
3. Model checking by testing whether the estimated model conforms to the specifications of a stationary univariate process. In particular, the residuals should be independent of each other and constant in mean and variance over time (plotting the mean and variance of residuals over time and performing a Ljung-Box test or plotting autocorrelation and partial autocorrelation of the residuals are helpful to identify misspecification). If the estimation is inadequate, we have to return to step one and attempt to build a better model.

3.8 Box-Jenkins Model Identification

3.8.1 Stationarity and Seasonality

The first step in developing a Box–Jenkins model is to determine if the time series is stationary and if there is any significant seasonality that needs to be modeled.

3.8.2 Detecting Stationarity

Stationarity can be assessed from a run sequence plot. The run sequence plot should show constant location and scale. It can also be detected from an autocorrelation plot. Specifically, non-stationarity is often indicated by an autocorrelation plot with very slow decay. Finally, unit root tests provide a more formal approach to determining the degree of differencing such as Kwiatkowski-Phillips-Schmidt-Shin (KPSS) and Phillips-Perron Unit Root Tests are carried out employing the unit root testing procedures of Hamilton (1994). The KPSS test for the null hypothesis of a level stationary against an alternative of unit root together with the Philips-Peron test for the null hypothesis of a unit root against the alternative of a stationary series. The decision rule is that for the KPSS test if the p-value of its test statistic is greater than the critical value, say 0.05, then reject the null hypothesis of having a level stationary series and therefore conclude the alternate hypothesis that it has a unit root. The Philips-Peron Test, on the other hand, test for the null hypothesis of unit root against an alternative hypothesis of stationarity by rejecting the null hypothesis if its *p*-value is less than the critical value chosen.

3.8.3 Differencing to achieve Stationarity

Box and Jenkins recommend the differencing approach to achieve stationarity. However, fitting a curve and subtracting the fitted values from the original data can also be used in the context of Box-Jenkins models.

3.8.4 Seasonal Differencing

At the model identification stage, the goal is to detect seasonality, if it exists, and to identify the order for the seasonal autoregressive and seasonal moving average terms. For many series, the period is known and a single seasonality term is sufficient. For example, for monthly data one would typically include either a seasonal AR 12 term or a seasonal MA 12 term. For Box-Jenkins models, one does not explicitly remove seasonality before fitting the model. Instead, one includes the order of the seasonal terms in the model specification to the ARIMA estimation software. However, it may be helpful to apply a seasonal difference to the data and regenerate the autocorrelation and partial autocorrelation plots. This may help in the model identification of the non-seasonal component of the model. In some cases, the seasonal differencing may remove most or all of the seasonality effect.

3.8.5 Identifying *p* and *q*

Once stationarity and seasonality have been addressed, the next step is to identify the order (i.e. the *p* and *q*) of the autoregressive and moving average terms. These are determined by examining the values of the autocorrelations and the partial autocorrelations with their corresponding plots as explained below.

3.8.6 Autocorrelation and Partial Autocorrelation Plots

The primary tools for doing this are the autocorrelation plot and the partial autocorrelation plot. The sample autocorrelation plot and the sample partial autocorrelation plot are compared to the theoretical behaviour of these plots when the order is known.

3.9 Best Model Identification and Selection

The Ljung-Box statistic would be used to identify the best model. The Ljung-Box statistic, also called the modified Box-Pierce statistic, is a function of the accumulated sample autocorrelations, r_j, up to any specified time lag m. As a function of m, it is determined as

$$Q(m) = n(n+1)\sum_{j=1}^{m}\frac{r_j^2}{n-j}$$

where n = number of usable data points after any differencing operations.

The Ljung-Box test can be defined as follows:

> **H_0:** The data are independently distributed (i.e. the correlations in the population from which the sample is taken are 0, so that any observed correlations in the data result from randomness of the sampling process).
>
> **H_a:** The data are not independently distributed.

The choice of a plausible model depends on its p-value for the modified Box-Pierce if is well above 0.05, indicating "non-significance." In other words, the bigger the p-value, the better the model.

3.10 Some Applications of ARIMA Model in Real Life Situations

ARIMA modeling techniques have been applied in many fields of research. For example, Aidoo (2010) applied ARIMA model on the monthly inflationary rates in Ghana. He indicated that Ghana faces a macroeconomic problem of inflation for a long period of time. The problem in somehow slows the economic growth in this country. Using monthly inflation data from July 1991 to December 2009, we find that ARIMA $(1,1,1)(0,0,1)_{12}$ can represent the data behavior of inflation rate in Ghana well. Based on the selected model, we forecast seven (7) months inflation rates of Ghana outside the sample period (i.e. from January 2010 to July 2010). The observed inflation rate from January to April which was published by Ghana Statistical Service Department fall within the 95% confidence interval obtained from the designed model. The forecasted results show a decreasing pattern and a turning point of Ghana inflation in the month of July.

Again, Cui (2011) researched on the topic: “ARIMA Models for Bank Failures: Prediction and Comparison.” Cui said that the number of bank failures has increased dramatically over the last twenty-two years. A common notion in economics is that some banks can become “too big to fail.” Is this still a true statement? What is the relationship, if any, between bank sizes and bank failures? In this thesis, the proposed modeling techniques are applied to real bank failure data from the FDIC. In particular, quarterly data from 1989:Q1 to 2010:Q4 are used in the data analysis, which includes three major parts: 1) pairwise bank failure rate comparisons using the conditional test (Przyborowski & Wilenski, 1940), 2) development of the empirical recurrence rate (Ho, 2008) and the empirical recurrence rates ratio time series; and 3) the Autoregressive Integrated Moving Average (ARIMA) model selection, validation, and forecasting for the bank failures classified by the total assets.

3.10.1 ARIMA Modeling: Advantages and Disadvantages

According to Reis and Mandl (2003), one of the primary benefits of ARIMA models is their ability to correct for local trends in the data – what has happened on the previous day is incorporated into the forecast of what will happen today. This works well, for example, during a particularly severe flu season, where prolonged periods of high visit rates are adjusted to by the ARIMA model, thus preventing the alarm from being triggered every day throughout the flu season. However, if the ARIMA model “adjusts” to an actual outbreak instead of detecting it, a slowly spreading outbreak or attack might be missed because of this correction. This correction is most likely to affect detection of outbreaks occurring over several days, rather than those that occur suddenly. It is therefore also important to rely on the non-ARIMA or non-classical ARIMA models for outbreak detection.

ARMA models in Reis and Mandl (2003) and Reis et al. (2003) require large historic records of patient visits in order to begin surveillance. This is a substantial disadvantage. As can be seen from Moore et al. (2002), in some cases long historical data are not available and not necessary. Also, combining both historical and recent trends is quite realistic. Another disadvantage of ARMA is that the corresponding detector is not sensitive to the slow growth. According to Rizzo et al., (2005), outbreaks that evolve over a matter of days, for example, can often be detected with ARMA models that generate single-day predictions based on historical data. More gradually developing outbreaks are generally easier to detect by using such techniques as CuSum (Hawkins & Olwell, 1998).

CHAPTER 4

DATA COLLECTION, ANALYSIS AND RESULTS

4.0 Introduction

This chapter introduces the analysis of the various models and discussion of findings. It compromises of the preliminary analysis and modelling sections. The AR, MA, ARMA and ARIMA model shall be used in the modelling. The data covered water consumption from 2009 to 2012 gathered on fortnight bases.

4.1 Preliminary Analysis

It is recommended that a lengthy time series data is required for univariate time series forecasting. Meyler et al., (1988) recommended that at least 50 observations should be used for such a univariate time series forecasting. This could be problematic if few observations are used. However, when using a long time series data, it could be possible that the series contains a structural break which may necessitate only examining a sub-section of the entire data series or alternatively using intervention analysis or dummy variables. This is because there may be some conflict between the 60 needed for sufficient degrees of freedom for statistical robustness and having a shorter data sample to avoid structural breaks. The series should be plotted against time to assess whether any structural breaks, outliers or data errors occurred. This step may also reveal whether there is significant seasonal pattern in the times series or not. A dimension of the preliminary analysis for examining non-stationarity of the data is by considering the trend analysis plot of 106 fortnight water consumption between 2009 and 2012 as shown in Figure 4.1.

