ASSESSING THE IMPACT OF THE ANGLOGOLD ASHANTI INTEGRATED MALARIA CONTROL PROGRAMME ON THE NUMBER OF MALARIA REPORTED CASES IN THE OBUASI MUNICIPALITY

BY

BONIFACE AABEYIR

(MSC. INDUSTRIAL MATHEMATICS)

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

Master of Science

Mathematics Department,

Institute of Distance Learning

August 2010

DECLARATION

I, hereby declare that this submission is my own work towards the award of MSc and that, to the best of my knowledge, it contains no materials 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.

……	……	……
Student‘s Name and ID	Signature	Date

Certified by:

……	……	……
Supervisor’s Name	Signature	Date

……..	……..	……
Head of Department’s Name	Signature	Date

DEDICATION

This work is dedicated to the Almighty God who is my source of wisdom, knowledge and power.

ABSTRACT

AngloGold Ashanti, a mining company in Obuasi realized the threat the malaria posed to its operation so it launched an intervention programme called integrated malaria programme. This research was aimed at assessing the impact of the programme on the number of reported cases over the five-year period of its implementation, formulating an autoregressive integrated moving average model for checking the trend of reported cases and finding ways of improving on the effectiveness of the programme. Secondary data on reported cases of the diseases was collected from the municipal health directorate and questionnaires were given to the four main hospitals in the municipality to seek their opinions on the effectiveness or otherwise of the programme. The management of the programme was given a questionnaire to find out the challenges encountered by the programme. Method of comparing two population means was used to compare the average monthly reported and statistical package was used to formulate an integrated autoregressive moving average model for forecasting the number of malaria reported cases. The results indicated that there is decreasing trend in the number of reported cases. Hence, the programme is having a positive impact on the number of reported cases. The model developed for forecasting the number of reported cases is $y_t = -0.052 + \varepsilon_t - 1.914\varepsilon_{t-1} + 0.914\varepsilon_{t-2}$. This model is recommended to the malaria control programme, municipal health directorate and researchers who want to monitor the malaria reported cases in the municipality. The programme is having positive impact on the number of malaria reported cases bin the municipality. Health facilities in the municipality should adopt the surveillance system put in place by the malaria control programme which is aimed at tracking communities with high number of reported cases.

ACKNOWLEDGEMENT

I am grateful to the Almighty God for he is my source of wisdom, knowledge and power.

I wish to express my profound gratitude Dr F. T. Oduro, my supervisor, whose contributions towards the completion of this work are immeasurable.

Finally, I am grateful to Raymond Aabeyir and Frank Owusu for the encouragement and assistance given to when this material was being prepared.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS/ABBREVIATIONS

ACF……. Autocorrelation function

ACT……Artemisinin combined therapy

ADF …….Augmented Dickey-Fuller

AGA…….AngloGold Ashanti

AIC…….Akaike information criterion

AR ……..Autoregressive

ARMA……..Autoregressive moving average

ARIMA…..Autoregressive integrated moving average

BIC…….Bayesian information criterion

CDC……Center for Disease Control and Prevention

DF……..Degrees of freedom

GDP…….Gross domestic product

GIS……..Geographical Information System

IPTI……..Intermittent Preventive Treatment in Infant

MAPE……Maximum absolute percentage error

PACF……Partial autocorrelation function

RBM……..Roll back malaria

SE……..Standard Error

UNICEF….United Nations Children's Fund

WHO……World Health Organization

CHAPTER ONE

INTRODUCTION

1.1 BACKGROUND TO THE STUDY

Malaria is a disease caused by a protozoon called plasmodium. Children under five years and pregnant women are vulnerable to this disease in Ghana and Africa as a whole. One million people die of malaria each year, mostly in sub-Sahara Africa (Wiseman et al., 2000). The disease is transmitted to humans through the bite of mosquitoes.

The effects of malaria go beyond mortality and morbidity as malaria endemic areas suffer dearly in terms of human productivity and economic loss. AngloGold Ashanti is a mining company operating in Obuasi and in some other countries in Africa. It realized that malaria posed threats to its operations in East and West Africa. This was clearly shown in increased in morbidity, mortality and absenteeism in the workforce as well as decrease in productivity (AngloGold Ashanti Report to Society 2007).

The situation was not different in the company's operational area in Obuasi. In April 2004, the company realized that an estimated average of 11,000 malaria cases per month was recorded according to the municipal health authority. An additional 6,800 cases were reported by the mine medical service. At any point in time, 20% of the workforce had malaria and the average time off work due to malaria ailment was between two and three days. Extending the trend to the whole Obuasi community clearly indicated that the disease had gained ground and therefore called for actions to be taken to reverse the increasing trend of the disease (AngloGold Report to Society 2004).

The company then decided to implement a multi-million dollar malaria control campaign to combat the menace. Prior to the onset of the campaign, Professor Maureen Coetzee of the National Institute of Communicable Diseases in South Africa and Professor Richard Hunt, an entomologist, were contracted to identify the resident mosquito vector species and possible insecticide resistant patterns to the species. Two dominant anopheles mosquito species were identified, namely *funestus* and *gambiae*. Laboratory investigation revealed that these species were infected with the malaria parasite, *plasmodium falciparum.* The study also revealed that the species were completely or partially resistant to three of the standard insecticides approved by the World Health Organization (WHO) for use in controlling malaria. However they were susceptible to organophosphate insecticide. The company therefore, adopted an integrated malaria control programme since the use of any of the WHO measures would not be effective in controlling the species. Just before the programme was implemented, a baseline community prevalent study was done at the Noguchi Research Institute in Accra. This study turned out to be a follow-up to assess the success or failure of the programme (AngloGold Report to Society 2004).

The fully implemented programme consisted of the following activities; vector control, disease management, disease surveillance and monitoring, information dissemination, education and communication (AngloGold Ashanti Report to Society, 2004).

To control the vector, houses were sprayed with organophosphate. The indoor residual spraying involved the spraying of every item on every inch of space including interior walls, ceiling and underside of furniture at homes, offices, churches, schools and village huts. Every room within 110mile radius of the municipality was sprayed. The use of insecticide treated nets was promoted.

Figure 1.1: Spraying Men at Work

(Source: AngloGold Ashanti Report to Society 2007)

The nets were given out to orphanages, maternity homes and the children's ward of the two main hospitals in the municipality. On the disease management, effective treatment protocols such as the mandatory prescription of artemisinin combined therapy which comply with the national guidelines were introduced since the parasite has developed resistant to chloroquine and has made it ineffective in the treatment of the disease (AngloGold Ashanti Report to Society, 2007).

To monitor the programme, Geographical Information System (GIS) was established. The database contained houses that were covered by the spraying exercise, insecticide usage, insecticide treated nets distribution and usage, case detection breeding sites, drug resistance, disease outbreak foci, insecticide resistance, house screening, and larval surveys.

Medical staff and spraying teams were also trained to disseminate information to people on malaria prevention, diagnosis and treatment. This was supported by the provision of educational material such as pamphlets, posters and videos on malaria (AngloGold Ashanti Report to Society, 2007).

There is therefore the need for finding a suitable time series model for forecasting the monthly reported cases malaria in the municipality. The developed model will help management of the programme to allocate adequate resources to efficiently control the disease (AngloGold Ashanti Report to Society, 2007).

1.2 STATEMENT OF THE PROBLEM

Malaria has become a financial burden on the health sector in Ghana. During the launched of the Nationwide Malaria Control Programme by Zoomlion Ghana Limited at Koforidua in 2009, the minister of health, Dr Sipa Yankey, said that it cost the country $760 million to treat malaria and the disease accounted for 10% of the country's GDP.

There is evidence that North America and Europe have succeeded in eradicating the disease completely because they implemented effective programmes.

Several control programmes – Intermittent Preventive Treatment in Infants (IPTI) by UNICEF, Roll Back Malaria (RBM) and Integrated Malaria Control Programme by AngloGold Ashanti have been initiated in Ghana to treat, control and eradicate malaria but its prevalent rate is still high. The current high prevalent rate of malaria puts the success of these programmes in doubt.

Several control programmes have been initiated to control and eradicate the disease but its prevalent rate is still high in Ghana putting the success of these programmes in doubt. Malaria control is still ineffective in the endemic areas in spite of combined anti-malaria therapies and insecticide-treated materials. The AngloGold Ashanti Integrated Malaria programme has been running since 2006 without any rigorous analysis of its impact on the reported cases of malaria in its operational area. In the assessment of the programmes like this, very often, policy makers rely on the difference between the base year and any other year within the operational period. The magnitude of the difference alone is deceptive because it fails to tell the seriousness with which the impact should be treated.
This study aims to contribute to a better understanding of the impact of the AngloGold Integrated Malaria Control Programme on the reported cases of malaria and to provide a scientific evidence of its success or failure for policy makers, researchers and health workers in the development of public health strategies in order to control the disease effectively.

A mathematical tool is used to find out whether there is significant impact of the programme on the reported malaria cases over the years (i.e. from January 2005 to May

2010). The average monthly reported cases of the disease before and after the intervention would be compared to find the effectiveness of the intervention. A suitable time series model would be formulated. The model which would be formulated and would help in finding the general trend of the malaria reported cases. Modeling the reported cases through correlated models, such as time series would allow greater part of data behavioural dynamics to be adjusted into a single equation and future malaria reported cases would be estimated based on this. This would benefit the municipality in areas of planning, reserving resources and performing more efficient and timely control of the disease.

1.3 OBJECTIVES OF THE STUDY

Based on the statement of the problem, the objectives of the study are as follows;

1.3.1 General Objective

To assess the performance of the AngloGold Ashanti Malaria Control Programme in reducing malaria within the Obuasi Municipality and to develop a model for predicting reported cases of malaria in the Obuasi municipality

1.3.2 Specific objectives

i) To investigate effect of AngloGold Ashanti Integrated Malaria Control Programme on reported malaria cases in Obuasi municipality.

ii) To develop a time series model for predicting monthly reported malaria cases in Obuasi municipality.

iii) To identify ways of improving on the activities of the malaria control programme.

1.4 RESEARCH QUESTIONS

The study sought to find answers to the following questions;

i) What is the effect of the AngloGold Ashanti Integrated Malaria Control Programme on the number of reported malaria cases?

ii) How best can the monthly reported malaria cases in the Obuasi Municipality be predicted?

iii) What must be done to improve on the effectiveness of the programme?

1.5 JUSTIFICATION OF THE RESEARCH

Several programmes have been initiated in this country to combat this disease of the poor but its prevalent rate is still high and accounts for 40% out of the 70% communicable disease in Ghana in 2008. The National Malaria Control Programme (NMCP) is the mother agent for controlling the disease in this country. Other companies like Zoomlion Company Limited and AngloGold Ashanti have also joined in the fight against the disease.

Considering the concomitant loss of lives, cost in the medication and loss of productive hours, it requires renewed commitment from the government, non-governmental organizations and all and sundry to fight for complete eradication of the disease.

The programme was launched by AngloGold Ashanti (AGA) on 28th April 2006. The indoor-residual spraying exercise has been replicated seven times, covering the infrastructures of the mining communities and their environs. Since the launching of the

programme, the indoor spraying exercise is carried out twice a year, February to April and July to December. The company is also replicating the programme at its other operational areas with Obuasi as its headquarters. The programme has a centre in Sansu, a suburb of Obuasi, with insectary and laboratory, planning and strategy centre and training facilities (AngloGold Ashanti Report to Society, 2007).

The company is claiming that it has succeeded in achieving its objective of reducing the malaria reported cases in the municipality. If the claim of the company is true, then the National Malaria Control Programme must adopt this approach and replicate it in the other parts of the country since it cost less to prevent the disease than to treat it. It therefore calls for verification of the effectiveness of the Integrated Malaria Control Programme adopted by the company.

The ability to predict would provide a mechanism for the authority of the programme, government and health-care services to respond to outbreak in timely fashion and minimizing its effect. At the end of the research, the programme would have to be repackaged or recommended for adoption by the National Malaria Control Programme and other non-governmental organizations to help control the disease.

This study is also aimed to find out whether the malaria control programme has yielded good results or not. The average monthly reported cases of the disease before and after the launching programme would be compared. A model will be developed for predicting the time series monthly reported cases of malaria in the municipality.

1.6 METHODOLOGY

The secondary data of the reported malaria cases from January 2005 to June 2010 is obtained from the municipal health directorate for the analysis. Data from January 2005 to January 2009 would be used for comparing the mean monthly reported cases. The malaria programme was launched after 2005 so that year is taken as the based year and its mean reported cases would be compared with the mean reported cases of the other years (2006, 2007, 2008 and 2009) after the implementation of the programme to measure the effect. The two-sample t-test would be used for comparing the means of the reported cases because the sample size is less than 25. Minitab 14.1.0.0 is used to compare average monthly reported cases.

Data from January 2006 to April 2010 would be used for developing a time series ARIMA model for predicting the number of reported cases in the municipality. Values for the months May and June 2010 would be reserved for post time series forecast. An SPSS version 16.0 is used in the formulation of the ARIMA model. Minitab is used to plot the graphs. The least square regression analysis is used to determine the trend equation of the reported cases after the 2005. Minitab software is used to determine the trend equation.

Questionnaires were issued to the four main hospitals in the municipality namely, Government Hospital, St Jude Hospital, Bryant Mission Hospital and AGA Hospital to get their views on the effectiveness of the programme. A questionnaire would be issued to the authority of the malaria programme to find out the challenges they are facing.

1.7 LIMITATION

Most of health facilities could not provide data on the monthly reported cases of malaria as far back as 2004. Some complained of lost data because their systems had been formatted and they had no back-ups. The data was obtained from the municipal health directorate. Though the research area is the entire municipality, only the health facilities which sent their morbidity reports to the directorate were covered because most of the other health facilities did not have data on malaria even as far back as 2004. Even those who had it, a long bureaucratic process had to be followed in order to get it.

1.8 ORGANZATION OF THE THESIS

Chapter one is the introduction which comprises the background to the problem, statement of the problem, objectives of the study, research questions, justification of the research, methodology and limitation.

Chapter two basically deals with the review of literature. Literature on malaria and some malaria interventions programme are reviewed. Some people's literature on time series and time series modelling.

Chapter three deals with the methods used in the analysis. It composes of the introduction, study area, describing basic statistics, method and formula of the two-sample t-test and confidence interval, concept of time series and Box-Jenkins methodology.

Chapter four is data analysis and results. It comprises the introduction, descriptive

statistics of the reported cases, the two-sample t-test and confidence interval, trend analysis, data preparation, model selection, parameter estimation, model diagnosis, forecasting with the ARIMA model, in-sample period forecasts and post sample forecast.

Chapter five deals discussion of the results which consists of summary of results, findings, conclusion and recommendations.

CHAPTER TWO

LITERATURE REVIEW

2.0 INTRODUCTION

A lot of people have done research on malaria. This chapter deals with the review of some literature works on malaria. Other peoples' work on autoregressive model would also be reviewed.

2.1 MALARIA

Malaria has been one of the most prominent and ancient diseases which has been profiled and studied. It has been one of the greatest burdens to mankind, with a mortality rate that is unmatched by any other modern disease other than tuberculosis (Sudhakar et at., 2007). It remains the leading cause of death in children under five years in Africa (Houeto et al., 2007).

