International Journal of Statistics and Applied Mathematics 2018; 3(6): 98-105

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2018; 3(6): 98-105

www.mathsjournal.com
Received: 13-09-2018
Accepted: 14-10-2018

Nkpordee Lekia
Department of Mathematics/ Statistics, Ignatius Ajuru University of Education, Port Harcourt, Nigeria

Nduka Wonu
Department of Mathematics/ Statistics, Ignatius Ajuru University of Education, Port Harcourt, Nigeria

Correspondence
Nkpordee Lekia
Department of Mathematics/ Statistics, Ignatius Ajuru University of Education, Port Harcourt, Nigeria

Application of time series analysis on the forecasting of the outbreak of malaria epidemic in Nigeria

Nkpordee Lekia and Nduka Wonu

Abstract
This study examined the application of time series analysis in forecasting of the monthly outbreak of malaria epidemic in Nigeria. The secondary data was obtained from the National Bureau of Statistics (NBS). The model; ARIMA (1,0,1) was used to forecast the monthly reported cases of malaria with 16 months lead, which shows a gradual increase and decrease in the series. It was recommended among others that the Government should ensure the provision of insecticide-treated nets, insecticides and anti-malaria drugs in the rural areas in Nigeria.

Keywords: Application, time series analysis, forecasting, outbreak, malaria

Introduction

Today, one of the most serious health challenges facing the world is malaria. A recent report from the World Health Organization (2018) [12] shows that irrespective of the efforts to control malaria, data from 2015-2017 indicated that there appears to be no significant progress in minimizing the cases of malaria globally during this period. The report further shows that there was an estimated 219 million cases and 435,000 related deaths worldwide in 2017 with 3.1 billion dollars in funding to reduce the scourge. Specifically, WHO (2018) [12] also reported that most, 90% of the malaria-related deaths in 2015 were in the African region with South-East Asia and Eastern Mediterranean WHO regions represented with 7%and 2% respectively. A huge proportion of the Nigerian population lives in abject poverty in the rural areas, where there is neither access to adequate healthcare nor drinkable water. Nigeria is faced with massive foreign debt issues amidst low income. There is the probability of Nigeria sinking into more debt while struggling with sick residents because its economic growth seriously depends on the good health of the populace (Carrington, 2001) [5].

The prevalence of malaria in Africa is influenced by poverty. According to Pattanayak *et al*., (2006) [10] many of the world's poorest people reside in locations of high rates of malaria. Financial constraint hinders the access of these people to effective healthcare. The financial status of a susceptible nation plays an added part in ascertaining the adequate facility-based preparation and control measures in case of epidemics (Sudhakar & Subramani 2007) [11]. It is, therefore, imperative to forecast the monthly outbreak of malaria using time series analysis in an attempt to identify the current trend in its prevalence in Nigeria and make recommendations where necessary.

Mathematically, a time series is defined by the functional relationship $Y = F(t)$, where Y is the value of the variable under consideration at time t. Lin *et al*., (2009) [8] used time series analysis to investigate the relationship between Falciparum malaria in the endemic provinces and the imported malaria in the non- endemic provinces of China. An autoregressive integrated moving average model was first fit to the predictor variable. Out of all the models tested, the seasonal ARIMA (1,1,1,) (0,1,1,) 12 model for malaria cadence, fits the data best according to the AIC and goodness-of-it criteria. Briet *et al*., (2008) [3], formulated a model for short-term malaria prediction in Sri Lanka.

Contreras *et al*., (2003) [6] developed a model for predicting the next-day electricity prices in mainland span and California markets using an ARIMA model.

Their developed model was able to forecast the 24 market clearing prices of tomorrow. The ARIMA model is an effective tool for forecasting time series. A good model is to be developed for forecasting the malaria cases. A model fitting quality is defined as the sum of the residuals squares divided by the sample size. Its objective is to measure the model's capacity to produce the sample data (Guerrero, 2003) [7].