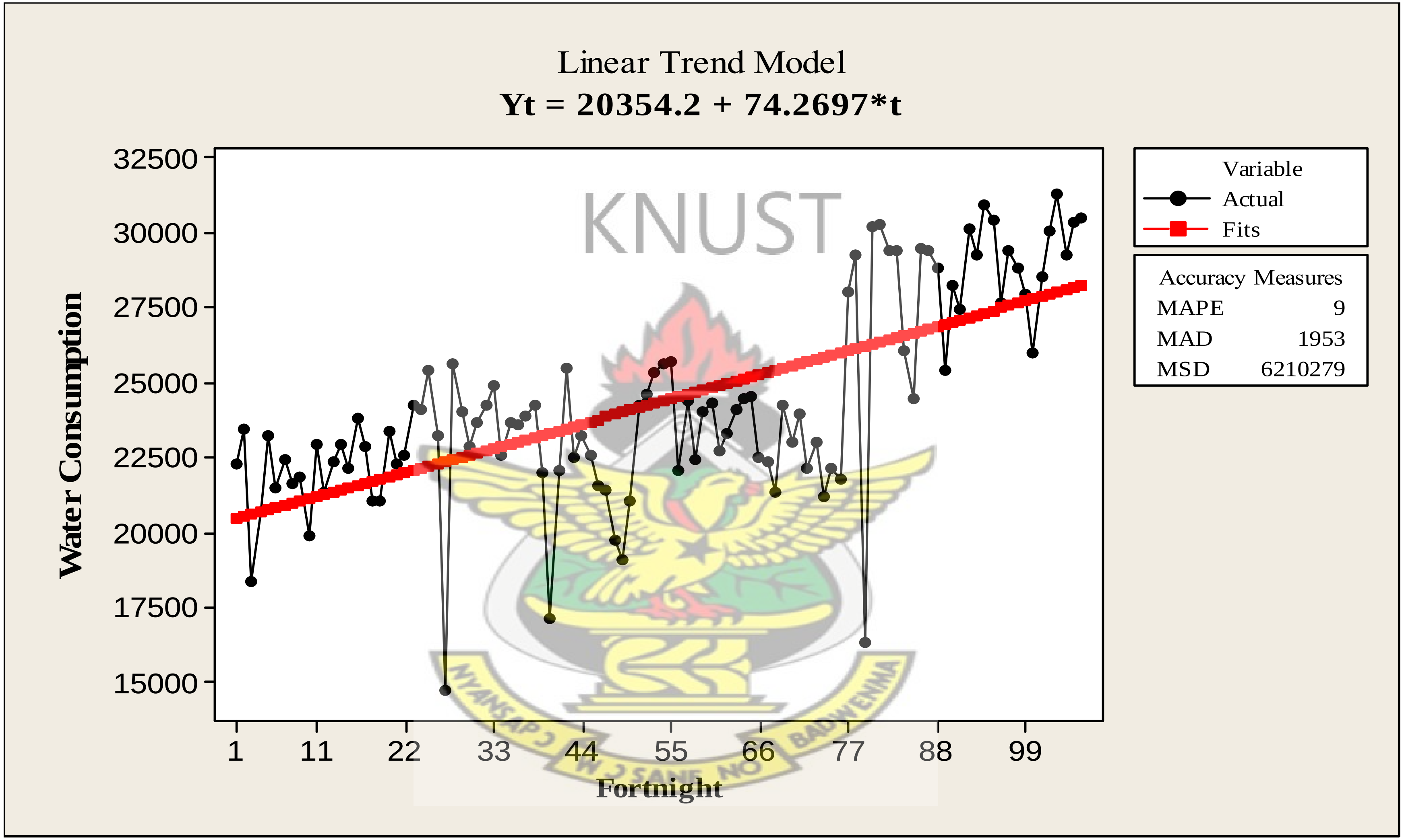

Figure 4.1: Trend in Fortnight Water Consumption between 2002 and 2012

It is revealed from Figure 4.1 that water consumption in the Hohoe Municipality between 2009 and 2012 has been largely non-stationary. The mean is not constant throughout the series as it assumes a fairly stable mean till 26th fortnight. The 27th, 40th and 79th fortnights recorded significantly low water consumption, perhaps due to insufficient water provision in the municipality. Furthermore, Moving Average (MA) analyses for lags 2, 4 and 8 are in Figures 4.2, 4.3, and 4.4 below. A comparison of their respective accuracy measures indicates that MA (8) better fits the data.