Malaria is one of the leading killer diseases in the tropical and subtropical countries. It therefore poses a serious health problem to these countries including Ghana. This disease is frequently called disease of the poor because its prevalent rate is very high in the poorest continent and in the poorest countries (Worral et al., 2003).

Malaria is caused by a tiny parasite called plasmodium found in female mosquitoes called anopheles. Symptoms of malaria are feverishness accompanied by chills, headache, shivering and pains in the joints. There are four species of plasmodium. These are *plasmodium falciparum, plasmodium vivax, plasmodium malariae* and *plasmodium*

ovale. Infection with the *plasmodium falciparum* can lead to life threatening complications. This parasite is usually resistant to chloroquine and must be treated with other medications and the other three types of malaria are not life threatening and a person infected with any of them may recover in a month without treatment. The vector breeds in water habitat and any change in the environment affects it.

According to Houeto et al., (2007) successes have been made in the areas of prevention and treatment through the adoption of artemisinin combined therapy (ACT). The use of insecticide treated nets has helped to reduce morbidity and mortality rates. But the disease still poses a threat to our part of the world.

According to Nchinda, (2005) sub-Sahara Africa was never part of the global malaria eradication programme because the period coincided with colonial and immediate postcolonial period and so the indigenous had little or no power to initiate and sustain an eradication programme.

This review looks at the factors affecting the prevalent rate of this dreadful disease and some of the control measures which have already been initiated to control the disease. There are combination of factors which contribute to the resurgence of this disease in Ghana and Africa as a whole.

Poverty contributes to the prevalence of malaria in Africa. According to Pattanayak et al., (2003) many of the world's poorest people live in areas of high rates of malaria. These

people do not have access to effective health care due to financial constraint. Worral et al, 2003 called malaria as a disease of poverty. The economic status of a vulnerable country plays another role in determining the equippedness and control measures in case of epidemics (Sudhakar et al., 2007).

A survey in Zambia found a substantially higher prevalence of malaria infections among the poorest population group. Poverty compels people to move from non-endemic areas to endemic areas in their quest to search for jobs. As with other diseases, malaria has unequal effect on different members of the population; pregnant women and children are most susceptible. Given that the intensity of malaria transmission and therefore the likelihood of control depends on the relative abundance of and contact patterns among susceptible, infected, infectious and immune individuals, it is essential to target mothers and children in treatment (Guerin et al., 2002). The immediate economic burden on households from losing mothers is devastating, whereas childhood malaria imposes future burdens (Pattanayak et al., 2003).

Another contributing factor to the high morbidity of malaria is altitude. Research shows that there is high prevalent rate of malaria in low altitude areas and the prevalent rate is low in high altitude areas. According to Wiwanitkit 2006, a previous research on altitude and malarial mosquito prevalence in Thailand indicated high prevalence of the disease in low altitude. However people living in high altitude areas may experience high prevalent rate of malaria if they create the enabling environment for the breeding of the vector.

Environmental factors hinder efforts to control the disease. Inhabitants of houses surrounded by bushes or garbage heaps and swamps or stagnant water showed higher malaria prevalence and densities as compared with those from cleaner surroundings (Nkuo-Akenji et al, 2006). Irrigation and deforestation have affected the transmission of the disease. Irrigation requires the construction of dams which serve as fertile breeding grounds for the parasite. In Sri Lanka, the construction of hydropower dams on Mahaweli River created pools with sandy and rocky nature, which are suitable for the breeding of anopheles culicifacies, the primary malaria vector for the country (Nkuo-Akenji et al., 2006). Deforestation changes the ecology of the vector and its option for the host (Pattanayak et al., 2003). Whereas the forest floor in primary growth tends to be heavily shaded and littered with a thick layer of organic matter that absorbs water and renders acidic, clear lands are generally more sunlit and prone to the formation of puddles with more neutral pH, which can favour specific anopheline larvae development (Patz et al., 2000).

According Lindsay et al., (2004), deforestation is one of the most potent factors at work in emerging and re-emerging of infectious diseases. Mining causes deforestation and environmental degradation. Mining pits dug during land dredging mining creates stagnant water pools serving as breeding ground for mosquitoes and other water-borne diseases (Wiwanitkit, 2009). In Kanchanaburi, Thailand the primary forest malaria vector, An. dirus increased mainly because breeding places were created by excavation work (Wiwanitkit, 2006).

In sub-Sahara Africa, climate change has several features that could influence the prevalent rate of malaria. Rising temperature can extend the habitat of mosquitoes, shifting the boundaries of latitude and altitude for malaria transmission. Highland areas in Burundi, Kenya and Uganda which initially were malaria-free are now experiencing epidemics (Sulaiman, 2007). Floods and drought also have impact on the incidence of malaria. Drought leads to the formation of pool of stagnant water which creates a favourable habitat for the parasite. Relative humidity affects the transmission of malaria. It affects the survival of the vector.

The resistance of the parasite to antimalaria drugs especially chloroquine is a major cause in the reemerging of the disease. Resistance evolves through fundamental principles of natural selection and evolution, including diverse factors such as extent of treatment, nature and site of antibiotic action or genomic complexity of the parasite (Wilson, 2001). Perhaps the biggest threat to malaria control- be it prevention or treatment- is the increasing resistance to pesticides and drugs. Optimal control and treatment maximize the useful life span of insecticides and drugs. Resistance is more likely to emerge when background immunity is weak, parasite numbers in individuals are high, transmission is low, and insecticides and drug pressure is intense. Plasmodium falciparum has become variably resistant to all drug classes except the artemisinin derivatives. Multiple economic factors influence the inappropriate use of drugs and insecticides (Reed et al., 2002).

Since the discovery of the disease about 4,000 years ago, several control measures have been put in place to curb it but the incidence of the disease is still high in sub-Sahara Africa. The United States and some Europeans countries have been able to eradicate it through the use of insecticide and manipulation of the environment (Nkuo-Akenji et al., 2006). United States launched the National Malaria Eradication Program on 1st July 1947. Over 4,650,000 houses were sprayed by the end of 1949. In 1947, the malaria cases reported were 15,000 and reduced to 2,000 in 1950. The disease was considered eradicated in 1951.

The World Health Organization launched the global eradication of malaria in 1955. Unfortunately, this coincided with the struggle for independence in sub-Sahara Africa. Countries with temperate climates succeeded in eradicating the disease. Countries like India and Sri Lanka had sharp reduction in morbidity. However countries like Indonesia, Afghanistan, Haiti and Nicaragua made negligible progress (CDC, 2004).

The Roll Back Malaria (RBM) initiative launched in 1998 has the ambitious target of decreasing malaria mortality by 50% by the year 2010. Although several control and preventive measures will contribute to the achievement of this target, an essential contribution needs to come from a substantial reduction of the case-fatality rate for the disease (WHO, 2002). African leaders met in Abuja in 2000 to reaffirm their commitment to the RBM. The goals of the Abuja Declaration include ensuring that 60% of those with malaria have access to treatment within 24 hours of the onset of the symptoms; at least 60% of the at-risk pregnant women receive preventive drugs and at

least 60% of the at-risk sleep under bed nets (The African Summit on Roll Back Malaria, 2005). After nine years of implementing the Abuja Declaration, it appears Ghana is making a negligible success since the morbidity rate is still high and in Obuasi malaria is still the most common disease recorded daily at the health centres. The ability to prevent and treat the disease is a function of one's income so the issue of affordability in terms of treatment and the acquisition of the nets should not be downplayed if modest gain is to be achieved different from the others.

2.2 TIME SERIES AND MALARIA STUDIES

A mathematical model can help response to the increasing threat of malaria in the municipality. The man who first discovered that malaria is transmitted though mosquitoes, Sir Roland Ross, developed the first mathematical model for malaria transmission in 1911. In presenting his model Ross pointed out that "the mathematical method of treatment is really nothing but the application of careful reasoning to the problem at issue".

AR, MA, ARMA and ARIMA models are some of the fundamental models for forecasting time series data. The MA model assumes that the series depends linearly on its previous values and normally distributed error term. The autoregressive, AR part of the model attempts to forecast the time series values based on historical data. The moving average component models the error terms which are correlated. Non-stationarity can be decreased by differencing the series with specific time lag.

The main objective of a time series is to develop statistical model explaining the behavior of a random variable changing over time which allows making future estimations of the said random variable (Suarez et al., 2009). Analysis of time series may serve a number of purposes. Often the main interest lies in the regressive model, for example relating infection incidence to staffing levels or antibiotic usage data. Other applications include forecasting and the development of the alert systems to detect periods or places where transmission exceeds some threshold (Brown et al., 2002).

In the United States, mathematical models are familiar, everyday tools in engineering, business and military applications in most sciences. They represent hypotheses about underlying mechanism that generate observed phenomena or options for action and potential consequence (McKenzie, 2004).

Regarding planning in future needs in any system, traffic prediction accuracy is really important when defining required future capacity and planning any changes. A fairly accurate time series model could predict several years in the future, this being an advantageous skill when planning future requirements (Fillatre et al., 2003).

Lin et al., (2009) used time series analysis to investigate the relationship between the falciparum malaria in the endemic provinces and the imported malaria in the non-endemic provinces of China. An autoregressive integrated moving average model was first fit to the predictor variable. Of all the models tested, the seasonal ARIMA (1, 1, 1)

$(0, 1, 1)_{12}$ model for malaria incidence fits the data best according to the according to AIC and goodness-of -it criteria.

Briet et al., (2008) formulated a model for short term malaria prediction Sri Lanka. Exponentially moving average models, autoregressive integrated moving average models with seasonal components and seasonal multiplicative autoregressive integrated moving average (ARIMA) models were compared on monthly time series of district malaria cases for their ability to predict the number of malaria cases one to four months ahead. The best model for forecasting and forecasting error varied strongly among the districts for instance, for the district of Ampara, for a one month forecasting horizon, the best model was an ARIMA (2, 1, 1) with seasonality through a harmonic with a period of one year and a harmonic with period of six months. For further forecasting horizons, the ARIMA (0, 1, 2) model with seasonality through a first order seasonal autoregressive and a first order seasonal moving average component was best for the district of Ampara.

Contreras et al., (2003) developed a model for predicting the next-day electricity prices in mainland Spain and California markets using an ARIMA model. Their developed model was able to forecast the 24 market clearing prices of tomorrow. The ARIMA model is an effective tool for forecasting time series.

A good model is to be developed for forecasting the malaria cases. A model fitting quality is defined as the sum of the residuals' squares divided by the sample size. Its

objective is to measure the model's capacity to produce the sample data (i.e. to verify how similar the modelled series and the actual series really are) (Guerrero, 2003).

CHAPTER 3

METHODOLOGY

3.1 INTRODUCTION

This chapter deals with the study area, method and formula of two-sample t-test, concept of time series, preliminary test and the Box-Jenkins methodology for modelling time series

3.2 STUDY AREA

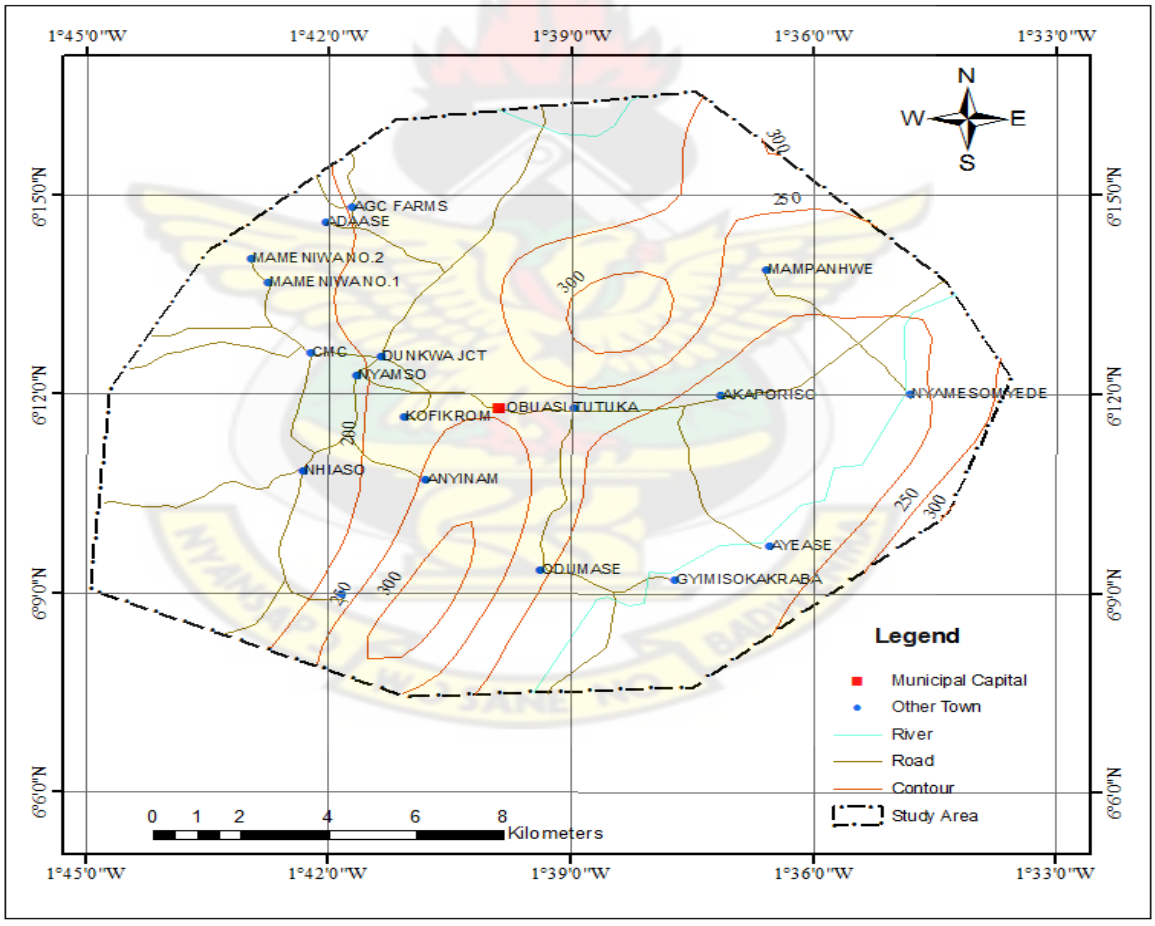

Figure 3.1: Map of study area

The study area covers the Obuasi township and some nearby communities surrounding

it as shown in figure 3.1 above**.**

Questionnaires were given out to the four main hospitals (Government Hospital, AGA Hospital, St Jude Hospital and Brilliant Mission Hospital) to seek their views about the effectiveness or otherwise of the malaria control programme. A questionnaire was also sent to the authority of the AngloGold Integrated Malaria Programme to obtain information about the activities of the programme.

3.3 DESCRIBING BASIC STATISTICS OF REPORTED MALARIA CASES FROM 2005 TO 2009

The data from January 2005 (base year) to January 2010 is entered into a Minitab spread sheet and analysed to compare the means of the monthly reported cases. The descriptive statistics give information about the mean monthly reported cases of each year, the minimum and maximum number of reported cases each year. This gives a foreknowledge of the impact of the programme.

