FORECASTING MALARIA CASE ADMISSIONS IN THREE KENYAN HEALTH FACILITIES

BY

JOSEPHINE KALUNDE MALINGA

School of Mathematics

College of Biological and Physical Sciences

University Of Nairobi

Research project submitted in partial fulfillment of the requirements for the degree of Master in Science in Biometry

2015

Declaration

Candidate

I declare that this research project is my original work and has not been presented for a degree in any other university for any other award

Signature: .....

Date: .....

Josephine Kalunde Malinga

Reg No: I56/67778/2013

Supervisor

This research project has been submitted with my approval as university supervisor

Signature: .....

Date: .....

Dr. Nelson Owuor Onyango

School of Mathematics, University of Nairobi

P.O Box 301197-00100, Nairobi

Acknowledgement

I acknowledge the support of my supervisor Dr. Nelson Owuor Onyango, colleagues, and other lecturers for their unwavering support and immense contribution and guidance.

My family and closest friends whose help and emotional support has seen me through my lowest moments.

And above all God, the creator, giver of all that is good for seeing me through this process.

Dedication

To Mallingah W, Mumbi R, Mwatha A, Mailu D and Chemutai P

ABSTRACT

Malaria is one of the major causes of morbidity and mortality, especially in sub-Saharan Africa. In Kenya, it accounts for almost 35% of all outpatient consultations and is a major cause of infant and child mortality. The aim of the study was to fit appropriate time series models and compare and assess the accuracy of two different methods; Exponential Smoothing Method versus the Box Jenkins Method to forecast malaria case admissions for children under 15 years of age from different epidemiological zones in Kenya. The different methods were tested using data from three health facilities located in the "Coast Endemic", "Lake Endemic" and "Highland Epidemic" regions of Kenya. The model performances were evaluated through data splitting using various error measures namely: mean error, mean percentage error and the mean absolute squared error. The statistical forecast accuracy of the models showed that the methods were effective in predicting malaria case admissions in Kenya. However, no particular model emerged superior to the other as the simpler decomposition and exponential smoothing method performed equally as well and in some instances even better than the more complex Box Jenkins model. The Box Jenkins method was more accurate in long term projections due to its statistical underlying theory while the exponential smoothing method performed better in short and medium term forecasts and in data which experienced recent abrupt level shifts. The study also showed the limitations of forecasting malaria case admissions only from historical patterns with the need to develop improved models by incorporating external predictor's such as climate-related variables. In conclusion, although different methods can be applied to malaria forecasting, they should be tailored to the specific malaria transmission setting to avoid misleading results due to the underlying disease transmission dynamics.

ABBREVIATIONS

AIC	Akaike Information Criterion
AL	Artemether-Lumefanthrine
AR	Autoregressive
ARIMA	Autoregressive Integrated Moving Average
ASAL	Arid and Semi-Arid
BIC	Bayesian Information Criterion
CQ	Chloroquine
DHS	Demographic Health Surveys
DIBD	Division of Insect-Borne Diseases
GLM	Generalized Linear Models
GMAP	Global Malaria Action Plan
HMIS	Health Management Information Systems
IRS	Indoor Residual Spraying
ITN	Insecticide Treated Nets
LLIN	Long Lasting Insecticide bed net
MA	Moving Average
MAE	Mean Absolute Error
MAPE	Mean Percentage Error
MARE	Mean Absolute Relative Error
MASE	Mean Squared Error
MDG	Millennium Development Goal 6
MIS	Malaria Indicator Surveys
MLE	Maximum Likelihood Estimation
MSE	Mean Squared Error
NMS	National Malaria Strategy
PA*Pf*	*Plasmodium falciparum* parasite rate
RBM	Roll Back Malaria
SAF	Seasonal Adjusted Factor

SARIMA	Seasonal Auto-Regressive Integrated Moving Average
SP	Sulphadoxine Pyrimethamine
SSA	Sub-Saharan Africa
SVM	Support Vector Machine
WHO	World Health Organization
WMR	World Malaria Report

LIST OF FIGURES

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

Malaria has been one of the major causes of morbidity and mortality globally, but by a bigger extend in the African Region. In 2013, 82% of the 198 million cases (uncertainty range 124–283 million) and 90% of the 584 000 million malaria specific deaths (uncertainty range 367 000–755 000) were estimated to be from this region [1]. Even with renewed calls to eradicate the disease through increased international donor assistance, country-specific government involvement and adoption of varied malaria prevention, control and monitoring interventions including; increased insecticide treated nets (ITN) and long lasting insecticide nets (LLIN) distribution and reported use, increased indoor residual spraying (IRS), campaigns on the use of first line antimalarial drugs (ACTs) and setting up of monitoring and evaluation systems, malaria still remains a cause of worry in malaria endemic regions and more so for the high risk populations comprising of children under 5 years and pregnant women [2].

Since the World Health Organization (WHO) set a call in 1955 for malaria eradication in the world, various initiatives have been launched over time with an ambitious goal for a malaria free world and the reduction of malaria specific deaths to zero by 2015. These include:

- The Roll Back Malaria (RBM) movement set up in October 1998, which led to the signing of the Abuja Declaration and Plan of Action in 2001 by African heads of State and Government with a goal to halve malaria mortality by 2010 by ensuring at least 60% (redefined later to 80%) of the at-risk population was protected or treated with appropriate methods [3].
- The Millennium Development Goal (MDG) 6 of 2003 seeking to stop and reverse the incidence of malaria by 2015 [4, 5].
- The Global Malaria Action Plan (GMAP) of 2008 which called for universal coverage of at-risk populations with some form of vector control with a goal of a 75% decline in malaria morbidity between 2000 and 2015 [6, 7].

In the past decade, billions of dollars have been committed and directed to this fight to reduce the burden of malaria in Africa even after a slow start in 2003. In 2014 however, the evaluation of trends over time showed that only 12 countries in Africa with sufficient data, were projected by the WHO to observe a 50 - 75% reduction in malaria cases by 2015 given the 2000 baseline [1].

The WHO and RBM have continuously advocated for development of forecasting and early warning systems to inform disease control and prevention measures, with additional use in diagnostics and drugs supply chain management [8]. As malaria is more rampant in tropical and sub-tropical regions especially in low and middle income settings with limited resources [9], appropriate early warning

signals and accurate disease predictions can inform policy, and provide public health services with information required for a targeted approach to the control and prevention of malaria through the effective use of the available resources [10].

1.1 Background Information: Kenya

1.1.1 Country Profile

Kenya covers a total area of 582,646 km^2, sprawls the Equator in eastern Africa and lies along longitudes 34^0 West to 42^0 East and latitudes 5^0 North to 5^0 South. Approximately, only 20% of Kenya's land area is arable in the Lake Victoria and highlands regions while 80% is arid and semi-arid, mostly in the Northern and Northeast parts of the country. Merely 1.9% of the total surface area is occupied by stagnant water. The climate of Kenya ranges from very dry in the north and northeast, temperate in the interior, to hot and humid at the coastal shoreline. This heterogeneity of climate across the country is due to the variations in terrain and altitude. There are two rainy seasons; the short rains occurring from October to December and the long rains from April to June. On the other hand, the temperature remains relatively constant throughout the year with the coldest months from July to August and the hottest from February to March [11].

Kenya's population stood at 38.6 million from the Population and Housing census conducted in 2009. Following the population growth rate estimated at 2.9% annually over the past 30 years, was projected to be around 45 million in 2015 [12]. The census reports Kenya's population to be characterized by the "very young" with 43% of the population being under 15 years of age and only 4% aged 65 years and older. This might be explained by the high fertility rates and declining mortality observed in the last few decades due to improved health care awareness, access and delivery. In addition, overall declining trends for most diseases have been observed in the country [12-14].

1.1.2 Epidemiology of Malaria in Kenya

The epidemiology of malaria has been studied over many decades, but the complexities of the disease, for instance; the climatic suitability for transmission which is heterogeneous even within countries, the variations in vector distribution, human behavior that might hinder or promote infection, effects of control interventions at different levels on mortality and morbidity, governments policies and donor involvement, and the influence of past trends, have made it harder to control the disease leaving a number of regions still at risk while some other regions have successfully eradicated malaria [15-25].

In Kenya, the major malaria vectors are from the *Anopheles Gambiae* and *Anopheles funestus* family, and disease transmission is by all four species of human Plasmodium: *P. falciparum, P. malariae, P. ovale,* and *P. vivax. P. falciparum* accounts for almost all malaria infections (>95%) with around 16% comprising of mixed infections with *P. malariae, P. ovale* or both [26]. The survival of the malaria vectors is largely depended on a number of factors such as the prevailing environmental conditions and human factors especially in the stable transmission regions. The main climate-related variables include rainfall, which has been shown to provide the breeding sites for vectors by increasing water availability, while warmer air due to a relative rise in temperature holds more moisture hastening the vector development and prolonging the risk of infection. In addition, human factors like rural to urban migration and other varied activities have also been linked to the malaria infection spectrum, with additional baseline demographics such as age and sex being able to explain some variations in disease transmission [17, 27-29].

The epidemiological zones in Kenya are mainly four;

- The Endemic zones around the coastal and the Lake Victoria regions with stable malaria transmission occurring throughout the year. The Entomological Inoculation Rate (EIR) ranges from <10% to >100% while the parasite prevalence rate (*Pf*PR) from community surveys has been reported to be between <5% and <40%. The suitability of the climate in these regions allows for the survival of malaria vectors as rainfall, humidity and temperature are the main determinants for malaria infection;
- The Highland epidemic zone found mainly in the Western highlands which is prone to outbreaks but often characterized by seasonal malaria comprising of yearly cycles and variations. The increase in the minimum temperature to at least 18^0C during the long rains is optimal for vector development and survival which drastically increases the rates of malaria transmission and case fatality rates rising to even ten times higher than in areas of stable transmission;
- The Seasonal malaria transmission zones which are composed of arid and semi-arid areas in the north, northeast, east and southeast parts of the country. Malaria transmission is highly seasonal and intense only during the rainy season due to lower temperatures, increased rainfall and vegetative cover for vector breeding. The *Pf*PR is usually <1% in this areas;
- The Low risk malaria zones in Nairobi and the central highlands of the country where temperatures are too low to support vector survival and malaria transmission. Most cases from these regions are usually imported from malaria-prone areas and this, together with the climatic changes are likely to widen areas suitable for disease transmission. ***[Figure]***.

In 2014, Noor [30] reported a transition in the malaria epidemiology and risk of infection in Africa over time, by making a comparison of maps on malaria prevalence between 2000 and 2010. Malaria epidemic zones were reported to have shrank over the 10 years with an increase in the number of people living in low transmission zones and malaria free areas. In Kenya, there was a >20% to 40% reduction in the *Plasmodium falciparum* parasite rate which had been standardized to the 2-10 years age group (PA*Pf*PR_{2-10}). In conclusion, it was estimated that more than 50% of the Kenyan population presently lives in low-transmission or malaria-free areas [30].

1.1.3 Malaria Control in Kenya

Malaria control in Kenya began during the colonial era, to protect the European settlers and later the government and farms labor force who needed protection in a bid to safeguard the economic viability of the Colony [31]. The initial activities included environmental control activities referred to as "mosquito brigades" such as cutting down bushes, filling of empty pits and stagnant water draining. By 1913, the Colony had begun the promotion of weekly doses of quinine prophylaxis and free bed nets distributed to European settlers, the police, Indians and railway workers. Continued efforts led to the establishment of the Division of Insect-Borne Diseases (DIBD) in 1944 to coordinate control efforts for all insect borne diseases such as malaria which had become a major cause worry following an epidemic in 1940 and the acceptance of the notion that all febrile patients in country were infected with malaria [2].

IRS was adopted as the main form of protection for the mass populations with targeted approaches to highly endemic and epidemic regions of the country beginning in the early 1950's. Over time, varied malaria control measures have been deployed including; use of traditional prevention methods, sleeping under bed nets, continued targeted indoor spraying and appropriate prophylaxis for disease management. Quinine (QN) was the first anti-malarial to be adopted in the early 1900 and in the mid-1950's was changed to chloroquine (CQ). CQ resistance escalated in the 1980's, eventually leading to malaria epidemic's in the country which occurred in the mid-1990's. This lead to the adoption of sulphadoxine-pyrimethamine (SP) in 1998 which was abandoned in 2004 after emergence of resistance, and replaced with the more efficacious Artemether-Lumefanthrine (AL) implemented in 2006 [32-34]. The scale up of these measures has been attributed to the observed reduction in the malaria burden but a direct causality link has not been established [18, 23, 25, 35, 36].

Kenya's first ten year National Malaria Strategy was launched in 2001 following the formation of the RBM movement and a second strategy, the National Malaria Strategy (NMS) 2009 – 2017, was established in 2009 [37, 38]. In line with the MDG's, RBM and WHO goals for malaria control,

prevention and case management, the strategy aimed at facilitating the reduction of the malaria burden by two-thirds given the 2007-2008 case and attributable deaths numbers in Kenya [38]. As of July 2013 the following milestones had been reported; approximately 80% of all households in malaria endemic regions owned at least an ITN obtained either through purchase at subsidized prices, free mass distributions, and targeted dissemination for children and pregnant women; capacity building efforts had been observed through training of health workers on malaria diagnostics and treatment; formulation of plans to implement community strategies for malaria control and case management; and targeted IRS activities being carried out in specific parts of the country. Additionally, monitoring and evaluation systems for disease trends and intervention impacts have been developed with more than 30% of the 75 arid and semi-arid (ASAL) and 45 epidemic prone districts having functional sentinel sites for epidemic detection and response[39].

To maintain favorable malaria trends however, a need was identified during the development of the NMS to tailor make malaria interventions specific to a transmission intensity setting, target approaches towards particular sub-populations with higher risks of infection whilst strengthening the monitoring, evaluation and disease surveillance systems to achieve effective and sustainable malaria control [38].

1.1.4 Malaria Trends in Kenya

Researchers have utilized different methods in scientific studies, using different sources of data to model malaria trends worldwide [16, 19, 20, 40-43]. In Kenya, the easily available and accessible sources of data for malaria indicators which involve population-based surveys such as Demographic Health Surveys (DHS) and Malaria Indicator Surveys (MIS), are burdened with many limitations including data incompleteness, inaccuracies and are unavailable in actionable timeframes. Health facility-based data has been an alternative to these cross-sectional surveys due to its longitudinal nature and hence has been applied in many scenarios to determine short and long-term disease trends. However, health facility data should be utilized with the full recognition of its limitations [44, 45]. The application of various mathematical and statistical methods on this data has been instrumental and has enabled the modelling of disease trends over time, with additional methods that account for the shortcomings of the available data.

Malaria consultations, admissions and attributable deaths have been reported to be on a declining trend in most parts of Africa, beginning in the late 1990s when malaria was declared a disaster as it was one of the main causes of morbidity and mortality. The global community and country-specific efforts led to the adoption of various interventions to curb the disease burden, efforts which have

been sustained and improved over time [2, 22, 46-51]. This is evident in the transition that has been observed both in malaria transmission intensities and disease incidence which have shifted favorably in the last decade [30].

Studies from Kenya have generally reported declining trends of malaria cases and attributable deaths observed from community surveys and health facility records. However, some areas still report increasing trends while in others, trends have been stable over time. These variations could be due to the epidemiology of transmission dynamics and the level of coverage and application of control interventions. Between 1998 and 2010, malaria cases and deaths declined in the coastal region, the highland region and in some low-land regions, with more pronounced declines being observed in regions that have a historical low parasite prevalence rate. In the Western high transmission parts of the country however, malaria trends are reported to have increased over time often with yearly variations [20, 21, 30, 34, 36, 40, 52, 53].

Overall, scaling up of malaria preventive and curative interventions has been seen to contribute to the declines in the malaria burden where according to the Malaria Indicator Survey carried out in 2010, clinical malaria accounted for 34% of all outpatient visits in Kenya with malaria prevalence being highest for children between 5-14 years of age at 13% [13]. However, a direct causality link between interventions and trends over time has not yet been established which could be due to the unavailability of consistent data to measure indicators, the possibility of intervention supply information redundancy from different stakeholders whilst in some regions a shift in trends was observed even before application of any control measure. Moreover, most studies on disease trends have only been carried out in specific sites and regions making it impractical to aggregate results countrywide [21, 36, 40, 54].

1.2 Statement of the Problem

Reliable, consistent and timely data is a prerequisite for effective planning and implementation of malaria prevention, control and case management interventions. The major shortcomings hindering this process in many low and middle income countries include; poor vital registration systems, lack of comprehensive Health Management Information Systems (HMIS) and disease surveillance systems, unavailability of longitudinal data over time and if available, of low quality, incomplete and/or inaccessible in actionable timeframes [44, 45, 55, 56]. For instance, in the World Malaria Report (WMR) 2014, Kenya was one of the malaria endemic countries in Sub-Saharan Africa (SSA) that lacked sufficient data reported to WHO to monitor malaria trends between 2000 and 2013 [1].

In addition to the shortcomings of the health systems, the insufficiency of the data on malaria indicators hinders the monitoring of temporal trends, be it in control of the disease or evaluation of intervention impacts. Furthermore, malaria cases are still treated presumptively in some regions even after changes in the global and national guidelines recommending confirmatory treatment of malaria, leading to overestimation of the malaria burden. Additionally, the only data that can be captured from routine health records is for people who have access and actually utilize the formal health system which might lead to underestimation of the disease burden [2, 22].

To accommodate this, statistical modelling has been deployed often to create tools which are able to obtain more accurate estimates for the required data which are closer to the truth. The use of these methods has evolved over time, from merely studying underlying forces and structures that produced particular sets of observed data to modelling and forecasting into the future [42].

The ability to predict malaria incidence accurately is a major milestone in the control and management of the disease. The results obtained could facilitate optimal distribution of resources, enabling the adoption of appropriate control interventions tailor made to the county, region or transmission setting under consideration. This will in the long run lead to the reduction in the number of new and resurgence malaria cases and malaria-attributable deaths on the path to fulfilling the global and country-specific targets for the disease at large.

1.3 Study Objectives

General Objective

1. Forecast malaria case admissions in Kenyan Health Facilities for children under 15 years of age.

Specific Objectives

1. Develop malaria forecasting models for three Kenyan health facilities located in different epidemiological settings with emphasis on the Box Jenkins Methodology.
2. Compare the applicability and assess the accuracy of the different forecasting methods in predicting malaria case admissions in three Kenyan health facilities.

1.4 Justification/Significance of the Study

Malaria forecasting models have been developed in many malaria endemic countries, using mainly historical data with additional data on environmental risk factors like weather-related variables, to predict future malaria cases over certain periods of time. Different methods have been used in varied geographical and transmission settings, utilizing different approaches to identify the best mathematical or statistical models, and significant malaria case predictors. However, the variability in

the computational methods and lack of common forecast accuracy measures does not allow for straightforward comparisons across studies and sites [10].

Kenya recently adopted the county system of governance, where policy making, health related decisions and the distribution of resources is at the county level [57]. Efforts towards malaria prevention, control and elimination require that resources are optimally distributed, to enable timely adoption and application of control interventions. Since malaria data availability, consistency and quality is questionable, there is need to establish functional and appropriate statistical models for validating existing data, monitoring disease trends over time and predicting malaria incidence in the Kenyan context, especially at the county or epidemiological zone level. This is due to the already documented health system shortcomings bundled with too many reports by health workers to different stakeholders all with one goal of reducing the malaria morbidity and mortality [38, 58].

