ISSN: 0975-833X

Available online at http://www.journalcra.com

International Journal of Current Research
Vol. 11, Issue, 12, pp.8650-8655, December, 2019

DOI: https://doi.org/10.24941/ijcr.34264.12.2019

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

RESEARCH ARTICLE

TIME SERIES ANALAYSIS OF MONTHLY PREVALENCE OF MALARIA IN KEMISSIE TOWN

***Getahun Worku Babolet**

Department of Statistics, Wollo University

ARTICLE INFO

Article History:
Received 14th September, 2019
Received in revised form
28th October, 2019
Accepted 15th November, 2019
Published online 30th December, 2019

Key Words:
Seasonal, SARIMA,
ARIMA,
Decomposition.

ABSTRACT

Malaria is worldwide and very common in the developing countries because it spread due to low social and economic conditions. This study was conducted in Kemissie health center, Kemissie town. The objective of the study was identifying the pattern of the prevalence of malaria infection in Kemissie town. Moreover, the study identifies the most appropriate timeseries model and estimate the parameters of the model, and forecast the expected numbers of malaria infected for 24 months Kemissie town. The data used for this study was obtained from Kemissie Health center administrative record office. These data recorded for 72 months of observations collected form reported documents on the prevalence of malaria. In order to analyze the dataset, descriptive statistics and inferential statistics is used. The result of this study shows that relatively large numbers of malaria infected on the month December, January, February, March and April, on the other hand, June, July, and August small numbers of malaria infected seasonal. Furthermore, the appropriate timeseries model for monthly prevalence of malaria in Kemissie town SARIMA $(0,1,1)x(1,1,1)[12]$.

Citation: Getahun Worku Babolet. **2019.** "Time series analaysis of monthly prevalence of malaria in kemissie town", *International Journal of Current Research*, 11, (12), 8650-8655.

INTRODUCTION

Malaria is worldwide and very common in the developing countries because it spread due to low social and economic conditions. The occurrence of malaria epidemics has been more different races ethnic and age groups. Malaria infects more geographic areas, which are sustainable for existence of the anopheles' mosquito. Globally, over 40% of the world's population are infected by malaria. In addition, over 2000 million people exposed varying degree of malaria risk in some 100 countries. The tropical Africa, India and south-east Asia are the parts of the world where malaria species in the world (Breman, 2001). Plasmodium falciparum the deadliest of the malaria parasite and the principal cause of malarial infection in tropical Africa kills about 1 million infection and young children each year, 90% of whom are from tropical Africa. Those who survive childhood plasmodium falciparum malaria acquire partial on full immunity infection that is maintained through inoculation, without which acquired immunity is soon compromised *(*Gallup and Sachs, 2001). In 2010 there were an estimated 216 million cases of malaria worldwide, of which 91% were due to Plasmodium falciparum. The last majority of cases 81% were in the African region followed by south East Asia (13%) and eastern Mediterranean regions 5% *(Mota MM et al, 2001).*

***Corresponding author: Getahun Worku Babolet,**
Department of Statistics, Wollo University.

In Ethiopia 74% of the land which is below 2000meter above sea level favored the transmission of malaria and 66% of the total population living in malaria area *(Gallup and Sachs, 2001).* The prevalence of malaria in Kemissie town was found to be high. The prevalence was strongly associated with proximity of residence potential mosquito breeding site. The occurrence of the disease among fewer than 5 years old children and would indicate that malaria is indigenous to the area. Use of personal protection methods such as insecticide treated. Mosquito nets should be scaled up and malaria a control intervention should target resident who are at closer proximity to mosquito breeding sites *(EMoH, 2008).* According to Ethiopian Minister of Health (EMoH, 2008) reports, Malaria control interventions should target are within 2002-2011 a total of 59,208 blood films were requested for malaria diagnosis in kola up, and malaria control interventions should target and 23,477(39.6%) microscopically confirmed malaria cases were reported in the town with a fluctuating trend. Regarding the identified plasmodium species, plasmodium falciparum and plasmodium vivax accounted for 75% of malaria morbidity, respectively. Malaria was reported in all age groups and sex, but the 15-44-year age group and males were more affected. Despite the apparent fluctuation of malaria attends in area, the highest peak of malaria case was reported during spring seasons (EMoH, 2008). Even if malariais the most common disease in Kemissie, its distribution among the age groups or sex are not well known in Kemissie town. Similarly, there is no clear statistical

information whether their prevalence has been decreasing or not. The major objective of this study was to identify the pattern of the prevalence of malaria infection in Kemissie town. Specifically, this study identifies appropriate model and estimate the parameters of the model, forecast number of malaria infection for 24 months.