3.4 METHOD AND FORMULA OF TWO-SAMPLE T-TEST AND CONFIDENCE INTERVAL

The two-sample t-test is used for comparing two population means when the small size is less than 25. The differences between mean of reported cases before the commencement of the Integrated Malaria Control Programme (i.e. mean of the monthly reported cases in 2005) and means of the reported cases after the commencement of the programme (i.e. 2006, 2007, 2008 and 2009) are compared to find out whether the programme is having positive, negative or no effect on the reported cases.

Let $\mu_1, \mu_2, \mu_3, \mu_4$ and μ_5 denote the monthly reported cases in 2005, 2006, 2007, 2008 and 2009 respectively. Thus $\mu_1 - \mu_2, \mu_1 - \mu_3, \mu_1 - \mu_4, \mu_1 - \mu_5$. The differences between the means over the periods of the implementation of the programme are computed using Minitab package. Minitab uses the method illustrated below to compare two population means and construct a confidence interval.

The null and alternative hypotheses for a two-sample t-test are:

$H_0 : \mu_1 - \mu_2 = 0$

$H_1 : \mu_1 - \mu_2 \neq 0$

Where μ_1 = the mean for the first population and

μ_2 = the mean for the second population

The test statistic is $t = \dfrac{(\bar{x}_1 - \bar{x}_2)}{s}$,

Where $\bar{x}_1$ and $\bar{x}_2$ are sample means. The sample standard deviation s depends on the variance assumption. Assuming equal variance, the common variance is estimated by the pooled variance:

$$s_p = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

The standard deviation of $(\bar{x}_1 - \bar{x}_2)$ is estimated by

$$s = s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

The test statistic degrees of freedom are $n_1 + n_2 - 2$. The null hypothesis is rejected if the

calculated value is greater than the table value or for smaller p-value, the null hypothesis is rejected at α=5% level of significance.

(100-α)% confidence interval on μ_1-μ_2 is given by

$$\left(\bar{x}_1 - \bar{x}_2\right) - t_{\frac{\alpha}{2}} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \le \mu_1 - \mu_2 \le \left(\bar{x}_1 - \bar{x}_2\right) + t_{\frac{\alpha}{2}} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}.$$

with $n_1 + n_2 - 2$ degrees of freedom.

Minitab is used in analysing the data.

3.5 CONCEPT OF TIME SERIES

Time series is the set of observations on a variable of interest that has been collected in time order that is, daily, weekly, monthly, etc. If the time series can be predicted exactly, it is said to be deterministic for example a person's salary may be determined according to the number of years worked but most time series are stochastic in nature in that the future values are determined based on the past values.

3.5.1 COMPONENTS OF TIME SERIES

Traditional time series are mainly concerned with decomposition. All time series contain at least one of the following four components: trend, cyclical, seasonal and irregular variations.

The trend refers to the upward and downward movement that characterizes the time series over a period of time. That is, the trend reflects the long term growth or decline in the time series.

The seasonal variations of a time series are the periodic patterns and they complete themselves within one calendar year are then repeated on yearly bases. Seasonal variations are usually caused by factors such as custom and weather. Examples are the abundance of fish between July and August every year in Ghana. Also the population of the Kwahu's areas increases during Easter period every year.

The cyclical component is a wavelike variation in the general level of business activities over a relatively long period say 2-10 years. A cycle is characterized by growth or expansion, peak, recession (contraction) and depression.

The irregular component of a time series is the residual factor that accounts for the deviations of the actual time series value from what we would expect if the trend, cyclical and seasonal components completely explain the time series . It is caused by a short term unanticipated and non-recurring factors such as wars, earthquakes, floods and so that affect the time series. Since this component accounts for the random variability in the time series, it is unpredictable that is we cannot predict its impact on the time series in advance.

Time series can be expressed as a combination of the four components. Two types of models are usually associated with time series namely additive model and the multiplicative model. The additive model is

$Y_t = T_t + S_t + C_t + I_t$,

where Y_t = the value of the series at time t,

T_t = the value of the trend at time t,

S_t = the value of the cyclical component at time t,

C_t = the value of the seasonal component at time t,

I_t = the value of the irregular component at time t.

All values are usually expressed in the original units.

The drawback of the additive model is the unrealistic assumption of independent of components.

The multiplicative model is

$Y_t = T_t \times S_t \times C_t \times I_t$

T_t is in the original unit while S_t, C_t and I_t are in percentages.

3.5.2 TREND ANALYSIS

A trend line can be estimated by using a regression technique where the response variable is the time series and the predictor is the time. A forecast is calculated by inserting a time value into the regression equation. The regression equation is determined from the time series data using the least squares method.

The general model is given by

$Y_t = TR_t + \varepsilon_t$, where

Y_t = the value of time series in period t,

TR_t = the trend in time period t, and

ε_t = the error term in time period t

3.5.3 STATIONARITY OF TIME SERIES

A time series is said to be stationary if there is no systematic change in mean, variance and strictly periodic variations have been removed. There are two types of stationarity. These are strict stationary time series and weakly stationary time series. If a time series is strictly stationary, it means that the distribution of the process is invariant against the shift of time.

The concept of stationarity is important when analysing time series. The random variables' joint density function must usually be known to fully characterise a stochastic process; however, in practice, it is not realistic to think that this can be achieved with a time series. As previously mentioned regarding covariance, there is no dependence on time but on separation (k) between variables. This led to thinking that the series would display the same general behaviour, irrespective of observation time. This meant that if a number of a series of contiguous observations were to be plotted, the graph obtained would be quite similar to the graph obtained when plotting the same number of contiguous observations but k periods forward or backward respecting the initially considered terms (Brillinger, 2002).

Strictly stationary is also called first order stationarity. For weakly stationary time series, the mean and variance are constant and finite. Not all time series are stationary. Most of the probability theorems of time series is concerned with stationarity and for this reason, time series analysis often requires that we turn non-stationary time series into stationary time series. Lagging and differencing are employed to make non-stationary time series stationary. They remove the element of variation. The lag operator L is defined for a time series $\{Y_t\}$ by $LY_t = Y_{t-1}$

The operator can be defined for linear combinations

$$L(c_1 Y_{t_1} + c_2 Y_{t_2}) = c_1 Y_{t_1 - 1} + c_2 Y_{t_2 - 1}$$

In general,

$$L^k Y_t = Y_{t-1}$$

The differencing operator is defined by

$$\nabla Y_t = (1 - L) Y_t = Y_t - Y_{t-1}$$

$$\nabla^2 Y_t = \nabla(\nabla Y_t)$$

$$\nabla^2 Y_t = \nabla(Y_t - Y_{t-1})$$

$$\nabla^2 Y_t = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2})$$

This operator is very useful when dealing with non-stationary time series.

3.5.4 AUTOCORRELATION FUNCTION

The Autocorrelation function (ACF) of a stationary stochastic process is an important tool for assessing its properties. The autocorrelation coefficient is a measure of the correlation between observations at different distance apart. The ordinary correlation coefficient between two variables x and y is given by

$$r = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{\sqrt{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2\left(y_i - \bar{y}\right)^2}} \qquad (3.1)$$

given n observations , $x_1, x_2, ..., x_n$, we can form n-1 pairs of observations namely,

(x_1, x_2), (x_2, x_3), …, (x_{n-1}, x_n). Regarding each observation in each pair as one variable, the correlation coefficient between x_t and x_{t+1} is given

$$r = \frac{\sum_{i=1}^{n-1}\left(x_t - \bar{x}_1\right)\left(x_{t+1} - \bar{x}_2\right)}{\sqrt{\sum_{t=1}^{n-1}\left(x_t - \bar{x}_1\right)^2 \sum_{t=1}^{n-1}\left(x_t - \bar{x}_2\right)^2}} \qquad (3.2)$$

Where $\bar{x}_1 = \frac{1}{n-1}\sum_{t=1}^{n-1} x_t$ and $\bar{x}_2 = \frac{1}{n-1}\sum_{t=1}^{n-1} x_{t+1}$

As the coefficient given by (3.2) measures correlation between successive observations, it is called autocorrelation coefficient. Assuming that $\bar{x}_1 \approx \bar{x}_2 \approx \bar{x} = \frac{1}{n}\sum_{t=1}^{n} x_t$ because of stationarity then (3.2) becomes

$$r = \frac{\sum_{t=1}^{n-1}\left(x_t - \bar{x}_1\right)\left(x_{t+1} - \bar{x}_2\right)}{\sqrt{\sum_{t=1}^{n-1}\left(x_t - \bar{x}\right)^2 \sum_{t=1}^{n-1}\left(x_t - \bar{x}\right)^2}}$$

$$r = \frac{\sum_{t=1}^{n-1}\left(x_t - \bar{x}_1\right)\left(x_{t+1} - \bar{x}_2\right)}{\sqrt{\left(\sum_{t=1}^{n-1}\left(x_t - \bar{x}\right)^2\right)^2}}$$

$$r = \frac{\sum_{t=1}^{n-1}\left(x_t - \bar{x}_1\right)\left(x_{t+1} - \bar{x}_2\right)}{\sum_{t=1}^{n-1}\left(x_t - \bar{x}\right)^2} \qquad (3.3)$$

Taking the denominator and simplifying,

$$\sum_{t=1}^{n-1}\left(x_t - \bar{x}\right)^2 = \sum_{t=1}^{n-1}\left(x_t^2 - 2x_t\bar{x} + \bar{x}^2\right)$$

$$\sum_{t=1}^{n-1}\left(x_t - \bar{x}\right)^2 = \sum_{t=1}^{n-1}x_t^2 - (n-1)\bar{x}^2$$

$$\sum_{t=1}^{n-1}\left(x_t - \bar{x}\right)^2 = (n-1)\left(\frac{\sum_{t=1}^{n-1}x_t^2}{n-1} - \bar{x}^2\right)$$

Replacing $(n-1)$ in the brackets with n,

$$\sum_{t=1}^{n-1}\left(x_t - \bar{x}\right)^2 = (n-1)\left(\frac{\sum_{t=1}^{n}x_t^2}{n} - \bar{x}^2\right)$$

$$\sum_{t=1}^{n-1}\left(x_t - \bar{x}\right)^2 = \frac{(n-1)}{n}\left(\sum_{t=1}^{n}x_t^2 - n\bar{x}^2\right)$$

Equation (3.3) thus becomes

$$r = \frac{\sum_{t=1}^{n-1}\left(x_t - \bar{x}_1\right)\left(x_{t+1} - \bar{x}_2\right)}{\frac{n-1}{n}\sum_{t=1}^{n}\left(x_t - \bar{x}\right)^2}$$

As $n \to \infty$, this implies $\frac{n-1}{n} = 1$. Hence

$$r = \frac{\sum_{t=1}^{n-1}\left(x_t - \bar{x}_1\right)\left(x_{t+1} - \bar{x}_2\right)}{\sum_{t=1}^{n}\left(x_t - \bar{x}\right)^2} \quad (3.3)$$

Correlogram is graph of autocorrelation coefficients r_k plotted against the lag k. It is useful in interpreting the correlation coefficient. The autocorrelation coefficients of a stationary time series drop to zero after the second or third time lag whilst for a non-stationary series or data they significantly differ from zero for several time periods. When represented graphically, the autocorrelation of a stationary data show a trend going diagonally from right to left as the number of time lag increases.

The autocorrelation function is used for model identification. If the autocorrelation function is exponential, decaying to zero the then the model is autoregressive and the partial autocorrelation plot is used to identify the order. If it is alternating positive and negative, decaying to zero then the model is autoregressive and the partial autocorrelation plot is used to identify the order. If one or more spikes and the rest are essentially zero, then the model is moving average and the partial autocorrelation plot is used to identify the model. The model is a mixed autoregressive and moving average model if the autocorrelation function decays after a few lags.

3.5.5 MOVING AVERAGE (MA) PROCESS

Suppose $\{\varepsilon_t\}$ is a purely random process with mean zero and variance σ^2, a process $\{y_t\}$ is said to be a moving average process of order denoted by MA (q) if

$$y_t = \theta_0\varepsilon_t + \theta_1\varepsilon_{t-1} + ... + \theta_q\varepsilon_{t-q},$$

Assume $\theta_0 = 1$,

$$y_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + ... + \theta_q\varepsilon_{t-q},$$

where θ_i are constants, the mean $E(y_t) = 0$ and $\text{var}(y_t) = \sigma^2\sum_{i=0}^{q}\theta_i^2$

The auto covariance function r (k) for the MA(q) process is

$$r(k) = \text{cov}(y_t, y_{t+k})$$

$$r(k) = \begin{cases} \left(1+\theta_1^2 + ... + \theta_q^2\right)\sigma^2, k = 0 \\ (\theta_k + \theta_1\theta_{k+1} + ... + \theta_{q-k}\theta_k)\sigma^2, u = 1, ..., q \\ 0, otherwise \end{cases}$$

The autocorrelation function P(k) is given by

$$P(k) = \begin{cases} 1, k > 0 \\ \dfrac{\sum_{i=0}^{q-k}\theta_i\theta_{i+k}}{\sum_{i=0}^{k}\theta_i^2}, k = 1, 2, ..., q \\ 0, k > q \\ P_{-k}, k < 0 \end{cases}$$

For the lag operator $L^k y_t = y_{t-1}$,

MA (q): $y_t = \theta_0\varepsilon_t + \theta_1\varepsilon_{t-1} + ... + \theta_q\varepsilon_{t-q}$

$$y_t = \theta_0 L^0 \varepsilon_t + \theta_1 L^2 \varepsilon_t + ... + \theta_q L^q \varepsilon_t$$

$$y_t = \left(\theta_0 L^0 + \theta_1 L^2 + ... + \theta_q L^q\right) \varepsilon_t$$

If the roots of the equation $0 = (\theta_0 + \theta_1 L^2 + ... + \theta_p L^p)$ lie outside a unit circle, the MA process is invertible else it is not invertible.

3.5.6 AUTOREGRESSIVE (AR) PROCESS

Suppose that ε_t is a purely random process with mean zero and variance σ^2, then a process y_t is said to be an autoregressive process with order p if

AR (p): $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + \varepsilon_t$.

The autocorrelation function is given by the Yule-Walker equations.

These are sets of differential equations whose general solution is given by

$P_k = A_1 \pi_1^{|k|} + A_2 \pi_2^{|k|} + ... + A_p \pi_p^{|k|}$, where $\{\pi_i\}$ are the roots of

$y^p - \phi_1 y^{p-1} - ... - \phi_p = 0$,

It is clear that $P_k \to 0$ as $k \to \infty$ provided $|\pi_i| < 1$, $\forall i$ and this is necessary and sufficient. If the roots are real, then the autocorrelation function decay exponentially with k. If the roots are complex, the graph of the autocorrelation function exhibits oscillatory behaviour. An equivalent way of expressing the stationary condition is check if the roots of the equation

$\phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p = 0$, lie outside a unit circle

3.5.7 ARMA PROCESS

It is a class of models for time series formed from the combination of MA and AR processes. A mixed autoregressive moving average process containing MA(q) terms and AR(p) terms is said to be of the order (p, q). The is given by

$$y = \phi_1 y_{t-1} + ...\phi_p y_{t-p} + \theta_1 \varepsilon_{t-1} + ... + \theta_q \varepsilon_{t-q},$$

Using the backward shift operator,

$\phi(L) y_t = \theta(L)\varepsilon_t$ where, $\phi(L)$ and $\theta(L)$ are polynomial of order p and q respectively. Thus

$$\phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p,$$

$$\theta(L) = 1 - \theta_1 L - \cdots - \theta_q L^q,$$

If the roots of the equation, $\theta(L) = 1 - \theta_1 L - \cdots - \theta_q L^q$, lie outside the unit circle, then the ARMA (p, q) process is stationary.