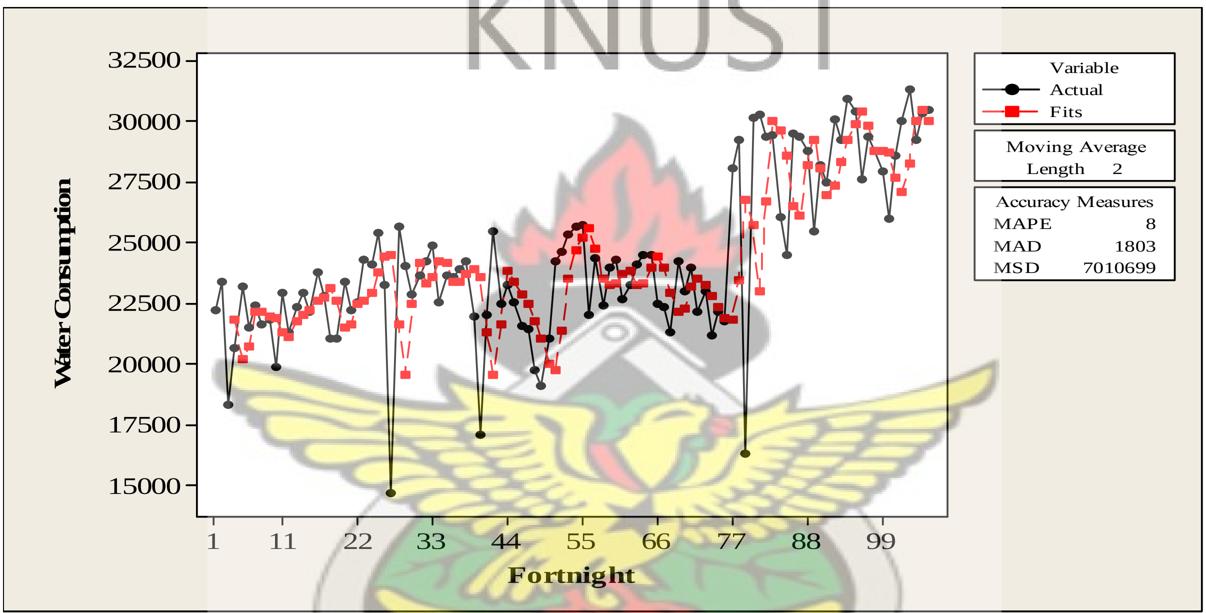

Figure 4.2: Moving Average (MA) with 2 Averages

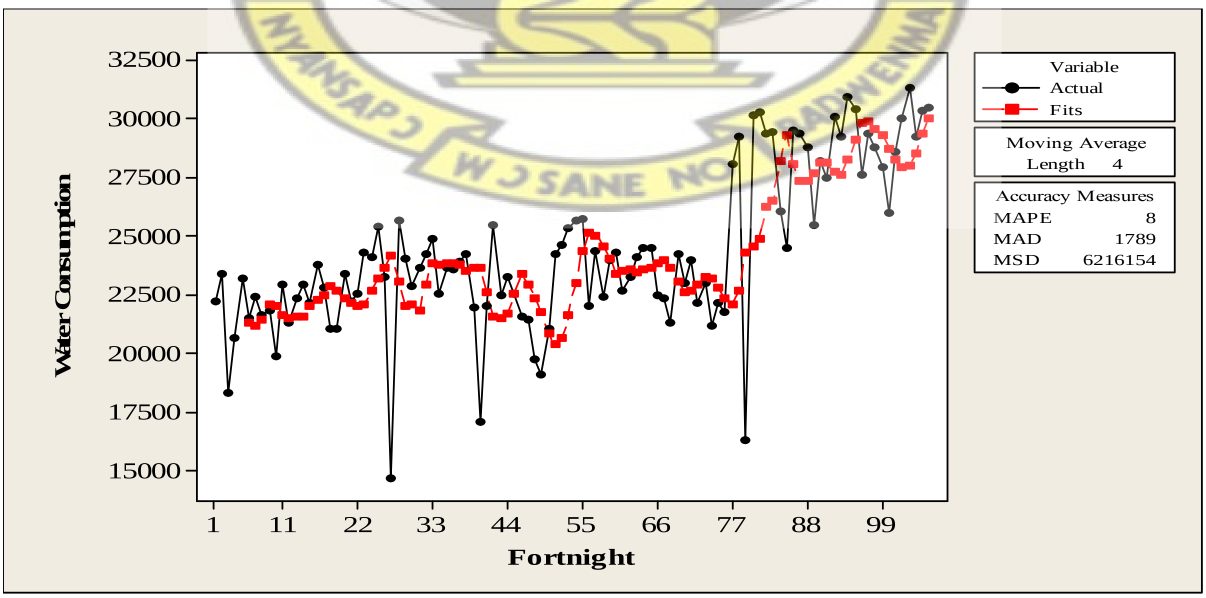

Figure 4.3: Moving Average (MA) with 4 Averages

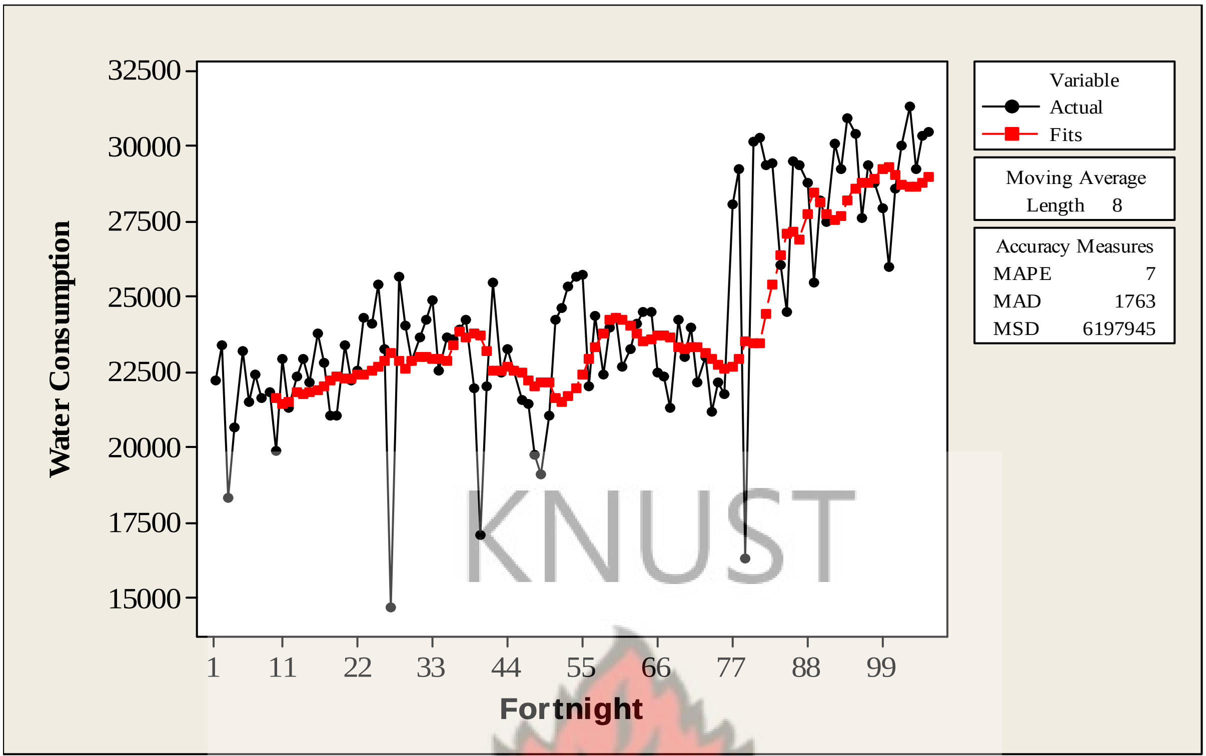

Figure 4.4: Moving Average (MA) with 8 Averages

The next step in the model building procedure is to determine the order of the AR and MA for both seasonal and non-seasonal components. This was suggested by the sample ACF and PACF plots based on the Box-Jenkins approach. From Figure 4.2, the correlations are significant for a large number of lags but perhaps the autocorrelations at lags 2 or and above are merely due to the propagation of the autocorrelation at lag 1. This is confirmed by the PACF plot.

The ACF and PACF plots in Figures 4.5 and 4.6 respectively suggest that q = 2 or 3, and p = 2 would be needed to describe this data set as coming from a non-seasonal moving average and autoregressive process respectively.

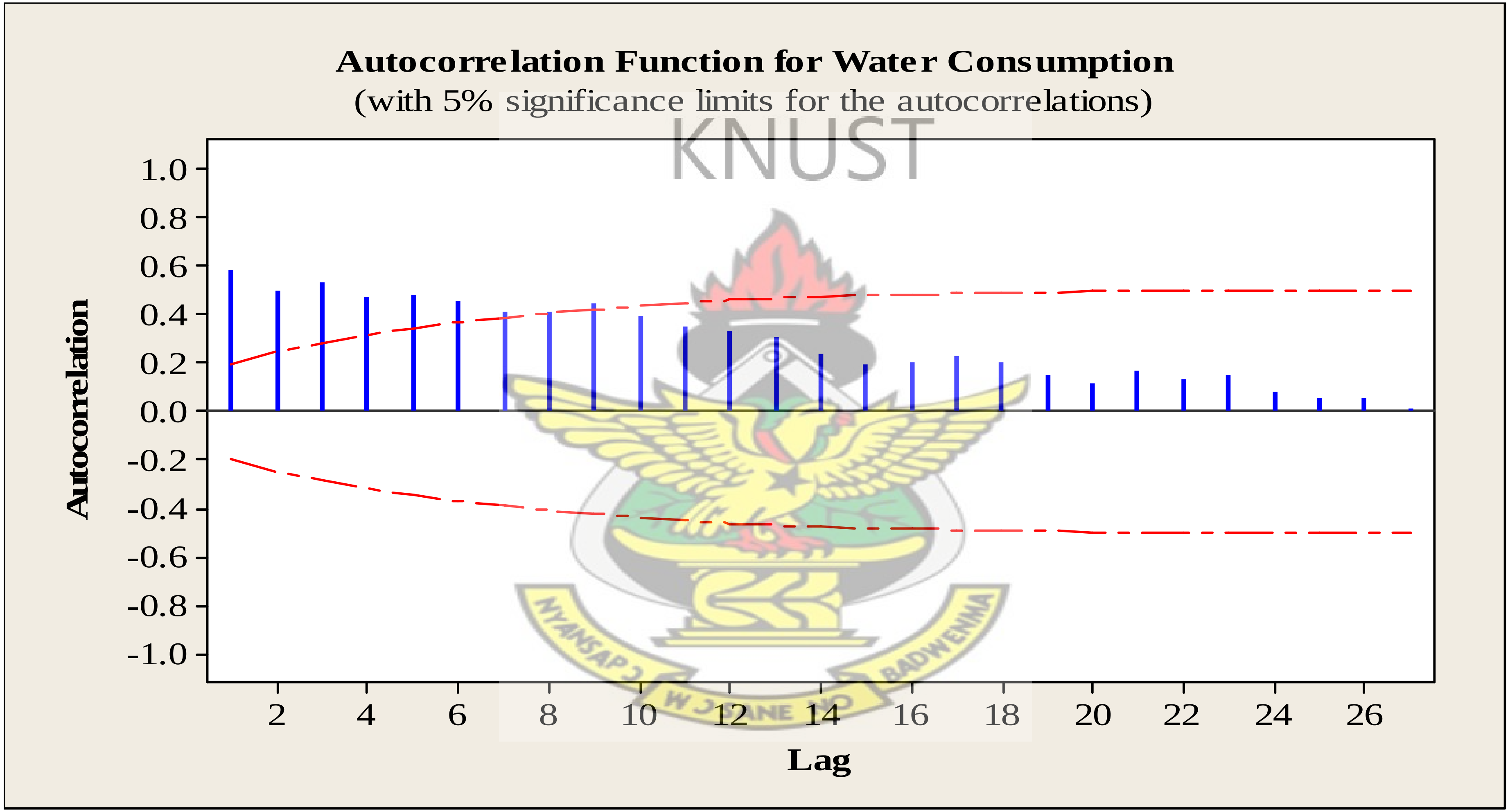

Figure 4.5: ACF for First Order Differencing

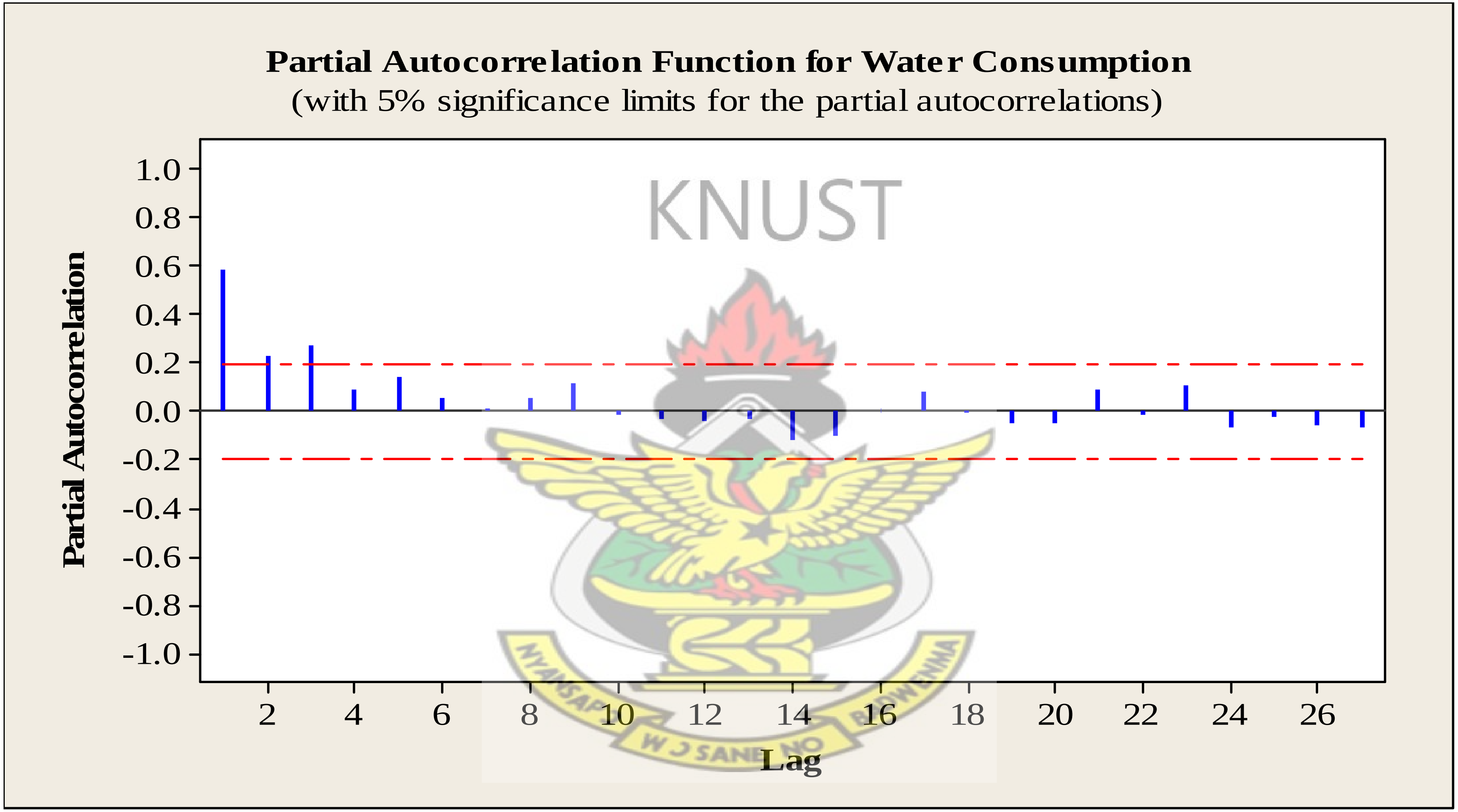

Figure 4.6: PACF for First Order Differencing

4.2 Non-seasonal ARIMA Model Estimations

Several non-seasonal ARIMA models are constructed as follows:

Table 4.1: ARIMA (1, 1, 0)

Type	Coefficient	SE	t	p
Constant	106.0	269.5	0.39	0.695
AR 1	-0.4055	0.0901	-4.50	0.000

$\chi^2 = 22.3$; $p = 0.013$; df = 10

Table 4.2: ARIMA (1, 1, 1)

Type	Coefficient	SE	t	p
Constant	73.97	31.15	2.37	0.019
AR 1	0.0839	0.1158	0.72	0.470
MA 1	0.8696	0.0569	15.27	0.000

$\chi^2 = 4.1$; $p = 0.901$; df = 9

Table 4.3: ARIMA (1, 1, 2)

Type	Coefficient	SE	t	p
Constant	152.80	58.63	2.61	0.011
AR 1	-0.8832	0.0609	-14.50	0.000
MA 1	-0.1156	0.0667	-1.73	0.086
MA 2	0.8599	0.0411	20.92	0.000

$\chi^2 = 3.3$; $p = 0.917$; df = 8

Table 4.4: ARIMA (1, 1, 3)

Type	Coefficient	SE	t	p
Constant	153.20	63.37	2.42	0.017
AR 1	-0.8891	0.0756	-11.76	0.000
MA 1	-0.1413	0.1355	-1.04	0.299
MA 2	0.8461	0.0545	15.52	0.000
MA 3	0.0205	0.1109	0.19	0.854

$\chi^2 = 3.1$; $p = 0.877$; df = 7

Table 4.5: ARIMA (2, 1, 2)

Type	Coefficient	SE	t	p
Constant	126.59	56.37	2.25	0.027
AR 1	-0.4754	0.5375	-0.88	0.379
AR 2	-0.0800	0.1508	-0.53	0.597
MA 1	0.2946	0.5340	0.55	0.582
MA 2	0.4659	0.4789	0.97	0.333

$\chi^2 = 2.1$; $p = 0.955$; df = 7

Table 4.6: ARIMA (2, 1, 4)

Type	Coefficient	SE	t	p
Constant	152.84	64.41	2.37	0.020
AR 1	0.0177	0.0660	0.27	0.789
AR 2	-0.9191	0.0640	-14.37	0.000
MA 1	0.7815	0.1057	7.39	0.000
MA 2	-0.8968	0.0922	-9.73	0.000
MA 3	0.8148	0.0983	8.29	0.000
MA 4	0.0227	0.1016	0.22	0.824

$\chi^2 = 4.2$; $p = 0.514$; df = 5

Table 4.7: ARIMA (3, 1, 3)

Type	Coefficient	SE	*t*	*p*
Constant	116.68	53.790	2.17	0.032
AR 1	-0.5735	1.5748	-0.36	0.716
AR 2	0.2157	1.3598	0.16	0.874
AR 3	-0.0696	0.1258	-0.55	0.581
MA 1	0.1736	1.5828	0.11	0.913
MA 2	0.8636	0.1712	5.05	0.000
MA 3	-0.2686	1.2911	-0.21	0.836

$\chi^2 = 2.4$; $p = 0.788$; df = 5

Table 4.8: ARIMA (2, 1, 0)

Type	Coefficient	SE	*t*	*p*
Constant	159.3	249.0	0.64	0.524
AR 1	-0.5639	0.0909	-6.20	0.000
AR 2	-0.3965	0.0909	-4.36	0.000

$\chi^2 = 124$; $p = 0.191$; df = 9

Table 4.9: ARIMA (1, 2, 3)

Type	Coefficient	SE	*t*	*p*
Constant	-1.831	5.425	-0.34	0.736
AR 1	-0.8567	0.1498	-5.72	0.000
MA 1	0.5238	0.2006	2.61	0.010
MA 2	0.8690	0.1384	6.28	0.000
MA 3	-0.4019	0.1168	-3.34	0.001

$\chi^2 = 12.2$; $p = 0.096$; df = 7

Table 4.10: ARIMA (3, 2, 3)

Type	Coefficient	SE	*t*	*p*
Constant	2.465	5.954	0.41	0.680
AR 1	-0.9321	0.2465	-3.78	0.000
AR 2	-0.3516	0.1768	-1.99	0.050
AR 3	-0.1892	0.1089	-1.74	0.085
MA 1	0.7768	0.2022	3.84	0.000
MA 2	0.6917	0.4553	1.52	0.132
MA 3	-0.4790	0.2571	-1.86	0.065

$\chi^2 = 5.4$; $p = 0.369$; df = 5

In comparing the *p*-values and Chi-square values of the four likely non-seasonal ARIMA, it can be concluded that model ARIMA (2, 1, 2) for 27 lags (because there 27 fortnights in a year) has the highest *p* and a relatively low Chi-square values of 0.955 and 2.2. This indicates that it is the best non-seasonal model for the data.

Table 4.11: Comparison of Non-Seasonal Models

Model	***p*-value**	**Chi-Square**	**Df**
ARIMA $(1, 1, 1)_{27}$	0.901	4.1	9
ARIMA $(1, 1, 2)_{27}$	0.917	3.3	8
ARIMA $(1, 1, 3)_{27}$	0.877	3.1	7
ARIMA $(2, 1, 2)_{27}$	**0.955**	**2.2**	**7**

4.3 Seasonal ARIMA Model Estimations

At this stage, it is important also to consider the seasonality of the data by adopting the seasonal ARIMA models. Looking at the seasonal lags, both ACF and PACF spike at seasonal lag 27 (because there are 27 fortnights in a year) and drop to zero for other seasonal lags suggesting that Q = 1 or 2 and P = 0 or 1 with $d = 1$ would be needed to describe these data as coming from a seasonal moving average and autoregressive process. Therefore, 15 proposed SARIMA models are presented in Table 4.12 with their corresponding p-values, Chi-square values and degree of freedom.

Table 4.12: Suggested SARIMA Models

Model	***p*-value**	**Chi-Square**	**Df**
ARIMA (0, 1, 1) $(1, 0, 1)_{27}$	0.557	7.8	9
ARIMA (1, 1, 1) $(0, 1, 1)_{27}$	0.798	3.8	7
ARIMA (1, 1, 3) $(1, 0, 1)_{27}$	0.793	2.4	5
ARIMA (1, 1, 3) $(1, 0, 1)_{27}$	0.699	2.2	4
ARIMA (1 1, 3) $(0, 1, 0)_{27}$	0.427	6.0	6
ARIMA (2, 1, 2) $(0, 1, 0)_{27}$	0.859	3.3	7
ARIMA (2, 1, 1) $(1, 1, 1)_{27}$	0.534	5.1	6
ARIMA (2, 1, 2) $(1, 0, 1)_{27}$	0.774	2.5	5
ARIMA (2, 1, 2) $(0, 0, 1)_{27}$	0.846	2.7	6
ARIMA (2, 1, 3) $(1, 0, 1)_{27}$	0.699	2.2	4
ARIMA (2, 1, 2) $(1, 0, 0)_{27}$	**0.915**	**2.0**	**6**
ARIMA (4, 1, 3) $(0, 0, 1)_{27}$	0.614	1.8	3
ARIMA (3, 1, 1) $(1, 1, 1)_{27}$	0.011	15.0	5
ARIMA (2, 2, 3) $(0, 0, 1)_{27}$	0.186	7.5	5
ARIMA (2, 2, 2) $(1, 0, 1)_{27}$	0.006	16.2	5

A critical comparison of the models based on their respective p-values and Chi-Square values shows that seasonal ARIMA (2, 1, 2)$(1, 0, 0)_{27}$ is the appropriate model that best fitted the fortnight water consumption data in the Hohoe Municipality in the Volta Region of Ghana. This would, however, be compared with the non-seasonal ARIMA model for final selection of the most adequate and parsimonious model.

4.4 Model Evaluation and Selection

From the aforementioned, we have identified two good models, namely, a non-seasonal and seasonal ARIMA models as shown in Tables 4.13 and 4.14 respectively for comparison and selection. We used the conditional-sum-of-squares to find starting values of parameters, then do the Maximum Likelihood Estimate (MLE) for the proposed models. The procedure for choosing these models relied on choosing the model with the maximum *p*-values for the Ljung-Box statistic (more than 5% as a rule of thumb) and minimum Chi-square values.

Comparing the non-seasonal ARIMA and the seasonal ARIMA models, it can be concluded that the non-seasonal model of (2, 1, 2) is somewhat adequate than the seasonal ARIMA model of (2, 1, 2) $(1, 0, 0)_{27}$. Hence, ARIMA (2, 1, 2) is the best model and plausible time series model for the fortnight water consumption because of its high *p* and least Chi-Square values of 0.955 and 2.0 respectively.