Purpose of the Study

This study investigated the applicability of time series analysis in forecasting the outbreak of malaria epidemic in Nigeria. (i.e reported monthly cases of malaria epidemic in Nigeria January 2009 – April 2013)

Model Formation

This research work used ARIMA models to analyse the data. ARIMA model consists of autoregressive integrated and moving average models. The ARIMA model incorporates both non-seasonal and seasonal factors in a multiplicative model. A forecast was calculated by carefully inserting a time value of malaria cases reported in Nigeria into the Minitab statistical package. Box-Jenkins (1976) [1] methodology was used to construct a suitable mathematical model by putting the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) correlogram into consideration.

Mathematical Formation

One shorthand notation for the ARIMA model is given by ARIMA (P,d, q,)X (P,D,Q,) S,
With P = non-seasonal AR order,
d = non- seasonal differencing
q = non- seasonal MA order
P = monthly AR order
D = monthly differencing
Q= monthly MA order, and
S = time span of repeating monthly pattern.

Without differencing operations, the model could be written more formally as:
$\Phi(B^s)\ \Psi(B)\ (X_{t-\mu}) = \Theta(B^s)\ \theta(B)\ \omega_t$
The non-seasonal components are:
AR: $\Psi(B) = 1-\Psi_1 B-\ldots-\Psi_p B^p$
MX: $\theta(B) = 1+\theta_1 B+\ldots+\theta_q B^q$
The seasonal components are:
AR: $\Phi(B^s) = 1-\Phi_1 B^s-\ldots-\Phi_p B^{ps}$
MA: $\Theta(B^s) = 1+\Theta_1 B^s+\text{----}+\Theta_Q B^{Qs}$.
Hence, the forecasting model is $(1-\Phi_1 B)(1-\Psi_1 B)(X_{t-\mu}) = \omega_t$

Method of Solution

A time series is said to follow autoregressive moving average model of p,q [ARMA (p,d,q)] if it satisfies the difference equation;

$X_t - \phi 1 X_{t-1} - \ldots - \phi_p X_{t-p} = e_t - \theta_1 e_{t-1} - \ldots - \theta_q e_{t-q}$ ------------ (1)
Or
$\phi(B)X_t = \theta(B)e_t$ ------------ (2)
Where $\phi(B) = 1-\phi_1(B) - \ldots - \phi_p B^p$ ------------ (3)
$\theta(B) = 1-\theta_1(B)-\ldots-\theta_q B^q$ ------------ (4)

The error term is generally assumed to be independently and identically distributed random variable from a normal distribution with mean zero and variance, $\sigma^2 < \infty$.
For the process (2) to be stationary, we required that the root of $\phi(B) = 0$ lie outside the unit circle. To be invertible we required that the roots of $\theta(B) = 0$ lies outside the unit circle. We assume that $\phi(B) = 0$ and $\theta(B) = 0$ share no common roots.
Procedure for choosing p and q often known as model identification as well as estimation of time series models can be found in Box, Jenkins and Reinsel (1994) [2], Chartfield (2004) [4].

Trend Analysis

From figure 1, the time series has a linear trend model given by; Yt=9439.41-30.2194t, where Yt is the number of monthly reported cases and t is the given month.