3.5.8 INTEGRATED MODEL

In practice, most time series are non-stationary. In order to fit a stationary model it is necessary to remove non-stationary source of variation. If the observed time series is non-stationary in the mean, then we can difference the series. The differenced model is called an integrated model becomes the stationary model which is fitted to the difference data has to be summed or integrated to provide a model for non-stationary data. If X_t is the time series, then the differenced series can be written as $W_t = \nabla^d X_t$. The integrated model is of the form ARIMA (p, d, q)

$$W_t = \phi_1 w_{t-1} + ... + \phi_p w_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + ... + \theta_q \varepsilon_{t-q}$$

3.5.9 PARTIAL AUTOCORRELATION COEFFICIENT

Partial autocorrelation coefficient is used to measure the degrees of association between y_t and y_{t-k}. The singular purpose of partial autocorrelation coefficient in time series analysis is to identify an appropriate ARIMA model for forecasting. The autocorrelation function of an MA series exhibits different behaviour from that of AR and general ARMA series.

3.6 BOX-JENKINS METHODOLOGY

The Box-Jenkins methodology is used in modelling the time series. The pioneers who popularized an approach which combines the moving average and autoregressive models were Box and Jenkins. Although both autoregressive and moving average approaches were known (and were original investigated by Yule), the contribution of Box and Jenkins was in developing a systematic methodology for identifying and estimating models that could incorporate both approaches and this makes Box-Jenkins models a powerful class of models (Dobre et al., 2008).

There are five primary stages in building a Box-Jenkins time series model. These are model data preparation, model selection, parameter estimation, model checking and forecasting.

3.6.1 DATA PREPARATION

This involves inputting the secondary data of the reported malaria cases from January 2006 to April 2010, which is extracted form the municipal health directorate's morbidity report, into an SPSS 16.00 spread sheet for the analysis and performing a preliminary test. The data entered is cross-checked to avoid omissions and typographical error.

The preliminary test involves transformations and differencing. The first step after the data entry is to check whether the series is stationary or not. If the series is non-stationary, transforming the series by using square root, logarithm and so on may help stabilize the variance. The data can also be differenced to achieve stationarity. Differencing means taking the difference between consecutive observations. Box and Jenkins recommend differencing non-stationary one or more times to make it stationary. Original series over-differentiation must be prevented as well as deletion of valuable information that may arise from the autocorrelation function because in an over-differentiation case, autocorrelations become more complicated, the model loses parsimony, variance increases and observations are lost (Brockwell et al., 2002). The series plot and its autocorrelations and partial autocorrelations are used to determine whether the series is stationary or not. But to be sure of the stationarity, the Augmented Dickey-Fuller (ADF) test is used.

3.6.2 MODEL SELECTION

Model selection is based on the use of graphs of the autocorrelations and partial autocorrelations after transforming or differencing the data to identify potentially

ARIMA processes. Akaike Information Criterion (AIC) and the Bayesian Information Criterion are also used in selecting the best model. The model that produces smaller values of the AIC and Normalized BIC is selected as the best model. The idea of parsimony is that a good model has a few parameters as it has captured the properties inherent to the analyzed series; likewise, a complicated model with too many parameters is a model lacking parsimony (Suarez et al., 2009).

3.6.3 PARAMETER ESTIMATION

Parameter estimation means finding the values of the model coefficient that gives the best fit to the data. The SPSS is used to estimate the values of the coefficients or the parameters.

3.6.4 MODEL DIAGNOSIS

Model checking or diagnosis involves testing to find out if there is no violation of the assumptions of the developed model. If there is a violation in any of the assumptions, then it calls for selecting a new model.

3.6.5 FORECASTING

Forecasting is the main reason of formulating the time series model. The developed model should be able to forecasting the time series with minimal forecasting errors. According to Guerrero (2003), a model fitting is defined as the sum of the residuals squares divided by the sample size. Its object is to measure the model's capacity to

reproduce the sample data (i.e. to verify how similar the modeled series and the actual series are).

.

CHAPTER 4

DATA ANALYSIS AND RESULTS

4.1 INTRODUCTION

This chapter deals with the analysis of the data and display of results of the two-sample t-test and the results of the modelling process.

4.2 DESCRIPTIVE STATISTICS OF MONTHLY REPORTED CASES

The monthly reported cases from January 2005 to January 2009 were fed into a Minitab spread sheet and analysed to obtain the descriptive statistics. Let $\mu_1, \mu_2, \mu_3, \mu_4$ and μ_5 denote the monthly reported cases in 2005, 2006, 2007, 2008 and 2009 respectively. The table below showed the descriptive statistics of the reported cases.

Table 4.1: Descriptive Statistics of Reported Cases

Year	No. of Months	Mean	S.E. Mean	StDev	Min	Q1	Median	Q3	Maximum
2005	12	10073	444	1540	7443	9199	10189	10839	13510
2006	12	8634	428	1483	5954	7215	9217	9966	10010
2007	12	8714	378	1311	5681	7912	9041	9713	10014
2008	12	7898	233	804	6123	7437	8211	8514	8760
2009	12	6627	330	1110	4789	5916	6329	7906	8090

In 2005, the average monthly reported case was 10073; the least number was recorded as 7443 in September and the highest number was recorded in June as 13510. In 2006, the average monthly reported case was 8634; the least number was recorded in February as 5954. In 2007, 2008 and 2009, the average monthly reported cases were 8714, 7898 and

6627 respectively. Their respective highest reported cases were 10014 (in April), 8760 (in January) and 8090 (in March).

4.3 TWO-SAMPLE T-TEST AND CONFIDENCE INTERVAL

The two-sample t-test was used to compare the sample means of reported cases of malaria between the base year (the year the programme was not begun) and years of implementation of the programme. Minitab package was used in this analysis and the results are displayed in appendix B.

4.3.1 Two-Sample T-Test and Confidence Interval for Monthly Reported Cases in 2005 and 2006

Assuming equal variances, the null hypothesis and the alternative hypotheses stated are as follows;

$H_0 : \mu_1 - \mu_2 = 0$

$H_1 : \mu_1 - \mu_2 \neq 0$

From the Minitab output in appendix B, the difference between the average number of reported cases in 2005 and 2006 was 1439. The difference was then tested to find out whether it was significant or not. The T-value (calculated value) was found to be 2.330 and the p-value=0.029. Since the p-value was less than 0.05 level of significance, the null hypothesis was then rejected and concluded that the difference in the average monthly reported cases between 2005 and 2006 was statistically not the same at 5% level of significance. This indicated a decrease in the number of reported cases. Also from the

output, the 95% confidence interval of the difference between the average monthly reported cases in 2005 and 2006 was between 159 and 2718.

4.3.2 Two-Sample T-Test and Confidence Interval for Monthly Reported Cases in 2005 and 2007

The null and alternative hypotheses for comparing the annual reported cases have been stated below.

$H_0 : \mu_1 - \mu_3 = 0$

$H_1 : \mu_1 - \mu_3 \neq 0$

From the Minitab output in appendix B, the difference between the average number of reported cases in 2005 and 2007 is 1359. The difference was then tested to find out whether it was significant or not. The T-value (calculated value) was found to be 2.330 and the p-value=0.030. For smaller p-value, the null hypothesis is rejected. So the null hypothesis was rejected at 5% and this indicated that there was significant decrease in the number of malaria reported cases. The 95% confidence interval for the means was between 148 and 2569.

4.3.3 Two-Sample T-Test and Confidence Interval for Monthly Reported Cases in 2005 and 2008

The null and alternative hypotheses for comparing the annual reported cases have been stated below.

$H_0 : \mu_1 - \mu_4 = 0$

$H_1 : \mu_1 - \mu_4 \neq 0$

From the Minitab output in appendix B, the difference between the average number of reported cases in 2005 and 2008 is 2175. The difference was then tested to determine whether it was significant or not. The T-value (calculated value) was found to be 4.340 and the p-value=0.000 at 22 degrees of freedom. For smaller p-value, the null hypothesis is rejected. So the null hypothesis was rejected at 5% and this indicated that there was significant reduction in the number of malaria reported cases in 2005 and 2008. The 95% confidence interval of the means was between 1135 and 3215.

4.3.4 Two-Sample T-Test and Confidence Interval for Monthly Reported Cases in 2005 and 2009

The null and alternative hypotheses for comparing the annual reported cases have been stated below.

$H_0 : \mu_1 - \mu_5 = 0$

$H_1 : \mu_1 - \mu_5 \neq 0$

From the Minitab output in appendix B, the difference between the means was 3446 indicating decrease in the number of reported cases. The difference was then tested to determine whether it was significant or not. The T-value (calculated value) was found to be 6.29 and the p-value=0.000 at 22 degrees of freedom. For smaller p-value, the null hypothesis is rejected. So the null hypothesis was rejected at 5% and this indicated that there was significant decrease in the malaria reported cases comparing the number in 2005 with that of 2008. The 95% confidence interval of the difference between the monthly reported cases in 2005 and 2008 was between 2310 and 4583.

4.3.5 Two-Sample T-Test and Confidence Interval for the Monthly Reported Cases in 2006 and 2007

The monthly reported cases of the disease in 2006 and 2007 were compared. The null and alternative hypotheses were

$H_0 : \mu_2 - \mu_3 = 0$

$H_1 : \mu_2 - \mu_3 \neq 0$

The Minitab output in appendix B indicated that difference between the means is -80. This implied that the reported cases increased between 2006 and 2007. The test statistic value, T-value calculated is -0.410 and the p-value=0.890. Since the p-value is greater than 0.05 level of significance, the null hypothesis is accepted and concluded that the difference in the increase in the monthly reported cases of malaria between 2006 and 2007. The 95% confidence interval of the difference between the average monthly reported cases in 2006 and 2007 was between -1265 to 1105.

4.3.6 Two-Sample T-Test and Confidence Interval for the Monthly Reported Cases in 2006 and 2008

The mean monthly reported cases of the disease in 2006 and 2008 were compared. The null and alternative hypotheses were

$H_0 : \mu_2 - \mu_4 = 0$

$H_1 : \mu_2 - \mu_4 \neq 0$

The Minitab output indicated there was a drop in the number by 734. It was the tested to find whether is significant or not. The T-value (calculated value) was 1.150 and the p-value at 22 degrees of freedom was found to be 0.145. Since the p-value was greater than

the 0.05 level of significance, the null hypothesis was accepted and concluded that there was no significant decease in the monthly reported cases of malaria between 2006 and 2008. The 95% confidence interval of the difference between the average monthly reported cases in 2006 and 2008 was between -274 and 1747.

4.3.7 Two-Sample T-Test and Confidence Interval for the Monthly Reported Cases in 2006 and 2009

The monthly reported cases of the disease in 2006 and 2009 were compared. The null and alternative hypotheses were

$H_0 : \mu_2 - \mu_5 = 0$

$H_1 : \mu_2 - \mu_5 \neq 0$

The Minitab output indicated a drop in the number of the reported cases between 2006 and 2009 by 2008. The T-value was found (calculated value) to be 3.750 and the p-value at 22 degrees of freedom was found to be 0.001. Since the p-value was less than the 0.05 level of significance, the null hypothesis was rejected and concluded that there was significant decrease in the monthly reported cases of malaria between 2006 and 2009. The 95% confidence interval of the difference between the average monthly reported cases in 2006 and 2009 was between 899 and 3117.

4.3.8 Two-Sample T-Test and Confidence Interval for the Monthly Reported Cases in 2007 and 2008

The monthly reported cases of the disease in 2007 and 2008 were compared. The null and alternative hypotheses were

$H_0 : \mu_3 - \mu_4 = 0$

$H_1 : \mu_3 - \mu_4 \neq 0$

The Minitab output indicated a drop in the number of reported cases between 2007 and 2008 by 816. The T-value was found (calculated value) to be 1.84 and the p-value at 22 degrees of freedom was found to be 0.080. Since the p-value was greater than the 0.05 level of significance, the null hypothesis was accepted and concluded that there was no significant decrease in the monthly reported cases of malaria between 2007 and 2008. The 95% confidence interval of the difference between the average monthly reported cases in 2007 and 2008 was between -105 to 1737.

4.3.9 Two-Sample T-Test and Confidence Interval for the Monthly Reported Cases in 2007 and 2009

The monthly reported cases of the disease in 2007 and 2009 were compared. The null and alternative hypotheses were

$H_0 : \mu_3 - \mu_5 = 0$

$H_1 : \mu_3 - \mu_5 \neq 0$

The Minitab output indicated decrease in the number of reported cases by 2088. The T-value calculated was 4.21 and the p-value at 22 degrees of freedom was found to be 0.000. Since the p-value was less than the 0.05 level of significance, the null hypothesis was rejected and concluded that there was significant decrease in the monthly reported cases of malaria between 2007 and 2009. The 95% confidence interval of the difference between the average monthly reported cases in 2007 and 2009 was between 1059 and 3116.

4.3.10 Two-Sample T-Test and Confidence Interval for the Monthly Reported Cases in 2008 and 2009

The monthly reported cases of the disease in 2008 and 2009 were compared. The null and alternative hypotheses were

$H_0 : \mu_4 - \mu_5 = 0$

$H_1 : \mu_4 - \mu_5 \neq 0$

The Minitab output indicated decrease the number of reported cases by 1272. The T-value calculated was 3.21 and the p-value at 22 degrees of freedom was found to be 0.004. Since the p-value was less than the 0.05 level of significance, the null hypothesis was rejected and concluded that there was significant difference in the monthly reported cases of malaria between 2008 and 2009. The 95% confidence interval of the difference between the average monthly reported cases in 2008 and 2009 was between 450 and 2093.

4.4 Trend Analysis

From table 4.1 and figure, the time series has a linear trend model is given by $y_t = 9.313 - 0.055t$, where y_t is the number of monthly reported cases and t is the given month.