Table 4.13: ARIMA (2, 1, 2)

Type	Coefficient	SE	*t*	*p*
Constant	126.59	56.37	2.25	0.027
AR 1	-0.4754	0.5375	-0.88	0.379
AR 2	-0.0800	0.1508	-0.53	0.597
MA 1	0.2946	0.5340	0.55	0.582
MA 2	0.4659	0.4789	0.97	0.333

$\chi^2 = 2.1$; $p = 0.955$; df = 7

Table 4.14: Estimates of Parameters for SARIMA (2, 1, 2) (1, 0, 0)$_{27}$

Variable	**Coefficients**	**SE**	***t***	***p***
Constant	122.55	60.72	2.02	0.046
AR 1	-0.5697	0.4053	-1.41	0.163
AR 2	-0.1006	0.1410	-0.71	0.477
SAR 27	0.1811	0.1199	1.51	0.134
MA 1	0.2128	0.3988	0.53	0.595
MA 2	0.5279	0.3600	1.47	0.146

***p=0.735*, χ^2=4.4, df=6**

4.5 Diagnostic Analysis

The diagnostic analyses using the ACF of residuals, PACF residuals, and the normal probability plot of the residuals as shown in Figures 4.7, 4.8, 4.9 and 4.10 reveal that the residuals of the model have zero mean and constant variance. The ACF of the residuals depicts that the autocorrelation of the residuals are all zero, that is to say they are uncorrelated. Hence, it can be concluded that there is a constant variance among residuals of the selected model and the true mean of the residuals is approximately equal to zero. Thus, the selected model satisfies all the model assumptions. Since the ARIMA (2, 1, 2) satisfies all the necessary assumptions, it can be inferred that the model provides an adequate representation of the data. Hence, the predictive model would be formulated from the parameter estimates in Table 4.13.