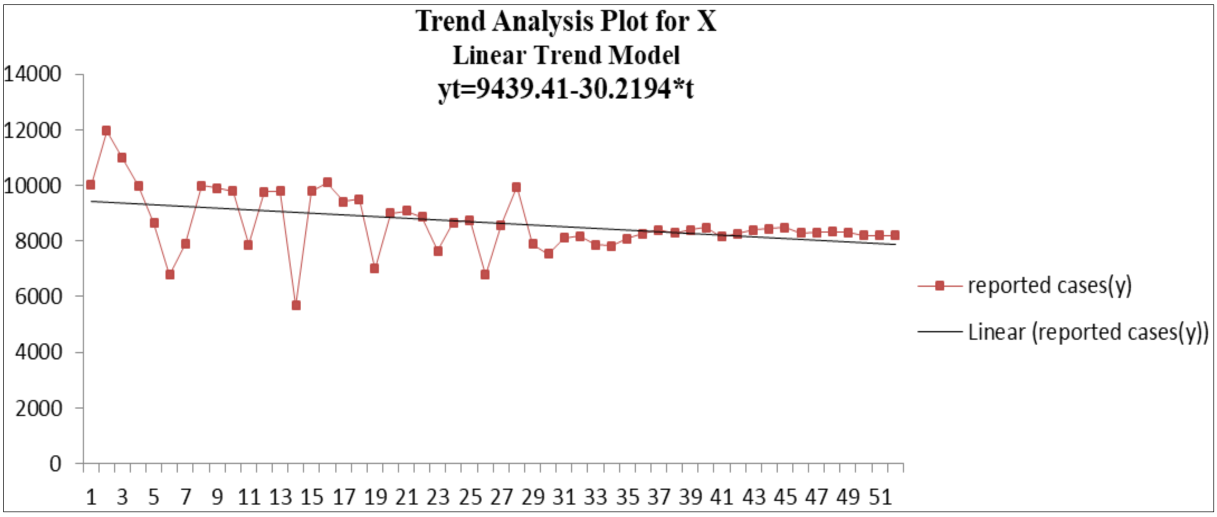

Fig 1: Trend Analysis Plot For X

Preliminary Test for Stationarity

A stationary series has a constant mean, constant variance and constant autocorrelation structure. The time series plot of the monthly reported cases from January 2009 to April 2013 is shown in Figure 1. The figure showed that the time series was non-stationary. The values had irregular swings and hence had irregular variability. Where X is the reported cases of malaria epidemic.

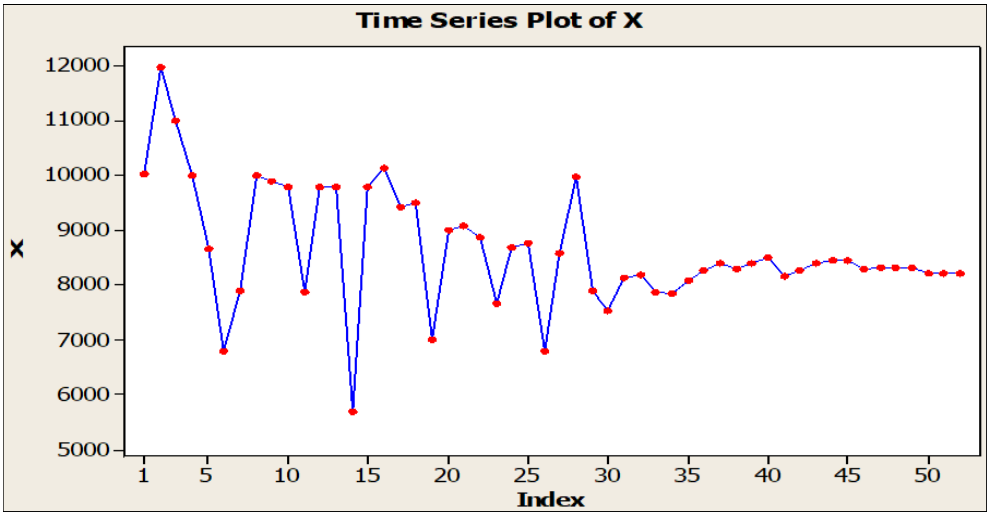

Fig 2: Time Series Plot of Reported Cases.

But the fact that there were irregular fluctuations in the plot was not a clear indication that the series was non-stationary. This called for verification to confirm the non-stationarity as indicated by the plotting of autocorrelation function and partial autocorrelation function.