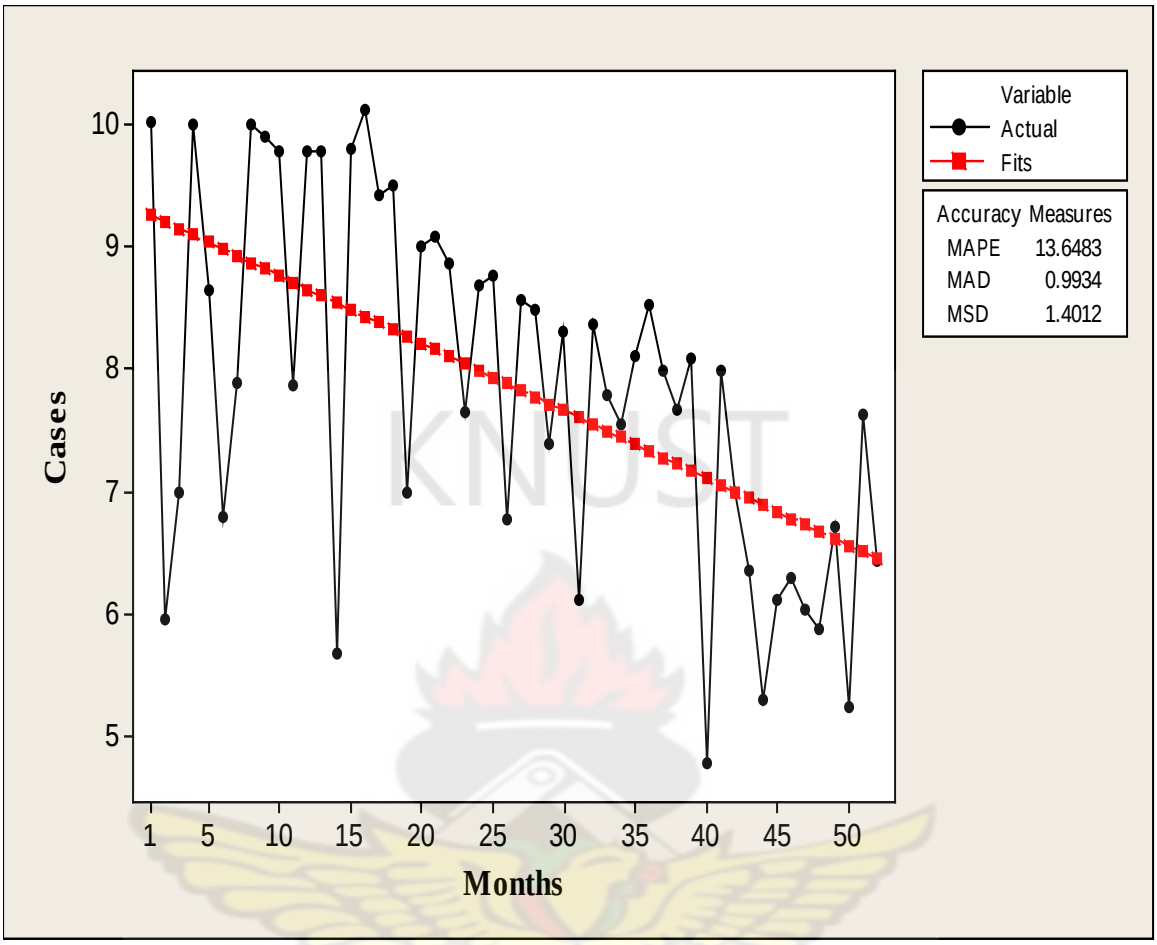

Figure 4.1: Linear Trend Model

Table 4.2: Coefficients of Linear Trend

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	9.313	.340		27.414	.000
	Months	-.055	.011	-.572	-4.928	.000

Table 4.3: ANOVA Table for Trend Model

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	35.397	1	35.397	24.290	.000[a]
	Residual	72.864	50	1.457		
	Total	108.261	51			

The equation indicates that the number of reported cases decreases by 55 each month in a linear fashion. From table 4.2, the constant and the coefficient are different from zero at any level of significance. The test for the model in table 4.3 also indicates that the model is significant at 1%, 5% and10%.

4.5 DATA PREPARATION FOR MODELLING

The secondary data on malaria reported cases from January 2005 to April 2010 was collected from the Municipal Health Directorate. To formulate a model for predicting the reported cases of malaria, the data on the reported cases of malaria from January 2006 to April 2010 (Shown in table A.1 in the appendix A) was used. The data was carefully entered into the SPSS statistical package.

4.5.1 Preliminary Test for Stationarity

A stationary series has constant mean, constant variance and constant autocorrelation structure. The time series plot of the monthly reported cases from January 2006 to April is shown in figure 4.2. The plot showed that the time series was non-stationary. The values had irregular swings and hence had irregular variability.

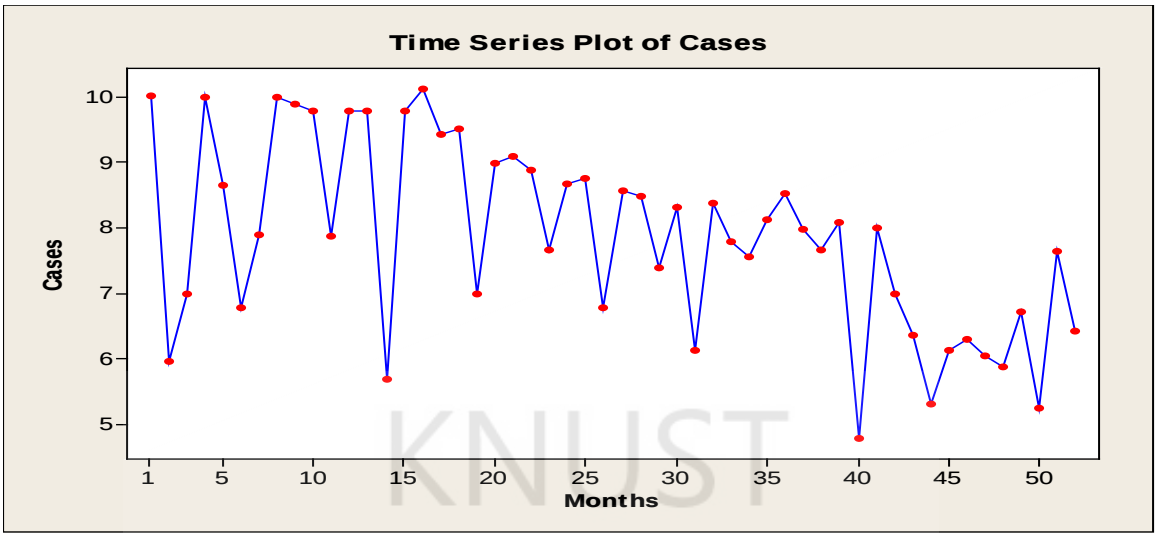

Figure 4.2: Time Series Plot of Malaria Cases

But the fact the there were irregular fluctuations in the plot was not a clear indication that the series was non-stationary. This called for verification to confirm the non-stationarity as indicated by the plotted graph below. The Augmented Dickey-Fuller (ADF) test and the ACF were used to verify the non-stationarity of the series. Table 4.4 below shows the result of the ADF test.

Table 4.4: Augmented Dickey-Fuller Test

Null Hypothesis: t series has a unit root				
Exogenous: None				
Lag Length: 4 (Automatic Based on AIC, MAXLAG=10)				
			t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic			-0.913220	0.315592
Test critical values:	1% level		-2.615187	
	5% level		-1.947969	
	10% level		-1.612427	
*MacKinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(t series)				
Method: Least Squares				
Included observations: 47 after adjusting endpoints				
Variable	Coefficient	Std. Error	t-Statistic	Prob
tseries(-1)	-0.019457	0.021306	-0.913220	0.366340
D(tseries(-1))	-0.955864	0.145992	-6.547389	0.000000
D(tseries(-2))	-0.887213	0.185121	-4.792621	0.000021
D(tseries(-3))	-0.541307	0.176807	-3.061575	0.003831
D(tseries(-4))	-0.278871	0.138139	-2.018772	0.049925
R-squared	0.522807	Mean dependent var		-0.047234
Adjusted R-squared	0.141052	S.D. dependent var		1.604100
S.E. of regression	1.160179	Akaike info criterion		3.235314
Sum squared resid	56.532689	Schwarz criterion		3.432139
Log likelihood	-71.029890	F-statistic		9.202932
Durbin-Watson stat	1.913795	Prob(F-statistic)		0.000020

From the table above, the ADF test-statistic -0.913220) is greater than the critical values (-2.615187, -1.947969 and -1.612427) at 1%, 5% and 10% respectively. This showed that the series had a unit root problem thus indicating that the series was non-stationary and must be differenced.

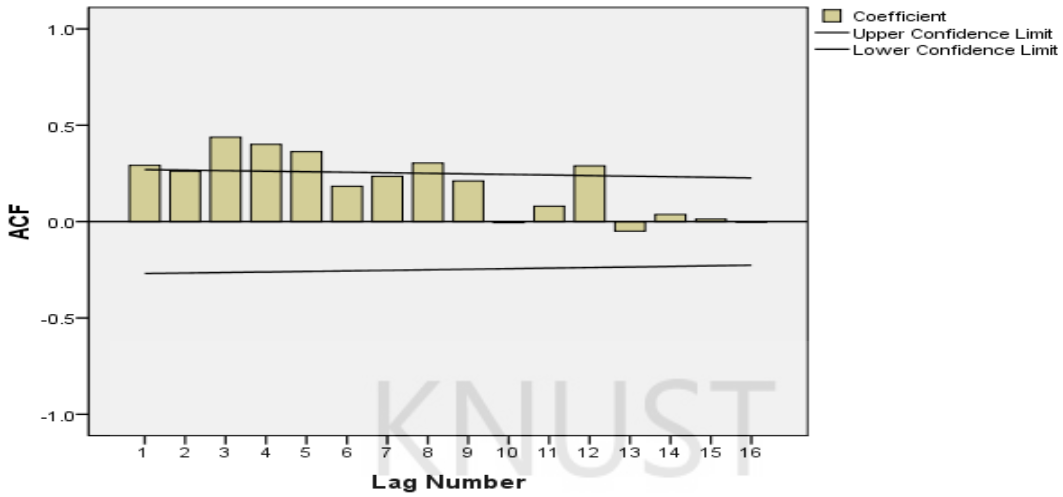

Figure 4.3: Graph of Autocorrelations of Malaria Cases

Table 4.5: Autocorrelation Function of Cases

Lag	Autocorrelation	Std. Error[a]	Box-Ljung Statistic		
			Value	Df	Sig.[b]
1	.291	.135	4.676	1	.031
2	.262	.133	8.517	2	.014
3	.438	.132	19.510	3	.000
4	.401	.131	28.900	4	.000
5	.363	.129	36.760	5	.000
6	.183	.128	38.805	6	.000
7	.235	.127	42.246	7	.000
8	.303	.125	48.122	8	.000
9	.210	.124	51.014	9	.000
10	-.006	.122	51.016	10	.000
11	.080	.121	51.451	11	.000
12	.289	.119	57.315	12	.000
13	-.050	.118	57.493	13	.000
14	.037	.116	57.594	14	.000
15	.013	.115	57.606	15	.000
16	-.002	.113	57.606	16	.000

Transforming the data by using natural logarithm did not also portray stationarity. So the series had to be differenced. The figure below showed the plotted time series after it had been differenced. The differenced data is provided in table 2 in the appendix.

The graph in figure 4.02 appeared to exhibit stationarity more than the one in figure 4.1.

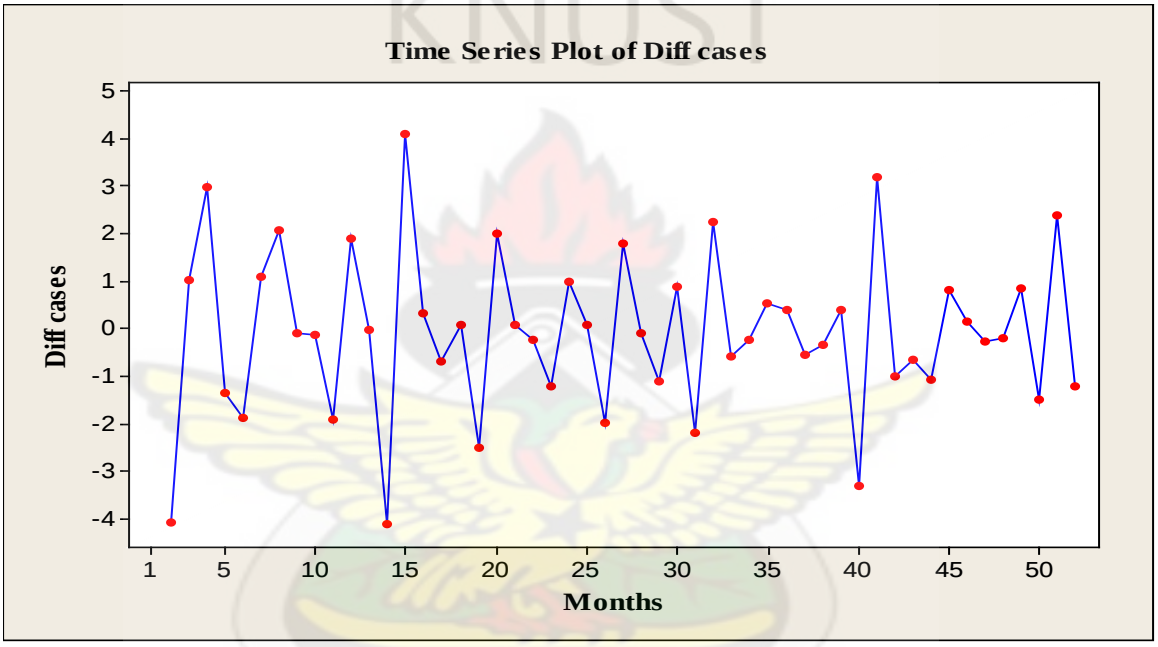

Figure 4.4: Time Series Plot of Differenced Data

The ADF test was also performed on the differenced series to confirm its stationarity. The table 4.6 showed the result of the ADF test performed on the differenced data. From the result below, since the computed ADF test-statistic (-6.729818) is less than the critical values (-2.615187, -1.947969, and -1.612427) at 1%, 5% and 10% significance level respectively, it was concluded that the series had no unit root problem and hence it became stationary after first differencing. This then led to the next stage which was the

selection of the appropriate model. The Normalized Bayesian Information Criterion (Normalized BIC), autoregressive correlation function (ACF) and the partial autocorrelation function (PACF) played an important in the selection and estimation of the parameters. Smaller values of the Normalized BIC produced more reliable model.

Table 4.6: ADF Test on Differenced Data

Lag Length: 3 (Automatic Based on AIC, MAXLAG=10)				
			t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic			-6.729818	0.000000
Test critical values:	1% level		-2.615187	
	5% level		-1.947969	
	10% level		-1.612427	
*MacKinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(tseries)				
Method: Least Squares				
Included observations: 47 after adjusting endpoints				
Variable	Coefficient	Std. Error	t-Statistic	Prob
tseries(-1)	-3.653066	0.542818	-6.729818	0.000000
D(tseries(-1))	1.691698	0.442439	3.823572	0.000420
D(tseries(-2))	0.807369	0.287257	2.810615	0.007415
D(tseries(-3))	0.270906	0.137597	1.968840	0.055438
R-squared	0.836730		Mean dependent var	0.002915
Adjusted R-squared	0.714278		S.D. dependent var	2.770687
S.E. of regression	1.157937		Akaike info criterion	3.212423
Sum squared resid	57.655227		Schwarz criterion	3.369883
Log likelihood	-71.491944		F-statistic	55.092017
Durbin-Watson stat	1.905997		Prob(F-statistic)	0.000000

4.6 MODEL SELECTION

ARIMA models are usually estimated after transforming the variable under forecasting into a stationary series. A stationary series is one whose values vary over time only around a constant mean and variance. As figure 4.4 clearly indicated, the graph of the newly constructed variable indicated stationarity in both the mean and the variance and this assertion was supported with the result of ADF test.