In this study, various time series methods will be compared using the same datasets, to determine their applicability and accuracy in the prediction of malaria case admissions. Models that are suitable and applicable to various local transmission settings will be established, enabling targeted approaches in regards to optimal resource distribution, implementation of appropriate control interventions and strengthening of disease surveillance, monitoring and evaluation systems. This data will be invaluable for malaria control and elimination efforts to the national and county governments, health facilities, policy makers, researchers', the donor community and the population at large.

CHAPTER 2: LITERATURE REVIEW

The first malaria early warning system was developed by Christopher's in 1911 to predict malaria epidemics that were common in Punjab. Since then, more apt malaria foresting models have been developed over time, enabling the use of varied statistical, mathematical, machine learning and grey methods to estimate past, current and even forecast future disease burdens [59].

There are three major categories of forecasting methods which include; qualitative/judgmental methods, quantitative methods and technological methods. Qualitative forecasting methods are generally subjective in nature and are based on expert opinions in formulating relationships. Conversely, quantitative forecasting methods involve statistical procedures in analysis of past values or historical data to establish true mathematical relationships or approximate associations which are reasonably closer to the truth [60, 61]. The three sub-categories of quantitative methods include; time series methods which seek to identify historical patterns using time as a reference point and then forecast future values using time-based extrapolation procedures; explanatory methods which seek to identify past relationships producing observed outcomes and forecasting by applying the established relationships in the future; and monitoring methods which seek to identify the changes in relationships and patterns. Technological methods address societal, economic, political or technological long-term issues using expert-based methods, or historical relationships, patterns and analogies to define and "forecast" pre-determined future values [62].

There are a number of factors to consider before selecting the appropriate technique to use in forecasting. Some of these factors include; the type of data being analyzed, the time horizon which can be classified as either short, medium or long term; the level of detail or frequency required which increases with need; the number of series' and parameters involved and the historical patterns and constancy [61]. When data is presented with time as the reference point, time series models are often adopted. These are unique methods whose usage first involves the understanding of the structure of an observed set of data and the underlying forces producing it through estimation and fitting of appropriate models, and second, using these established relationships for forecasting, monitoring and evaluation, or even feedback and feedforward control. The time series approach has often been used to model disease patterns, explaining what will happen but not why [63].

Appropriate mathematical and statistical methodologies have been developed to measure the malaria burden in many malaria endemic countries such as those in SSA, with malaria forecasting being a critical part of many malaria control programmes and research organizations worldwide. Using malaria case numbers, and additional data on coverage of both preventive and curative interventions,

environmental factors and patient demographics as risk factors, various models have been used to predict future malaria cases over certain periods of time in different geographical settings [59, 64-70].

In a scoping review published by Zinszer (2012), the history and future directions of malaria forecasting methods were reviewed. The authors included studies that forecasted malaria prevalence, incidence and epidemics over time with almost all studies using patient records in health facilities and the general population as the main data sources. While some studies used mathematical modelling and machine-learning methods as the forecasting approaches, majority of the studies used statistical methods including Autoregressive Integrated Moving Average (ARIMA) models, Generalized Linear Models (GLM's) and Holt Winter's methods. Most studies also included climate-related covariates like rainfall, temperature, relative humidity, evaporation rates and vegetation cover as covariates, applying lagging to account for the delayed effects of weather on malaria transmission and infection. To evaluate the models, different approaches were used where most authors divided the data into two parts, with one part for model building and the remaining bit for model validation. Conversely, some authors still used all available data for model building and made out-of-sample forecasts. The mean-squared error (MASE), mean percentage error (MAPE), mean absolute relative error (MARE), 95% confidence intervals (CI), paired t-tests, correlation coefficients and data visualization techniques were the main methods used to evaluate model performance and forecast accuracy [10].

In conclusion, the authors reported that the accuracy of the forecasting methods from different studies could not be compared due to the lack of common measures applied and gave recommendations on the same. The authors established that many researchers only incorporate one method of forecasting in any given data set and report results based on a single model. They also noted the availability and variability of various statistical models as one of the strengths of forecasting the malaria burden, but recommended that an effort should be put to tailor-make and adopt forecasting approaches that are applicable and appropriate to the local transmission settings to avoid misleading results. In addition, the use of common accuracy measures in forecasting is encouraged, to allow for comparisons between studies. Multiple methods could also be applied to the same data, giving attention to the assumptions, advantages and disadvantages of any given model, allowing the identification of the most parsimonious forecasting model, and establish which predictors would be significant for malaria forecasting [10].

In 2002, while comparing different forecasting approaches, Abeku [64] used five statistical methods to forecast malaria incidence in epidemic-prone areas of Ethiopia with unstable transmission, where a simple seasonal adjustment with three past observations method performed best. The authors first log-transformed the data due to the nature of its original distribution and applied different methods including; averaging methods some with adjustments of past values' mean deviation from expected

values and ARIMA models where data was first differenced to obtain stationarity. Part of the data was used for model building and the last twelve (12) observations used for model validation. The forecasting accuracy was assessed by calculating errors using the differences between the forecasted and observed values. The accuracy of the methods, the forecasted values and their ability to accommodate seasonality and changes in the trend differed with respect to the length of the historical time series and the post-sample length of the forecasts [64].

In Burundi, Gomez-Elipe (2007) used ARIMA models to model and predict malaria incidence rate by studying the association between weather-related variables and the malaria burden in an unstable transmission area of the country. The number of malaria cases in the preceding month, temperature, rainfall and the normalized vegetative index were used as explanatory variables, where they all were included in the best forecasting model which had a 93% forecasting accuracy. The differences between observed and forecasted values were assessed for normality, confirmed whether they fell within 5% of the 95% confidence interval, and whether the forecast precision depended on the magnitude of the incidence rate change over time, a process to check the reliability of the forecasting model. Malaria incidence was significantly associated with lagged rainfall, vegetative index data and maximum temperature while there was no association between malaria cases and minimum temperature [71].

Briet [65] compared the exponentially weighted moving average models, ARIMA models with seasonal components, and seasonal multiplicative autoregressive integrated moving average (SARIMA) models on their ability to predict monthly malaria cases four months ahead in Sri Lanka. Rainfall (which was lagged) and the number of malaria cases in neighboring districts were used as additional covariates and assessed on their ability to improve the predictive power of the models. Data was first log-transformed, pre-whitened and then segmented for model building and validation. The MARE was used to assess forecast accuracy of the models. The best models varied by district and forecasting horizon, suggesting that models should be tailored to specific settings, but overall, the SARIMA model followed by the ARIMA model that modelled seasonality through second order harmonics were often adopted, while the seasonal adjustment method performed worst. The inclusion of additional covariates did not always improve the model fit, but in many districts and horizons more accurate forecasts were obtained afterwards [65].

Malaria cases in Bhutan have been on a declining trend over time, and in 2010, Wangdi [68] carried out a study to develop prediction and forecasting models for malaria incidence in seven of the twenty malaria endemic districts in the country. A multiplicative SARIMA model (2, 1, 1) $(0, 1, 1)^{12}$ was identified as the best model fit for the overall data with slight parameter variations for different districts. 85% of the data points were used for model building and the remaining 15% for model validation to forecast malaria incidence 2 years into the future. The transfer function method of the

ARIMAX models was used to determine significant predictors of malaria transmission for each district and overall while the MAPE was used to evaluate the forecast accuracy. Rainfall, mean maximum temperature and the number of malaria cases in the preceding month were the only factors that were associated with malaria cases, but the presence and strength of the predictive power varied within districts [68].

In 2014, Kumar [72] used climatic factors including; mean monthly rainfall, mean maximum temperature and relative humidity, as risk factors to predict monthly malaria slide positives in Delhi, India. Seasonal ARIMA models were used with an ARIMA (0, 1, 1) $(0, 1, 0)^{12}$ being identified as the best fit model and a stationary R-squared statistic was employed to evaluate the model's goodness of fit. Only rainfall and relative humidity lagged at one month were found to be significant predictors of malaria cases in the study region. To determine the peak seasonal variation, the seasonal adjusted factor (SAF) was used with the Ljung-Box statistics to evaluate model fit [72].

Ezekie (2014) used SARIMA models to model and forecast malaria mortality rate in Nigeria. The data was first differenced to obtain stationarity with the model parameters being estimated through the maximum likelihood method (MLE). The AIC and BIC were used for model selection where the model with the lowest statistic being chosen as the best fit. Past malaria mortality data was used to forecast out of sample future values. Hence, other than fulfilling the model assumptions for selection of the best model fit, there was no way to assess the forecast accuracy as no observed values could be compared with those predicted by the model. A SARIMA (1, 1, 1) $(0, 0, 1)^{12}$ was selected as the best model to predict malaria mortality given the data available in Nigeria [73].

Comparing four time series models for epidemiological surveillance, Zhang (2014) evaluated the performance of four time series methods including ARIMA models, support vector machine (SVM) models and two decomposition methods (exponential smoothing and regression) in forecasting future disease burdens. The MAE, MAPE and MSE error measures were used to evaluate the models' forecast efficacy. In conclusion, the authors noted that no single method was superior to the others, as they all performed differently given different diseases. However, in the presence of a level shift (sudden changes in the time series) SARIMA models were observed to perform the worst in comparison with the simpler decomposition methods and the more complex SVM methods. Although none of the forecasting methods was "fully appropriate" for disease forecasting, the authors advice that care should be taken in selection of the forecasting model to use given the structure of the data and the underlying model assumptions [74].

According to a number of the studies reported above, variations in the best fit models were observed, in regards to the disease being forecasted, the study sites, and regions [65, 68, 74]. This phenomenon,

they concluded, could be due to the heterogeneity of malaria transmission, difference in the level of intervention adoption and impact, variations in human activities and health seeking behavior in different countries and transmission settings. Additionally, they emphasized on the need to select the most parsimonious model for the most accurate estimates.

CHAPTER 3: METHODOLOGY

This chapter focuses in detail on the understanding of the Box-Jenkins Modelling Approach and the Time Series Methods used in the analysis and forecasting of malaria case admissions. Methodologies used in modelling, software specifications and features incorporated into each model will be some of the main aspects of the chapter.

3.1 Data

The study uses health facility-based data obtained from three county hospitals located in different epidemiological zones across the country which include; Siaya County Hospital (located in the Lake endemic region of Western Kenya bordering Lake Victoria), Kitale County Hospital (from the highland epidemic region where populations live 1500 meters above sea level) and Malindi County Hospital (along the endemic region of the Kenyan Coast). The primary data on case admissions for children aged 0 and <15 years was collected over time from health facility records for the period 1999 – 2011. A total of 156 possible monthly data points were obtained.

3.2 The Concept of Time Series

A time series is an ordered sequence of values of a variable at equally spaced time intervals, for instance, daily, weekly, monthly or yearly. Mathematically, it is a time dependent sequence $\boldsymbol{Y_t}$: where $\boldsymbol{t}$ denotes the time points/steps on a set of integers. A time series is said to be deterministic if it can be expressed as a known function $Y_t = f(t)$; while if it is expressed as $Y_t = X(t)$ where *X* is a random variable, then is a stochastic time series. The usage of time series analysis can either be description, prediction and/or even feedback for control measures [60].

3.2.1 Components of a Time Series

A feature of most time series is that they can be decomposed into four components:

Trend

This is the general inclination the graph of a time series appears to be directed towards over a long interval of time.

Seasonality

This refers to the tendency of the series to vary, with regular periodic changes usually in the course of the year, following identical or almost identical patterns recurring consistently. It is important to investigate how the trend and the seasonal component interact to determine the appropriate smoothing and forecasting technique.

Cycle

This refers to the long-term oscillations about a trend curve, are periodic but not necessarily at equal intervals of time. The length and magnitude of the cycle are not constant as in the seasonal component and vary from one cycle to the other

Irregular/Idiosyncratic component

This is the error component of the time series. It is the residual component of the series that accounts for the deviation of the time series from what would have been obtained had the trend, seasonal and cyclical components explained the series fully. It is mostly caused by external factors, is unpredictable and can be said to account for the random variability in the series.

3.2.2 Types of time series models

Time series models can generally be classified into two types:

- The additive model assumes that the data is the sum of the time series components. If the data do not contain any of the components, the value for that missing value is assumed to be zero. It is appropriate if the magnitude of the seasonal fluctuation does not vary with the series.

$$Y_t = l_t + s_t + c_t + e_t; \text{ the additive form} \quad (1)$$

- The multiplicative model assumes that the data is the product of the various time series components, and if a component is missing, the value is assumed to be 1.

$$Y_t = l_t s_t c_t e_t \text{ ; the multiplicative form} \quad (2)$$

Where $l_t = l(t)$ is a function of time − the trend;
$s_t = s(t)$ is a periodical function of time − the seasonal component;
$c_t =$ cyclic variations;
$e_t =$ error term and,
$t = 0,1, \ldots,$

3.2.3 Decomposition of Time Series models

Decomposition is the process of breaking down the underlying pattern of a time series to identify the component factors. The trend and seasonal components are the two main components of the basic underlying pattern. The decomposition model assumes the data has the following form:

Data = Pattern + Error

= *f* (trend, cycle, seasonality, error)

Seasonal adjustment (deseasonalizing) is one of the main products of decomposition where for an additive model:

$$Y_t - s_t = l_t + c_t + e_t$$

And a multiplicative model:

$$\frac{Y_t}{l_t} = s_t c_t e_t$$

The seasonalized data allows us to better observe the underlying pattern of the series and provides measures of the magnitude of seasonality in the data. This process also allows reliable comparison of values at different time points.

Let us assume we have an additive time series.

$$Y_t = l_t + s_t + e_t$$

Classical Decomposition

In classical decomposition, first the series is de-trended using smoothing techniques and the trend component is subtracted from the original data series to obtain:

$$Y_t - \widehat{l_t} = s_t + e_t$$

Since it is assumed that the seasonal component is constant from year to year, we obtain the seasonal index for each month by averaging the detrended values for each particular month and adjusting them to ensure that they add to zero. This gives $\hat{s}_t$.

However, the classical decomposition has its own shortcomings which include: unavailability of the trend estimates for the first few and last few observations, the assumption that the seasonal component is constant within year to year and the lack of method robustness in taking care of outlier values. We will thus utilize the STL method to decompose our series.

Seasonal and Trend Decomposition using Loess (STL)

The STL method was developed by Cleveland (1990) for time series decomposition with the Loess method which is used for estimating nonlinear relationships [75]. Loess, “locally weighted scatterplot smoothing” uses regression to remove the “bumpy” nature of the data by replacing values with a “locally weighted” robust regression estimate to estimate the trend and seasonal effects. First, we place a window of specified width over the data (smoother loess curves require wider windows); then we fit a regression curve to the observations that fall within the window with reducing weights as points move further from the line; the regression is re-run and the weighting is re-calculated. This process is repeated several times and the point at the center of the window is obtained. The loess

curve is obtained by moving the window across the data with each point on the resulting loess curve being the intersection of a regression line and a vertical line at the center of any such window.

Unlike the classical decomposition method, STL is not limited in the frequency of data it can handle, the seasonality is allowed to vary over time and can be controlled by the user and the method is robust to outliers. The trend window and the seasonal window parameters control how rapidly the trend and seasonal components can change with small values allowing more rapid change. However, the STL method is only recommended for additive models.

3.2.4 Lag

The lag is a difference in time between an observation at time ***t*** and a previous observation at time ***(t-i).*** Therefore; Y_{t-i} lags Y_t by ***i*** periods.

The Backshift (Lag) Operator

The backshift operator ß operating on Y_t has the effect of shifting the data back any *i* periods;

For instance to shift the data back one period gives;

$$ßY_t = Y_{t-1}; \text{while}$$

$$ß^i Y_t = Y_{t-i}; \text{ shifts the data back i periods} \quad (3)$$

When the series is composed of monthly data, and we want to shift data to "the same month last year" we use $ß^{12}$.

$$ß^{12} Y_t = Y_{t-12}$$

3.2.5 Stationarity and Non-stationarity

A time series is said to be stationary if it satisfies the following conditions;

$$1.\ E(Y_t) = \mu_y; \text{for all t} \quad (4)$$

$$2.\ Var(Y_t) = E\left[\left(Y_t - \mu_y\right)^2\right] = \sigma^2; \text{for all t} \quad (5)$$

$$3.\ Cov\,(Y_t, Y_{t+i}) = \tau_i; \text{ for all t} \quad (6)$$

In simple terms;

1. If the mean of the plotted time series varies over time, the series is considered non-stationary in mean. In such a case, the concept of "differencing" is used to make the series stationary. Only one or two orders of differences are suitable in most time series.
2. If the variance of the plotted time series shows no obvious changes over time, then the series is considered stationary in variance. If not, there is need to transform (log or cubic or square root) the data to make it stationary in variance.

3. The autocorrelation structure is constant over time.

The mean and/or variance of the non-stationary series changes with time unlike for the stationary series, and differencing the series once or more makes the series stationary removing the observed heterogeneity. The condition of stationarity ensures that the moving average parameters are invertible and the autoregressive parameters are stable within a particular range in the estimated model [Hamilton 1994].

Consider a time-invariant and stable linear filter and a stationary input time series X_t with $\mu_x = E(X_t)\ and\ \tau_x(v) = Cov\ (X_t, X_{t+v})$ the output time series (Y_t) is also a stationary time series with:

$$E(Y_t) = \mu_y = \sum_{-\infty}^{\infty} \psi_i \mu_x$$

$$Cov\ (Y_t, Y_{t+i}) = \tau_y(i) = \sum_{v=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \psi_i\, \psi_j \tau_x\ (v - j + i)$$

Thus, we can show that;

$$Y_t = \mu + \sum_{i=0}^{\infty} \psi_i\, e_{t-i}; \tag{7}$$

is also a stationary stable linear process with white noise

Writing this in terms of the backshift operator;

$$Y_t = \mu + \psi_0 e_t + \psi_1 e_{t-1} + \psi_2 e_{t-2} + \cdots \quad \xrightarrow{yields} \quad Y_t = \mu + \sum_{i=0}^{\infty} \psi_i\, ß^i e_t$$

Where $\Psi(ß) = \sum_{i=0}^{\infty} \psi_i\, ß^i$

$$Y_t = \mu + \Psi(ß) e_t \tag{8}$$

This is called the infinite moving average which is a general class for any stationary time series.

Testing for stationarity requires that we test for the existence or the inexistence of a unit root. The unit root test determines the presence of either a deterministic or stochastic trend in the time series and must be conducted for both the seasonal and non-seasonal parts of the series. Examples of unit root tests include ADF and the KPSS tests [76, 77].

1. Unit Root Testing

The equations for the stationary process and the unit root process are:

$$\varphi_t = \Omega_1 \rho_{t-1} + e_t\ ;\ \text{unit process} \tag{9}$$

$$\varphi_t = \Omega_0 + \Omega_1 \rho_{t-1} + e_t\ ; \text{stationary process} \tag{10}$$

If the $\Omega_1 = 1$, then the series is said to have a unit root

In 1979, Dickey and Fuller (ADF) derived the unit root test which tests the presence of a unit root in a time series versus the stationary process [76]. The hypothesis was formulated as below:

H_0: $\Omega_1 =$ the time series has a unit root
H_1 : $\Omega_1 =$ the time series is stationary

We reject the null hypothesis if the test statistic of the ADF is less than the critical value.