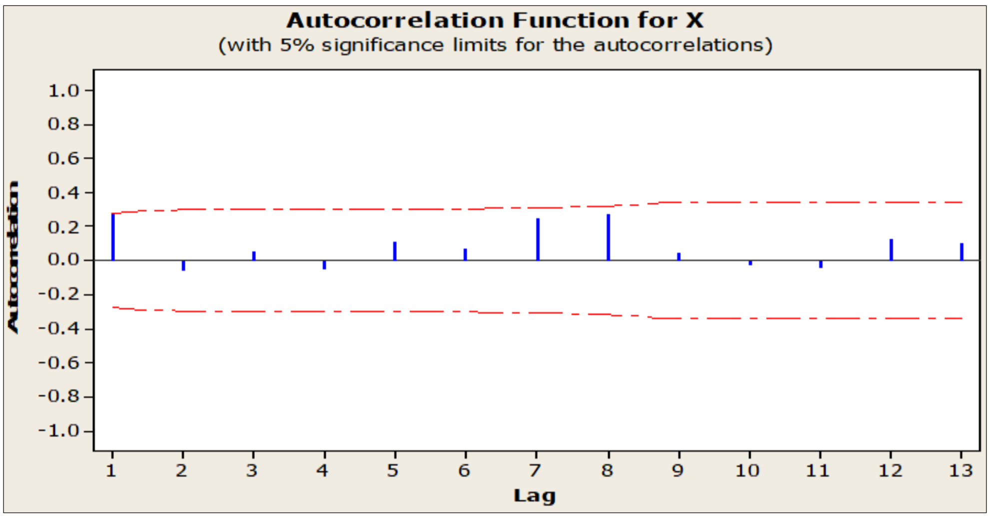

Fig 3: Autocorrelation Function

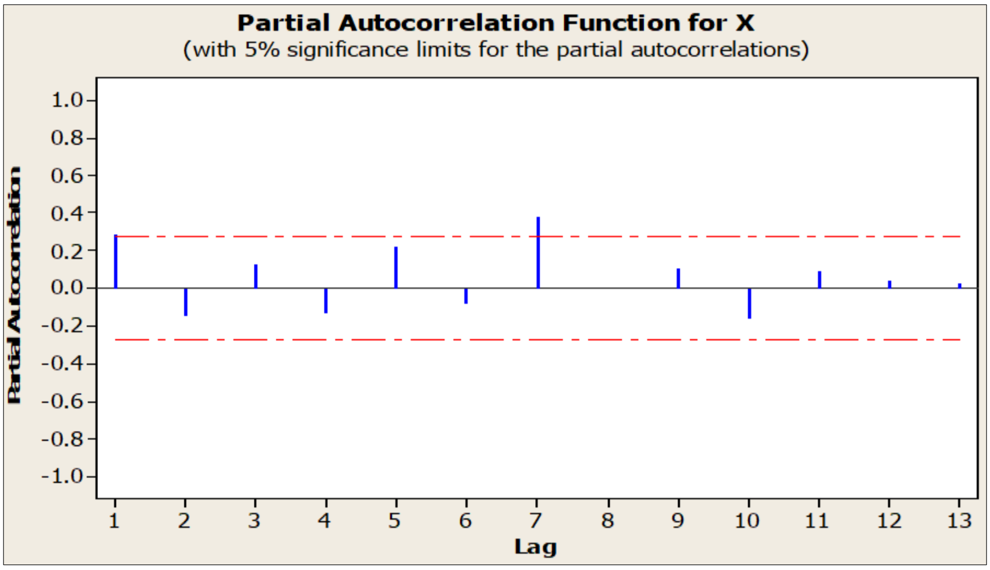

Fig 4: Partial Autocorrelation Function

Model Selection

ARIMA models are usually estimated after transforming the variable under forecasting into a stationary series. A stationary series is one whose values vary over time only around a constant mean and variance. The selection of the appropriate model depends on the values of Normalized BIC and the ACF together with the PACF. The graphs of the ACF and PACF are shown in Figures 3 and 4 respectively. Three tentative models were entertained and the model with the minimum Normalized BIC and AIC was chosen.