MATERIALS AND METHODOS

Description of The Study Area: The study was conducted in Kemissie town at Kemissie Health Center. Kemissie town was located in Ethiopia, Amhara region which is special administrative center of the Oromia Zone far from 326 km to Addis Ababa, capital city of Ethiopia to the North direction. This town has a latitude and longitude of 10°43′N 39°52′E with an elevation of 1424 meters above sea level.

Source of Data: In Kemissie town, prevalence of malaria was recorded in number unit per person daily basis. The data provide information on monthly and yearly malaria prevalence recorded in number unit per mans. This study was employeda secondary data obtained from the health center of Kemissie town. The data were recorded for 6years from years June 20011-May 2018.

Box-Jenkins ARIMA Process of Model Analysis: Box-Jenkins forecasting models consist of a four-step iterative procedure as follows; Model Identification, Model Estimation, Model Checking (Goodness of fit) and Model Forecasting. The four iterative steps are not straight forward but are embodied in a continuous flow chart depending on the set of data under study.

Model Identification (Selecting an Initial Model): The first task to determine whether the series is stationary or not by dealing plots, ACF and PACF graphs. If a time series is plotted looks like no evidence of a change in a mean and variance over time then we expected to the time series is stationary. If a graph of *ACF* of the time series values either cuts off fairly quickly or dies down fairly quickly, then the time series values should be well-thought-out stationary. If a graph of *ACF* dies down extremely slowly, then the time series values should be considered non-stationary. The autocorrelation of stationary data drops to zero relatively quickly, while for non-stationary data they are significantly different from zero for several and *PACF* will have a large pike close to 1 at lag 1. If the series is not stationary; it would then be converted to a stationary series by differencing. That is, the original series is replaced by a series of differences. An *ARIMA* model is then specified for the differenced series. Differencing is done until a plot of the data indicates the series varies about a fixed level, and the graph of *ACF* either cuts off fairly quickly or dies down fairly quickly. Once a stationary series has been obtained, then the form of the model to be used can be identified.

Augmented Dickey–Fuller test (AD): Augmented Dickey–Fuller test (AD) is a test for a unit root in a time series sample. It is an augmented version of the Dickey–Fuller test for a larger and more complicated set of time series models. The augmented Dickey–Fuller (ADF) statistic, used in the test, is a negative number. The more negative it is, the stronger the rejection of the hypothesis that there is a unit roots at some level of confidence.

Phillips–Peron (pp.) test: The Phillips–Peron test is a unit root test. That is, it is used in time series analysis to test the null hypothesis that a time series is integrated of orders one. It builds on the Dickey–Fuller test of the null hypothesis in where is the first difference operator. Like the augmented Dickey–Fuller test, the Phillips–Peron test addresses the issue that the process generating data for might have a higher order of autocorrelation than is admitted in the test equation - making endogenous and thus invalidating the Dickey–Fuller t-test. Whilst the augmented Dickey–Fuller test addresses this issue by introducing lags of as regressors in the test equation, the Phillips–Peron test makes a non-parametric correction to the t-test statistic.

White Noise (Pure Random) Series: This model is purely random process with sequence of independent and identically distributed variables. The white noise random model is given as with constant mean and variance

Methods of Data Analysis: The data was analyzed by using R softwar statistical packages and information of time series analysis of trend, ACF, PACF and Box-Jenkins analysis computed (*Hastie T, Tibshirani R & Friedman J, 2009*).