To conduct the ADF test, fit a regression model;

$$\dot{Y}_t = aY_{t-1} + b_1\dot{Y}_{t-1} + b_2\dot{Y}_{t-2} + \cdots + b_p\dot{Y}_{t-p} \quad (11)$$

Where ; $\dot{\mathrm{Y}}_{\mathrm{t}}$ represents the differenced series $\mathrm{Y_t} - \mathrm{Y_{t-1}}$; and
; the number of lagged terms p is usually set to 3

The value of a is estimated using ordinary least squares regressions (OLS). If the original series needs to be differenced, the estimated value of a will be close to zero. If Y_t is stationary, the estimated value of a will be negative.

According to Kwiatkowski-Phillips-Schmidt-Shin (KPSS), the hypothesis for the unit root test [77]:

H_0: $\Omega_1 =$ the time series is stationary
H_1 : $\Omega_1 =$ the time series has a unit root

We fail to reject the null hypothesis if the test statistic of the KPSS is less than the critical value.

The KPSS test was developed to complement other unit root tests as they are assumed to have low power with respect to long-run trend and near unit-root processes. To conduct the KPSS test, we consider a three component time series representation $Y_1, Y_{2,} \ldots, Y_n$ as the sum of a random walk, a deterministic time trend and a stationary residual.

$$Y_t = bt + (r_t + \alpha) + e_t \quad (12)$$

Where; $r_t = r_{t-1} + u_t$ is a random walk, and the initial value $r_0 = \alpha$ is the intercept
; u_t are $iid \sim N(0, {\sigma_u}^2)$

2. Differencing

This is process of making a non-stationary time series stationary. The *order of differencing* denoted by ***d*** is the number of times the original series must be differenced in order to achieve stationarity. For instance, after the first difference, the series now has *(n-1)* values since each value of the time series is subtracted from the immediate previous value and so on [78].

$$\omega_t = \Delta Y_t = Y_t - Y_{t-1}\,; \quad (13)$$

a new time series ω_{t} having $(\mathrm{n}-1)$values

A difference of order two means that the first order differenced series is differenced again;

$$\omega_t{}' = \omega_t - \omega_{t-1} = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) = Y_t - 2Y_{t-1} + Y_{t-2}$$

Which is a new time series $\omega_t{}'$

The backward shift operator ß is convenient for describing the process of differencing. For example;

$$\Delta Y_t = Y_t - Y_{t-1} = Y_t - ßY_t = (1 - ß)Y_t;$$

for the first order difference

In general, the ***d***th-order difference can be expressed as:

$$\Delta^d Y_t = (1 - ß)^d Y_t \tag{14}$$

When the time series has the seasonality component, seasonal differences should be applied. A seasonal difference is represented as: $(1 - ß^s)$; while a seasonal difference followed by a first difference is: $(1 - ß)(1 - ß^s)$

Hence:

$$(1 - ß)(1 - ß^s)Y_t = (1 - ß - ß^s + ß^{s+1})Y_t \tag{15}$$
$$= Y_t - Y_{t-1} - Y_{t-s} + Y_{t-s-1}$$

3.2.5 White Noise

The concept of white noise was first coined in engineering applications, with the analogy of the presence of all possible white light periodic oscillations having an equal strength. The collection of uncorrelated random variables ω_t with a mean 0 and finite variance ${\delta_\omega}^2$ was used to model noise. In most cases the noise is required to be identically and independently distributed (iid), with a particularly useful series, the Gaussian white noise, where the $\omega_t(s)$ are independent normal random variables with a mean 0 and finite variance ${\delta_\omega}^2$. In the analysis of time series data, the random error which is the inexplicable part of the series, is referred to as white noise.

3.2.6 Autocorrelation Function (ACF)

The ACF is very useful in describing the procedures involving in model development and model checking. It measures the degree of correlation between neighboring observations in a time series. The covariance between Y_t and Y_{t-i} is called the auto covariance at lag *i* and is defined by;

$$\tau_i = COV[(Y_t, Y_{t-i})] \tag{16}$$

The auto covariance at lag i = 0 is just the variance of the time series; $\tau_0 = {\delta_Y}^2$

The autocorrelation at any lag *i* is defined as; $CORR(Y_t, Y_{t-i})$ and is measured by:

$$CORR(Y_t, Y_{t-i}) = \rho_i = \frac{\tau_i}{\tau_0} = \frac{COV[(Y_t, Y_{t-i})]}{\delta_{Y_t}\,\delta_{Y_{t-i}}} = \frac{E[(Y_t - \mu_y)(Y_{t-i} - \mu_y)]}{[E(Y_t - \mu_y)^2\,(Y_{t-i} - \mu_y)^2]} \tag{17}$$

Where Y_t is the observations at time t,
Y_{t-i} observation at time $(t - i)$,
τ_i the the covariance between Y_t and Y_{t-i} $\{COV[(Y_t, Y_{t-i})]\}$,
$\delta_{Y_t}\,\delta_{Y_{t-i}}$ the variance of the time series and μ_y the mean

The collection of ρ_i values, $where\ i = 0,1,2 \ldots$, is the Auto Correlation function (ACF). If the data are non-stationary, then $\rho_i = 1$, for all values of *i.* But, if the series is stationary $|\rho_i| < 1$. Thus for any defined process where $\rho_0 = 1$, it is also true that $\rho_i = \rho_{-i}$, which means, the ACF is symmetric around zero.

The ACF is estimated by the sample ACF where the autocorrelation and auto covariance functions are estimated from a time series of finite length say $Y_t = y_1, y_2, \ldots y_n$. Sample auto covariance:

$$c_i = \hat{\tau_i} = \frac{1}{n}\sum_{t=1}^{n-i}(Y_t - \mu_y)(Y_{t-i} - \mu_y) \text{ where } i = 0,1,2 \ldots n \tag{18}$$

And the Sample auto correlation:

$$r_i = \hat{\rho}_i = \frac{c_i}{c_o} \text{ where } i = 0,1,2 \ldots n \tag{19}$$

In general, all stationary AR and ARMA processes exhibit ACF patterns that die down to zero as ***i*** increases while the ACF of a non-stationary AR process is always 1 for all values of *i.* MA processes are always stationary with ACF's that cut off after certain lags. The theoretical behavior of the ACF and PACF is explained in the next section.

According to Yule *et al*, a stationary time series can be seen as the weighted sum of the present and past random "disturbances" [79]. Given a stationary time series model Y_t that is an AR (p) process, the Yule-Walker equations for the ACF of an AR (p) process:

$$\rho_1 = \phi_1 + \rho_1\phi_2 \tag{20}$$

$$\rho_2 = \rho_1\phi_1 + \phi_2 \tag{21}$$

Replacing $\rho_1 = r_1$ and $\rho_2 = r_2$ we obtain

$$r_1 = \phi_1 + r_1\phi_2$$

$$r_2 = r_1\phi_1 + \phi_2$$

We solve the equations to obtain the estimates:

$$\hat{\phi}_1 = \frac{r_1(1 - r_2)}{1 - {r_1}^2} \; and \; \hat{\phi}_2 = \frac{r_2 - {r_1}^2}{1 - {r_1}^2} \tag{22}$$

3.2.7 Partial Autocorrelation Function (PACF)

The PACF measures the degree of association or correlation between Y_t and Y_{t-i} when the effects of the other lower-order lags are held constant. For instance, the partial autocorrelation between Y_t and Y_{t-3} is the amount of correlation not explained by their common correlation with Y_{t-1} and Y_{t-2}. It is considered when we are unaware of the appropriate autoregressive process to fit the time

series, and the partial autocorrelations at lags 1, 2, 3 form the PACF. Thus, for an AR (p) model, the PACF between Y_t and Y_{t-i} for *i > p* should be equal to zero.

Table 1: Behavior of Theoretical ACF and PACF patterns

Model	***ACF***	***PACF***
AR (p)	Dies down/exponential decay	Cut off after lag p
MA (q)	Cut off after lag q	Dies down/exponential decay
ARMA (p,q)	Dies down/exponential decay	Dies down/exponential decay

3.3 ARMA Models and the Box Jenkins Methodology

3.3.1 ARMA Models

In 1926, Yule introduced the Auto Regressive (AR) models which were later complemented in 1937 by Slutsky who introduced the Moving Average (MA) models [79, 80]. In 1938, Wold combined both the AR and MA models into ARMA process showing that ARMA models could be used in the analysis of stationary time series by ensuring that appropriate number AR terms, order of p, and MA terms, order of q were clearly specified [81]. This meant that any time series Y_t could be modelled as a function of its past values Y_{t-1}, Y_{t-2}, … and the past error terms e_{t-1}, e_{t-2} , …. However, to use the ARMA models; the original series Y_t must first be transformed to be stationary around the mean and variance; second, there should be a specification of the appropriate orders of p and q; third, the estimation of the parameters $\emptyset_1, \emptyset_2 \dots \emptyset_p \text{ and/or } \theta_1, \theta_2 \dots \theta_q$ using non-linear optimization procedures to minimize the sum of squares or maximize the likelihood function.

In the 1970's, Box Jenkins popularized the use of ARMA modelling through an iterative procedure for Model Identification, Model Estimation, Model Checking and Diagnosis and Model Forecasting [78]. This involved i). Providing a guideline for making a time series stationary around the mean and variance, ii). Provision of visual inspection guides and computer programs for identification of appropriate orders of p and q, and parameter estimation, iii). Procedural diagnostic checks for residual analysis in which case process (i), (ii), and (iii) are repeated if the residuals are not white noise. The model is only considered final and used for the purposes of forecasting or control if the residuals are random and conform to the Gaussian distribution.

The Box Jenkins approach proposed differencing as the procedure of making a series stationary and introduced the "integrated" part of the ARMA process making it an ARIMA process. The ARIMA models were popularized due to their wide applicability in modelling time series data as they account for all the components if the time series'. Differencing is a convenient way of eliminating the seasonal component of the time series. However, fitting an appropriate ARMA model to the differenced series

gives rise to a special case of general seasonal ARIMA (SARIMA) which models seasonal time series data [78].

SARIMA models have recently become popular due to their capability to account for the seasonal and non-stationary behavior observed in many time series'. In many cases, the dependence of the past tends to occur most strongly at multiples of some underlying seasonal lag ***s***. With monthly data for instance, there is a strong yearly component occurring at lags that are multiples of *s* = 12 due to activity connection to the calendar year. Many physical and biological processes are likely to be disposed to seasonal fluctuations hence the need to introduce AR and MA polynomials that take care of the seasonal lags [78, 82].

These methods are described below:

1. **AR Model**

AR models are pegged on the idea that the current time series value Y_t can be expressed as a function of ***p*** past values in the series $Y_{t-1}, Y_{t-2}, \ldots\ Y_{t-p}$. *P* determines the number of steps taken back into the past to forecast the current value.

Consider the general class model from *Eq. 8* for a stationary time series:

$$Y_t = \mu + \Psi(ß)e_t = \mu + \sum_{i=0}^{\infty} \psi_i ß^i e_t$$

Following an exponential decay pattern, we will set $\psi_i = \emptyset^i$ where $|\emptyset| < 1$ guarantees the "exponential decay". Thus the weights for the current random term going back will be 1, $\emptyset^1, \emptyset^2$ …

$$Y_t = \mu + \sum_{i=0}^{\infty} \emptyset^i e_{t-i} \tag{23}$$

$$Y_t = \mu + e_t + \emptyset e_{t-1} + \emptyset^2 e_{t-2} + \cdots$$

$$Y_{t-1} = \mu + e_{t-1} + \emptyset e_{t-2} + \emptyset^2 e_{t-3} + \cdots$$

Combining the two equations and equating;

$$Y_{t-1} - \mu = e_{t-1} + \emptyset e_{t-2} + \emptyset^2 e_{t-3} + \cdots$$

Where; $\emptyset Y_{t-1} - \emptyset\mu = \emptyset e_{t-1} + \emptyset^2 e_{t-2} + \cdots$

$$Y_t = (\mu - \emptyset\mu) + \emptyset Y_{t-1} + e_t$$

Setting; $\delta = (\mu - \mu\emptyset)$

$$Y_t = \delta + \emptyset_1 Y_{t-1} + e_t\,; \tag{24}$$

Which is the first-order autoregressive process, AR (1)

Therefore, the general AR (p) model is:

$$Y_t = \delta + \emptyset_1 Y_{t-1} + \emptyset_2 Y_{t-2} + \cdots + \emptyset_p Y_{t-p} + e_t = e_t + \delta; \tag{25}$$

$$\hat{Y}_t - \sum_{i=1}^{p} \emptyset_i Y_{t-i} = e_t + \delta$$

Or using the backshift operator:

$$Y_t - \emptyset_1 Y_{t-1} - \emptyset_2 Y_{t-2} - \cdots - \emptyset_p Y_{t-p} = e_t + \delta;$$

$$Y_t - \emptyset_1(ß)Y_t - \emptyset_2(ß^2)Y_t - \cdots - \emptyset_p(ß^p)Y_t = e_t + \delta;$$

$$\left(1 - \emptyset_1(ß) - \emptyset_2(ß^2) - \cdots - \emptyset_p(ß^p)\right) Y_t = e_t + \delta;$$

$$\emptyset\,(ß)Y_t = e_t + \delta \tag{26}$$

Where ; e_t is the error term
; $\delta = (\mu - \mu\emptyset)$ is the mean of the AR process
; μ is the mean of the time series
; $\emptyset\,(ß)$ is the autoregressive operator of order p defined by

$$\emptyset(ß) = 1 - \emptyset_1(ß) - \emptyset_2(ß^2) - \cdots - \emptyset_p(ß^p)$$

Where Y_t is stationary and $\emptyset_1 \ldots \emptyset_p$ are constants. The AR (p) is a polynomial of degree *p.*

2. **MA Model**

The MA model assumes that the time series value Y_t can be expressed as a function of ***q*** past values of the random error terms in the series. A MA (q) process is always stationary regardless of values of the weight [61].

The general MA (q) model:

$$Y_t = \mu + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q} = \theta\,(ß)e_t + \mu; \tag{27}$$

$$Y_t = \mu + e_t - \sum_{i=1}^{q} \theta_i e_{t-i}$$

Or using the backshift operator:

$$Y_t - \mu = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q} ;$$

$$Y_t - \mu = e_t - \theta_1(ß)e_t - \theta_2(ß^2)e_t - \cdots - \theta_q(ß^q)e_t ;$$

$$Y_t - \mu = \left(1 - \theta_1(ß) - \theta_2(ß^2) - \cdots - \theta_q(ß^q)\right) e_t ;$$

$$Y_t - \mu = \theta(ß)e_t ; \tag{28}$$

$Where$; $(Y_t - \mu)$ is the white noise;
; μ is the mean of the time series
; $\theta(ß)$ is the moving average operator of order q defined by

$$\theta(ß) = 1 - \theta_1(ß) - \theta_2(ß^2) - \cdots - \theta_q(ß^q)$$

3. **ARMA Model**

ARMA models represent the linear relationship between the current and past values in a time series through the combination of two processes: AR process expressing $\boldsymbol{Y_t}$ as a function of the past values and a MA process expressing $\boldsymbol{Y_t}$ as a function of past values of the error term ***e*.**

The general ARMA (p, q) model:

$$\hat{Y}_t = \delta + \emptyset_1 Y_{t-1} + \emptyset_2 Y_{t-2} + \cdots + \emptyset_p Y_{t-p} + e_1 - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q} \quad (29)$$

$$\hat{Y}_t = \delta + \sum_{i=1}^{p} \emptyset_i Y_{t-i} + e_t - \sum_{i=1}^{q} \theta_i e_{t-i}$$

Or using the backshift operator:

$$\emptyset\ (ß)\ Y_t = \delta + \theta(ß)\ e_t \quad (30)$$

$Where$; $e_t \sim (\mu, \sigma^2)$ the white noise process;
; p and q are autoregressive and moving average orders

4. **ARIMA Model**

A process Y_t is said to be ARIMA (p, d, q) if the stationary series obtained after differencing is an ARMA (p, q) shown below:

$$Z_t = Ð^d Y_t = (1 - ß)^d Y_t \quad (31)$$

In ARIMA models the random error term e_t is assumed to be the white noise which is identically and independently distributed with a mean of 0 and common variance σ^2; $e_t \sim iid\ (0,\ \sigma^2)$

The general ARIMA (p, d, q) model; with an AR part of order *p*, a MA part of order *q* and with a *d* order differencing is given by:

$$[1 - \emptyset_1(ß) - \emptyset_2(ß^2) - \cdots - \emptyset_p(ß^p)]\ (1 - ß)^d\ Y_t = [1 - \theta_1(ß) - \theta_2(ß^2) - \cdots - \theta_q(ß^q)\ e_t]$$

$$\emptyset\ (ß)(1 - ß)^d\ Y_t = \delta + \theta(ß)\ e_t \quad (32)$$

OR

$$Z_t = \sum_{i=1}^{p} \emptyset_i Z_{t-i} - \sum_{i=1}^{q} \theta_i e_{t-i} + \delta + e_t \quad (33)$$

Where $\emptyset s$ and θs are coefficients of the AR and MA processes;
p and q the number of past values of Y_t and the error term.

5. **Seasonal ARIMA (SARIMA) Models**

SARIMA models are a form of ARIMA models that incorporate the seasonality component (periodicity) of the time series. These models rely on seasonal lags and seasonal differences to fit seasonal patterns. Consider a time series with monthly observations. The current value Y_t might depend on the previous month's value (local in-time trend) and the previous similar month from a year ago (long range trend). For instance the value for March might depend on the values for February and January of the same year and the value for March the previous year.

A pure seasonal time series only has seasonal AR and/or MA parameters. Seasonal parameters represent the autoregressive and moving average relationships between the time series data segmented by multiples of the number of periods per season. The auto-covariance between any two adjacent values is zero except when $\mathrm{s} = 1$ at $\mathrm{i} = 12$, where the auto-covariance overlaps and is significantly different from zero. We ignore all the lags between $\mathrm{i} = 1$ and $\mathrm{i} = 11$ and only consider lags at multiples of the period.

A general SAR (p) model is given by:

$$Y_t = \sum_{j=1}^{P} \Phi_{js} Y_{t-js} \tag{34}$$

A general Seasonal MA (q) model is given by:

$$Y_t = e_t + \sum_{j=1}^{Q} \Theta_{js} e_{t-js} \tag{35}$$

While the mixed SAR and SMA is given by:

$$Y_t = \sum_{j=1}^{P} \Phi_{js} Y_{t-js} + \sum_{j=1}^{Q} \Theta_{js} e_{t-js} + e_t \tag{36}$$

Or

$$\Phi_p\,(ß^s)\,Y_t = \Theta_q(ß^s)\,e_t \tag{37}$$

Where Y_{t-s} is of order s, Y_{t-2s} *is* of order $2s$ … Y_{t-ps} is of order ps; and
P and Q are the orders of the ARMA process

When the series does not only have the seasonal AR and/or MA parameters but also displays patterns for the local trend, we combine the local model which models the most recent months and the seasonal model which models what happened the previous year and the year before. When the seasonal aspects are combined with the regular ARMA model, we obtain a seasonal ARMA (p, q)*(P, Q)^s. Which is:

$$\Phi_p\,(ß^s)\emptyset\,(ß)Y_t = \Theta_q(ß^s)\theta(ß)e_t \tag{38}$$

With both regular and seasonal AR and MA terms.

For example an ARMA (1, 1) (1, 1)^{12} becomes:

$$(1 - \Phi_1 ß^{12})(1 - \emptyset_1 ß)Y_t = (1 - \Theta_1 ß^{12})(1 - \theta_1 ß)e_t.$$

Expanding this leads to a complex equation with too many terms and hence we only focus on the seasonal lags and the influence of the local trends (lags near zero). The SARMA model accounts for the lags neighboring the seasonal peaks.