Model

ARIMA (1, 0, 1)

ARIMA (0, 1, 1)

ARIMA (1, 1, 1)

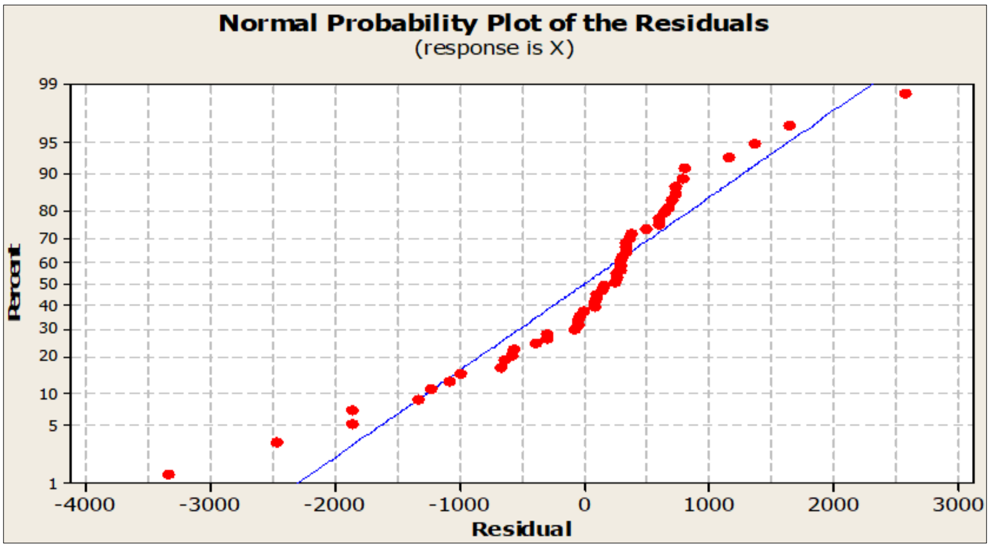

Fig 5: Normal Probability Plot of the Residuals

Table 1: Sample ACF and PACF for Malaria Epidemic in Nigeria

lag	ACF	T	LBQ	PACF	T
1	0.281049	2.03	4.35	0.281049	2.03
2	-0.058797	-0.39	4.54	-0.149603	-1.08
3	0.052323	0.35	4.70	0.125906	0.91
4	-0.051723	-0.34	4.86	-0.131118	-0.95
5	0.109094	0.73	5.57	0.213327	1.54
6	0.070864	0.47	5.87	-0.081567	-0.59
7	0.250205	1.64	9.78	0.378171	2.73
8	0.274648	1.71	14.59	0.002753	0.02
9	0.044501	0.26	14.72	0.111489	0.80
10	-0.028569	-0.17	14.78	-0.157328	-1.13
11	-0.042513	-0.25	14.90	0.093847	0.68
12	0.122194	0.72	15.95	0.040778	0.29
13	0.101928	0.60	16.70	0.022707	0.16

The ACF and PACF for X_t are given in Table 1. It can be seen from Table 1 that the ACF and PACF cut off after lag 1, indicating that X_t is stationary. A close examination of the ACF and PACF of X_t suggesting an ARMA (1, 0, 1) model which is given by.

$$x_t = \mu + \phi_1 x_{t-1} - \theta_1 \ell_{t-1} + \ell_t$$

Model Parameter Estimation for Reported Cases

The ARIMA model for the Malaria Epidemic in Nigeria (Jan 2009-April 2013)
AR1: ϕ_1 = -0.4471 with standard error of 0.1528
MA1: Θ_1 = -0.9423 with standard error of 0.0616