Box-Jenkins ARIMA Process of Model Analysis: Autoregressive, moving average (ARIMA) time series models is appropriate for time series of medium to long length at least 50 observations is needed (Box and Jenkins, 1976). Model autoregressive, AR process with the series is linear aggregate of a finite number plus random shock at parameter.

$$AR(p): Y = \Phi_0 + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \cdots + \Phi_p Y_{t-p} + \varepsilon_t$$

That is Moving average, MA modelis the current time series is linear and finite number of variable with random shock and at order. That is

$$MA\ (q):\ X_t = \theta_0 + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

ARM Amodel is the combination of the above two models with the respective to orders.

$$ARMA(p,q): Yt = \Phi_0 + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \cdots + \Phi_p Y_{t-p} + \theta_0 + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

ARIMA model is a model for non-stationary processes of data which is needed differencing to be stationary. That is

$$ARIMA(p,d,q): Yt = \nabla^d y_t = (1-\theta)^d y_t$$

$$Y_t = \Phi_0 + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \cdots + \Phi_p Y_{t-p} + \theta_0 + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots\cdots + \theta_q \varepsilon_{t-q}$$

$$Y_t - \Phi_0 - Y_{t-1} - \Phi_2 Y_{t-2} + \cdots + \Phi_p Y_{t-p} = \theta_0 + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots\cdots + \theta_q \varepsilon_{t-q}$$

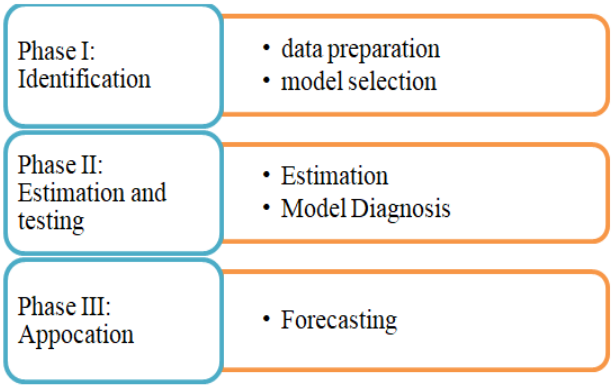
Seasonal ARIMA Model

This model, however, only takes in to account the autocorrelation at seasonal lags s, 2s, 3s …*(Hamilton, 1994).* Hence amore general seasonal *ARIMA* model of order$(p,d,q) \times (P,D,Q)$ with period s is $SARIMA(p,q,q)x(P,D,Q)$

$$Y_t = \Phi_0 + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \cdots + \Phi_p Y_{t-12} + \theta_0 + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-12}$$

Where $p - autoregressive\ parameter$
$d - Order\ of\ difference$
$q - Moving\ average\ parameter$
$P - Autoregressive\ parameter\ for\ seasonal$
$D - Order\ of\ difference\ for\ seasonal$
$Q - Moving\ average\ parameter\ for\ seasonal$
$\theta_1, \theta_2 \ldots \theta_q\ and\ \Phi_1, \Phi_2 \ldots \Phi_p are\ coefficient\ of\ the\ model$
$\Phi_0\ and\ \theta_0\ are\ constant$

Box-Jenkins forecasting models consist the following three phases;



Akaike Information Criterion (AIC): The final model after estimation can be selected using a penalty function statistic such as the Akaike Information Criterion (AIC), a measure of the goodness of fit an estimated statistical model. Malaria data set, several competing models may be ranked according to their AIC with one having the lowest information criterion value being the best. Generally, the AIC is calculated using the relation,

$$AIC = 2k - 2log(L)\ (10)$$

Where k is the number of parameters in the statistical model

Model Checking (Goodness of Fit): In this step, the model must be checked for adequacy by considering the properties of the residuals whether there siduals from an ARIMA model must have the normal distribution or should be random *(Shumway and Stoffer, 2010)*.