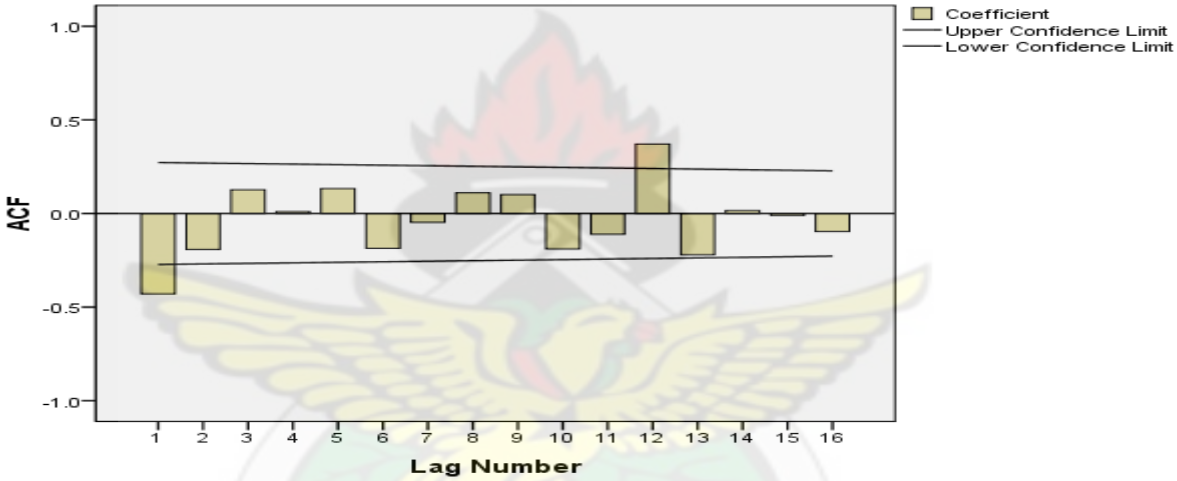

Figure 4.5: Autocorrelations Plot of Differenced Data

Table 1.7: Autocorrelations of Differenced Data

Lag	Autocorrelation	Std. Error[a]	Box-Ljung Statistic		
			Value	df	Sig.[b]
1	-.429	.136	9.970	1	.002
2	-.192	.135	12.002	2	.002
3	.127	.133	12.906	3	.005
4	.010	.132	12.913	4	.012
5	.133	.130	13.958	5	.016
6	-.186	.129	16.037	6	.014
7	-.047	.128	16.172	7	.024
8	.110	.126	16.939	8	.031
9	.101	.125	17.597	9	.040
10	-.190	.123	19.965	10	.030
11	-.111	.122	20.799	11	.036
12	.371	.120	30.343	12	.002
13	-.221	.119	33.823	13	.001
14	.016	.117	33.842	14	.002
15	-.010	.115	33.849	15	.004
16	-.096	.114	34.568	16	.005

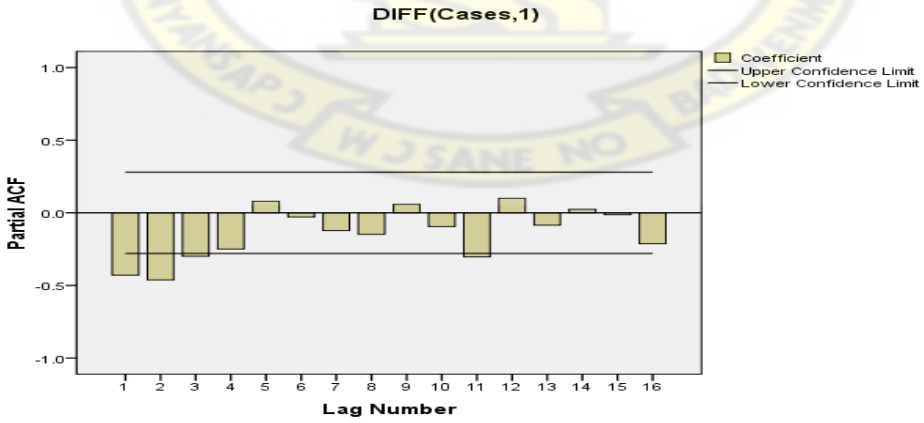

Figure 4.6: Graph of Partial Autocorrelations of Differenced Data

Table 4.8: Partial Autocorrelations

Lag	Partial Autocorrelation	Std. Error
1	-.429	.140
2	-.461	.140
3	-.299	.140
4	-.249	.140
5	.080	.140
6	-.028	.140
7	-.122	.140
8	-.148	.140
9	.059	.140
10	-.095	.140
11	-.303	.140
12	.100	.140
13	-.085	.140
14	.024	.140
15	-.013	.140
16	-.213	.140

The selection of the appropriate model depended on the values of Normalized BIC and the ACF together with the PACF. The graphs of the ACF and PACF are shown above in figures 4.5 and 4.6 respectively. Three tentative models were entertained and the model with the minimum Normalized BIC was chosen. The models and their corresponding Normalized BIC values had been illustrated below in table 4.9.

Table 4.9: Tentative ARIMA Models and their corresponding normalized BIC.

Model	Normalized BIC
ARIMA(0, 1, 1)	0.713
ARIMA(1, 1, 0)	1.045
ARIMA(1, 1, 1)	0.820

So the most suitable model selected was ARIMA (0, 1, 1) because this model had the minimum normalized BIC of 0.713. Hence the model for predicting the malaria reported cases is

$$y_t = a + \varepsilon_t - (1+\theta)\varepsilon_{t-1} + \theta\varepsilon_{t-2} \qquad (4.1)$$

Where, a is a constant, θ is a parameter and ε_t is the residual term.

4.7 PARAMETER ESTIMATION

Model parameters were estimated using SPSS 16.00. The results of the estimation had been displayed in table 4.5. Equation (4.1) thus becomes

$$y_t = -0.052 + \varepsilon_t - 1.914\varepsilon_{t-1} + 0.914\varepsilon_{t-2} \qquad (4.2)$$

From table 4.10, the test statistic value of the constant is -2.625 and p-value is 0.012, indicating that the constant is significant at 5% and 10% level of significance. The parameter is 11.313 and p-value is zero. This indicates that the parameter is significant at 1%, 5% and 10% levels of significance.

Table 4.10: ARIMA Model Parameters

					Estimate	SE	t	Sig.
Cases-Model_1	Cases	No Transformation	Constant		-.052	.020	-2.625	.012
			Difference		1			
			MA	Lag 1	.914	.081	1.131E1	.000

Table 4.11: Model Statistics

Model	Number of Predictors	Model Fit statistics		Ljung-Box Q(18)			Number of Outliers
		Stationary R-squared	Normalized BIC	Statistics	DF	Sig.	
Cases-Model_1	0	.415	.713	22.672	17	.160	0

Table 4.12: Model Fit

Fit Statistic	Mean	SE	Minimum	Maximum	Percentile						
					5	10	25	50	75	90	95
Stationary R-squared	.415	.	.415	.415	.415	.415	.415	.415	.415	.415	.415
R-squared	.172	.	.172	.172	.172	.172	.172	.172	.172	.172	.172
RMSE	1.322	.	1.322	1.322	1.322	1.322E0	1.322E0	1.322E0	1.322E0	1.322E0	1.322E0
MAPE	13.983	.	13.983	1.398E1	1.398E1	1.398E1	1.398E1	1.398E1	1.398E1	1.398E1	1.398E1
MaxAPE	67.248	.	67.248	6.725E1	6.725E1	6.725E1	6.725E1	6.725E1	6.725E1	6.725E1	6.725E1
MAE	1.022	.	1.022	1.022	1.022	1.022E0	1.022E0	1.022E0	1.022E0	1.022E0	1.022E0
MaxAE	4.004	.	4.004	4.004	4.004	4.004E0	4.004E0	4.004E0	4.004E0	4.004E0	4.004E0
Normalized BIC	.713	.	.713	.713	.713	.713	.713	.713	.713	7.131E-1	.713

4.8 MODEL DIAGNOSIS

The model verification is concerned with verification of the residuals of the models to see if they contain any systematic pattern which still can be removed to improve on the chosen ARIMA. This is done through careful examination of the autocorrelations and partial autocorrelations of the residuals of various orders. The residuals should be white noise (or independent when their distributions are normal) drawing from a fixed

distribution with a constant mean and variance. From table 4.13, as the Ljung-Box statistic indicated none of the autocorrelations (from 1 to 16 lags computed) is significantly different from zero at any reasonable level of significance. The ACF in figure 4.6 and PACF in figure 4.7 of the residuals also supported this assertion. This proves that the ARIMA model is an appropriate model. The spike at lag 11 could be attributed to randomness.

Table 4.13: Autocorrelations of Residuals

Lag	Autocorrelation	Std. Error[a]	Box-Ljung Statistic		
			Value	df	Sig.[b]
1	-.003	.136	.000	1	.983
2	-.170	.135	1.586	2	.453
3	.104	.133	2.195	3	.533
4	.158	.132	3.630	4	.458
5	.155	.130	5.036	5	.411
6	-.142	.129	6.246	6	.396
7	-.083	.128	6.670	7	.464
8	.085	.126	7.124	8	.523
9	.064	.125	7.385	9	.597
10	-.212	.123	10.351	10	.410
11	-.097	.122	10.990	11	.444
12	.302	.120	17.314	12	.138
13	-.087	.119	17.856	13	.163

14	-.070	.117	18.217	14	.197
15	-.081	.115	18.705	15	.227
16	-.059	.114	18.975	16	.270

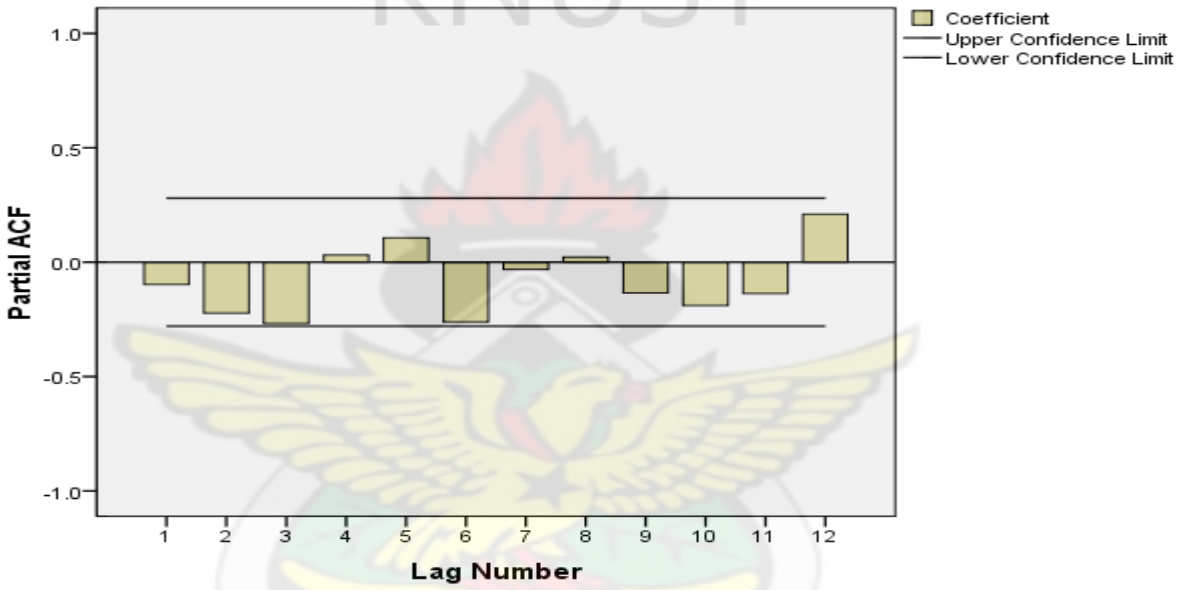

Figure 4.2: Graph of PACF of Residuals

Table 4.14: Partial Autocorrelations of Residuals

Lag	Partial Autocorrelation	Std. Error
1	-.003	.140
2	-.170	.140
3	.106	.140
4	.133	.140
5	.199	.140
6	-.113	.140
7	-.070	.140
8	-.020	.140
9	.030	.140
10	-.197	.140
11	-.039	.140
12	.274	.140
13	-.106	.140
14	.060	.140
15	-.102	.140
16	-.139	.140

4.9 FORECASTING WITH ARIMA MODEL

ARIMA models are basically developed to forecast the corresponding variable. There are two types of forecasts: sample period forecasts and post sample period forecasts. The former is used to develop confidence interval in the model and the latter to generate genuine forecasts for planning and other purposes.

4.9.1 In-Sample Period Forecasts

The sample period forecasts were obtained by simply plugging the actual values of the explanatory variables in the estimated equation (4.2). The explanatory variables were the lagged values of Z_t and the estimated lagged errors. The estimated values and the actual values has been shown in table 4.15. To judge the forecasting ability, of the fitted model, important measures of the sample period forecast were computed. From table 4.15, the mean absolute percentage error (MAPE) for the malaria reported cases calculated is 13.983. This measure indicated low value in the forecasting inaccuracy. The mean absolute error (MAE) is 1.022 and the maximum absolute error (MaxAE) is 4.004.

4.9.1 Post Sample Forecast

The main objective of developing an ARIMA model for a variable is to generate post sample period forecast. This done using equation (4.2). From table 4.12, the forecast for malaria reported cases for May 2010 is 6281 and the actual number of reported cases for that same month is 6110. Also comparing the forecast value of June (6229) and its actual reported value one it realized that the difference between the two values is not great.

These predicted values are very close to the actual values and this indicates the reliability of the model.

Table 4.15: Actual and estimated values of malaria cases at 95% limit.