$$X_t = \mu - 0.4471 X_{t-1} + 0.9423 e_{t-1} + e_t$$

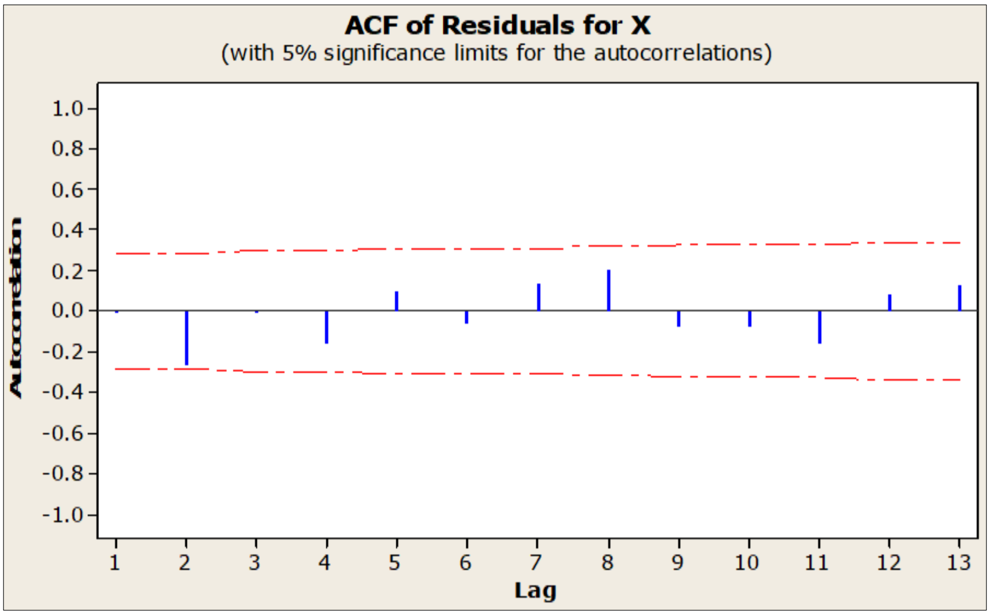

Fig 6: ACF of Residuals for Reported Cases

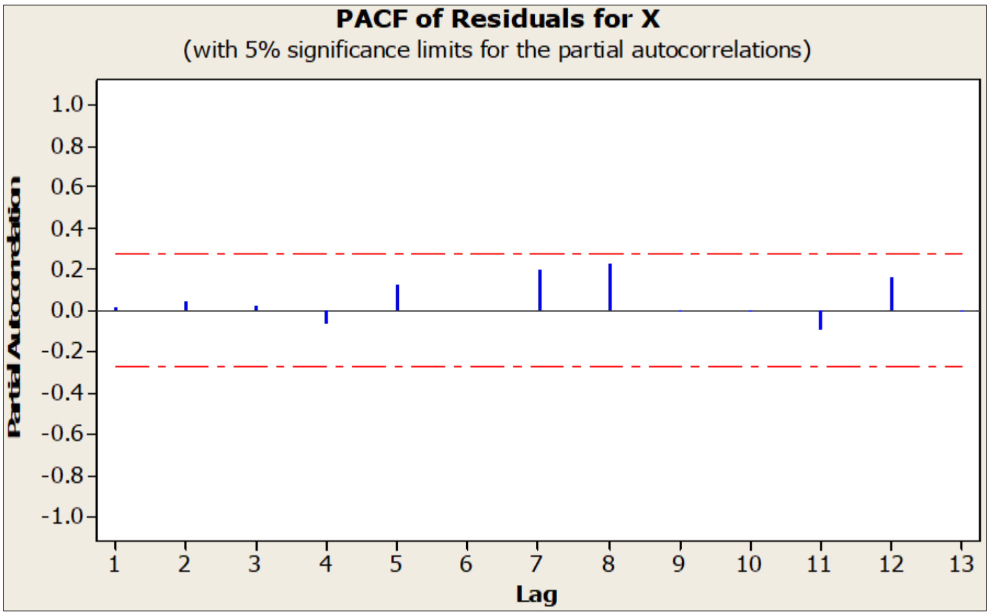

Fig 7: PACF of Residuals for Reported Cases

To determine the adequacy of the model, a diagnostic check of the residual ACF and PACF given in Table 1 was done. Following Chatfield (2004), the ACF and PACF are close to zero and lie within $\pm\frac{2}{\sqrt{n}} = \frac{2}{\sqrt{52}} = \pm 0.277$ indicating that the model is adequate or fit the data. Similarly, the values of the Q statistic (Ljung and Box (1978) [9] are not as significant as given above confirming that the model is adequate.