Use the Ljung-Box q statistic to test whether a series of observations over time are random and independent. In general, the Box-Ljung test is defined as Ho: the model does not exhibit lack of fit. The test statistic Q is given as;

$$Q_m = n(n+2)\sum_{k=1}^{m}\frac{\hat{r}_k^2}{n-k} \sim x_{m-r}^2$$

where $\hat{r}_k$ is the estimated autocorrelation of the series at lag k, and m is the number of lags being tested, n is thenumber of residuals, and m is the number of times lags is included in the test.

Forecasting: Once a model has been identified and its parameters have been estimated, the next procedure is forecast for the future value of a time series. Forecasting with this system is the expected values. Confidence intervals may also be easily derived from the standard errors of the residuals (Brockwell *et al.,* 2000).

RESULTS AND DISCUSSION

This section is about analysis and discussion about monthly data gained from Kemissie health center from June, 2004 to May,2010 as Ethiopian calendar, using R statistical software.

Time series plot of data

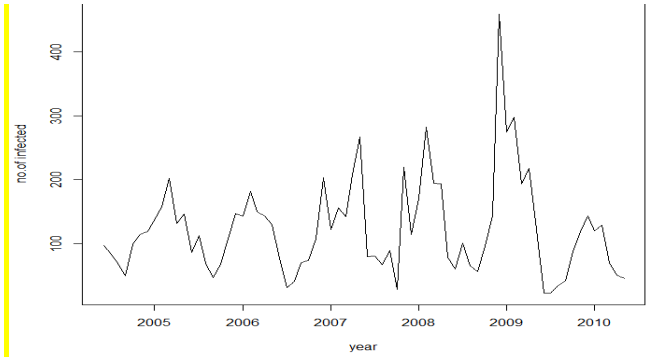


Figure 1. Time series plot of malaria infected for original data

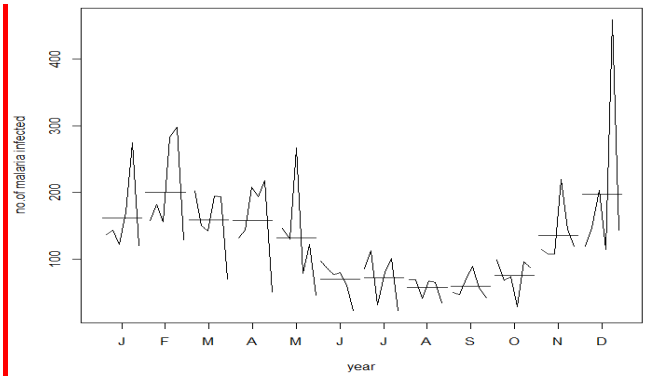


Figure 1. Monthly plot of malaria infected people

ACF (Autocorrelation Function): From the figure 4 the autocorrelation function rapidly decreases from 1 to 0 and increase rapidly from 0 to 1at lag 1. This **i**mplying that the data is stationary but there is some seasonality**.** Regarding to partial autocorrelation function the graph rapidly decrease, indicate the data is stationary but there is some seasonality because the lags cutoff after some lag.

Model Checking (Goodness of Fit): In this step, the model must be checked for adequacy by considering the properties of the residuals whether the residuals from an SARIMA model must have the normal distribution or should be random. An overall check of the model adequacy is provided by using the Ljung-Box statistic.

Forecasting: We are going to forecast the number of people who infected by malaria for the next 24 months. Based on the overall results time series model for monthly prevalence of malaria of Kemissie town was adjusted, processed, diagnostically checked and lastly SARIMA model is established with a 95% prediction interval that can adequately be used to forecast 24 monthly malaria prevalence values. The number of malarial infected distribution in Kemissie is found to be relatively high from November to May and relatively from June to October. The appropriate time series model for the change of malaria is *SARIMA (0, 1, 1) x (1,1,1) [12]*.