Months	Actual Reported Cases (Thousands)	Estimated Reported Cases (Thousands)	Residual	Lower Confidence Limit	Upper Confidence Limit
1.0	10.01	-	-	-	-
2.0	5.954	9.958	-4.004	6.570	13.346
3.0	6.995	7.896	-0.901	4.957	10.834
4.0	9.989	7.539	2.450	4.765	10.313
5.0	8.652	8.117	0.535	5.427	10.806
6.0	6.788	8.177	-1.389	5.538	10.816
7.0	7.898	7.877	0.021	5.272	10.482
8.0	9.991	7.828	2.163	5.247	10.409
9.0	9.895	8.083	1.812	5.519	10.647
10.0	9.786	8.267	1.519	5.716	10.819
11.0	7.873	8.400	-0.527	5.859	10.942
12.0	9.782	8.288	1.494	5.754	10.822
13.0	9.784	8.399	1.385	5.871	10.927
14.0	5.681	8.493	-2.812	5.970	11.016
15.0	9.788	8.155	1.634	5.636	10.674
16.0	10.114	8.264	1.850	5.748	10.780
17.0	9.425	8.391	1.034	5.878	10.904
18.0	9.5	8.437	1.063	5.926	10.948
19.0	6.995	8.484	-1.489	5.975	10.994
20.0	8.994	8.295	0.699	5.787	10.803
21.0	9.087	8.307	0.780	5.710	10.814
22.0	8.87	8.325	0.545	5.819	10.831
23.0	7.657	8.322	-0.665	5.817	10.827
24.0	8.675	8.211	0.464	5.706	10.715
25.0	8.76	8.200	0.560	5.696	10.703
26.0	6.784	8.197	-1.413	5.693	10.700
27.0	8.574	8.021	0.553	5.518	10.524

28.0	8.485	8.017	0.468	5.514	10.520
29.0	7.396	8.006	-0.610	5.503	10.508
30.0	8.308	7.901	0.407	5.399	10.403
31.0	6.123	7.884	-1.761	5.382	10.386
32.0	8.369	7.679	0.690	5.178	10.181
33.0	7.782	7.687	0.095	5.185	10.189
34.0	7.559	7.643	-0.084	5.142	10.145
35.0	8.113	7.584	0.529	5.082	10.085
36.0	8.523	7.577	0.946	5.076	10.079
37.0	7.985	7.607	0.378	5.105	10.108
38.0	7.669	7.588	0.081	5.086	10.089
39.0	8.09	7.542	0.548	5.041	10.044
40.0	4.789	7.538	-2.749	5.036	10.039
41.0	7.992	7.249	0.743	4.748	9.750
42.0	6.997	7.261	-0.264	4.760	9.762
43.0	6.367	7.186	-0.819	4.685	9.687
44.0	5.301	7.063	-1.762	4.562	9.565
45.0	6.119	6.860	-0.741	4.359	9.361
46.0	6.291	6.744	-0.453	4.243	9.245
47.0	6.046	6.653	-0.607	4.152	9.154
48.0	5.872	6.549	-0.677	4.047	9.050
49.0	6.723	6.438	0.285	3.937	8.939
50.0	5.242	6.411	-1.169	3.910	8.912
51.0	7.632	6.258	1.374	3.757	8.759
52.0	6.432	6.324	0.108	3.823	8.825
53		6.281		3.780	8.783
54		6.229		3.719	8.740
55		6.177		3.658	8.697
56		6.125		3.597	8.654
57		6.073		3.535	8.611
58		6.0212		3.474	8.568
59		5.969		3.413	8.525
60		5.917		3.352	8.482

CHAPTER 5

SUMMARY, CONCLUSION AND RECOMMENDATIONS

5.1 INTRODUCTION

This chapter deals with summary of results, discussion, findings, conclusion and recommendations.

5.2 SUMMARY OF RESULTS

Comparing the average monthly reported cases of 2005 (before the commencement of the programme) with each year after the implementation of the programme, it is realized that the number of reported cases has been decreasing significantly over the years. There is also a clear indication that the number of reported cases is decreasing after the implementation of the programme. However, the number of reported cases in 2006 is higher than that of 2007, though the difference is not significant.

The trend equation, $y_t = 9.313 - 0.055t$, indicates that with the implementation of the programme, the number of reported cases on the average decreases by 55 each month. This shows a positive sign about the programme.

The ARIMA model developed for predicting the monthly reported malaria cases is ARIMA (0, 1, 1): $y_t = -0.052 + \varepsilon_t - 1.914\varepsilon_{t-1} + 0.914\varepsilon_{t-2}$. The model was used to predict an eight-month lead period of the reported cases.

DISCUSSION OF RESULTS

The study shows that in the first quarter of 2005, the month January records the highest prevalence of malaria cases (11127). In the second quarter of that same year, the highest reported cases are in June with a total number of 13510. The highest number of reported cases in third quarter is in August with a total number of 10588. In the last quarter of that same year, the highest number of reported cases is in December. In all, the highest number of reported cases in the whole year is in January and the least number is in the month of September with a total number of 7443. The number of reported cases per month is 10073.

In the year 2006, the highest number of reported cases is in the first quarter is in the month of January with a number of 10010. March has the highest number of cases in the second quarter with a total number of 9898. In the third quarter, the highest number is 9991 and it is recorded in August. In the fourth quarter, the highest number is 9786 in October. In all, the highest number of reported cases in 2006 is in the month of January. The least number of reported cases is 5954 and it is recorded in February. The average monthly number of reported cases in the year is 8634.

In the year 2007, the study indicates that the highest number of reported cases in the first quarter is in the month of March with a total of 9788 cases. The highest number in the second quarter is 10114 cases in the month of April. The third quarter recorded the highest number in September with a total of 9087 cases. The highest number of cases in the fourth quarter is in October with a total of 8870. The highest number of reported

cases in the whole year is in April and the least number of reported cases is 5681. The reported cases per of the year is 8714.

The study also shows that the highest number of reported cases in 2008 for the first quarter is in the month of January with a total of 8760 cases. The second quarter has the highest value in April with a total of 8485 cases. In the third quarter, the highest number of reported cases is 8369 in September. The fourth has a total of 8523 cases in December. The highest number of reported cases for the whole year is in January with a total of 8760. The least number of reported cases in the whole year is 6123 in the month of July. The average monthly reported cases are 7898.

In 2009, the highest value in the third quarter is in the month of March with a total of 8090 cases. The second quarter has its highest value in the month of May with a total of 7992 cases. The highest number of reported cases for the third quarter is in the month of July with a total of 6367 cases. The fourth quarter has a record of 6291 cases in the month of October as the highest number of reported cases. The highest number of reported cases for the whole year of 2009 is 8090 in the month of March. The number of reported cases per month in that year is 6627.

Comparatively, the number of reported cases per month (10073) in 2005 indicates the highest among all the years. This figure indicates that on the average, 10073 people had malaria in each month in 2005. The average monthly reported cases in 2006 are 8634.This implies that 8634 had malaria every month in 2006. This indicates a drop in

the average reported cases. This probably could be attributed to the commencement of the Malaria Control Programme. In 2007, the reported cases per month increased slightly to 8714. This figure indicates the second highest. In 2008 the number of reported cases per month is 7898. This figure however indicates a drop in the number of reported. In 2009, an average of 6627 cases is recorded. Comparing the number of reported cases per month in 2009 with 2008, there is a further drop in the number of reported cases. The least number of reported cases per month is in 2009. Observing the trend of the reported cases from 2005 to 2009, it is most likely that the number of reported cases of the disease will drop in 2010 with the continuation of the programme.

Comparing the number of reported cases per month in 2005 with that of 2006, the result indicates that there is a significant difference in the number of reported cases. The number of reported cases in 2005 is higher than that of 2006. One is 95confident that the difference in the means of 2005 and 2006 is between 908.91 and 3139.75. The interval does not contain zero and this confirms the fact that the number of reported cases per month between these years are significantly different from zero. Comparing the number of reported cases in 2005 with that of 2007 indicates that the number of reported cases in 2005 is significantly different from that of 2007. One can say there is 95% confidence that the difference in the number of reported cases per month in 2005 and 2007 is between1456.89 and 3738.78. Also comparing the number of reported cases in 2005 with 2008 and 2009 indicate the average number of monthly reported cases in 2005 is significantly different from those of 2008 and 2009. The number of reported cases per month in 2005 is higher than those of 2008 and 2009. There is 95% confidence that the

difference in the number of reported cases per month in 2005 and 2008 is between 2400.83 and 4773.00. There is 95% confidence that the difference in the number of reported cases per month in 2005 and 2009 is between 2636.90 and 5142.10.

However, comparing the number of monthly reported cases per month in 2006 and 2007, it is realized that there is no significant difference in their means. This implies that the average number of reported cases in 2006 and 2007 are almost the same. There is 95% confidence that the average number of reported cases per month is between -346.793 and 1493.793. This interval contains zero so it confirms the fact that there is no significant difference in their means. Considering the number of reported cases per month in 2006 and 2008, there is statistical evidence that the average number of reported cases in 2006 is different from that of 2008. The average number of reported cases in 2006 is 8634 and that of 2008 is 7898. So in 2006, the average number of reported cases is higher than that of 2006. There is 95% confidence that the difference in means is between 586.89 and 2538.28. There is evidence that the difference in the average number of reported cases in 2006 and 2009 is significantly different from zero. The average number of reported cases in 2009 is 6627. There is 95% confidence that the difference in the average number of cases in 2006 and 2009 is between 808.61 and 2920.72. There is no evidence that the difference in the number of reported cases per month in 2007 and 2008 is different from zero. There is 95% confidence that the difference in the number of reported cases per month is between -15.593 to 1993.857. In 2007 and 2009, there is evidence that the difference in the number of reported cases per month is significantly different from zero. There is 95% confidence that the difference in the number of reported cases in the years

2007 and 2008 is between 209.17 and 2374.16. There is no statistical evidence that the number of reported cases per month in 2008 and 2009 is significant different from zero. There is 95% confidence that the difference in the number of reported cases per month in 2008 and 2009 is between -827.386 and 1432.552.

Four questionnaires were issued to the four main hospitals in the municipality (St Jude Hospital, AGA Hospital, Bryant Mission Hospital and Government Hospital). They all agreed that the malaria control programme is yielding dividend. They want the programme to be sustained

ARIMA model is a good technique for predicting the values of time series variables. Its strength lies in the fact that the method is suitable for any time series with any pattern of change. The model developed can be used for short term prediction of the malaria reported cases in the municipality if the malaria control programme is still put in place.

The time series model developed for predicting the number of reported cases of malaria in the Obuasi municipality is ARIMA (0, 1, 1). This model can be used by researchers for forecasting malaria reported cases in the municipality. However, it should be updated from time to time with the incorporation of current data.

5.3 FINDINGS

The study reveals that the integrated malaria control programme launched by A.G.A is having positive impact on the malaria reported cases in the municipality. Since the launch of the programme in 2005, the reported cases per month in subsequent years (2006, 2007, 2008 and 2009) have not exceeded the reported cases per month in 2005 (when the programme was not put in place). It is realized from the results that the programme has been able to reduce the alarming number of reported cases per month from 10,073 in 2005 to 8,634 in 2006. In 2007 number of reported cases per month however increased slightly but not statistically significantly from that of 2006. Significant improvement is seen in 2008 with the number of reported cases decreasing to 7,898. In the year 2009, there is also improvement in the number of reported cases per month in 2009. In all, the programme has made a modest gain in the area of controlling the disease.

Some health workers do not make use of the surveillance mechanism put in place to track communities with high prevalent rate of the disease. Some patients also shun it because much time is required to complete.

Despite the fact that people have been complaining about the odour of the insecticide used, no one has complained of any ailments as a result of the insecticide used by the spraying team to any of the four main hospitals in the municipality.

There are still some people who do not allow their rooms or houses to be sprayed despite the fact that the programme is very helpful in that it reduces one's risk of contracting malaria. This is because they do not trust the spraying team and will not want them to go into their rooms or houses.

5.4 CONCLUSION

The malaria control programme has reduced the number of reported cases of malaria. The number of reported cases per month in 2005 is 10073 and the number of reported cases per month in 2006 is 8634, which shows an improvement over that of 2005. The number of reported cases per month in 2007 is 8714 which is not statistically significant different from that of 2006 but indicates an improvement as compared to that of 2005. In 2008, the number of reported cases per month is 7898. This value shows reduction in the number of reported cases although not significantly different from that of 2007. The number of reported cases per month in 2009 is 6627. This figure indicates reduction in the number of reported cases.

The model developed for forecasting the monthly reported cases of malaria in the municipality is $y_t = -0.052 + \varepsilon_t - 1.914\varepsilon_{t-1} + 0.914\varepsilon_{t-2}$.

All the health facilities must embrace the surveillance system to help track the communities with high number of reported case. The management of the Integrated Malaria Control Programme must also revise the surveillance system to make people

spend less time in filling it. People employed as spraying men should be trustworthy so that people can entrust their rooms or houses to them with fear of theft.

In conclusion the malaria control programme launched AngloGold Ashanti (AGA) is having a positive impact on the number of reported cases of malaria in the Obuasi municipality.

5.5 RECOMMENDATIONS

Since in totality, the programme has made modest gains in reducing the number of reported cases in the municipality, it is recommended for adoption by other districts, municipalities, metropolis and other bodies who want to help fight malaria. The central government should also help raise funds for the programme to be replicated in other parts of the country.

However to make the Integrated Malaria Control Programme more effective, all health facilities in the municipality should adopt the surveillance system put in place by the management of the programme to track communities with high number of reported cases. Management of the programme should revise the surveillance system (which is a form to be filled by patients) so that patients and health workers will spend less time in filling it.

Some people perceived that the spraying team is still using the insecticide used in the first round of spraying in 2006 which a lot of people complained about its odour and so do not

want their rooms to be sprayed. These people must be made aware that the current insecticide used is different and does not produce odour.

The spraying men should be people the community members know to be trust worthy so that they can entrust their rooms to them. If this done, the fear of theft will be minimal.

The ARIMA model developed for forecasting the number of reported cases is ARIMA (0, 1, 1): $y_t = -0.052 + \varepsilon_t - 1.914\varepsilon_{t-1} + 0.914\varepsilon_{t-2}$. This model is recommended for the management of the malaria control programme for adoption. It will help them forecast the number of reported cases each month. Having a foreknowledge of what the number of reported cases each month in the future will help management strategize how to carry out their activities to reduce the number of reported cases. The model is also recommended for health facilities in the municipality. It will help the municipal health directorate to plan and make maximum use of its scarce resources. Other researchers can also use this piece of work as a reference material.

Researchers who will like to do further work to assess the impact of the programme should consider using spectral analysis and autoregressive exogenous time series analysis to compare the impact of the programme on each month. They should also consider interviewing the people in beneficiary communities to find out their opinions about the effectiveness of the programme. In this study, only the opinions of the health facilities and management of the programme were sought.

BIBLIOGRAPHY

AngloGold Ashanti, AngloGold Ashanti Report to Society Case Studies-Regional Health Threats, 2007. http//www.agareports.com/07/malaria-Obuasi.htm, accessed 24 February 2009.

AngloGold Ashanti, AngloGold Ashanti Report to Society 2004. http//www.agareports.com/04/malaria-Obuasi.htm. accessed 4th December 2010.

Briet O. J.T., Vounatsou P., Amerasinghe, P. H., (2008). "Malaria seasonality and rainfall seasonality in Sri Lanka are correlated in space", Geospatial Health 2(2), pp. 183-190

Brillinger, D. R., (2002). "Time series: data analysis and theory", Universidadde California, Holden-Day, SIAM.

Brockwell, P.J. and Davis, R. A. (2002). "Introduction to Time Series and Forecasting", 2nd. ed., Springer, Verlag.

Brown, V., Issak, M.A., Rossi, M., Barboza, P. and Paugam, A. (2002). "Epidemic of malaria in North Eastern Kenya", Lancet 352, 1356–1357.

Center for Disease Control and Prevention (CDC), National Center for infectious Disease, Malaria Epideminology Branch, (2004). "Global Malaria Prevention and Control Programme". http//www. cdc.gov/ncidod/dpd/parasites/malaria. Accessed

Contreras, J., Espínola, R., Nogales, F. J. andAntonio J. Conejo, A. J., (2003). "IEEE Transactions on Power Systems", Vol. 18, No. 3, pp.1014-1020.

Dobre, I., Alexandru, A. A., (2008). "Quantitative Methods Inquires", Journal of Applied Quantitative Methods vol. 40, pp.165-166.