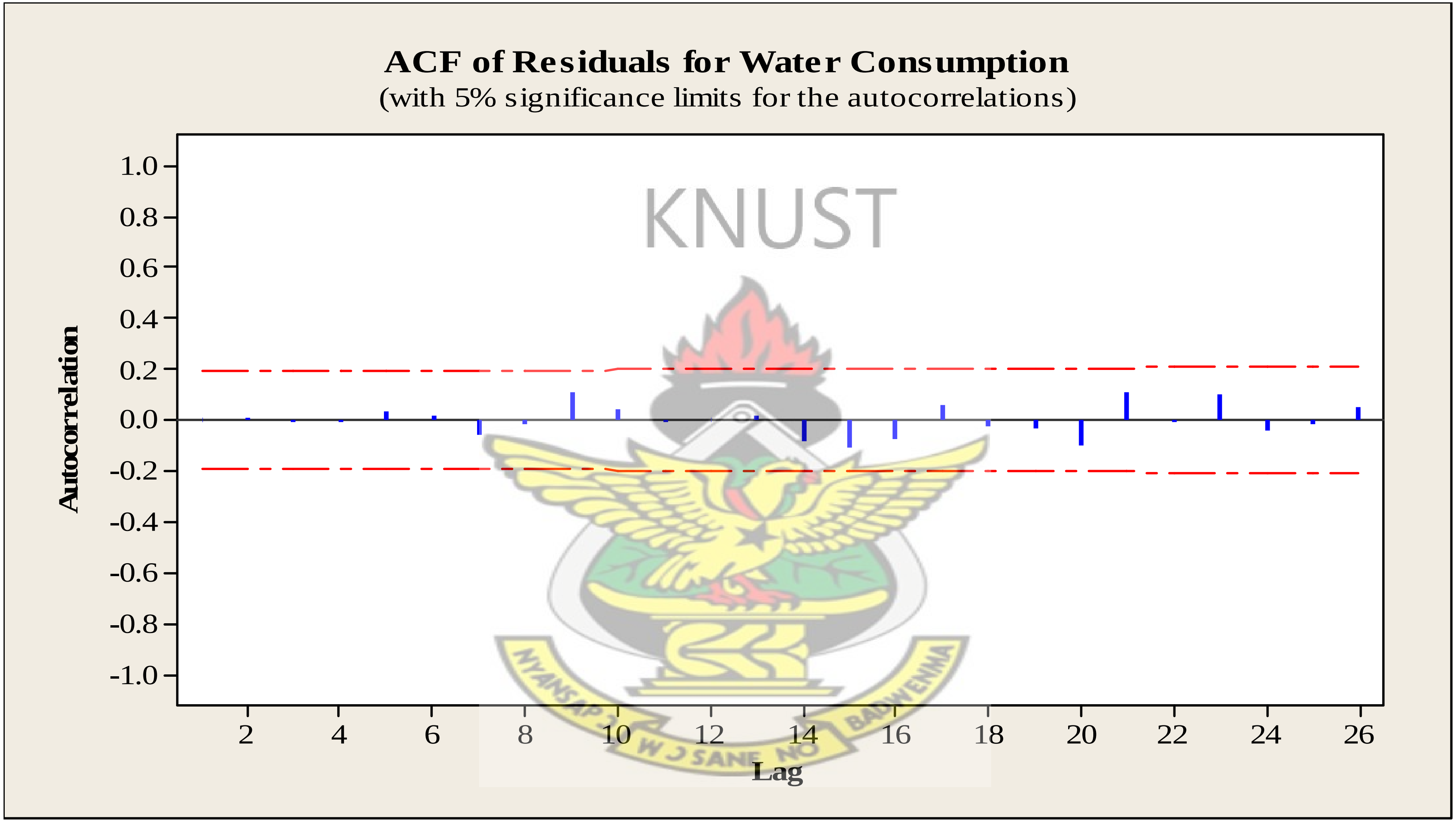

Figure 4.7: ACF Diagnostic Plot of the Residuals for ARIMA (2, 1, 2) Model

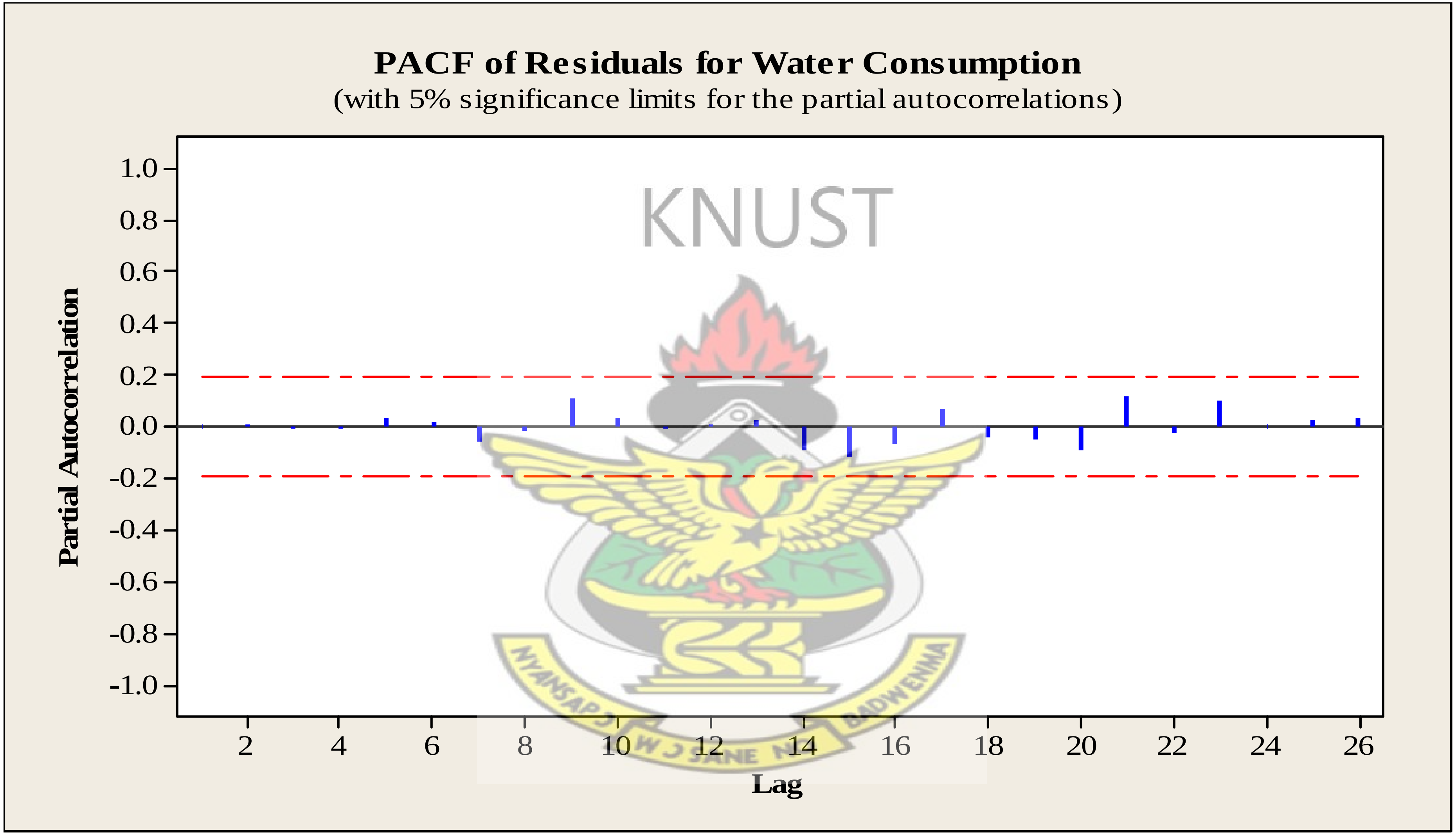

Figure 4.8: PACF Diagnostic Plot of the Residuals for ARIMA (2, 1, 2) Model

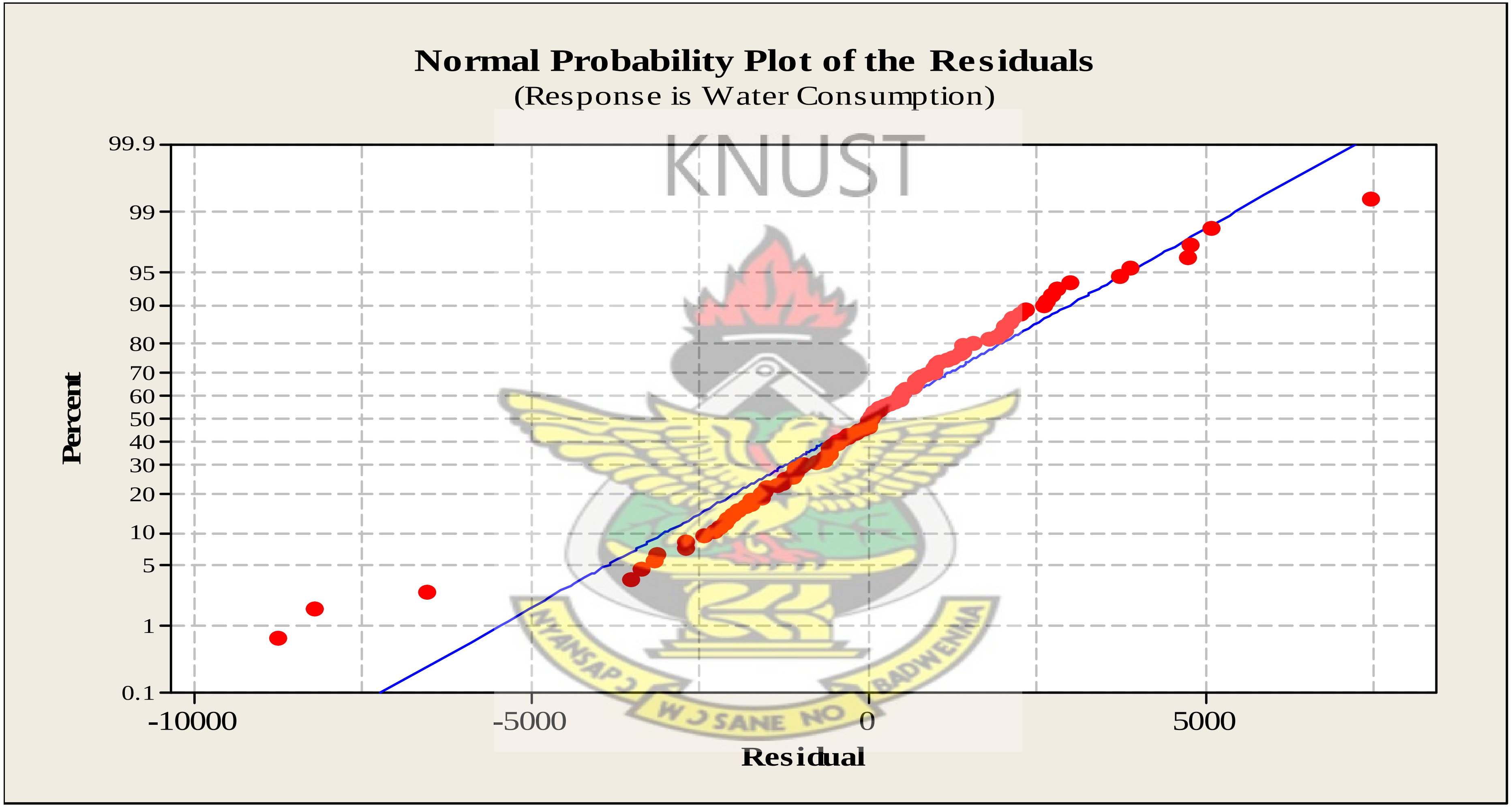

Figure 4.9: Normal Probability Plot of the Residuals for ARIMA (2, 1, 2) Model

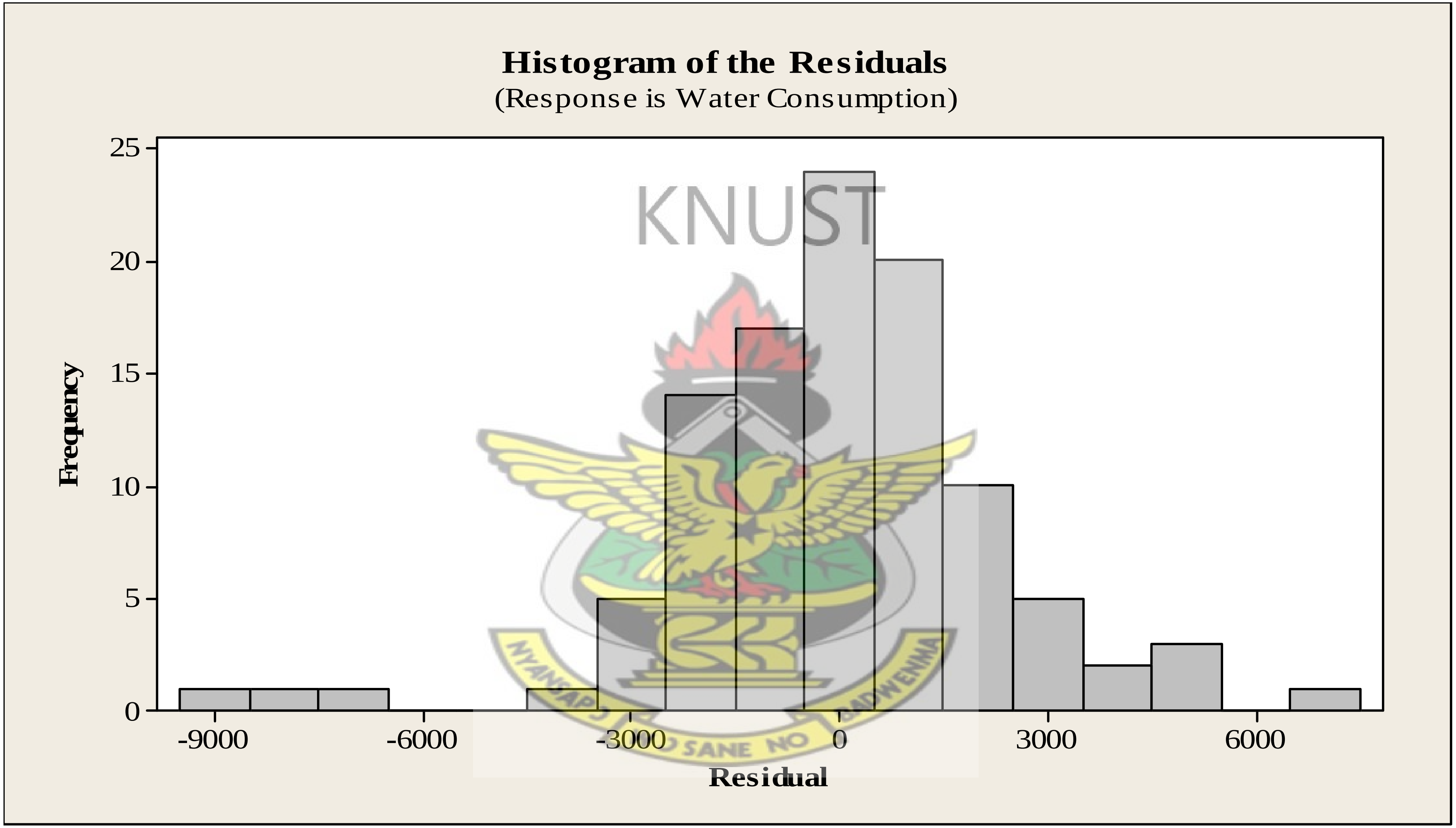

Figure 4.10: Histogram of the Residuals for ARIMA (2, 1, 2) Model

4.6 Model validation

In order to test the adequacy and predictive ability of the chosen model, ARIMA (2, 1, 2), the actual data set and its predicted values, lower and upper limits are plotted and displayed in Figure 4.11 below. It can be seen that the predicted values are well-fitted through the original data with the lower and upper limits containing majorities of the original data. This indicates that the model is the best fitted one for the data set.