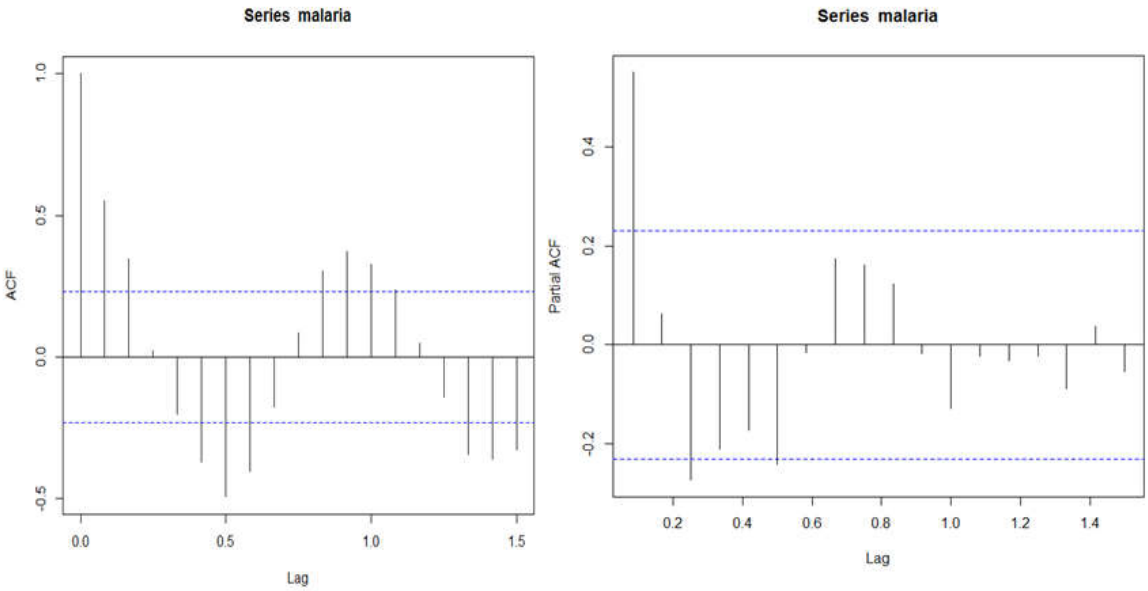


Figure 4. ACF and Partial ACF for malaria before seasonal differencing

Table 1. Tests the results for original data

Variables	Tests	Dickey-Fuller& Dickey-Fuller Z(alpha)	P-value
Malaria	Augmented Dickey-Fuller	-4.7051	0.01
	Phillips-Peron Unit Root	-33.967	0.01

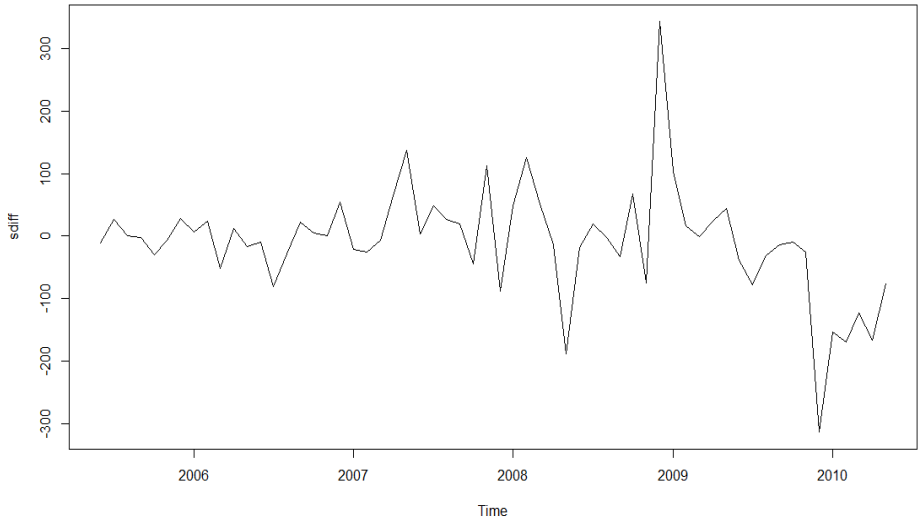


Figure 5. Plot of time series after seasonal differencing

Table 2.Tests the results for seasonal differencing data

Variables	Tests	Dickey-Fuller& Dickey-Fuller Z(alpha)	P-value
Malaria	Augmented Dickey-Fuller	-2.3199	0.0445
	Phillips-Peron Unit Root	-45.19	0.01