Fillatre, L., Marakov, D., Vaton, S. (2003). "Forecasting seasonal trafficflows", Computer Science Department, ENST Bretagne, Brest, Paris.

Guerin, P. J., Olliaro, P., Nosten, F., Druilhe, P., Laxminarayan, R., Binka, F., Kilama, W.L., Ford, N. and White N. J. (2002). "Malaria: current status of control, diagnosis, treatment, and a proposed agenda for research and development." The Lancet Infectious Diseases 2(9): 564–73.

Guerrero, V. M., (2003). "Análisis estadístico de series de tiempo económicas", Segunda edición, México.

Houeto, D., D'Hoore, W., Ouendo, E. M., Charlie, D. and Deccache, A., (2007). "Malaria control among children under five in sub-Saharan Africa: the role of empowerment and parents' participation besides the clinical strategies", ARHEN.

Kendall, M. G., and A. Stuart. 1966. “The advanced theory of statistics: Design and Analysis and Time-Series”, Charles Griffin & Co. Ltd., London, United Kingdom, Vol. 3.

Lin, H., Lu, L., Tian, L., Zhou, S., Wu, H., B, Y., Ho, S. C. and Liu, Q., (2009). “Spatial and temporal distribution of falciparum malaria in China”, Malaria Journal, 8:130 doi:10.1186/1475-2875-8-130

Lindsay W. S and Birley M., (2004). “Rural Development and Malaria Control in Sub-Sahara Africa”, EcoHealth, Springer New York, Volume 1 Number 2. pp 129-137.

McKenzie D, 2003. “Measuring Inequality with Asset Indicators”. Bureau for Research in Economic Analysis of Development Working Paper No. 042: Cambridge, MA: Center for International Development, Harvard University.

Nchinda, T.C. (2005). “Malaria: A Reemerging Disease in Africa”, World Health Organization, Geneva.

Nkuo-Akenji, T., Ntonifor, N. N, Ndukum M. B, Kimbi H. K, Abongwa E. L, Nkwescheu A, Anong D. N, Songmbe, M, Boyo M. G, Ndamukong, K. N. and Titanji, V. P. K. (2006). “Environmental Factors Affecting Malaria Parasite Prevalence in Rural Bolifamba, South-West Cameroon”, African 40 Journal of Health Sciences, Volume 13, Number 1-2.

Patz, J.A., Graczyk, T.K., Geller, N. and Vittor, A. Y., (2000). “Effects of environmental change on emerging parasitic diseases”, International Journal for Parasitology 30(12-13): 1395-1405.

Pattanayak, S.K., Mercer D.E., Sills, E. and Yang, J. C. (2003). “Taking Stock of the Agroforestry Adoption Studies”, Agroforestry Systems 57 (3): 173-186.

Reed, S.D., Laxminarayan, R., Black, D.J., and Sullivan, S.D., (2002). “Economic issues and antibiotic resistance in the community.” The Annals of Pharmacotherapy, 36(1): 148-54

Suarez, C.A.H, Parra, O. J. S. and Diaz, A. E. (2009). Revista Ingenieria Investigacion, Volume 29 Number 2. pp 65-67.

Sudhakar, P., and Subramani, P., (2007). “Literature Review: Insights into Formulating a Protective Malaria Medicine, Journal of Young Investigators”, Volume 16, Issue 1. Sulaiman, S., (2007). How is climate change shifting Africa’s malaria map? http//www.scidev.net.com. Accessed on Friday, 9 October 2009.
The African Summit on Roll Back Malaria (2005) Abuja, Nigeria, April 25 2000, World Health Organization, Geneva (document WHO/CDS/RBM/2000.17).

The African Summit on Roll Back Malaria (2005) Abuja, Nigeria, World Health Organization, Geneva (document WHO/CDS/RBM/2000.17)

WebMD. Malaria. http//www.wedmd.com/a-to-z-guides/malaria-cause. Accessed on 24th August 2009.

WHO (2005a) The Roll Back Malaria strategy for improving access to treatment through home management of malaria WHO, Geneva (document WHO/HTM/MAL/2005.1101)

Wilson, M.L., (200)1. "Ecology and Infectious Disease." In Ecosystem Change and Public Health, The Johns Hopkins University Press: Baltimore, MD: 285-291.

Wiseman, V., Scott, A., Mcelroy, B., Conteh, L. and Warren, S. (2007). "Determinants of bed net use in the Gambia: implications for malaria control", American Journal of Tropical Medicine and Hygiene, volume 76(5), pp. 830-836.

Wiwanitkit V. (2006). Hellenic Society of Haematology, "Correlation between Prevalence of Malaria and Altitude, a Case Study in a Rural Endemic area of Thailand", HAEMA 2006; 9(1). pp 56-58.

Worrall, E., Basu, S. and Hanson, K., (2003). "The relationship between socio-economic status and malaria", London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1E 7HT, UK.

APPENDICES

APPENDIX A

Table A.1 Malaria Reported Cases from January 2005 to April 2010.

Months	Reported Cases (in thousands)	Months	Reported Cases (in thousands)
1.0	10.01	27.0	27.0
2.0	5.954	28.0	28.0
3.0	6.995	29.0	29.0
4.0	9.989	30.0	30.0
5.0	8.652	31.0	31.0
6.0	6.788	32.0	32.0
7.0	7.898	33.0	33.0
8.0	9.991	34.0	34.0
9.0	9.895	35.0	35.0
10.0	9.786	36.0	36.0
11.0	7.873	37.0	37.0
12.0	9.782	38.0	38.0
13.0	9.784	39.0	39.0
14.0	5.681	40.0	40.0
15.0	9.788	41.0	41.0
16.0	10.114	42.0	42.0
17.0	9.425	43.0	43.0
18.0	9.5	44.0	44.0
19.0	6.995	45.0	45.0
20.0	8.994	46.0	46.0
21.0	9.087	47.0	47.0
22.0	8.87	48.0	48.0
23.0	7.657	49.0	49.0
24.0	8.675	50.0	50.0
25.0	8.76	51.0	51.0
26.0	6.784	52.0	52.0

Table A.2: Results of first Differencing

Month	Differenced values	Month	Differenced values
1.0	-	27.0	-1.976
2.0	-4.056	28.0	1.790
3.0	1.041	29.0	-0.089
4.0	2.994	30.0	-1.089
5.0	-1.337	31.0	0.912
6.0	-1.864	32.0	-2.185
7.0	1.120	33.0	2.246
8.0	2.093	34.0	-0.587
9.0	-0.096	35.0	-0.223
10.0	-0.109	36.0	0.554
11.0	-1.913	37.0	0.410
12.0	1.909	38.0	-0.538
13.0	0.002	39.0	-0.316
14.0	-4.103	40.0	0.421
15.0	4.107	41.0	-3.301
16.0	0.326	42.0	3.203
17.0	-0.689	43.0	-0.995
18.0	0.075	44.0	-0.630
19.0	-2.505	45.0	-1.066
20.0	1.999	46.0	0.818
21.0	0.093	47.0	0.172
22.0	-0.217	48.0	-0.245
23.0	-1.213	49.0	-0.174
24.0	1.018	50.0	0.851
25.0	0.085	51.0	-1.481
26.0		52.0	2.390
			-1.200

APPENDIX B

Minitab output of the comparison of the average monthly reported cases.

Two-Sample T-Test and CI: 2005, 2006

```
Two-sample T for 2005 vs 2006

       N   Mean  StDev  SE Mean
2005  12  10073   1540      444
2006  12   8634   1483      428


Difference = mu (2005) - mu (2006)
Estimate for difference:  1438.50
95% CI for difference:  (158.61, 2718.39)
T-Test of difference = 0 (vs not =): T-Value = 2.33  P-Value = 0.029  DF = 22
Both use Pooled StDev = 1511.6982
```

Two-Sample T-Test and CI: 2005, 2007

```
Two-sample T for 2005 vs 2007

       N   Mean  StDev  SE Mean
2005  12  10073   1540      444
2007  12   8714   1311      378


Difference = mu (2005) - mu (2007)
Estimate for difference:  1358.75
95% CI for difference:  (148.30, 2569.20)
T-Test of difference = 0 (vs not =): T-Value = 2.33  P-Value = 0.030  DF = 22
Both use Pooled StDev = 1429.6839
```

Two-Sample T-Test and CI: 2005, 2008

```
Two-sample T for 2005 vs 2008

       N   Mean  StDev  SE Mean
2005  12  10073   1540      444
2008  12   7898    806      233


Difference = mu (2005) - mu (2008)
Estimate for difference:  2174.92
95% CI for difference:  (1134.61, 3215.22)
T-Test of difference = 0 (vs not =): T-Value = 4.34  P-Value = 0.000  DF = 22
Both use Pooled StDev = 1228.7266
```

Two-Sample T-Test and CI: 2005, 2009

```
Two-sample T for 2005 vs 2009

       N   Mean  StDev  SE Mean
2005  12  10073   1540      444
2009  12   6627   1110      320
```

```
Difference = mu (2005) - mu (2009)
Estimate for difference:  3446.42
95% CI for difference:  (2310.16, 4582.68)
T-Test of difference = 0 (vs not =): T-Value = 6.29  P-Value = 0.000  DF = 22
Both use Pooled StDev = 1342.0578
```

Two-Sample T-Test and CI: 2006, 2007

```
Two-sample T for 2006 vs 2007

       N  Mean  StDev  SE Mean
2006  12  8634   1483      428
2007  12  8714   1311      378


Difference = mu (2006) - mu (2007)
Estimate for difference:  -79.7500
95% CI for difference:  (-1264.7913, 1105.2913)
T-Test of difference = 0 (vs not =): T-Value = -0.14  P-Value = 0.890  DF = 22
Both use Pooled StDev = 1399.6742
```

Two-Sample T-Test and CI: 2006, 2008

```
Two-sample T for 2006 vs 2008

       N  Mean  StDev  SE Mean
2006  12  8634   1483      428
2008  12  7898    806      233


Difference = mu (2006) - mu (2008)
Estimate for difference:  736.417
95% CI for difference:  (-274.215, 1747.048)
T-Test of difference = 0 (vs not =): T-Value = 1.51  P-Value = 0.145  DF = 22
Both use Pooled StDev = 1193.6753
```

Two-Sample T-Test and CI: 2006, 2009

```
Two-sample T for 2006 vs 2009

       N  Mean  StDev  SE Mean
2006  12  8634   1483      428
2009  12  6627   1110      320


Difference = mu (2006) - mu (2009)
Estimate for difference:  2007.92
95% CI for difference:  (898.76, 3117.07)
T-Test of difference = 0 (vs not =): T-Value = 3.75  P-Value = 0.001  DF = 22
Both use Pooled StDev = 1310.0423
```

Two-Sample T-Test and CI: 2007, 2008

```
Two-sample T for 2007 vs 2008

       N  Mean  StDev  SE Mean
2007  12  8714   1311      378
2008  12  7898    806      233


Difference = mu (2007) - mu (2008)
Estimate for difference:  816.167
95% CI for difference:  (-104.948, 1737.281)
T-Test of difference = 0 (vs not =): T-Value = 1.84  P-Value = 0.080  DF = 22
Both use Pooled StDev = 1087.9455
```

Two-Sample T-Test and CI: 2007, 2009

```
Two-sample T for 2007 vs 2009

       N  Mean  StDev  SE Mean
2007  12  8714   1311      378
2009  12  6627   1110      320


Difference = mu (2007) - mu (2009)
Estimate for difference:  2087.67
95% CI for difference:  (1059.42, 3115.92)
T-Test of difference = 0 (vs not =): T-Value = 4.21  P-Value = 0.000  DF = 22
Both use Pooled StDev = 1214.4857
```

Two-Sample T-Test and CI: 2008, 2009

```
Two-sample T for 2008 vs 2009

       N  Mean  StDev  SE Mean
2008  12  7898    806      233
2009  12  6627   1110      320


Difference = mu (2008) - mu (2009)
Estimate for difference:  1271.50
95% CI for difference:  (450.31, 2092.69)
T-Test of difference = 0 (vs not =): T-Value = 3.21  P-Value = 0.004  DF = 22
Both use Pooled StDev = 969.9219
```

APPENDIX C

QUESTIONNAIRE FOR HEALTH CENTRES IN THE OBUASI MUNICIPALITY

Introduction

A research is being carried out to assess the impact of the AngloGold Ashanti Integrated Malaria Control Programme on the morbidity of malaria in the Obuasi Municipality.

This questionnaire is designed to elicit your candid views to help find out the effectiveness of the programme. Any piece of information provided will be treated as confidential.

Name of Hospital: ……

Location:……

Date:……..

1. Do you distribute bed nets to pregnant woman without any fee?
Yes ( ) No ( )

2. If you sell bed nets, what is the cost of a bed net in your hospital?
……

3. Give the three most common diseases/cases recorded over the past four years in your hospital.
2008……
2007……
2006……
2005……

4. Give the number of recorded cases and the cost of treatments of malaria over the past four years.

Year	No. of recorded cases	Cost of treatment
2008	……	……
2007	……	……
2006	……	……
2005	……	……

5. Does the AngloGold Ashanti Malaria Control Programme have positive impact on the prevalent rate of malaria in the municipality?
Yes ( ) No ( )

6. If your answer is no, why?
……
……
……
……

7. Has anyone ever complained of any ailments as a result of the insecticide used in spraying his/her room(s)?
Yes ( ) No ( )

8. Suggest ways of improving on the malaria control programme.
……
……
……
……
……
…….

Thanks for your co-operation.

APPENDIX D

QUESTIONNAIRE FOR ANGLOGOLD MALARIA CONTROL PROGRAMME

Introduction

A research is being carried out to assess the impact of the AngloGold Ashanti Integrated Malaria Control Programme on the morbidity of malaria in the Obuasi Municipality.

This questionnaire is designed to elicit your candid views to help find out the effectiveness of the programme. I hope the outcome of this research will be of much interest to your outfit. Any piece of information provided will be treated as confidential.

1. When did you launch the Malaria Control Programme?
……

2. When did you begin the spraying exercise?
……

3. Give the insecticide(s) used in the spraying exercise.
……

4. How long does it take the insecticide to become inefficacious when a room is sprayed?
……

5. Does the team inform the tenants and landlords the time it takes the insecticide to become inefficacious when a room is sprayed?
Yes ( ) No ( )

6. Does the spraying exercise cover the entire municipality?
……

7. How many times do you carry out the indoor residual spraying exercise in a year?
……

8. Does the team go back to spray rooms/houses which are not covered in each round?
Yes ( ) No ( )

9. If your answer is yes to no. 8, how do you identify rooms/houses which have not been sprayed in each round?
……
……

10. Have you put in place any measures to sustain the programme?
Yes ( ) No ( )

11. If your answer is yes, what are these measures?
……
……

12. Apart from the indoor residual spraying exercise, what other measures have you put in place to help control the disease?
……
……

13. What measures have you put in place to monitor the susceptibility of the vector to the insecticide?
……
……

14. Give the cost of carrying out the programme each year.
……
……
……
……

15. Give the challenges facing the programme.
……
……
……

THANKS FOR YOUR CO-OPERATION.