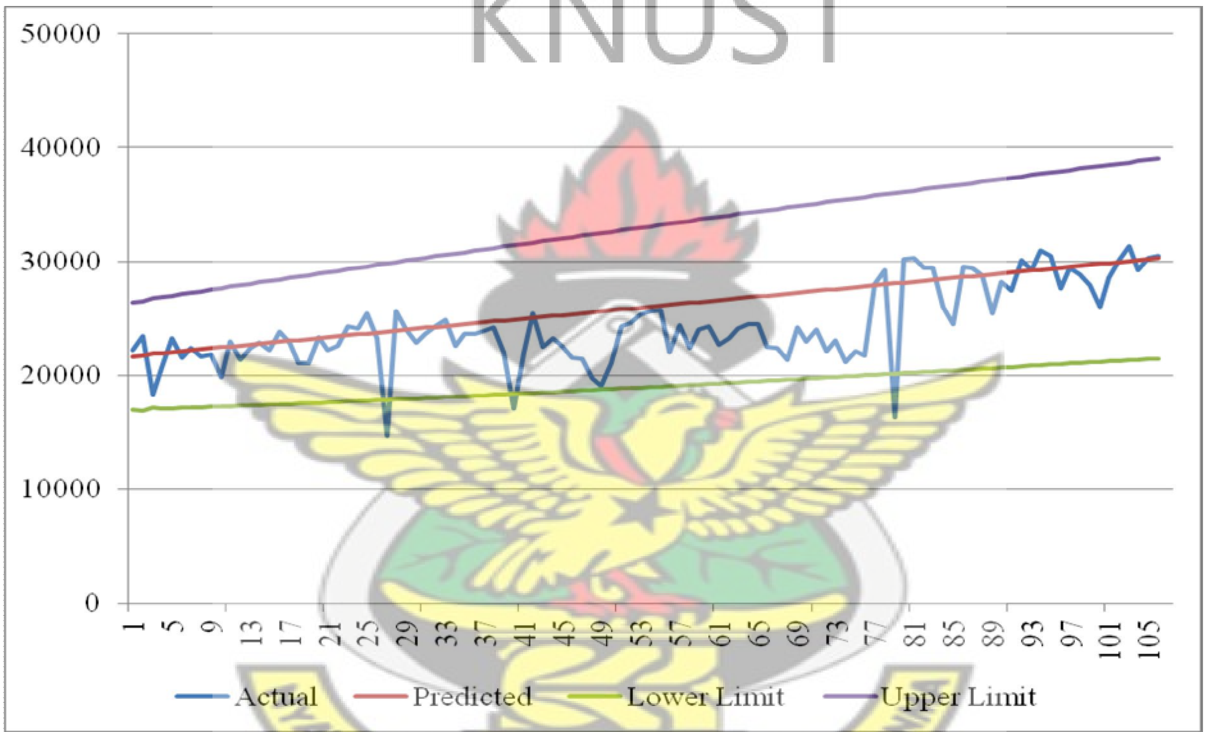

Figure 4.11: Actual and Predicted Water Consumption using ARIMA (2, 1, 2)

4.6 Forecasting

Forecasting plays an important role in decision making process. It is a planning tool which helps decision makers to foresee the future uncertainty based on the behaviour of past and current observations. Forecasting as describe by Box and Jenkins (1976), provide basis for economic and business planning, inventory and production control and control and optimisation of industrial processes. Forecasting is the process of predicting some unknown quantities. From previous studies, most research work has found that the selected model is not necessary the model that provides best forecasting. In this sense, further forecasting accuracy test such as ME, RMSE and MAE must be performed on the model. Table 4.15 summarises the forecasted values of water consumption in the Municipality over the period of January 2013 to December 2016 fortnightly with 95% confidence level using the ARIMA (2, 1, 2) model which has higher *p*-value of 0.955 (thus, greater than alpha value of 0.05) indicating that it is the best model according to Modified Box-Pierce (Ljung-Box) Chi-Square statistic.

Table 4.15: Forecasted Fortnight Water Consumption for the Next 4 Years

			95% CI	
Year	Fortnight	Forecast	Lower Limit	Upper Limit
2013	**1st**	**29809.6**	**25129.4**	**34489.7**
2013	2nd	29995.4	25193.0	34797.7
2013	3rd	30089.4	25281.3	34897.5
2013	4th	30156.4	25260.5	35052.4
2013	5th	30243.6	25303.0	35184.3
2013	6th	30323.4	25328.8	35318.0
2013	7th	30405.1	25358.8	35451.3
2013	8th	30486.5	25389.1	35583.8
2013	9th	30567.8	25419.7	35716.0
2013	10th	30649.2	25450.9	35847.5
2013	11th	30730.6	25482.5	35978.7

2013	12th	30812.0	25514.6	36109.3
2013	13th	30893.4	25547.2	36239.5
2013	14th	30974.7	25580.2	36369.3
2013	15th	31056.1	25613.7	36498.6
2013	16th	31137.5	25647.5	36627.5
2013	17th	31218.9	25681.8	36756.0
2013	18th	31300.3	25716.5	36884.1
2013	19th	31381.7	25751.5	37011.8
2013	20th	31463.0	25787.0	37139.1
2013	21st	31544.4	25822.8	37266.1
2013	22nd	31625.8	25858.9	37392.7
2013	23rd	31707.2	25895.4	37518.9
2013	24th	31788.6	25932.3	37644.9
2013	25th	31870.0	25969.5	37770.4
2013	26th	31951.3	26007.0	37895.7
2013	27th	32032.7	26044.8	38020.6
2014	**1st**	**32114.1**	**26083.0**	**38145.2**
2014	2nd	32195.5	26121.5	38269.5
2014	3rd	32276.9	26160.2	38393.5
2014	4th	32358.3	26199.3	38517.2
2014	5th	32439.6	26238.6	38640.6
2014	6th	32521.0	26278.3	38763.8
2014	7th	32602.4	26318.2	38886.6

2014	8th	32683.8	26358.4	39009.2
2014	9th	32765.2	26398.8	39131.5
2014	10th	32846.5	26439.5	39253.6
2014	11th	32927.9	26480.5	39375.4
2014	12th	33009.3	26521.7	39496.9
2014	13th	33090.7	26563.2	39618.2
2014	14th	33172.1	26604.9	39739.3
2014	15th	33253.5	26646.8	39860.1
2014	16th	33334.8	26689.0	39980.7
2014	17th	33416.2	26731.4	40101.0
2014	18th	33497.6	26774.1	40221.2
2014	19th	33579.0	26816.9	40341.1
2014	20th	33660.4	26860.0	40460.7
2014	21st	33741.8	26903.3	40580.2
2014	22nd	33823.1	26946.8	40699.5
2014	23rd	33904.5	26990.5	40818.5
2014	24th	33985.9	27034.4	40937.4
2014	25th	34067.3	27078.5	41056.0
2014	26th	34148.7	27122.8	41174.5
2014	27th	34230.1	27167.3	41292.8
2015	**1st**	**34311.4**	**27212.0**	**41410.8**
2015	2nd	34392.8	27256.9	41528.7
2015	3rd	34474.2	27302.0	41646.4

2015	4th	34555.6	27347.3	41763.9
2015	5th	34637.0	27392.7	41881.2
2015	6th	34718.3	27438.3	41998.4
2015	7th	34799.7	27484.1	42115.4
2015	8th	34881.1	27530.1	42232.2
2015	9th	34962.5	27576.2	42348.8
2015	10th	35043.9	27622.5	42465.3
2015	11th	35125.3	27669.0	42581.6
2015	12th	35206.6	27715.6	42697.7
2015	13th	35288.0	27762.4	42813.7
2015	14th	35369.4	27809.3	42929.5
2015	15th	35450.8	27856.4	43045.2
2015	16th	35532.2	27903.7	43160.7
2015	17th	35613.6	27951.1	43276.0
2015	18th	35694.9	27998.6	43391.3
2015	19th	35776.3	28046.3	43506.3
2015	20th	35857.7	28094.2	43621.2
2015	21st	35939.1	28142.2	43736.0
2015	22nd	36020.5	28190.3	43850.6
2015	23rd	36101.9	28238.6	43965.1
2015	24th	36183.2	28287.0	44079.5
2015	25th	36264.6	28335.6	44193.7
2015	26th	36346.0	28384.3	44307.8

2015	27th	36427.4	28433.1	44421.7
2016	**1st**	**36508.8**	**28482.0**	**44535.5**
2016	2nd	36590.1	28531.1	44649.2
2016	3rd	36671.5	28580.3	44762.7
2016	4th	36752.9	28629.7	44876.2
2016	5th	36834.3	28679.1	44989.5
2016	6th	36915.7	28728.7	45102.6
2016	7th	36997.1	28778.4	45215.7
2016	8th	37078.4	28828.3	45328.6
2016	9th	37159.8	28878.2	45441.4
2016	10th	37241.2	28928.3	45554.1
2016	11th	37322.6	28978.5	45666.7
2016	12th	37404.0	29028.8	45779.1
2016	13th	37485.4	29079.2	45891.5
2016	14th	37566.7	29129.8	46003.7
2016	15th	37648.1	29180.4	46115.8
2016	16th	37729.5	29231.2	46227.8
2016	17th	37810.9	29282.1	46339.7
2016	18th	37892.3	29333.0	46451.5
2016	19th	37973.7	29384.1	46563.2
2016	20th	38055.0	29435.3	46674.8
2016	21st	38136.4	29486.6	46786.2
2016	22nd	38217.8	29538.0	46897.6

2016	23rd	38299.2	29589.5	47008.8
2016	24th	38380.6	29641.1	47120.0
2016	25th	38462.0	29692.9	47231.0
2016	26th	38543.3	29744.7	47342.0
2016	27th	38624.7	29796.6	47452.9

CHAPTER FIVE

CONCLUSIONS AND RECOMMENDATIONS

5.0 Introduction

This chapter presents conclusions drawn from the study and some recommendations made to inform policy at the Ghana Water Company Limited at Hohoe in the Volta Region of Ghana.