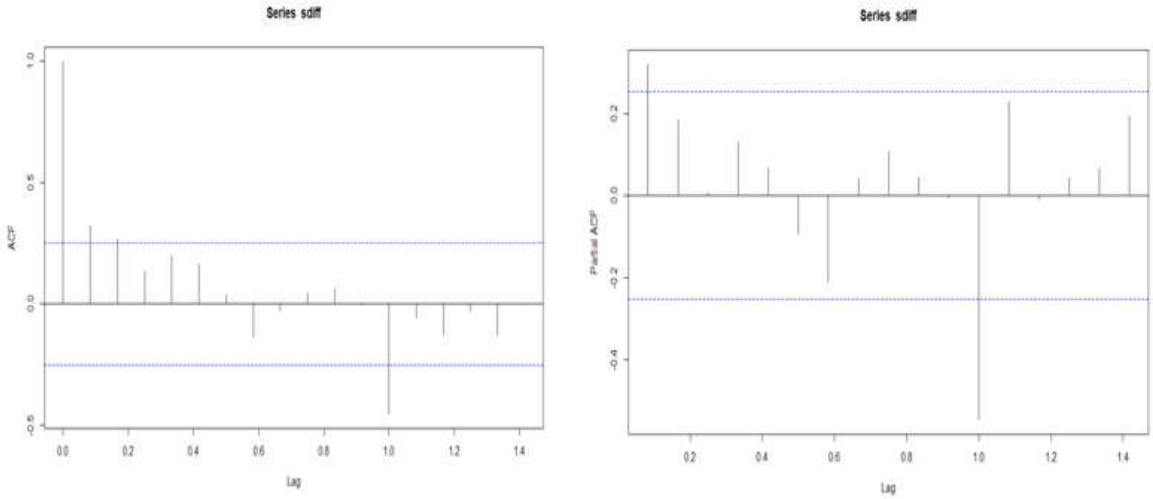
Generally using Augmented Dickey-Fuller Test and Phillips-Peron Unit Root Test the series is data stationary.

Table 3. Behavior of the ACF and PACF for ARMA Models

	AR(p)	MA(q)	ARMA(p, q)
ACF	Tails off	Cuts off after lags q	Tails off
PACF	Cuts off after lags p	Tails off	Tails off

Table 4. Behavior of the ACF and PACF for Pure SARIMA Models

	AR(p)s	MA(q)s	ARMA(p, q)s
ACF	Tails off at lags Ks, K=1, 2 ,…	Cuts off after lags Qs	Tails off at lags Ks
PACF	Cuts off after lags Ps	Tails off at lags Ks	Tails off at lags Ps



For ARIMA
$ACF\ cutoff\ after\ lag\ 1\ MA\ (p = 1)$
$PACF\ cutoff\ after\ lag\ 1\ AR\ (d = 1)$
$PACF\ cutoff\ after\ lag\ 2\ AR\ (d = 2)$

For Seasonal ARIMA

$ACF\ cutoff\ after\ lag\ 1(12)\ SMA\ (P = 1)$
$PACF\ cutoff\ after\ lag\ 1(12)\ SAR\ (D = 1)$