The general multiplicative seasonal ARIMA (SARIMA) model of orders (p, d, q)*(P, D, Q)^s obtained after applying both regular and seasonal differencing is represented by;

$$\Phi_p\ (ß^s)\ \emptyset\ (ß)(1-\ ß^s)^D\ (1-\ ß)^d\ \ Y_t = \delta\ +\ \Theta_Q(ß^s)\ \theta(ß)\ e_t \tag{39}$$

The polynomials $\emptyset\ (ß)$ and $\theta(ß)$ denote the ordinary AR and MA components of orders p and q respectively and $\Phi_P\ (ß^s) = \left(1-\ \Phi_1 ß^s - \cdots -\ \Phi_p ß^{sp}\right)$ and $\Theta_Q\ (ß^s) = (1-\ \Theta_1 ß^s - \cdots -\ \Theta_Q ß^s Q)$ the seasonal AR and MA components. The ***D*** in the equation is the seasonal differencing element.

Suppose we have a time series:

$$Y_t = \varsigma_t + X_t \tag{40}$$

Where: ς_t the seasonal trend with s is: X_t is a stationary process

We do a seasonal difference to get rid of the seasonal trend with *s =12*:

$$(1-\ ß^s)\ Y_t = (1-\ ß^s)\ \varsigma_t + (1-\ ß^s)\ X_t$$

$$Y_t -\ Y_{t-s} = (\varsigma_t -\ \varsigma_{t-s}) + (X_t -\ X_{t-s})$$

This means that $\varsigma_t -\ \varsigma_{t-s} = 0$ since *s =12,* removing the seasonal trend. We also consider the stationary process as white noise and substitute to:

$$Y_t -\ Y_{t-s} =\ e_t -\ e_{t-s} \tag{41}$$

However, getting rid of the seasonal trend introduces dependency at lags which are multiples of the period. Seasonal difference accounts for this by additionally getting rid of the seasonal random walk type of non-stationarity. According to Box-Jenkins, the maximum values of all the parameters is two (2), making the operator polynomials simple expressions.

3.3.2 Box-Jenkins Methodology

The Box-Jenkins approach refers to a set of procedures for identifying, estimating, checking and even forecasting a time series model within the class of ARIMA models. ARIMA models use historical or past values of a variable of interest, and/or the random error term as explanatory variables to forecast its future values. The variable of interest should be a time series with equally spaced time intervals.

Let's consider a discrete time series $Y_t = Y_1, Y_2, \ldots Y_n$.

The underlying principle of the Box-Jenkins procedure is that it considers the observed time series Y_t as an output of inputs from an unobservable random process. These inputs are a series of independent random shocks e_t , which are assumed to be normally distributed with a zero mean and a constant variance, and referred to as white noise. In simple terms, the approach views a time series as a result of the transformation of a white noise process through a "linear filter" to obtain a particular set of

outputs, which we now refer as the observed time series. ARIMA models of this form assume that the observed time series values may be dependent on;

1. The previous and current inputs (random shocks/white noise)
2. The previous output values of the time series under study [$Y_{t-1}, Y_{t-2}, \ldots$] in varying proportions

However, the Box-Jenkins approach assumes that the conditions at which the data is collected remain the same over time. If the assumption is not appropriate, a transfer function-noise model where a set of input variables which might have an effect on the time series are added to the model.

1. Model Building Strategy

Box-Jenkins defined a four step iterative procedure for Model Identification, Model Estimation, Model Checking and Diagnosis, and Model Forecasting. The steps are described below.

Step 1: Model Identification

This step involves the identification of a tentative model, whether multiplicative or additive, and establishing the number of parameters involved and their combinations. This is done through analysis of historical data.

Visual inspection of time series plots is the first assessment tool. The first step is to consider the ACF and the PACF graphs to determine whether the series is stationary or not. Additionally, unit root tests are performed on the data to confirm stationarity and/or to make sure that differencing is necessary. If the series is non-stationary the series has to be either differenced to make it stationary in mean, or transformed if the covariance between any two observations Y_t and Y_{t+i} is not constant over time. However, differencing should be done with care to avoid the issue of over-differencing which might introduce dependence where none exists.

Secondly, a proposed model is estimated by finding the initial values (p, q, d) of the model parameters. This is done through looking at the significant coefficients in the ACF and PACF plots. The AC's and PAC's are compared with theoretical values to investigate candidate models. This procedure leads to deployment of various diagnostic tests which are conducted to first confirm stationarity of the series and check model fit. If the residual analysis confirms inadequacy of any particular model, a new model is proposed. The process is repeated until potential models are identified.

When the time series data has the seasonality component, seasonal differencing is recommended to make the data stationary. A seasonal difference is the difference between an observation Y_t and the

corresponding observation from the previous year Y_{t-s}; where s is the length of the season. It is recommended that the seasonal differencing be done before the first difference of the whole series as this might make the data stationary.

Step 2: Model Estimation

This is the process of estimating the model parameters after selecting a tentative model. The parameter estimates should be significant, with each providing a substantial contribution to the model for the most accurate forecasts. There are a number of ways to estimate autoregressive and moving averages parameters in ARMA models such as:

- Least Squares Estimation Method
- Maximum Likelihood Estimation Method

1. Least Squares Estimation

Consider an AR (1) model:

$$Y_t - \mu = \emptyset[Y_{t-1} - \mu] + e_t \tag{42}$$

Which can be regarded as a regression with predictor $Y_{t-1}$and response variable Y_t.

We solve by minimizing the sum of squares of the differences: $(Y_t - \mu) - (\emptyset(Y_{t-1} - \mu))$ and summing from $t-2$ to $t-n$ since we only have $Y_1, Y_2, \dots, Y_n$ observations;

Let:

$$S_c\left(\emptyset, \mu\right) = \sum_{t=2}^{n}[(Y_t - \mu) - \emptyset(Y_{t-1} - \mu)]^2 \tag{43}$$

Be the *conditional sum of squares.*

We minimize $S_c(\emptyset, \mu)$ given $Y_1, Y_2, \dots, Y_n$ to estimate $\emptyset$ and μ

Consider with respect to μ

$$\frac{\partial S_c}{\partial \mu} = \sum_{t=2}^{n}[(Y_t - \mu) - \emptyset(Y_{t-1} - \mu)]\,(-1 + \emptyset) = 0$$

$$\mu = \frac{1}{(n-1)(1-\emptyset)}\left[\sum_{t=2}^{n} Y_t - \emptyset\sum_{t=2}^{n} Y_{t-1}\right] \tag{44}$$

For large values of *n*

$$\frac{1}{(n-1)}\sum_{t=2}^{n} Y_t \approx \frac{1}{(n-1)}\sum_{t=2}^{n} Y_{t-1} \approx \bar{Y}$$

Which reduces to: $\hat{\mu} = \frac{1}{(1-\emptyset)}(\bar{Y} - \emptyset\bar{Y}) = \bar{Y}$

Consider with respect to ϕ and minimize $S_c(\phi, \bar{Y})$

$$\frac{\partial S_c(\emptyset, \bar{Y})}{\partial \emptyset} = \sum_{t=2}^{n} [(Y_t - \bar{Y}) - \emptyset(Y_{t-1} - \bar{Y})]\,(Y_{t-1} - \bar{Y}) = 0$$

$$\hat{\phi} = \frac{\sum_{t=2}^{n} [(Y_t - \bar{Y})\,(Y_{t-1} - \bar{Y})]}{\sum_{t=2}^{n} (Y_{t-1} - \bar{Y})^2} \tag{45}$$

To estimate $\emptyset$ for an AR (p) model, we need to consider the second order AR (2) model and replace μ with $\bar{Y}$.

$$S_c(\emptyset_1, \emptyset_2, \bar{Y}) = \sum_{t=3}^{n} [(Y_t - \bar{Y}) - \emptyset_1(Y_{t-1} - \bar{Y}) - \emptyset_2(Y_{t-2} - \bar{Y})]^2 \tag{46}$$

Setting $\frac{\partial S_c}{\partial \emptyset_1} = 0$

We have: $-2\sum_{t=3}^{n} [(Y_t - \bar{Y}) - \emptyset_1(Y_{t-1} - \bar{Y}) - \emptyset_2(Y_{t-2} - \bar{Y})]\,(Y_{t-1} - \bar{Y}) = 0$

Re-writing the equation to

$$\sum_{t=3}^{n} (Y_t - \bar{Y})(Y_{t-1} - \bar{Y}) = \left[\sum_{t=3}^{n} (Y_{t-1} - \bar{Y})^2\right] \emptyset_1 + \left[\sum_{t=3}^{n} (Y_{t-1} - \bar{Y})(Y_{t-2} - \bar{Y})\right] \emptyset_2$$

And dividing both sides by $\sum_{t=3}^{n} (Y_t - \bar{Y})^2$

We obtain:

$$r_1 = \emptyset_1 + r_1 \emptyset_1 \tag{47}$$

Using $\frac{\partial S_c}{\partial \emptyset_2} = 0$

We have:

$$r_2 = r_1 \emptyset_1 + \emptyset_2 \tag{48}$$

Consider an MA (1) model: $Y_t = e_t - \theta e_{t-1}$

Invertible MA (1) models can be expressed as:

$$Y_t = e_t - \theta Y_{t-1} - \theta^2 Y_{t-2} - \cdots; \text{of infinite order} \tag{49}$$

The method of least squares can be carried out by choosing θ that minimizes

$$S_c(\theta) = \sum [e_t]^2 = \sum [Y_t + \theta Y_{t-1} + \theta^2 Y_{t-2} + \cdots]^2 \tag{50}$$

Where $e_t = e_t(\theta)$ a function of the observed series and unknown parameter θ

For moving averages, we result to numerical optimization, because first, the least squares is non-linear in the parameters and it is not possible to minimize $S_c(\theta)$ by taking a derivative with respect to θ and setting it to zero. We thus consider evaluating $S_c(\theta)$ for a single value of θ for our series.

We could re-write the MA (1) model to:

$$e_t = Y_t + \theta e_{t-1}$$

to calculate $e_1, e_2, \dots, e_n$ recursively if we have the initial value e_0. A common approach is to set $e_0 = 0$ the expected value for the mean of the random part.

Thus: $e_1 = Y_1, e_2 = Y_2 + \theta e_1, \dots, e_n = Y_n + \theta e_{n-1}$

We then calculate: $S_c(\theta) = \sum[e_t]^2$ conditioning on $e_0 = 0$

For higher order MA models, we compute $e_t = e_t(\theta_1, \theta_1, \dots, \theta_q)$ recursively from:

$$e_t = Y_t + \theta e_{t-1} + \cdots + \theta_q e_{t-q} \text{: with } e_0 = e_{-1} = \cdots = e_{-q} = 0 \qquad (51)$$

The sum of squares is minimized jointly in $\theta_1, \theta_1, \dots, \theta_q$ in a multivariate methods.

For Mixed AR and MA models

Consider ARMA (1, 1) case where

$$Y_t = \emptyset Y_{t-1} + e_t - \theta e_{t-1}$$

As in pure MA case, we consider: $e_t = e_t(\emptyset, \theta)$ and wish to minimize: $S_c(\emptyset, \theta) = \sum[e_t]^2$

Re-writing the equation to: $e_t = Y_t - \emptyset Y_{t-1} + \theta e_{t-1}$

To obtain $e_1$we have to have Y_0. One approach is to set $Y_0 = 0$ or to $Y_0 = \bar{Y}$ if the series has a non-zero mean. The other approach is to avoid Y_0 completely by begin recursion at $t - 2$ and minimize: $S_c(\emptyset, \theta) = \sum[e_t]^2$

For the general ARMA (p,q) model, consider the computation:

$$e_t = Y_1 - \emptyset_1 Y_{t-1} - \emptyset_2 Y_{t-2} - \cdots - \emptyset_p Y_{t-p} + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q} \qquad (52)$$

With $e_p = e_{p-1} = \cdots = e_{p+1-q} = 0$ and then minimize $S_c(\emptyset_1, \emptyset_2, \dots, \emptyset_p, \theta_1, \theta_2, \dots, \theta_q)$ numerically to obtain the conditional least squares parameter estimates

2. Maximum Likelihood Estimation

Given a time series $Y_1, Y_2, \dots, Y_n$ the likelihood function L is defined as the joint probability of obtaining the observed series data. It is considered a function of the unknown model parameters with the observed data held fixed. Maximum likelihood estimators are those parameter values for which the actual data observed are most likely, that is, values that maximize the likelihood function.

Consider an AR (1) model

We assume that the error terms are uncorrelated, normally distributed as $N(0, \sigma_e)$. The probability distribution function of each e_t follows that of a normal distribution and is:

$$(2\pi{\sigma_e}^2)^{-1/2}\, exp\left\{-\frac{{e_t}^2}{2{\sigma_e}^2}\right\}\text{ ; for } -\infty < \mathrm{e_t} < \infty \tag{53}$$

By independence, the joint probability distribution for $e_2, e_3, \dots, e_n$ is

$$(2\pi{\sigma_e}^2)^{-(n-1)/2} \exp\left\{-\frac{1}{2{\sigma_e}^2}\sum_{t=2}^{n}[e_t]^2\right\}$$

Consider the equations:

$$\begin{aligned} Y_2 - \mu &= \emptyset[Y_1 - \mu] + e_2 \\ Y_3 - \mu &= \emptyset[Y_2 - \mu] + e_3 \\ &\vdots \\ Y_n - \mu &= \emptyset[Y_{n-1} - \mu] + e_n \end{aligned}$$

Conditioning $Y_1 = y_1$ these equations define a linear transformation between $e_2, e_3, \dots, e_n$ and $Y_2, Y_3, \dots, Y_n$ (Jacobian Transformation). Thus the joint pdf of $Y_2, Y_3, \dots, Y_n$is obtained as:

$$f(y_2, y_3, \dots, y_n|y_1 = \ (2\pi{\sigma_e}^2)^{-\frac{n-1}{2}} * exp\left\{-\frac{1}{2{\sigma_e}^2}\sum_{t=2}^{n}[(y_t - \mu) - \emptyset(y_{t-1} - \mu)]^2\right\} \tag{54}$$

We now consider the marginal distribution of Y_1 which is a normal distribution with mean μ and variance ${\sigma_e}^2/(1-\emptyset^2)$. The joint probability distribution of $Y_1, Y_2, \dots, Y_n$is thus obtained by multiplying the conditional probability distribution above by the marginal distribution of Y_1.

The likelihood function for an AR (1) model defined as a function of the parameter $\emptyset, \mu, {\sigma_e}^2$ is given by:

$$L(\emptyset, \mu, {\sigma_e}^2) = \ (2\pi{\sigma_e}^2)^{-n/2}(1-\emptyset^2)^{1/2}\, exp\left\{-\frac{1}{2{\sigma_e}^2}S(\emptyset, \mu)\right\} \tag{55}$$

Where

$$S(\emptyset, \mu) = \sum_{t=2}^{n}[(Y_t - \mu) - \emptyset(Y_{t-1} - \mu)^2 + (1-\emptyset^2)(Y_t - \mu)$$

Which is referred to as the un*conditional sum of squares.*

Practically, we obtain the log of the likelihood function. For an AR (1) model, the log-likelihood function is:

$$l(\emptyset, \mu, {\sigma_e}^2) = -\frac{n}{2} log(2\pi) - \frac{n}{2} log({\sigma_e}^2) + \frac{1}{2} log(1 - \emptyset^2) - \frac{1}{2{\sigma_e}^2} S(\emptyset, \mu) \quad (56)$$

For given values of $\emptyset$ and μ we can maximize $l(\emptyset, \mu, {\sigma_e}^2)$ systematically with respect to ${\sigma_e}^2$ in terms of the unknown values of $\emptyset$ and μ to obtain:

$$\hat{\sigma}_e{}^2 = \frac{S_c(\hat{\emptyset}, \hat{\mu})}{n} \quad (57)$$

to obtain estimators with less bias, we divide by $n - 2$ instead of n since we are estimating two parameters.

To estimate $\emptyset$ and μ we compare the unconditional sum of squares $S(\emptyset, \mu)$ with the conditional sum of squares $S_c(\emptyset, \mu)$ to get:

$$S(\emptyset, \mu) = S_c(\emptyset, \mu) + (1 - \emptyset^2)(Y_t - \mu)^2$$

Hence:

$$S(\emptyset, \mu) \approx S_c(\emptyset, \mu)$$

Since $S_c(\emptyset, \mu)$ involves a sum of $n - 1$ with $(1 - \emptyset^2)(Y_t - \mu)^2$ not involving n.

We can also minimize $S(\emptyset, \mu)$ numerically as $\frac{\partial S}{\partial \emptyset} = 0$ and $\frac{\partial S}{\partial \mu} = 0$ are non-linear in $\emptyset$ and μ due to the term $(1 - \emptyset^2)(Y_t - \mu)^2$). This results in unconditional least squares estimates.

Following parameter estimation, the most parsimonious model is selected using appropriate criterion from the pool of potential models identified. The Akaike Information Criterion (AIC) and the Schwarz Bayesian Information Criterion (BIC) are the two common goodness-of-fit statistics that are often used for model selection. The model with the lowest AIC and BIC is usually selected as the best fit.

The formula for the AIC is:

$$AIC = -2 \ln L + 2m \quad (58)$$

The BIC is given as:

$$BIC = -2 \ln L + ln(n)m \quad (59)$$

$Where$: m is the number of model parameters ($m = p + P + Q + q$);
: n is the number of residuals that can be computed from the time series

Step 3: Model Diagnostics Checking

This step involves checking model adequacy, and if necessary incorporating potential improvements. Model checking is done through residual analysis. If the identified model is adequate, the residual observations should be transformed to a white noise process where the residuals are random and have the normal distribution. By studying the ACF plots of the residuals, we can establish whether the AC's and PAC's are small and significant enough to consider the model adequate. If the

autocorrelations are large, the values of *p* and/or *q* are adjusted and the model re-estimated until the best fit model is estimated.

The residuals of an ARMA (p, q) can be obtained as below:

$$\hat{e}_t = \hat{Y}_t - \left(\hat{\delta} + \sum_{i=1}^{p} \widehat{\emptyset}_i Y_{t-i} - \sum_{i=1}^{q} \widehat{\theta}_i \hat{e}_{t-i} \right) \tag{60}$$

Consider the sample ACF of the residuals denoted by $r_e(i)$. If the model is appropriate, the residual sample ACF should not have an identifiable structure. That is, the autocorrelation should not be significantly different from zero for all lags greater than 1.

Comparatively, other than considering the residual terms individually, we could obtain an indication if the first *i* residual correlations considered together are sufficient to indicate model adequacy. The approximate Box-Pierce chi-square test statistic is:

$$Q_{BP} = t \sum_{i=1}^{n} {r_e}^2(i) \text{ ; a chisquare dist with } (\mathrm{n} - \mathrm{p} - \mathrm{q})\mathrm{df} \tag{61}$$

If the model is not appropriate or is inadequate, the calculated *Q* will be too large and hence the null hypothesis of model adequacy should be rejected if it exceeds an approximate upper tail point of the chi-square distribution. A modification of the Box-Pierce statistic is the Ljung-Box goodness-of-fit statistics which works better with small samples [83]:

$$Q_{LP} = t(t+2) \sum_{i=1}^{n} \left[\frac{1}{t-i} \right] {r_e}^2(i); \text{ a chisquare dist with } (\mathrm{n} - \mathrm{p} - \mathrm{q})\mathrm{df} \tag{62}$$

Where the squared sample autocorrelation at lag i is weighted by $(t+2)/(t-i)$.