Forecasting

The fitted model generated the following forecast with lead time L =12

$$X_t = \varepsilon_t - 0.4471 X_{t-1} + 0.9423 e_{t-1} + e_t$$

Table 2: Forecasting

Period	Forecast
53	8601.2
54	8649.0
55	8627.6
56	8637.2
57	8632.9
58	8634.8
59	8634.0
60	8634.4
61	8638.4
62	8632.7
63	8635.2
64	8634.1
65	8634.6
66	8634.4
67	8634.5
68	8634.4

Discussion

The number of reported cases per month in 2009 is 9,468 and the number of reported cases per month in 2010 is 8,714, which shows an improvement over that of 2009. The number of reported cases per month in 2011 is 8,155 which are not statistically significantly different from that of 2010 but indicates an improvement as compared to that of 2009. In 2012, the number of reported cases per month is 8,353. This value shows an increase in the number of reported cases although not significantly different from that of 2011. The ARIMA model developed for predicting the monthly reported cases of malaria is ARIMA (1, 0, 1): $X_t = \varepsilon_t - 0.4471X_{t-1} + 0.9423e_{t-1} + e_t$ where e_t =0.

The model was used to predict a 16 months lead period of reported cases.

Appendix A: Reported monthly cases of malaria epidemic in Nigeria January 2009 – April 2013

	Monthly reported cases											
YEAR	**JAN**	**FEB**	**MAR**	**APR**	**MAY**	**JUN**	**JUL**	**AUG**	**SEP**	**OCT**	**NOV**	**DEC**
2009	10010	11954	10995	9989	8652	6788	7898	9991	9895	9786	7873	9782
2010	9784	5681	9788	10114	9425	9500	6995	8994	9087	8870	7657	8675
2011	8760	6784	8574	9958	7896	7539	8117	8177	7877	7828	8083	8267
2012	8400	8288	8399	8493	8155	8264	8391	8437	8484	8295	8307	8325
2013	8322	8211	8200	8197								

Appendix B
Arima Model: X

Estimates at each iteration

Iteration	SSE	Parameters
0	60701268	0.100 0.100 7774.273
1	55864418	0.218 -0.017 6751.753
2	55180669	0.097 -0.167 7803.683
3	54404288	-0.028 -0.317 8881.928
4	53405519	-0.156 -0.467 9988.655
5	51967997	-0.287 -0.617 11115.604
6	49844464	-0.415 -0.767 12220.975
7	47242834	-0.527 -0.917 13191.619
8	46428594	-0.448 -0.941 12506.995
9	46428122	-0.447 -0.942 12494.306

Unable to reduce the sum of squares any further

Final Estimates of Parameters

Type	Coef	SE Coef	T	P
AR 1	-0.4471	0.1528	-2.93	0.005
MA 1	-0.9423	0.0616	-15.29	0.000
Constant	12494.3	259.7	48.11	0.000
Mean	8634.2	179.5		

Differencing: 0 regular difference
Number of observations: 52, after differencing 51
Residuals: SS = 46177299 (backforecasts excluded)
MS = 942394 DF = 49

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	12	24	36	48
Chi-Square	9.7	13.9	19.5	24.6
DF	9	21	33	45
P-Value	0.372	0.872	0.970	0.994

Forecasts from period 52

95% Limits	Period Forecast	Lower Upper	Actual
53	8601.2	6698.1	10504.3
54	8649.0	6525.3	10772.7
55	8627.6	6462.5	10792.7
56	8637.2	6463.9	10810.5
57	8632.9	6458.0	10807.8
58	8634.8	6459.6	10810.1
59	8634.0	6458.7	10809.3
60	8634.4	6459.1	10809.7