Figure 6. ACF and Partial ACF for malaria of seasonal differencing

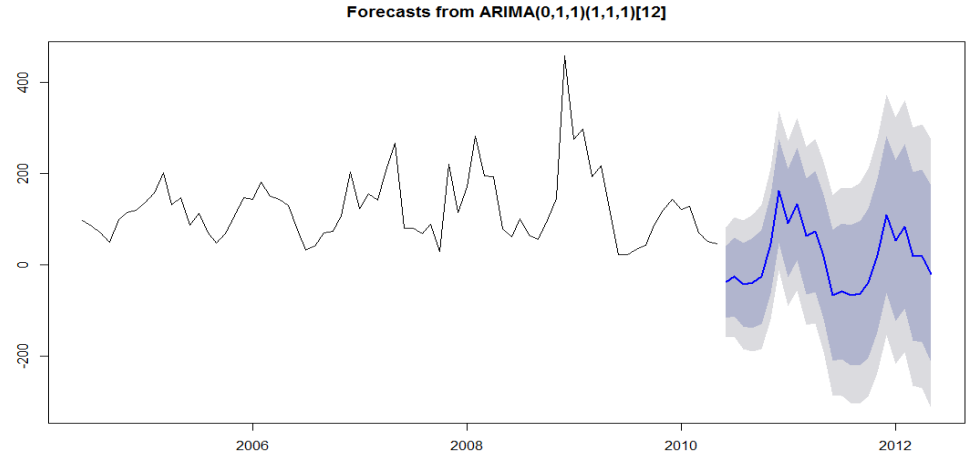
	Candidate Models	*df*	*AIC*	*BIC*
1.	$SARIMA\ (1,0,1)\ x\ (1,1,1)\ [12]$	5	677.7311	688.2028
2.	$SARIMA\ (1,0,2)x\ (1,1,1)\ [12]$	6	678.7488	691.3149
3.	$SARIMA\ (0,1,1)x(1,1,1)\ [12]$	4	670.7441	679.0543

Table 5. Final Estimates of Parameters of malaria

Type	*Coef*	*SE Coef*	*T*	*P-value*
MA1	-0.5758	0.1524	-12.92	0.000
SAR1	-0.2930	0.2128	-0.32	0.746
SMA1	-0.5464	02772	-12.66	0.000
Mean	-0.1630			
Constant	-1.0111	0.2119	-4.77	0.000

Table 6. Forecasted value of malaria

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2010						9.78	38.20	32.41				
2011	187.02	233.06	169.79	181.33	130	50	58.65	50.39	44.25	61.03	138.95	250.91
2012	170	203.87	141.60	145.35					55.78	77.23	138.92	222.91



Generally, we conclude from the above time series analysis of the last model of malaria was fluctuation of disease Kemissie health center.

REFERENCES

Box, G. E. P., and Jenkins, G.M. 1976. Time Series Analysis: Forecasting and Control; Holden day.

Breman J. G. 2001. The Ears of the Hippopotamus: Manifestations, Determinants, and Estimates of the Malaria Burden. *American Journal of Tropical Medicine and Hygiene*. 64 (Suppl. 12):1–11. https://www.ncbi.nlm.nih.gov/pubmed/11425172.

Brockwell, P.J., Davis, R.A. and Rockwell, P.J. 2000. Introduction to time series and forecasting, 2nd edition, Springer texts in statistics.

Chitnis 2001. Gordon Research Conference on Malaria will be held in Oxford, August 5 - 10, 2001 Chair Chris New bold.

Emanuel, Rubin 2010. Malaria Vision Award, PA; Nov 21. Healthcare, Guaranteed. (New York: Public Affairs Press 2008. Emanuel EJ, Crouch R, Grady C, Lie R, Miller F,

G20. 2011. Introduction of time series analysis

Gallup JL., Sachs JD. 2001. The economic burden of malaria. *Am J Trop Med Hyg.,* 64: 85–96

Hamilton, J. 1994. Time Series Analysis. Princeton University Press, New Jersey.

Hand DJ., Mannila H. & Smyth P. 2001. Principles of data mining. MIT Press, Cambridge, MA.80

Hastie T., Tibshirani R. & Friedman J. 2009. The elements of statistical learning : data mining,inference, and prediction. Springer Series in Statistics. Springer-Verlag, New York, NY, 2nd edition.

Marsland S. 2014. Machine learning: An algorithmic perspective. Chapman and Hall/CRC, BocaRaton, FL.

Mota, MM., Pradel, G., Vander berg, JP. 2001. Malaria vaccine developments. J Infect Dis. 2001.

Shumway, R.H. and Stoffer, D.S. 2010. Time series analysis and its applications with R Examples, Springer, 3nd Ed.

World Health Organization 2012. Global Malaria Disease surveillance for malaria control, Geneva, Switzerland, An operational manual. ISBN: 9789241503341.

****