Step 4: Forecasting

Forecasting is the last stage after the model has been identified and fitted. The model may be used to generate forecasts of future values. If we denote the current time as ***t***, the forecast for $\hat{Y}_{t+k}$ is the **k**-period-ahead forecast denoted by $\hat{Y}_{t+k}\,(t)$.

For an ARIMA (p, d, q) process at time $t + k$ (*k* periods in the future) the model is:

$$\hat{Y}_{t+k} = \hat{\delta} + \sum_{i=1}^{p+d} \emptyset_i Y_{t+k-i} + e_{t+k} - \sum_{i=1}^{q} \theta_i e_{t+k-i} \tag{63}$$

Considering the infinite MA representation of a stationary process;

$$Y_{t+k} = \mu + \sum_{i=1}^{\infty} \psi_i \, e_{t+k-i}$$

We can partition the equation as

$$Y_{t+k} = \mu + \sum_{i=1}^{t-1} \psi_i \, e_{t+k-i} + \sum_{i=k}^{\infty} \psi_i \, e_{t+k-i} \quad (64)$$

Where the $\sum_{i=1}^{t-1} \psi_i \, e_{t+k-i}$ component involves the present and past errors, and the $\sum_{i=k}^{\infty} \psi_i \, e_{t+k-i}$ component involves the future errors.

Hence, the best forecast in the sense of the mean square is;

$$\hat{Y}_{t+k}(t) = E(Y_{t+k-i}|Y_t, Y_{t-1} \ldots) = \mu + \sum_{i=k}^{\infty} \psi_i \, e_{t+k-i} \quad (65)$$

Since $E(Y_{t+k-i}|Y_t, Y_{t-1} \ldots) = \begin{cases} 0, & if\ i < k \\ e_{t+k-i}, & if\ i \geq k \end{cases}$

And the forecast error

$$e_t(k) = Y_{t+k} - \hat{Y}_{t+k}(t) = \sum_{i=0}^{k-1} \psi_i \, e_{t+k-i} \quad (66)$$

Where: $E[e_t(k)] = 0$;

$$Var[e_t(k)] = Var\left[\sum_{i=0}^{k-1} \psi_i \, e_{t+k-i}\right] = \sum_{i=0}^{k-1} {\psi_i}^2 \, Var(e_{t+k-i})$$

$$= \sigma^2 \sum_{i=0}^{k-1} {\psi_i}^2$$

$$= \sigma^2(k); \text{where } k = 1, 2, \ldots$$

The variance of the forecast error gets bigger with increasing forecast lead times *k.*

Since the random shocks are assumed to be normally distributed, then the forecast errors will be normally distributed with $N(0, \sigma^2(k))$. We can then obtain the 100(1-α) % prediction intervals for future observations as shown below;

$$P(\hat{Y}_{t+k}(t) - z_{\left[\frac{\propto}{2}\right]}\sigma(k) < Y_{t+k}(t) < (\hat{Y}_{t+k}(t) + z_{\left[\frac{\propto}{2}\right]}\sigma(k) = 1 - \alpha$$

Where $z_{\left[\frac{\propto}{2}\right]}$ is the upper $\left[\frac{\propto}{2}\right]$ percentile of the standard normal distribution $N(0, 1)$. Hence the 100(1-α) percent prediction interval for Y_{t+k} is

$$\hat{Y}_{t+k}(t) \mp z_{\left[\frac{\propto}{2}\right]}\sigma(k)$$

However, the forecast equation has two major issues; it involves infinitely many terms in the past but we only have a finite data amount practically and secondly, it requires knowledge of the strength of the random shocks in the past which is impractical.

The past random shocks are estimated through a one-step-ahead forecasts as below;

$$\hat{e}_t = Y_t - \left[\hat{\delta} + \sum_{i=1}^{p+d} \emptyset_i Y_{t-i} - \sum_{i=1}^{q} \theta_i e_{t-i}\right]; \quad (67)$$

Recursively by setting the initial values of the random shocks to zero for $t < p + d + 1$.

3.4. Forecasting with Decomposition and Exponential Smoothing Methods

The class of ARIMA models has been applied in varied fields for in and post sample forecasting. However, empirical studies have shown that simpler methods such as moving averages, exponential smoothing and regression were equally as good or performed better than the more complex ARIMA models [10, 64, 74]. This could be due to the high levels of randomness in most time series data and the un-definitiveness of the constancy of relationships or patterns. Hence, to investigate whether the ARIMA models adopted above are the most appropriate for model fitting and/or out-of-sample forecasting, the simpler methods are described in detail and compared with the Box-Jenkins method for malaria forecasting.

3.4.1 Seasonal Adjustment using STL

Consider an additive time series: $Y_t = l_t + s_t + e_t = s_t + d_t$

$$\text{Where } d_t = l_t + e_t \text{ is the seasonally adjusted component}$$

Seasonal adjustment involves decomposing a time series and separately forecasting the seasonal component $\hat{s}_t$ and the seasonally adjusted component $\hat{d}_t$. It is assumed that the seasonal component is constant from year to year or changing very slowly.

The seasonal component is forecast by setting each forecast to be equal to the last observed value from the same season (same month) in the previous year. The seasonal naive forecast for time $t + k$ is expressed as:

$$Y_{t+k-hm} \text{ ; where } m = \text{seasonal period and } h = \left[\frac{k-1}{m}\right] + 1$$

To forecast the seasonally adjusted component, exponential smoothing models may be used.

3.4.2 Exponential Smoothing Methods

While simple averages weight past observations equally, exponential smoothing assigns exponentially decreasing weights to older observations where recent observations are given relatively more weight than older observations. Different smoothing parameters determine the weights assigned to observations. It was first suggested by C.C Holt in 1957 to be used for non-seasonal time series with no trend (Simple Exponential Smoothing). In 1958, the methodology was updated to handle trends

(Holt's Linear Trend Method), and in 1965, winters generalized the method to include the seasonality component of a time series hence the "Holt-Winters Method" name [84].

The general class of models for modelling time series data as a function of time can be represented as;

$$Y_t = f(t{:}\,\beta) + e_t$$

Where β is a vector of unknown parameters and e_t the uncorrelated errors

In exponential smoothing, the smoothers are the estimates for the process level (model estimation parameters) in any given constant process. In our case, we need to estimate β_0 which is an estimator for μ, the general mean of the time series.

We introduce the sum of squared errors for the constant process;

$$SSE = \sum_{i=1}^{t} (Y_i - \mu)^2 \tag{68}$$

Assuming that different observations shouldn't have equal influence on the sum of squares, we introduce a set of weights geometrically decreasing in time with a discount factor ϑ on previous observations which is the weighting parameter;

$$SSE^* = \sum_{i=0}^{t-1} \vartheta^i (Y_{t-i} - \beta_0)^2 \text{ ; where } |\vartheta| < 1.$$

To find the least square estimate for β_0, we take the derivative with respect to $\beta_0$and set it to zero

$$\frac{dSSE^*}{d\beta_0}\Big]_{\beta_0} = -2\sum_{i=0}^{t-1} \vartheta^i \big(Y_{t-i} - \hat{\beta}_0\big) = 0$$

The solution thus becomes:

$$\hat{\beta}_0 \sum_{i=0}^{t-1} \vartheta^i = \sum_{i=0}^{t-1} \vartheta^i Y_{t-i} \qquad \xrightarrow{yields} \qquad \hat{\beta}_0 = \frac{1-\vartheta}{1-\vartheta^i} \sum_{i=0}^{t-1} \vartheta^i Y_{t-i}$$

And with large *i,* ϑ^i goes to zero;

$$\hat{\beta}_0 = (1-\vartheta) \sum_{i=0}^{t-1} \vartheta^i Y_{t-i} \tag{69}$$

It is clear that; $\hat{\beta}_0 \approx \hat{Y}_t$

a. Single Exponential smoothing

In a single exponential smoothing process, the discount factor on previous observations ϑ is introduced to obtain the exponentially weighted average:

$$\hat{Y}_t = (1-\vartheta) \sum_{i=0}^{t-1} \vartheta^i Y_{t-i} \xrightarrow{yields} \hat{Y}_t = (1-\vartheta)(Y_t + \vartheta Y_{t-1} + \cdots + \vartheta^{t-1} Y_1) \tag{70}$$

Which can also be expressed as a linear combination of the current observation and the smoothed observations at previous time points:

$$\hat{Y}_t = (1-\vartheta)Y_t + (1-\vartheta)(\vartheta Y_{t-1} + \vartheta^2 Y_{t-2} + \cdots + \vartheta^{t-1}Y_1)$$
$$\hat{Y}_t = (1-\vartheta)Y_t + \vartheta(1-\vartheta)(Y_{t-1} + \vartheta^1 Y_{t-2} \ldots + \vartheta^{t-2}Y_1)$$

Since: $\hat{Y}_{t-1} = (1-\vartheta)(Y_{t-1} + \vartheta^1 Y_{t-2} \ldots + \vartheta^{t-2}Y_1)$

$$\hat{Y}_t = (1-\vartheta)Y_t + \vartheta\hat{Y}_{t-1}$$

The simple exponential smoother is often expressed as: $\lambda = (1-\vartheta)$ with λ representing the weight on the last observation and $(1-\lambda)$ the weight put on the smoothed value of the previous observations.

$$\hat{Y}_t = \lambda Y_t + (1-\lambda)\hat{Y}_{t-1} \quad (71)$$

The initial value $\hat{Y}_0$ is needed in recursive calculations that start with $\hat{Y}_1 = \lambda Y_1 + (1-\lambda)\hat{Y}_0$ and the value needs to be estimated as shown;

$$\hat{Y}_1 = \lambda Y_1 + (1-\lambda)\hat{Y}_0$$
$$\hat{Y}_2 = \lambda Y_2 + (1-\lambda)\hat{Y}_1 = \lambda(Y_2 + (1-\lambda)Y_1) + (1-\lambda)^2\hat{Y}_0$$
.
$$\hat{Y}_t = \lambda[Y_t + (1-\lambda)Y_{t-1} + \cdots + (1-\lambda)^{t-1}Y_1] + (1-\lambda)^t\hat{Y}_0$$

Which means that as $\boldsymbol{t}$ gets large and $\boldsymbol{(1-\lambda)^t}$ gets small, the contribution of $\hat{Y}_t$ and $\hat{Y}_0$ becomes insignificant. For large datasets, two common estimates for $\hat{Y}_0$ are;

- Set $\hat{Y}_0 = \hat{Y}_1$, which is only reasonable if changes in the process are expected to occur early in the series and fast;
- Take the average of the available data $\bar{Y}$ and set $\hat{Y}_0 = \bar{Y}$, which is preferred if the process can be assumed to be locally constant at the beginning

The value of λ has been proposed in literature to vary between 0.1 and 0.4.

The single exponential smoothing technique is not suitable when there is a linear trend or seasonal pattern in the series data as the fitted values are often over or underestimating the actual values. If it is a biased estimator of the model parameters, a double exponential smoother is considered.

b. Double Exponential Smoothing

A second-order exponential smoothing is applied on $\widehat{Y}_t$ where:

$$\hat{Y}_t^{(2)} = \lambda\hat{Y}_t^{(1)} + (1-\lambda)\hat{Y}_{t-1}^{(2)} \quad (72)$$

Where $\hat{Y}_t^{(1)}$ and $Y_t^{(2)}$ denote the first and second-order smoothed exponentials respectively. Based on the Holt's methods, the "Double Exponential Smoothing" method divides the series into two parts: The level L_t and the trend, T_t; the two components are:

$$L_t = \alpha Y_t + (1-\alpha)(L_{t-1} + T_{t-1}) \quad (73)$$
$$T_t = \gamma(L_t + L_{t-1}) + (1-\gamma)T_{t-1}$$

For a given set of α and γ, both L_t and T_t are calculated. The level equation is composed of the weighted average of observation Y_t and the within sample one-step-ahead forecasts for time ***t.*** The trend equation is composed of the weighted average of the estimated trend at time ***t*** and the previous estimate of the trend T_{t-1}. The *k*-ahead forecast is thus equal to the last estimated level value plus *k* times the last estimated trend value; a linear function of *k.*

While the higher-order (polynomial) and double exponential smoothing techniques are less biased than the single exponential smoothing technique, they do not capture the periodic oscillations in data referred to as seasonal patterns, which are observed in time series modelling of most disease trends.

c. Holt-Winters Smoothing (Triple Exponential Smoothing)

The exponential smoothing methodology, *Holt Winters Smoothing*, introduced by Holt in 1957 and Winters in 1960 is used to model seasonal time series data [84]. It involves a seasonal adjustment made to the linear trend model.

For the Additive model where;

$$Y_t = l_t + s_t + e_t$$

l_t, the linear trend is a function of time; e_t the error term and; S_t represents the seasonal component with $S_t = S_{t+s} = S_{t+2s} \ldots for\ t = 1, \ldots, (s-1)$. *S* is the length of the period of the cycles. However, a restriction on the model is that seasonal adjustments add up to zero during one period;

$$\sum_{t=1}^{s} S_t = 0$$

To forecast future values, we first deploy the first-order exponential smoothers with different discount factors. The following procedure is for updating the parameter estimates once the current observation Y_t is obtained;

Step 1: Update L_t estimates using

$$\hat{L}_t = \lambda_1\left(Y_t - \hat{S}_{t-s}\right) + (1-\lambda_1)(\hat{L}_{t-1} + \beta_{1.t-1})$$

Where $0 < \lambda_1 < 1$. The first part in the equation is the "current value" for L_t and the second part is the forecast of L_t based on the estimates at $(t-1)$.

Step 2: Update the estimate of β_1 using

$$\hat{\beta}_{1.t} = \lambda_2\left(\hat{L}_t - \hat{L}_{t-1}\right) + (1-\lambda_2)\beta_{1.t-1}$$

Where $0 < \lambda_2 < 1$. The estimate of β_1 is a linear combination of the "current value" for β_1 and the "forecast" at $(t-1)$.

Step 3: Update the estimate of S_t using

$$\hat{S}_t = \lambda_3(Y_t - \hat{L}_t) + (1 - \lambda_3)\hat{S}_{t-s}$$

Where $0 < \lambda_3 < 1$.

Step 4: Finally the k-step-ahead forecast; $\hat{Y}_{t+k}(t)$ is

$$\hat{Y}_{t+k}(t) = \hat{L}_t + \beta_{1.t^k} + \hat{S}_t(k - S)$$

The problem is in the estimating the initial values of the exponential smoothers. For a given set of historic data with m seasons, thus *ms* observations, the least squares estimator can be used.

$$Y_t = \beta_0 + \beta_1 t + \sum_{i=1}^{s-1} \gamma_i(I_{t,i} - I_{t,s}) + e_t \quad (74)$$

Where $I_{t,i} = \begin{cases} 1, t = i, i+s, i+2s, \\ 0, \quad \text{otherwise} \end{cases}$

3.5 Evaluating Forecasting Model Performance

Often, one-step-ahead forecast errors are used to evaluate forecasting accuracy:

$$e_t(1) = Y_t - \hat{Y}_t(t-1); \quad (75)$$

$$where\ \hat{Y}_t(t-1)\ is\ the\ forecast\ of\ Y_t\ made\ one\ period\ prior$$

Suppose there are *n* observations for which forecasts have been made and *n* one-step-ahead forecasts $e_t(1)\ where\ t = 1,2,\dots,n$. The following are the standard measures of forecast accuracy which are scale-dependent [85, 86]:

The Mean Error (ME)

$$ME = \frac{1}{n}\sum_{t-1}^{n} e_t(1) \quad (76)$$

The Mean Squared Error (MSE) - *a direct estimator of the variance of the one-step-ahead-forecasts*

$$MSE = \frac{1}{n}\sum_{t-1}^{n} [e_t(1)]^2 \quad (77)$$

The Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{MSE} = \sqrt{\left(\frac{1}{n}\sum_{t-1}^{n} [e_t(1)]^2\right)} \quad (78)$$

The ME is an estimate of the expected values of forecast error which should be 0 from the residual analysis. Bias in the forecast is indicated if the mean forecast error significantly differs from zero. The

MAE, MSE and RMSE measure the variability in the forecast errors which is desired to be small though the MSE and RMSE are more sensitive to outliers as compared to the MAE.

However, accuracy measures that are scale dependent do not facilitate comparisons of a particular forecasting method across different data series. Hence, we adopt a measure of percentage forecast error:

$$pe_t(1) = \left\{\frac{Y_t - \hat{Y}_t(t-1)}{Y_t}\right\} 100 = \left\{\frac{e_t(1)}{Y_t}\right\} 100$$

The Mean Percent Forecast Error (MPE):

$$MPE = \frac{1}{n}\sum_{t-1}^{n} pe_t(1) \tag{79}$$

These percentage error measures tend to be undefined or indefinite when $Y_t = 0$ and having a skewed distribution when Y_t is close to zero. Additionally, they assume a meaningful zero even when measuring non-positive valued variables. Depending on the outcome of a forecasting procedure, any forecasting accuracy error measure can be used but with the full recognition of the underlying assumptions.

For comparison of forecast accuracy across series with/out different scales, the scaled errors which are less sensitive to outliers, are not subject to the degeneracy issues from the absolute and percentage error measures, have a meaningful scale and are easier to interpret are recommended. The Mean Absolute Scaled Error (MASE) was proposed in 2006 and is calculated by scaling the error based on the in-sample MAE from the naïve forecasts and assuming that the series has no more than one unit root [85].

$$Scaled\ Error_t = \varepsilon_t = \frac{e_t}{\frac{1}{n-1}\sum_{t-1}^{n} |Y_t - Y_{t-1}|};$$

$$MASE = mean\,|\varepsilon_t| \tag{80}$$

The $\varepsilon_t < 1$, if it arises from a forecast that is better that the average one step ahead naïve in-sample forecast while $\varepsilon_t > 1$ conversely when it arises from a worse forecast. However, scaled errors would be undefined and infinite when all historical observations are equal.

3.6 Choosing Between Competing Models

A particular time series can be forecasted by use of various competing models. Identifying and selecting the model that is the best fit to historical data does not always result in a forecasting model

producing the best forecasts [60-62]. When evaluating the fit of the model to historical data, various measures are applied. These include residual analysis and the error measures described in Section 3.5 which can be utilize to obtain the most accurate forecasts, compare forecast accuracy and choose between competing models with the ME, MSE, MAPE, RMSE, and MASE being the mostly commonly applied measures.

The best approach proposed by Montgomery is often to select the model with the smallest standard deviation (MSE) of the one-step-ahead forecast errors of the validation subset of the data series [61].

$$s^2 = MSE_{validation} = \frac{\sum_{t=1}^{n} e_t}{n-p} \tag{81}$$

; $n\ periods\ of\ data\ for\ model\ fitting\ with\ p\ parameters\ and\ e_t\ the\ model\ fit\ residuals$

The other criterion is the *R*-squared statistic:

$$R^2 = 1 - \frac{\sum_{t=1}^{n} e_t}{\sum_{t=1}^{n}(Y_t - \mu)^2} = \frac{Residual\ sum\ of\ squares}{Total\ sum\ of\ squares\ of\ the\ observations} \tag{82}$$

The model that maximizes the R^2 is equivalent to a model that minimizes the sum of squared residuals. Large values of *R*-squared suggest a good fit to historical data but relying on R^2 to select a forecasting model encourages over-fitting and extra parameters might be included to obtain future forecasts. The "adjusted" R^2 statistics is a better criterion:

$$R^2{}_{Adj} = 1 - \frac{\frac{\sum_{t=1}^{n} e_t}{T-p}}{\sum_{t=1}^{n}\frac{(Y_t - \mu)^2}{n-1}} = 1 - \frac{s^2}{\sum_{t=1}^{n}\frac{(Y_t - \mu)^2}{n-1}} \tag{83}$$

The model that maximizes the adjusted R^2 statistic is equivalent to a model that minimizes the residual mean square (RMSE).