5.1 Conclusions

The objective of this research was to develop a time series model and forecast fortnight water consumption by residents and businesses for the 4 years (2013 to 2016). Data from 2009 to 2012, on fortnight bases, were collated from the Ghana Water Company Limited at Hohoe. A time series trend analysis on the fortnight water consumption assumed stable mean except at the 27^{th}, 40^{th} and 79^{th} fortnights, which recorded significantly low water consumption due to low production.

Several time series models including AR, MA, ARMA, non-seasonal Autoregressive Integrated Moving Average (ARIMA) and seasonal ARIMA were used in modelling the data in the Minitab 14. The study identified several 'candidate' models which best fitted the data. However, with the use of the Modified Box-Pierce (Ljung-Box) Chi-square statistic criteria of the "largest *p*-value and minimum Chi-Square value," the best-fitted ARIMA model selected was ARIMA (2, 1, 2). After the estimation of the parameters of selected models, a series of diagnostic and forecasting accuracy tests were performed.

With reference to the findings of the research, it can be concluded that:

(i) The most adequate model for the data was ARIMA (2, 1, 2).

(ii) There will be no astronomical increases in water consumption in the Hohoe Municipality over the next 4 years as the forecasted values fell in the neighbourhood of 29,809.6 in the first fortnight in 2013 and 38,624.7 in the 27th fortnight in year 2016.

5.2 Recommendations

On the basis of the findings of the research, the following recommendations were made:

(i) The Ghana Water Company Limited authorities in the Hohoe Municipality should use the ARIMA (2, 1, 2) model in determining the water consumption levels by residents, institutions and businesses.

(ii) The predicted fortnight water consumption levels using the above model could greatly help the company in its operational activities.

(iii) In times of low water production, the company should prioritise areas and communities with high population and essential service providers such as hospitals, clinics and schools in the municipality.

(iv) Government through the Ghana Water Company Limited must put in effective and efficient water production and distribution strategies in the municipality since its population is steadily increasing.

REFERENCES

Aidoo, E. (2010). *Modeling and forecasting inflation rates in Ghana: An application of SARIMA models.* Högskolan Dalarna, School of Technology and Business Studies.

Anderson, J. M. (1996). Current water recycling initiatives in Australia: Scenarios for the 21st century. *Water Sci. Technol*, *33*, 37–43.

Australian Bureau of Statistics [ABS]. (2008). *Retail trade trends*. Retrieved on December 2012 from http://en.wikipedia.org/wiki/Australian_Bureau_of_Statistics

Arms, K. (2008). *Environmental science*. Austin, Texas: Holt, Rinehart and Winston.

Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (1994). *Time series analysis, forecasting and control* (3rd ed.). New Jersey: Prentice Hall, Englewood Clifs.

Caiado, J. (2009). *Performance of combined double seasonal univariate time series models for forecasting water consumption*. Munich: Munich University Personal RePEc Archive.

Cech, T. V. (2005). *Principles of water resources: History, development, management, and policy* (2nd ed.). Hoboken: John Wiley and Son, Inc,

Cheng, C. (2002). Study of the inter-relationship between water use and energy conservation for a building. *Energy and Buildings*, *34*, 261-266.

Chenoweth, J. (2008). *Looming water crisis simply a management problem*. Retrieved on November 20, 2012 from http://www.newscientist.com/ article/mg19926700.100-looming-water-crisis-simply-a-management-problem.html

Cui, F. (2011). *ARIMA models for bank failures: Prediction and comparison*. University of Nevada, Las Vegas.

Hagan, I. (2007). *Modeling the impact of small reservoirs in the Upper East Region of Ghana*. MSc thesis submitted to Lund University Press, Sweden.

Hardy, M. J. (2005). Integrated urban water cycle management: The urban cycle model. *Water Sci Technol, 52*, 1-9.

Hawkins, M. D., & Olwell, D. H. (1998). *Cumulative sum charts and charting for quality improvement*. Berlin/Heidelberg: SpringerVerlag.

Ho, C.-H. (2008). Empirical recurrent rate time series for Volcanism: Application to Avachinsky Volcano, Russia. *Volcanol Geotherm Res*, *173*, 15-25.

International Atomic Energy Agency (IAEA] (2004). *Nuclear desalination: Water supply and demand*. Retrieved on October 10, 2012 from http://www.iaea.org/ OurWork/index.html

Lasmiley, N. (2012). *Revamping Ghana's water sector: Lessons from South Africa*. Accra: Development Ghana.

Maher, M., & Lustig, T. (2003), Sustainable water cycle design for urban areas. *Water Sci Technol*, *47*, 25-31.

Marsalek, (2001). *Urban water as a part of integrated catchment management*, pp. 37–83.

Mitchell, G., Gray, S., Woolley, R., Erbacher, J., Egerton, G., & McKnoulty, J. (2003). Evaluating integrated urban water systems alternatives for Brisbane, Australia. *Water Sci Technol*, *47*, 1-9.

Mitchell, V. G., Mein, R. G., & McMahon, T. A. (2001). Modeling the urban water cycle. *Environmental Modelling and Software, 16*, 615-629.

Niemczynowicz, J. (1999). Urban hydrology and water management- present and future challenges. *Urban Water*, *1*, 1-14.

Przyborowski, J., & Wilenski, H. (1940). Homogeneity of results in testing samples from Poisson series. *Biometrika, 31*, 313–323.

Reis, B. Y., Pagano, M., & Mandl, K. D. (2003). Using temporal context to improve biosurveillance. *Proceedings National Academy of Sciences, 100*(4), 1961-1965.

Remirev-Beltran, N. (1993). *Prediction of water use in Puerto Rico, Phase 1: Mayaguez.* Washington, D.C.: Office of Water Resources Research.

Rizzo, S. L., Grigoryan, V. V., Wallstrom, G. L., Hogan, W. R., & Wagner M. M. (2005). *The use of case studies for the evaluation of surveillance systems.* RODS Technical Report Centre for Biomedical Informatics, University of Pittsburgh.

Udoh, E. J., & Etim, N. A. (2007). Analysis of domestic water consumption pattern by farming households in Itu, Akwa Ibom State, Nigeria. *European Journal of Social Sciences, 5*(2), 76-78.

UNESCO (2001). Marseille statement, the UNESCO Symposium on frontiers in Urban Water Management: Dead Lock or Hope. *Urban Water, 3*, 129-130.

United Nations World Water Development Report [UN WWDR] (2006). *World Water Assessment Programme (WWAP).* Paris: UNESCO.

University of Minnesota Extension Services [UMES] (2008). *Ensuring a safe water supply.* Retrieved on October 18, 2012 from http://www.extension.umn.edu/distribution/naturalresources/components/DD694d.html

WaterAid (2013). *The keeping promises: Why African leaders need now to deliver on their past water and sanitation commitments report.* Retrieved on March 15, 2013 from here: http://www.wateraid.org/documents/WaterAid_Keeping_Promises_Synthesis_sReport.pdf

WaterAid-Ghana (2013). *7 million more Ghanaians lack access to sanitation than in 1990. GhanaWeb.com*

WHO/UNICEF (2012). Global water supply coverage and sanitation assessment report. Retrieved on January, 24 2013 from http://www.wss info.org/en/22_wat_global.html

WHO/UNICEF Joint Monitoring Programme (2012). *Joint Monitoring Programme Water Supply and Sanitation.* Geneva: WHO/UNICEF.

World Health Organisation [WHO] (1997). *Guidelines for drinking-water quality: Surveillance and control of community supplies*. Geneva: WHO.

Zaizen M., Urakawa T., Matsumoto Y., & Takai H. (2000). The collection of rainwater from dome stadiums in Japan. *Urban Water, 1*, 355-359.

APPENDIX A

DATA ON FORTNIGHT WATER CONSUMPTION IN THE HOHOE MUNICIPALITY

2009		2010		2011		2012	
Fortnights	**Data**	**Fortnights**	**Data**	**Fortnights**	**Data**	**Fortnights**	**Data**
1	22245	28	25647	55	25721	81	30264
2	23415	29	24040	56	22025	82	29400
3	18319	30	22846	57	24366	83	29443
4	20675	31	23672	58	22419	84	26045
5	23220	32	24257	59	24010	85	24479
6	21514	33	24883	60	24326	86	29499
7	22398	34	22570	61	22700	87	29401
8	21639	35	23668	62	23267	88	28795
9	21852	36	23587	63	24127	89	25445
10	19862	37	23903	64	24480	90	28226
11	22953	38	24212	65	24519	91	27472
12	21338	39	21968	66	22486	92	30105
13	22356	40	17108	67	22368	93	29269
14	22904	41	22043	68	21314	94	30946
15	22161	42	25449	69	24226	95	30434
16	23786	43	22464	70	22994	96	27631
17	22827	44	23261	71	23972	97	29408
18	21072	45	22546	72	22151	98	28826
19	21044	46	21541	73	23032	99	27922
20	23374	47	21414	74	21173	100	26002
21	22245	48	19759	75	22133	101	28573
22	22551	49	19063	76	21766	102	30056
23	24274	50	21036	77	28054	103	31325
24	24086	51	24270	78	29265	104	29275
25	25394	52	24599	79	16325	105	30330
26	23241	53	25358	80	30179	106	30507

Source:* *Ghana Water Company Ltd., Hohoe.