Autocorrelation Function: X				Partial Autocorrelation Function: X		
Lag	ACF	T	LBQ	Lag	PACF	T
1	0.281049	2.03	4.35	1	0.281049	2.03
2	-0.058797	-0.39	4.54	2	-0.149603	-1.08
3	0.052323	0.35	4.70	3	0.125906	0.91
4	-0.051723	-0.34	4.86	4	-0.131118	-0.95
5	0.109094	0.73	5.57	5	0.213327	1.54
6	0.070864	0.47	5.87	6	-0.081567	-0.59
7	0.250205	1.64	9.78	7	0.378171	2.73
8	0.274648	1.71	14.59	8	0.002753	0.02
9	0.044501	0.26	14.72	9	0.111489	0.80
10	-0.028569	-0.17	14.78	10	-0.157328	-1.13
11	-0.042513	-0.25	14.90	11	0.093847	0.68
12	0.122194	0.72	15.95	12	0.040778	0.29
13	0.101928	0.60	16.70	13	0.022707	0.16

Conclusion

The forecast from May 2013 to August 2014 has a gradual increase and decrease in the series and constant infections. Malaria has endangered the lives of many people over the years. Education and sensitization should be taken seriously in order to avoid keeping stagnant water for mosquitoes to breed. The use of insecticide treated nets should be encouraged to minimize the spread. The anti-malaria drugs should be made available in rural areas to reduce the malaria related deaths among poor people.

Recommendations

In line with the finding of this work, it is recommended that:

1. The ARIMA model has been found to be useful in forecasting malaria epidemic and it should be used in future studies on the outbreak of malaria epidemic in Nigeria and other West African countries.
2. Parents/guardians should ensure they check the health condition of their children/wards at least once a month.
3. The Government should ensure the provision of insecticide-treated nets, insecticides and anti-malaria drugs in the rural areas in Nigeria

References

1. Box GEP, Jenkins GM. Time series analysis: Forecasting and control. Rev. ed. San Francisco. Holden-Day, 1976.
2. Box GEP, Jenkins GM, Reinsel GC. Time Series Analysis: Forecasting & Control, 3rd edition. Prentice Hall, New Jersey, 1994.
3. Briet OJT, Vounatsou P, Amerasinghe PH. Malaria seasonality and rainfall seasonality in Sri Lanka are correlated in space. Geospatial Health. 2008; 2(2):183-190.
4. Chartfield C. The Analysis of Series: An Introduction 6th edition, Chapman and Hall/CRC, 2004.
5. Carrington A. Malaria: its human impact, challenges, and control strategies in Nigeria. Harvard Health Policy Review. 2001; 2(2):1-3.
6. Contreras J, Espinola R, Nogales FJ, Antonio J, Conejo AJ. IEEE Transactions on power systems. 2003; 18(3):1014-1020.
7. Guerrero VM. Anålisis estadistico de series de tiempo económicas. Segunda edición, Mėxico, 2003.
8. Lin H, Lu L, Tian L, Zhou S, Wu HBY, Ho SC *et al*. Spatial and temporal distribution of falciparum malarial in China. Malaria Journal. 2009; 8(130). doi: 10.1186/1475-2675-8-130.
9. Ljung GM, Box GEP. On a measure of lack of fit in time series models, Biometrika. 1978; 65:297-303.
10. Pattanayak S, Dickinson K, Corey C, Murray B, Sills E, Kramer R. Deforestation, malaria and poverty: a call for transdisciplinary research to support the design of cross-sectoral policies. Sustainability: Science, Practice, & Policy. 2006; 2(2):45-56. http://sspp.proquest.com/archives/vol2iss2/0512-022.pattanayak.html
11. Sudhakar P, Subramani P. Literature review: insights into formulating a protective malaria medicine. Journal of Young Investigators. 2007; 16(1).

https://www.jyi.org/2007-january/2017/11/11/literature-review-insights-into-formulating-a-protective-malarial-medicine
12. World Health Organization. Global Health Observatory (GHO) data, 2018. https://www.who.int/gho/malaria/epidemic/deaths/en/