Additionally, the AIC and BIC [Section] which penalize the sum of squared residuals for inclusion of extra parameters in the model are two other important criterion for choosing between competing models, where the model with the lowest AIC and/or BIC is considered the best fit model.

3.7 Statistical Software to be used

Data management and analysis will be done in R-Console Statistical Software. The results will be presented in form of tables, graphs and context.

CHAPTER 4: DATA ANALYSIS AND RESULTS

This chapter presents the data analysis and results of forecasting malaria case admissions in three Kenyan health facilities. Data management and analysis was done in R-Gui Software. The data was segmented into two sets: Training Set (from 1999 to 2009) and the Test Set (from 2010 to 2011). The hold out set (test) provides the gold standard for measuring the model's true prediction error which refers to how well the model forecasts for new data. To note, the test data should only be used after a definitive model has been selected. This ensures unbiased estimates of the true forecast error.

Methods

Method 1: Box Jenkins Approach

Step 1: Plot the original data on a time plot and observe the structure, fit a linear trend and check for seasonal oscillations.

Step 2: Plot the ACF and PACF of the original time series to observe pattern of auto correlations and determine if the data contains an AR and/or an MA structure. Conduct stationarity tests using the ADF and KPSS statistics to investigate if the data is stationary in mean. If the variance is not stable, apply a transformation.

Step 3: If the series is non-stationary from Step 2, apply regular and/or seasonal differencing.

Step 4: Identify the values for the AR, MA, SMA and SAR parameter orders by observing the ACF and PACF plots and fit an ARIMA model

Step 5: Conduct diagnostic checks to evaluate the model fit. If through the residual analysis and parameter contribution to the model, a model is considered inadequate, repeat Steps 5 until potential models which fulfill the underlying assumptions are obtained

Step 6: Fit the different potential models to the data and estimate the appropriate parameters which either maximize the likelihood or minimize the sum of squared errors. Use appropriate criterion "AIC" and "BIC" to select the most parsimonious model and fit it to the data.

Step 7: Additionally, estimate forecast model performance through cross validation with holdout data subset by observing the various error measures to assist in choosing between competing models.

Step 7: The appropriate model chosen can now be used for forecasting to conduct the $k-\text{ahead}$ out-of-sample forecasts for the time series data as required

Method 2: Forecasting with Decomposition

Step 1: Decompose the data using the STL method to obtain the seasonal, trend and error components.

Step 2: Deseasonalize the data by subtracting the seasonal component from the original data.

Step 3: Conduct in-sample forecasts by forecasting the seasonally adjusted data using the Holt's linear exponential smoothing method and reseasonalize by adding the seasonal naïve forecasts of the seasonal component.

Step 4: Additionally, estimate forecast model performance through cross validation with holdout data subset by observing the various error measures to assist in choosing between competing models

Step 5: Conduct $k - ahead$ out-of-sample forecasts for the time series data as required.

4.1 Descriptive Data Analysis

a. **Malindi**

There was a general declining trend in malaria case admissions in Malindi between 1999 and 2011 with a peak in 2003. The mean malaria case admissions was highest in the month of July with 64 cases and lowest in September with 42 cases. The highest peak across the time period was in 2003 while the lowest was observed after 2010.

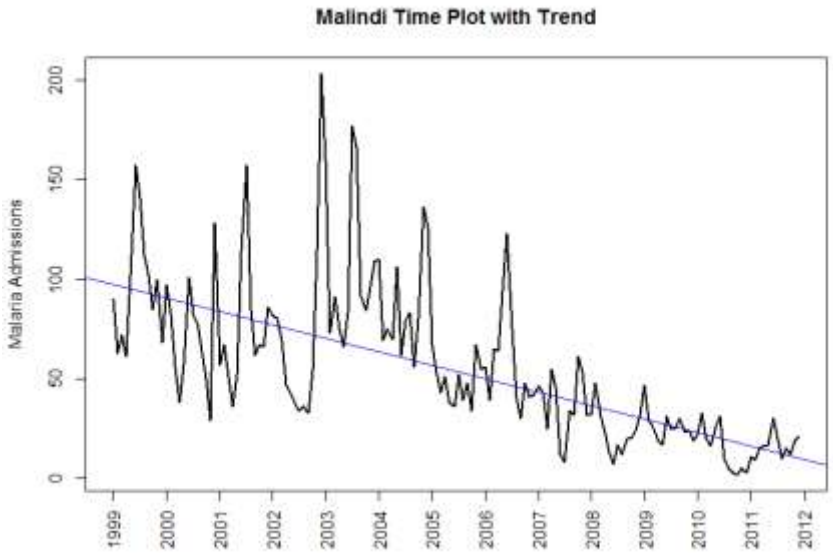


Figure 1: Malindi Malaria Case Admissions Time Plot

Kitale

In Kitale, malaria case admissions remained relatively stable within the same period with the highest peak observed in 2003 and the lowest in 2005. The mean number of cases was highest in the month of July at 332 cases and lowest in the month of September at 184 cases.

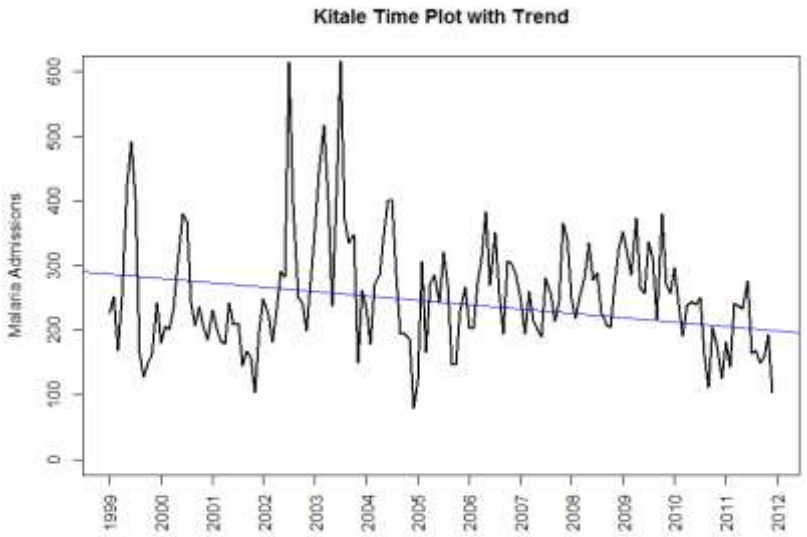


Figure 2: Kitale Malaria Case Admissions Time Plot

Siaya

There was a relative increase in the number of malaria case admissions between 1999 and 2010 in Siaya with declines beginning after 2010. The highest peak was observed between 2008 and 2010. The highest mean number of malaria case admissions was observed in the month June at 187 cases and the lowest in October at 112 cases.

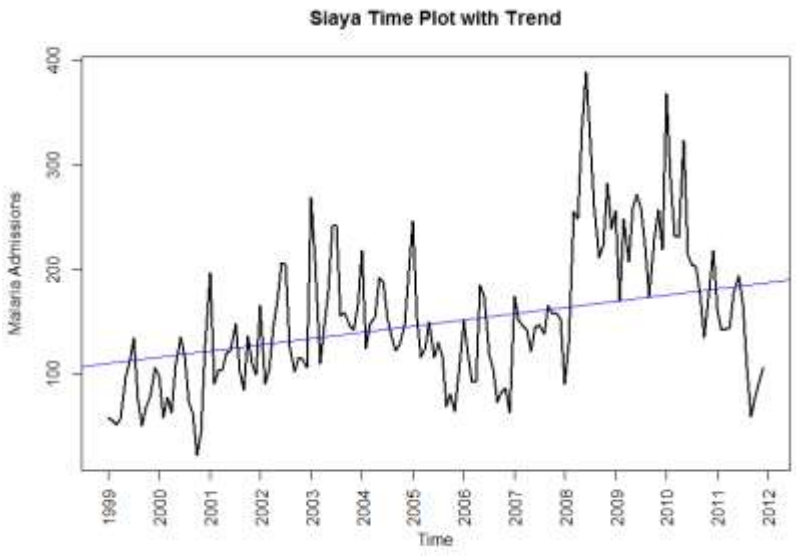


Figure 3: Siaya Malaria Case Admissions Time Plot

4.2 Exploratory Data Analysis

Malindi

Method 1: Box Jenkins Procedure

Plot the ACF and PACF plot of the data, conduct stationarity tests to check if data is stationary and identify the regular and seasonal parameters.

Table 2: Malindi Stationarity Test Statistics

	Statistic	P-Value		Statistic	P-Value
ADF Original	-3.1098	0.115	KPSS	1.966	0.01
ADF 1st Difference	-7.7356	0.01	KPSS 1st Difference	0.0112	0.10

According to ADF and KPSS tests, the data is non-stationary. Applying a regular differencing, the data becomes stationary as shown below in *Figure 4*.

Observing the ACF and PACF plots there are signs of over-differencing. We add MA (2) or MA (3) terms to the regular part of the model to cancel out these effects, and SMA (2) to cancel out the effects of over-differencing at the seasonal lags. In some instances, we apply a seasonal difference to preserve the seasonal pattern, lowering the amount of total differencing and increasing stability of trend projections.

The original plots and regular difference plots are as shown below.

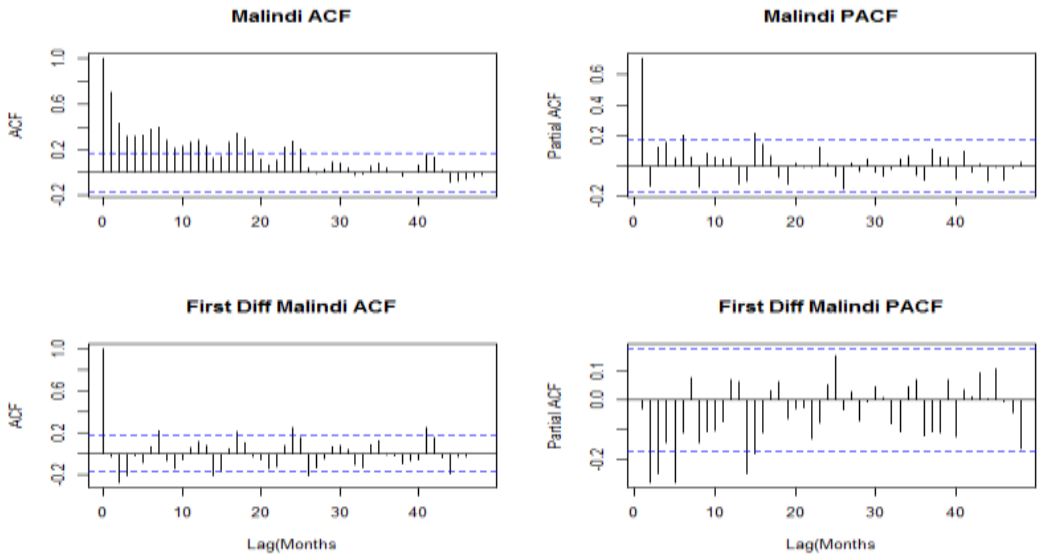


Figure 4: Malindi ACF and PACF Plots

Hence we compare between the following models:

Table 3: Malindi SARIMA Results

Model	**Log Likelihood**	**AIC**
SARIMA $(0,1,2)(0,0,2)^{12}$	-609.52	1229.04
SARIMA $(0,1,3)(0,0,2)^{12}$	-606.67	1225.33
SARIMA $(0,1,3)(0,0,1)^{12}$	-609.42	1228.83
SARIMA $(0,1,3)(0,1,2)^{12}$	-565.91	1143.82
SARIMA $(0,1,3)(0,1,1)^{12}$	**-566.02**	**1142.03**
SARIMA $(0,1,3)(0,0,2)^{12}$ auto-arima	-601	1216.00

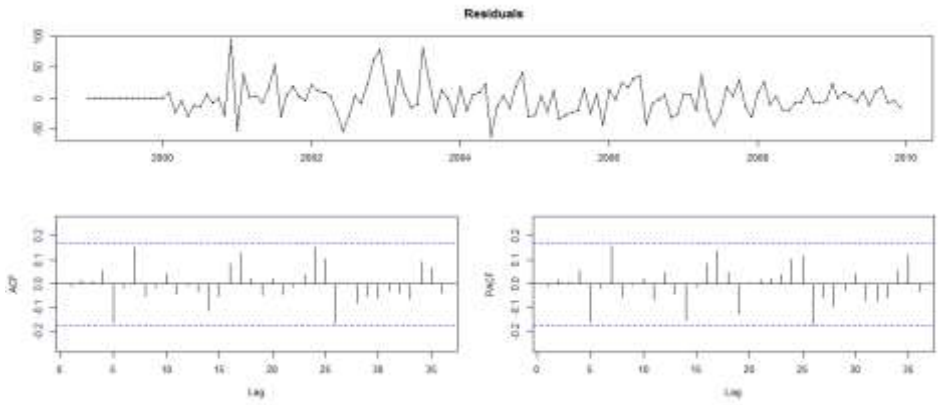
The best model selected was: **SARIMA $(0, 1, 3)\ (0, 1, 1)^{12}$**

The model parameters are:

Parameters	**Coefficient**	**Std Dev.**	**P-value**
ma1	-0.2805	0.0962	0.0036
ma2	-0.3671	0.0811	<0.0001
ma3	-0.2019	0.0919	0.0280
sma1	-0.8601	0.1343	<0.0001

We then conduct the residual analysis by observing the ACF and PACF, plotting a QQ plot and conducting the Ljung Box Statistic goodness of fit to check if the residuals conform to the normal distribution. The results are as shown in *Figure 5.*

Lag	**Lag 12**	**Lag 24**	**Lag 48**
Ljung Box Statistic	0.7622	0.0962	0.6729



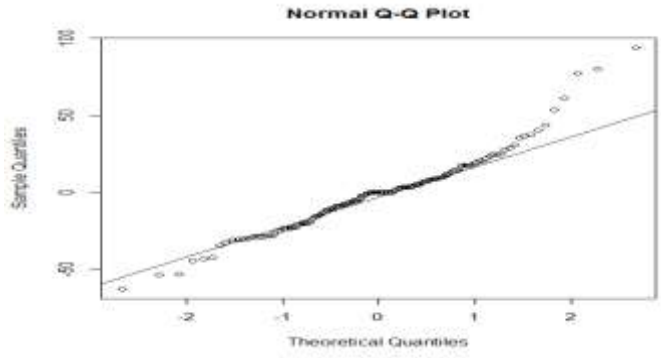


Figure 5: Malindi Residual Analysis Results

In addition, we cross-validate the forecast accuracy of the model by comparing model performance on data used for in-sample forecasts with data held out for validating the model (h=24 ahead out of sample forecast). The error measures are presented in *Table 5*.

Method 2: STL + ETS Method

The method is automated, and the de-seasonalization and forecasting occur concurrently. The parameter estimates obtained after we deseasonalize the series using STL are as below:

Table 4: Malindi STL + ETS Resulta

Smoothing Parameter	**Malindi**
Alpha (level)	0.4151
Beta (trend)	0.0000
AIC	1457.055
ETS Model	"MNN"

We conduct h=24 ahead out of sample and observe the error measures presented in *Table 5.* We then choose the best forecasting model to forecast malaria case admissions

Table 5: Malindi Error Measures

SARIMA (0,1,3)(0,1,1)^{12}					
	ME	***RMSE***	***MPE***	***MASE***	***ACF1***
Training set	-0.5751	24.8260	-15.8563	0.5660	-0.0100
Test set	6.0998	13.6208	-21.0994	0.3767	0.6202
STL + ETS					
Training set	-0.9760	22.1644	-13.7905	0.5182	0.3126
Test set	-5.7905	10.7494	-141.1876	0.2821	0.6174

Visual Comparison of SARIMA and STL + ETS Models

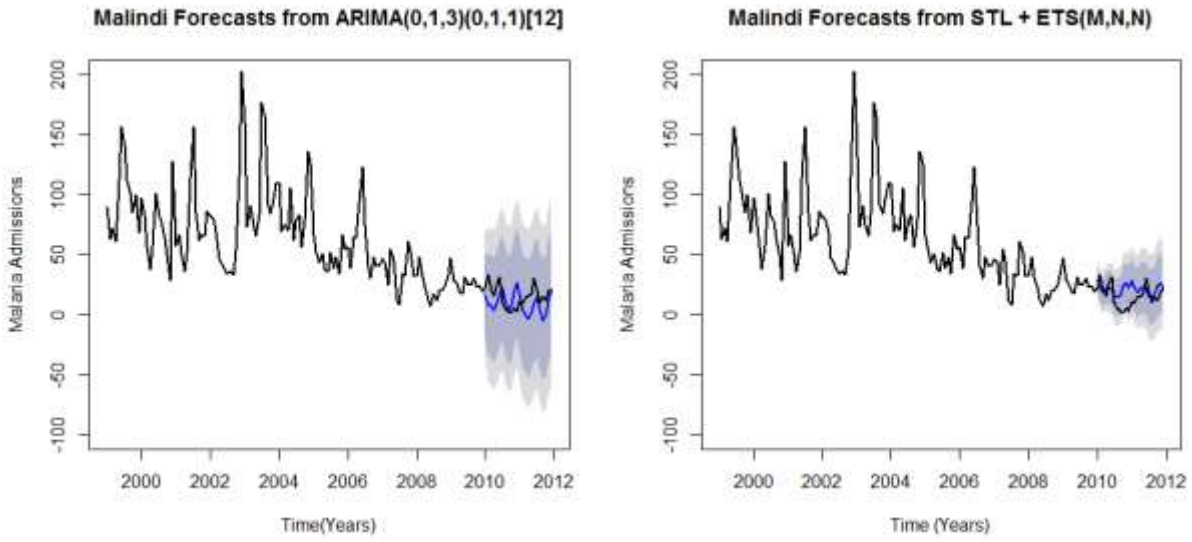


Figure 6: Malindi Cross-Validation 24-ahead Malaria Case Forecasts

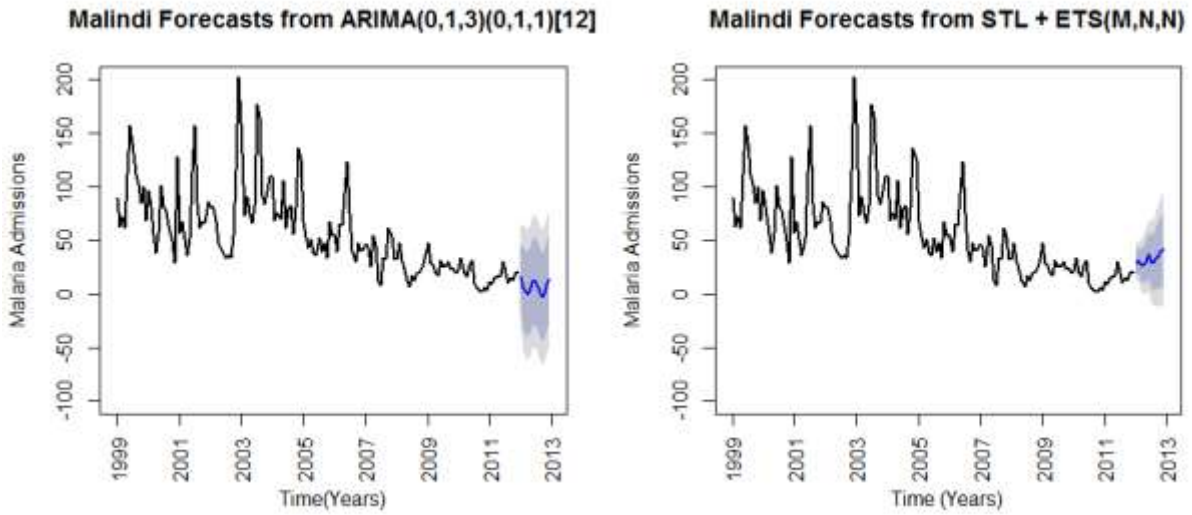


Figure 7: Malindi 12-ahead Malaria Case Admission Forecasts

Kitale

Method 1: Box Jenkins Procedure

Plotting the ACF and PACF plot of the data, and conducting stationarity tests the data is stationary but there are seasonal lags hence we conduct a seasonal difference.

Table 6: Kitale Stationarity Tests

	Statistic	*P-Value*		*Statistic*	*P-Value*
ADF	-4.0022	0.0113	KPSS	1.966	0.1

The Kitale Series is stationary. The ACF and PACF plots are as shown below.

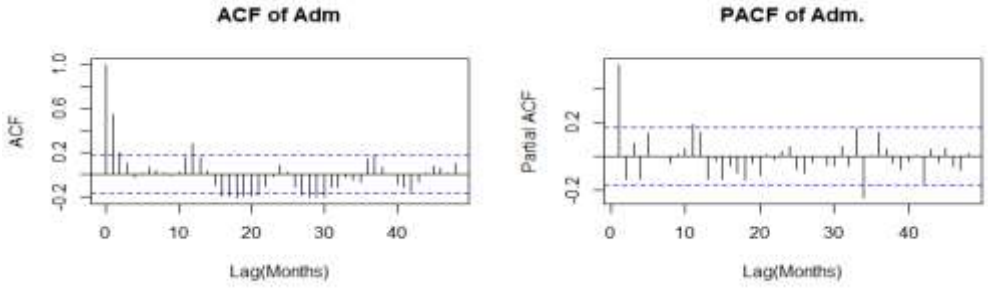


Figure 8: Kitale ACF and PACF plots

However, we replace the significant seasonal lag in the model by applying a seasonal difference and also to maintain the seasonal pattern over time and add an SMA(1) term to the model. In the regular component of the model we add either MA(1) or MA(2) or MA(3) and AR(1).

Hence we compare the following models:

Table 7: Kitale SARIMA Results

Model	**Log Likelihood**	**AIC**
SARIMA $(1,0,1)(0,1,1)^{12}$	-689.18	1386.36
SARIMA $(1,0,2)(0,1,1)^{12}$	-687.42	1384.85
SARIMA $(1,0,2)(0,1,0)^{12}$	-706.68	1421.36
SARIMA $(1,0,3)(0,1,1)^{12}$	-687.37	1386.74
SARIMA $(1,0,3)(1,0,1)^{12}$	-753.03	1520.06
SARIMA $(1,0,2)(1,0,1)^{12}$ auto-arima	**-753.03**	**1518.06**

The model selected is: **SARIMA $(1,0,2)(1,0,1)^{12}$**

However, the model selected does not have the lowest AIC. This is because while some of the other models contain parameters which are not significant, others do not fulfill the model diagnostic checks or have a higher standard deviation that the selected model.

We obtain the model parameters as:

Parameters	Coefficient	Std Dev.	P-value
ar1	0.9873	0.0129	<0.0001
ma1	-0.2966	0.0847	0.0005
ma2	-0.4589	0.0916	<0.0001
sar1	0.9484	0.0807	<0.0001
sma1	-0.8013	0.1727	<0.0001

Conducting the residual analysis and the goodness of fit statistics:

Lag	**Lag 12**	**Lag 24**	**Lag 48**
Ljung Box Statistic	0.9426	0.9787	0.6922

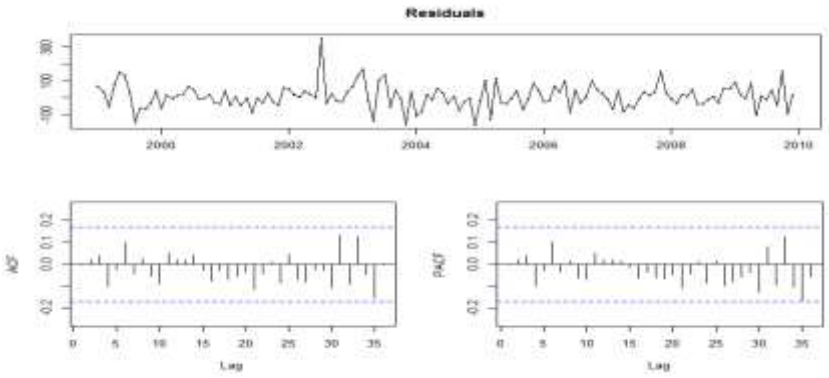


Figure 9: Kitale Residual Analysis

We compare model performance on data used for in-sample forecasts with data held out for validating the model (h=24 ahead out of sample forecast). The error measures are presented in *Table 9*.

Method 2: STL + ETS Method

The parameter estimates obtained after we deseasonalize the series using STL are as below:

Table 8: Kitale STL + ETS Results

Smoothing Parameter	**Kitale**
Alpha (level)	0.4173
Beta (trend)	0.0000
AIC	1720.65
ETS Model	"ANN"

We conduct h=24 ahead out of sample and observe the error measures presented in *Table 19*. We then choose the best forecasting model to forecast malaria case admissions

Table 9: Kitale Error Measures

SARIMA (1,0,2)(1,0,1)^{12}					
	ME	RMSE	MPE	MASE	ACF1
Training set	5.6180	70.3830	-4.0837	0.6378	-0.0027
Test set	-83.3431	94.4562	-51.8296	1.0554	
STL + ETS					
Training set	0.3232	56.7819	-3.6852	0.5537	0.1586
Test set	-74.9825	91.9682	-48.1289	0.9998	0.1601

Visual Comparison of SARIMA and STL+ETS Model

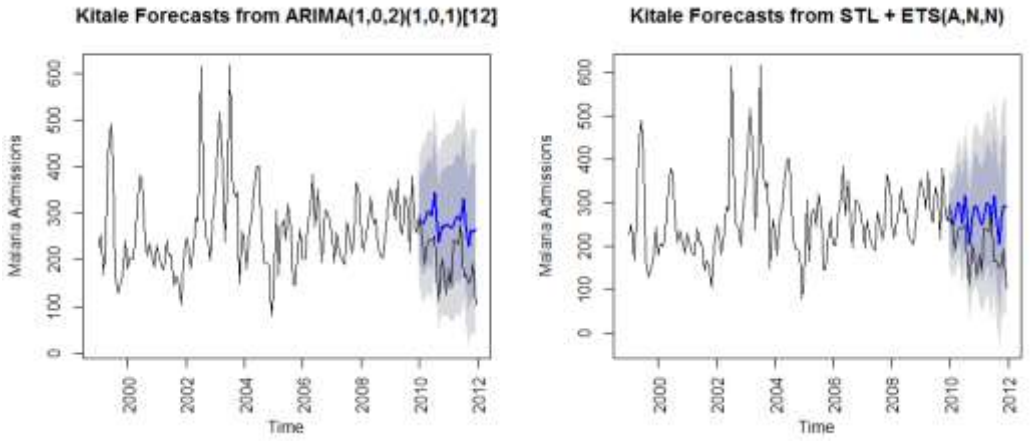


Figure 10: Kitale Cross-Validation 24-ahead Malaria case Forecasts

12-Ahead Forecasts

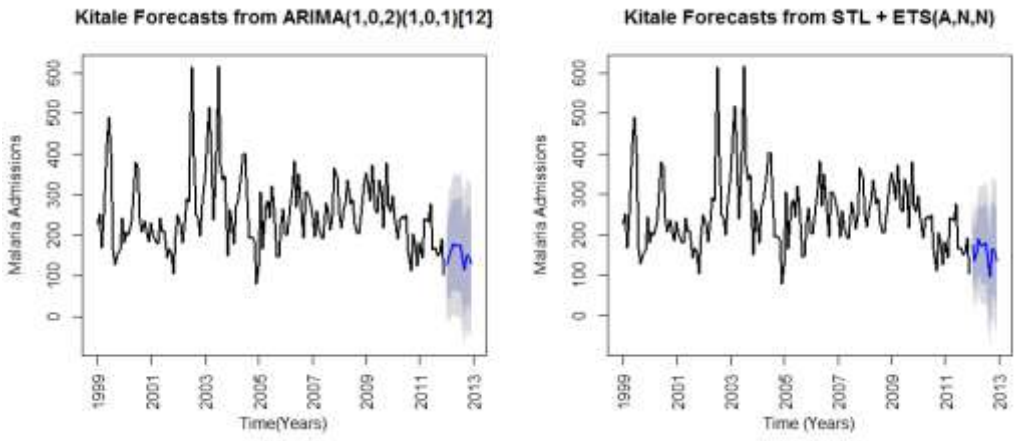


Figure 11: Kitale 12-ahead Malaria Case Admissions Forecasts

Siaya

Method 1: Box Jenkins Procedure

Plotting the ACF and PACF plot of the data and conduct stationarity checks.

Table 10: Siaya Stationarity Tests

	Statistic	*P-Value*		*Statistic*	*P-Value*
ADF Original	-2.4091	0.4064	KPSS	1.918	0.01
ADF 1st Difference	-6.8405	0.01	KPSS 1st Difference	0.0154	0.1

The data is non-stationary hence we apply a regular difference.

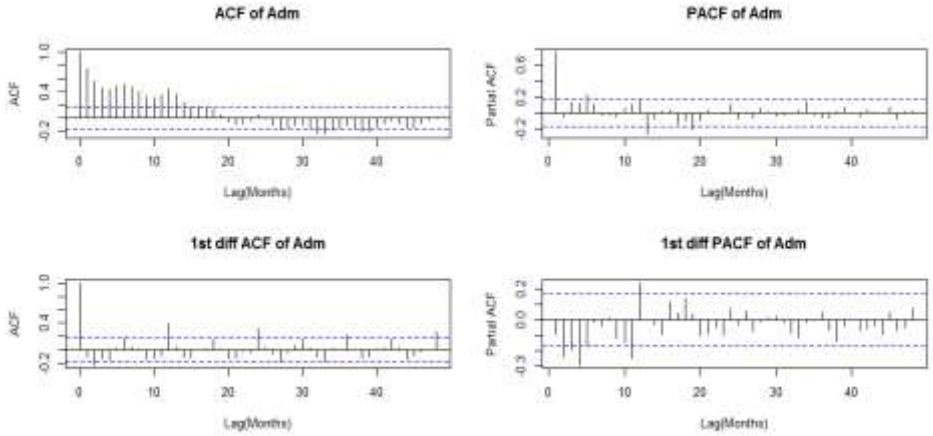


Figure 12: Siaya ACF and PACF graphs

The data is now stationary and we proceed to parameter estimation. However, there are still significant seasonal lags which are significant. We add either AR(1) or AR(2) to the model and MA(1) or MA(2) to the regular part of the model and SAR(1) and SMA(1) to the seasonal part of the model. We compare the models as below:

Table 11: Siaya SARIMA Results

Model	**Log Likelihood**	**AIC**
SARIMA $(1,1,1)(1,0,1)^{12}$	-659.42	1328.83
SARIMA $(1,0,2)(0,1,1)^{12}$	-604.5	1218.99
SARIMA $(2,1,1)(1,0,1)^{12}$	-657.94	1327.88
SARIMA $(0,1,1)(1,0,1)^{12}$	-666.1	1340.2
SARIMA $(1,1,1)(0,1,2)^{12}$	-603.4	1216.81
SARIMA $(0,1,2)(0,1,2)^{12}$	-600.06	1210.12
SARIMA $(0,1,1)(1,1,1)^{12}$	-609.09	1226.18
SARIMA $(0, 1, 2)(1,0,1)^{12}$ auto-arima	**-656.81**	**1323.62**

The parameters for the models with the lowest AIC's were mostly not significant while for some, the residuals were not uncorrelated.

The best model selected was: **SARIMA (0, 1, 2) (1, 0, 1)^{12}**

The model parameters are:

Parameters	Coefficient	Std Dev.	P-value
ma1	-0.2200	0.0835	0.0084
ma2	-0.4309	0.0862	<0.0001
sar1	0.9813	0.0491	<0.0001
sma1	-0.8568	0.1915	<0.0001

The p-value of the Ljung Box Statistic at lag 12, lag 24 and lag 48 is 0.6827, 0.9481 and 0.9900 respectively while from the ACF and PACF plots residuals are uncorrelated at all lags, confirming normality as shown:

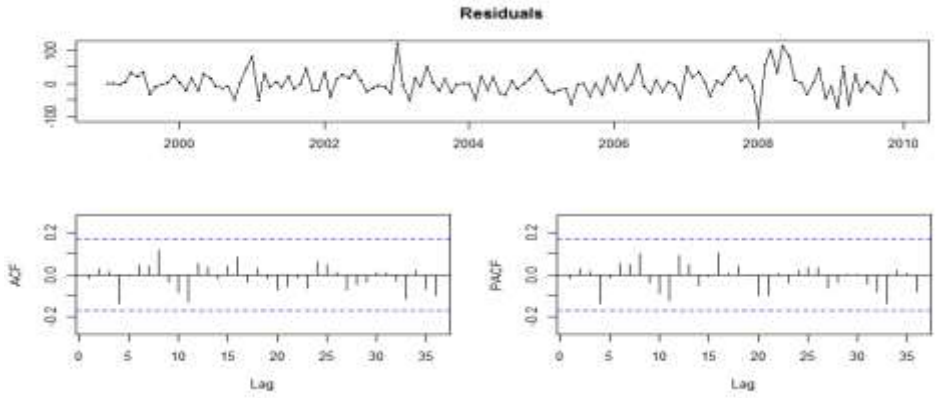


Figure 13: Siaya Residual Analysis

We compare model performance on data used for in-sample forecasts with data held out for validating the model (h=24 ahead out of sample forecast). The error measures are presented in *Table 13*.

Method 2: STL + ETS Method

The parameter estimates obtained after we deseasonalize the series using STL are as below:

Table 12: Siaya STL + ETS Results

Smoothing Parameter	**Siaya**
Alpha (level)	0.8010
Beta (trend)	0.0000
AIC	1540.53
ETS Model	"ANN"

We conduct h=24 ahead out of sample and observe the error measures presented in [Table xxx]. We then choose the best forecasting model to forecast malaria case admissions

Table 13: Siaya Error Measures

SARIMA (0, 1, 2)(1,0,1)^{12}					
	ME	RMSE	MPE	MASE	ACF1
Training set	1.1649	34.6960	-4.6702	0.5433	-0.0176
Test set	-74.1929	105.0948	-67.7307	1.9850	0.6805
STL + ETS					
Training set	2.0270	28.7014	-1.1349	0.4511	0.0408
Test set	-63.7880	93.8279	-58.3854	1.7773	0.6198

Visual Comparison of SARIMA and STL+ETS

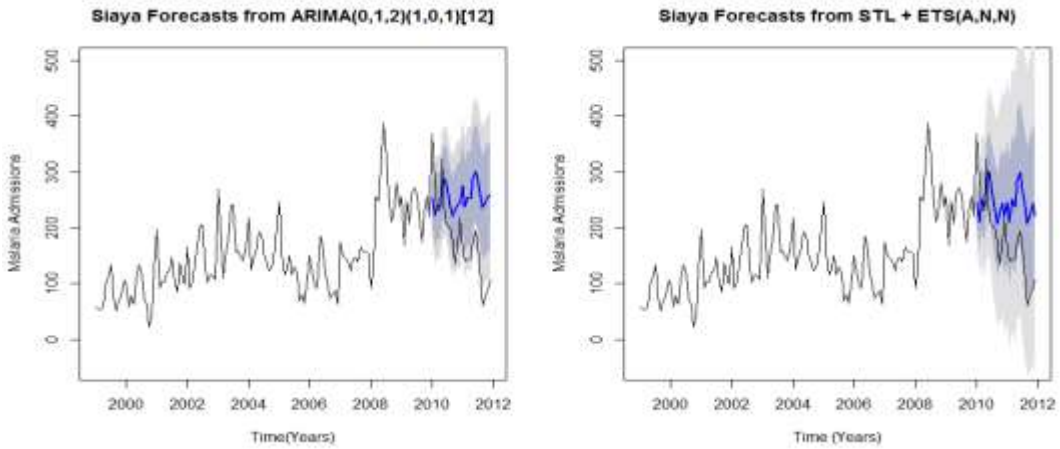


Figure 14: Siaya Cross-Validation 24-ahead Malaria Case Forecasts

12 Ahead Forecasts

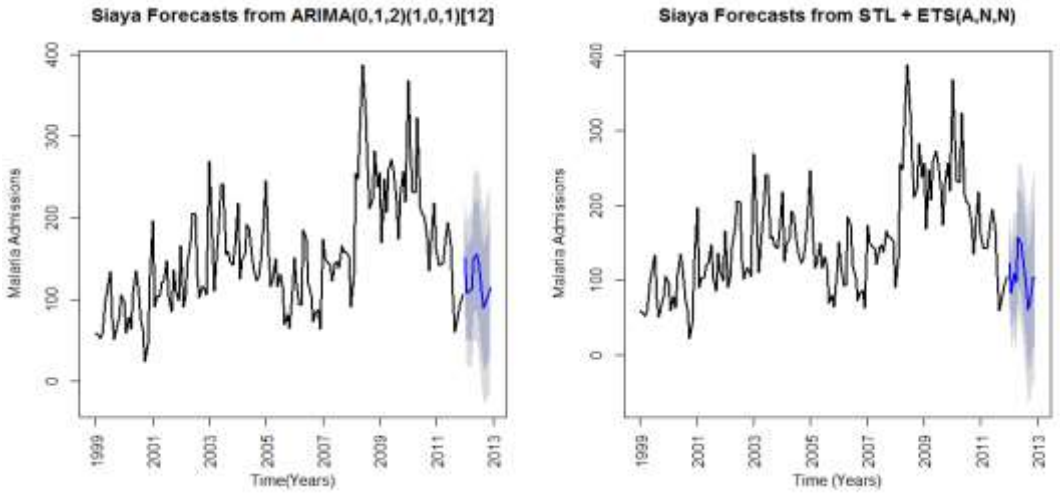


Figure 15: Siaya 12-ahead Malaria Case Admissions Forecasts

4.4 Results

In this study two statistical time series methods were compared and their accuracy to forecast malaria case admissions in different epidemiological zones in Kenya assessed. While the decomposition and ETS method is computationally simpler and easier to interpret to the end user, the SARIMA models have been popularized in recent decades due to their statistical sophistication and underlying theory and assumptions. However, based on the forecasting error measures and visualization of the forecasted case admission numbers, the study confirms empirical evidence as no one method was superior to the other. The simpler STL + ETS method performed equally as good and/or was in some instances even more accurate than the more complex Box Jenkins method. Other studies have also shown that the ARIMA class models perform better in model fitting the data but might be less robust in post-sample forecasts [62].

For the STL + ETS method, the STL technique was used in decomposing the series, and then the ETS method applied in modelling and forecasting the seasonally adjusted data. The seasonal naïve method was used to forecast the seasonal component and afterwards the series was re-seasonalized by adding the two forecasted data parts back together. This procedure enabled the model to accommodate both the seasonality component and recent trend changes concurrently. The SARIMA model was estimated by following the iterative Box Jenkins methodology and the best model which maximized the likelihood was chosen. Visual inspection of the forecasts shows that the SARIMA model [SARIMA (0, 1, 3) (0, 1, 1)^{12}] performed better in predicting malaria case admissions in Malindi where a general trend was observed across the series beginning after 2003. The method was able to capture both the trend and seasonal fluctuations over time. Conversely, the STL + ETS model performed better in Kitale and Siaya, where malaria case admissions were more stable over time characterized by abrupt trend and level shifts. However, the Dieblo-Mariano test statistics showed that both methods had the same forecast accuracy for all the sites [87].

The SARIMA model has been shown to perform better in long-range forecasting due to its ability to estimate a more accurate long-term average. However, since this study only incorporated 12-ahead forecasts the model seemed to capture the trend fluctuations poorly in Kitale and Siaya. On the other hand, the decomposition method was more accurate in short term forecasts due to its ability to accommodate recent changes in the trend and the seasonal oscillations. However, previous studies show that this method is less accurate in long-range forecasts due to the uncertainties in updating the smoothing parameters over long periods of time. Thus, the importance of considering the forecasting horizon while choosing a forecasting method should be among the initial steps, as proposed by Briet [65].

Many studies from malaria-endemic countries have shown the significant contribution of external factors such as climate-related variables to variations in malaria incidence levels. This is more noticeable in epidemic-prone areas where unexpected changes in the weather might lead to outbreaks culminating in very high case fatality rates. On the other hand, although transmission occurs throughout the year in the endemic regions, malaria case numbers are higher during the rainy season and lower during the dry months. However, the between year signals might be due to disease dynamics and not necessarily due to climatic changes [88]. Hence, the exclusion of external risk factors this study might have reduced the robustness of the models to capture year to year variations in malaria infection risk. This was one of the study limitations in addition to use of health facility based data which has been reported to be burdened by a number of shortcomings such as missing-ness, inconsistency and inaccuracies. However its longitudinal nature provides continuous data for monitoring temporal disease patterns.

4.5 Conclusion

Malaria poses a significant threat to public health and the growth and economic development of a country. The ability to forecast future malaria incidence is a major milestone as it will facilitate timely planning and implementation of control, prevention and case management interventions through optimal distribution of the available resources. While more research is required to develop accurate and efficient malaria forecasting methods, care should be taken to ensure that the models are tailor made to the specific transmission setting to avoid misleading results due to variations in malaria risk of infection and transmission dynamics. In addition, the pattern of the series, forecasting horizon, purpose to end user, and the expertise available should be some of the main considerations for the choice of a forecasting method.

CHAPTER 5: REFERENCES

1. WHO: **World Malaria Report 2014.** Geneva, Switzerland: WHO; 2014.
2. Snow RW, Amratia P, Kabaria CW, Noor AM, Marsh K: **The changing limits and incidence of malaria in Africa: 1939-2009.** *Adv Parasitol* 2012, **78:**169-262.
3. Nabarro DN, Tayler EM: **The "roll back malaria" campaign.** *Science* 1998, **280:**2067-2068.
4. Lomazzi M, Laaser U, Theisling M, Tapia L, Borisch B: **Millennium Development Goals: how public health professionals perceive the achievement of MDGs.** *Glob Health Action* 2014, **7:**24352.
5. Roll Back Malaria: **Saving Lives with Malaria Control: Counting Down to the Millennium Development Goals.** In *Progress & Impact Series, n° 3*. Geneva, Switzerland: World Health Organization, ; 2010.
6. Roll Back Malaria: **Global Malaria Action Plan for a free malaria world.** Roll Back Malaria Partnership; 2008.
7. Roll Back Malaria: **Key Facts, Figures and Strategies. The Global Malaria Action Plan. http://www.rollbackmalaria.org/gmap/GMAP_Advocacy-ENG-web.pdf.** 2008.
8. Roll Back Malaria: **The Global Partnership for a Malaria-free World. http://www.rollbackmalaria.org/mechanisms/merg.html**
9. Sachs JD: **A new global effort to control malaria.** *Science* 2002, **298:**122-124.
10. Zinszer K, Verma AD, Charland K, Brewer TF, Brownstein JS, Sun Z, et al: **A scoping review of malaria forecasting: past work and future directions.** *BMJ Open* 2012, **2**.
11. KNBS: **Kenya Facts and Figures 2012.** Kenya National Bireau of Statistics, 2014..
12. KNBS: **Kenya Population Census (2009). Preliminary Report.** Kenya National Bireau of Statistics, 2009.
13. DOMC: **Kenya National Malaria Indicator Survey.** Division of Malaria Control, Ministry of Public Health & Sanitation, Republic of Kenya, July 2011.
14. Noor AM, Mutheu JJ, Tatem AJ, Hay SI, Snow RW: **Insecticide-treated net coverage in Africa: mapping progress in 2000-07.** *Lancet* 2009, **373:**58-67.
15. Barat LM: **Four malaria success stories: how malaria burden was successfully reduced in Brazil, Eritrea, India, and Vietnam.** *Am J Trop Med Hyg* 2006, **74:**12-16.
16. Brasseur P, Badiane M, Cisse M, Agnamey P, Vaillant MT, Olliaro PL: **Changing patterns of malaria during 1996-2010 in an area of moderate transmission in southern Senegal.** *Malar J*, **10:**203.
17. Gething PW, Smith DL, Patil AP, Tatem AJ, Snow RW, Hay SI: **Climate change and the global malaria recession.** *Nature* 2010, **465:**342-345.

18. Karema C, Aregawi MW, Rukundo A, Kabayiza A, Mulindahabi M, Fall IS, et al: **Trends in malaria cases, hospital admissions and deaths following scale-up of anti-malarial interventions, 2000-2010, Rwanda.** *Malar J*, **11:**236.

19. Mmbando BP, Vestergaard LS, Kitua AY, Lemnge MM, Theander TG, Lusingu JP: **A progressive declining in the burden of malaria in north-eastern Tanzania.** *Malar J* 2010, **9:**216.

20. Okiro EA, Alegana VA, Noor AM, Mutheu JJ, Juma E, Snow RW: **Malaria paediatric hospitalization between 1999 and 2008 across Kenya.** *BMC Med* 2009, **7:**75.

21. Okiro EA, Hay SI, Gikandi PW, Sharif SK, Noor AM, Peshu N, et al: **The decline in paediatric malaria admissions on the coast of Kenya.** *Malar J* 2007, **6:**151.

22. O'Meara WP, Mangeni JN, Steketee R, Greenwood B: **Changes in the burden of malaria in sub-Saharan Africa.** *Lancet Infect Dis* 2010, **10:**545-555.

23. Otten M, Aregawi M, Were W, Karema C, Medin A, Bekele W, et al: **Initial evidence of reduction of malaria cases and deaths in Rwanda and Ethiopia due to rapid scale-up of malaria prevention and treatment.** *Malar J* 2009, **8:**14.

24. Vardo-Zalik AM, Zhou G, Zhong D, Afrane YA, Githeko AK, Yan G: **Alterations in Plasmodium falciparum genetic structure two years after increased malaria control efforts in western Kenya.** *Am J Trop Med Hyg* 2013, **88:**29-36.

25. Zhou G, Afrane YA, Dixit A, Atieli HE, Lee MC, Wanjala CL, et al: **Modest additive effects of integrated vector control measures on malaria prevalence and transmission in western Kenya.** *Malar J* 2013, **12:**256.

26. Okara RM, Sinka ME, Minakawa N, Mbogo CM, Hay SI, Snow RW: **Distribution of the main malaria vectors in Kenya.** *Malar J* 2010, **9:**69.

27. Manuel Ramos J, Reyes F, Tesfamariam A: **Change in epidemiology of malaria infections in a rural area in Ethiopia.** *J Travel Med* 2005, **12:**155-156.

28. Noor AM, Gething PW, Alegana VA, Patil AP, Hay SI, Muchiri E, et al: **The risks of malaria infection in Kenya in 2009.** *BMC Infect Dis* 2009, **9:**180.

29. Pullan RL, Bukirwa H, Staedke SG, Snow RW, Brooker S: **Plasmodium infection and its risk factors in eastern Uganda.** *Malar J* 2010, **9:**2.

30. Noor AM, Kinyoki DK, Mundia CW, Kabaria CW, Mutua JW, Alegana VA, et al: **The changing risk of Plasmodium falciparum malaria infection in Africa: 2000-10: a spatial and temporal analysis of transmission intensity.** *Lancet* 2014, **383:**1739-1747.

31. Snow RW, Craig MH, Deichmann U, le Sueur D: **A preliminary continental risk map for malaria mortality among African children.** *Parasitol Today* 1999, **15:**99-104.

32. Amin AA, Zurovac D, Kangwana BB, Greenfield J, Otieno DN, Akhwale WS, et al: **The challenges of changing national malaria drug policy to artemisinin-based combinations in Kenya.** *Malar J* 2007, **6:**72.

33. Fogh S, Jepsen S, Effersoe P: **Chloroquine-resistant Plasmodium falciparum malaria in Kenya.** *Trans R Soc Trop Med Hyg* 1979, **73:**228-229.

34. Shretta R, Omumbo J, Rapuoda B, Snow RW: **Using evidence to change antimalarial drug policy in Kenya.** *Trop Med Int Health* 2000, **5:**755-764.

35. Barnes KI, Chanda P, Ab Barnabas G: **Impact of the large-scale deployment of artemether/lumefantrine on the malaria disease burden in Africa: case studies of South Africa, Zambia and Ethiopia.** *Malar J* 2009, **8 Suppl 1:**S8.

36. Okech BA, Mwobobia IK, Kamau A, Muiruri S, Mutiso N, Nyambura J, et al: **Use of integrated malaria management reduces malaria in Kenya.** *PLoS One* 2008, **3:**e4050.

37. DOMC: **National Malaria Strategy 2001 - 2010.** Division of Malaria Control, Ministry of Public Health & Sanitation, Republic of Kenya, 2001.

38. DOMC: **National Malaria strategy 2009 - 2017** Division of Malaria Control, Ministry of Public Health & Sanitation, Republic of Kenya, November 2009.

39. KHPF: **Analysis of Performance, Analytical Review of Health Progress, and Systems Performance 1994-2010, Kenya.** 2010.

40. Van Hemelrijck MJ, Lindblade KA, Kubaje A, Hamel MJ, Odhiambo F, Phillips-Howard PA, et al: **Trends observed during a decade of paediatric sick visits to peripheral health facilities in rural western Kenya, 1997-2006.** *Trop Med Int Health* 2009, **14:**62-69.

41. Okiro EA, Bitira D, Mbabazi G, Mpimbaza A, Alegana VA, Talisuna AO, et al: **Increasing malaria hospital admissions in Uganda between 1999 and 2009.** *BMC Med*, **9:**37.

42. Allard R: **Use of time-series analysis in infectious disease surveillance.** *Bull World Health Organ* 1998, **76:**327-333.

43. Landoh ED, Tchamdja P, Saka B, Tint KS, Gitta SN, Wasswa P, et al: **Morbidity and mortality due to malaria in Est Mono district, Togo, from 2005 to 2010: a times series analysis.** *Malar J* 2012, **11:**389.

44. Afrane YA, Zhou G, Githeko AK, Yan G: **Utility of health facility-based malaria data for malaria surveillance.** *PLoS One* 2013, **8:**e54305.

45. Rowe AK, Kachur SP, Yoon SS, Lynch M, Slutsker L, Steketee RW: **Caution is required when using health facility-based data to evaluate the health impact of malaria control efforts in Africa.** *Malar J* 2009, **8:**209.

46. Amin AA, Kokwaro GO: **Antimalarial drug quality in Africa.** *J Clin Pharm Ther* 2007, **32:**429-440.

47. Marsh K: **Malaria disaster in Africa.** *Lancet* 1998, **352:**924.

48. Marsh K, Snow RW: **Malaria transmission and morbidity.** *Parassitologia* 1999, **41:**241-246.

49. Okiro EA, Alegana VA, Noor AM, Snow RW: **Changing malaria intervention coverage, transmission and hospitalization in Kenya.** *Malar J* 2010, **9:**285.

50. Snow RW, Marsh K: **Malaria in Africa: progress and prospects in the decade since the Abuja Declaration.** *Lancet* 2010, **376:**137-139.

51. White NJ, Nosten F, Looareesuwan S, Watkins WM, Marsh K, Snow RW, et al: **Averting a malaria disaster.** *Lancet* 1999, **353:**1965-1967.

52. Brooker S, Guyatt H, Omumbo J, Shretta R, Drake L, Ouma J: **Situation analysis of malaria in school-aged children in Kenya - what can be done?** *Parasitol Today* 2000, **16:**183-186.

53. Gitonga CW, Karanja PN, Kihara J, Mwanje M, Juma E, Snow RW, et al: **Implementing school malaria surveys in Kenya: towards a national surveillance system.** *Malar J* 2010, **9:**306.

54. O'Meara WP, Bejon P, Mwangi TW, Okiro EA, Peshu N, Snow RW, et al: **Effect of a fall in malaria transmission on morbidity and mortality in Kilifi, Kenya.** *Lancet* 2008, **372:**1555-1562.

55. Francis D, Gasasira A, Kigozi R, Kigozi S, Nasr S, Kamya MR, et al: **Health facility-based malaria surveillance: the effects of age, area of residence and diagnostics on test positivity rates.** *Malar J*, **11:**229.

56. Rowe AK: **Potential of integrated continuous surveys and quality management to support monitoring, evaluation, and the scale-up of health interventions in developing countries.** *Am J Trop Med Hyg* 2009, **80:**971-979.

57. GoK.: **Kenya Population Situation Analysis.** Kenya: Government of Kenya 2013.

58. DOMC: **Kenya Monitoring and Evaluation Plan 2009 - 2017.** Division of Malaria Control, Ministry of Public Health & Sanitation, Republic of Kenya, 2009.

59. Chistophers S: **Epidemic malaria of the Punjab: with a note of a method of predicting epidemic years.** *Trans Committee Stud Malaria Indi* 1911, **2:**17-26.

60. Chatfield C: ***The analysis of time series: an introduction***. London: Chapman & Hall; 2004.

61. Montgomery DC, Jenning, C.L, Kulahci, M.: ***Introduction to Time Series and Forecasting***. Wiley & Sons; 2008.

62. Makridakis S, Wheelwright, SC. & Hyndman, RJ.: ***Forecasting Methods and Applications***. New York: John Wiley & Sons, Inc; 1998.

63. Shumway RHaS, D.S.: ***Time Series Analysis and Its Applications***. Springer; 2009.

64. Abeku TA, de Vlas SJ, Borsboom G, Teklehaimanot A, Kebede A, Olana D, et al: **Forecasting malaria incidence from historical morbidity patterns in epidemic-prone areas of Ethiopia: a simple seasonal adjustment method performs best.** *Trop Med Int Health* 2002, **7:**851-857.

65. Briet OJ, Vounatsou P, Gunawardena DM, Galappaththy GN, Amerasinghe PH: **Models for short term malaria prediction in Sri Lanka.** *Malar J* 2008, **7:**76.

66. Cullen JR, Chitprarop U, Doberstyn EB, Sombatwattanangkul K: **An epidemiological early warning system for malaria control in northern Thailand.** *Bull World Health Organ* 1984, **62:**107-114.

67. Patil AP, Okiro EA, Gething PW, Guerra CA, Sharma SK, Snow RW, et al: **Defining the relationship between Plasmodium falciparum parasite rate and clinical disease: statistical models for disease burden estimation.** *Malar J* 2009, **8:**186.

68. Wangdi K, Singhasivanon P, Silawan T, Lawpoolsri S, White NJ, Kaewkungwal J: **Development of temporal modelling for forecasting and prediction of malaria infections using time-series and ARIMAX analyses: a case study in endemic districts of Bhutan.** *Malar J* 2010, **9:**251.

69. Zhou S HF, Shen Y. J Pathogen Biol ;2:. **Application of ARIMA model on prediction of malaria incidence.** *J Pathogen Biol* 2007, **2:**284–286.

70. Zhu JM TL, Zhou SS, et al. . 2007;25:232–6.: **Study on the feasibility for ARIMA model application to predict malaria incidence in an unstable malaria area.** *Chin J Parasitol Parasitic Dis* 2007, **25:**232-236.

71. Gomez-Elipe A, Otero A, van Herp M, Aguirre-Jaime A: **Forecasting malaria incidence based on monthly case reports and environmental factors in Karuzi, Burundi, 1997-2003.** *Malar J* 2007, **6:**129.

72. Kumar V, Mangal, A, Panestar, S., et al **Forecasting Malaria Cases Using Climatic Factors in Delhi, India: A Time Series Analysis.** *Malaria Research and Treatment* 2014, **2014**.

73. Ezekie D, Opara, J. & Idochi, O: **Modelling and Forecasting Malaria Mortality Rate using SARIMA Models (A Case Study of Aboh Mbaise General Hospital, Imo State Nigeria).** *Science Journal of Applied Mathematics and Statistics* 2014, **2:**31-41.

74. Zhang X, Zhang, T., Young, AA. & Li, X.: **Applications and Comparisons of Four Time Series Models in Epidemiological Surveillance Data.** *PLoS ONE* 2014, **9**.

75. Cleveland R, Cleveland WS., McRae, JE. & Terpenning, I: **STL: A Seasonal-Trend Decomposition Procedure Based on Loess.** *Journal of Official Statistics* 1990, **6:**3-73.

76. Dickey DF, WA: **Distribution of the Estimatore for Autoregressive Time Series With a Unit Root.** *Journal of the American Statistical Association* 1979, **74:**427-431.

77. Kwiatkowski D, Phillips, PCB., Schmidt, P. & Shin, Y: **Testing the null hypothesis of stationarity against the alternative of a unit root.** *Journal of Econometrics* 1992, **54:**159-178.

78. Box GEPJ, G.M: *Time Series Analysis: Forecasting and Control.* San Fransisco, Holden day1976.

79. Yule G: **On the method of investigating periodicities in disturbed series with special reference to Wôlfer's sunspot numbers.** *Philosophical Transactions* 1927, **A:**267-298.

80. Slutsky E: **The summation of random causes as the source of cyclic processes.** *Econometrica* 1937, **5:**105-146.

81. Wold H: ***A Study in the Analysis of Stationary Time Series***. 2nd. Ed. edn. Stockholm: Almqrist & Wiksell; 1954.

82. Pankratz A: ***Forecasting with Univariate Box–Jenkins Models: Concepts and Cases***. New York: John Wiley & Sons; 1983.

83. Ljung GB, GEP.: **On a measure of lack of fit in time series models.** *Biometrika* 1978, **2:**297-303.

84. Chatfield C: **The Holt-Winters forecasting procedure.** *J Roy Statist Soc* 1978, **27:**264–279.

85. Hyndman RK, AB. : **Another look at measures of forecast accuracy.** *Int J Forecasting* 2006, **22:**679–688.

86. Armstrong JS CF: **Error measures for generalizing about forecasting methods—empirical comparisons.** *Int J Forecasting* 1992, **8:**69-80.

87. Diebold FXM, R.S.: **Comparing Predictive Accuracy.** *Journal of Business and Economic Statistics* 1995, **13:**253 - 263.

88. Hay SI, Were EC, Renshaw M, Noor AM, Ochola SA, Olusanmi I, et al: **Forecasting, warning, and detection of malaria epidemics: a case study.** *Lancet* 2003, **361:**1705-1706.