American Journal of Theoretical and Applied Statistics
2020; 9(1): 1-7
http://www.sciencepublishinggroup.com/j/ajtas
doi: 10.11648/j.ajtas.20200901.11
ISSN: 2326-8999 (Print); ISSN: 2326-9006 (Online)

On the Fourier Residual Modification of Arima Models in Modeling Malaria Incidence Rates among Pregnant Women

Chinonso Micheal Eze[1], Oluchukwu Chukwuemeka Asogwa[2, *], Charity Uchenna Onwuamaeze[1, *], Nnaemeka Martin Eze[1], Chukwunenye Ifeanyi Okonkwo[2]

[1]Department of Statistics, Faculty of Physical Science, University of Nigeria, Nsukka, Nigeria

[2]Department of Maths, /Comp. Sc. /Stats. /Infor., Faculty of Science, Alex Ekwueme Federal University, Ndufu-Alike Ikwo, Ebonyi State, Nigeria

Email address:

chinonso.eze@unn.edu.ng (C. M. Eze), qackasoo@yahoo.com (O. C. Asogwa), uchenna.onwuamaeze@unn.edu.ng (C. U. Onwuamaeze), nnaemeka.eze@unn.edu.ng (N. M. Eze), okonkwochukwunenye@yahoo.com (C. I. Okonkwo)

[*]Corresponding author

To cite this article:

Chinonso Micheal Eze, Oluchukwu Chukwuemeka Asogwa, Charity Uchenna Onwuamaeze, Nnaemeka Martin Eze, Chukwunenye Ifeanyi Okonkwo. On the Fourier Residual Modification of Arima Models in Modeling Malaria Incidence Rates among Pregnant Women. *American Journal of Theoretical and Applied Statistics*. Vol. 9, No. 1, 2020, pp. 1-7. doi: 10.11648/j.ajtas.20200901.11

Received: February 10, 2020; **Accepted**: April 3, 2020; **Published**: April 13, 2020

Abstract: This work provides a general overview and consideration of Box-Jenkins models for temporal data and its extension known as Fourier residual autoregressive moving average models. We examined the modeling and forecasting of malaria incidence rate during pregnancy at Bishop Shannahan Hospital, Nsukka using Autoregressive Integrated Moving Average (ARIMA) Models propounded by Box and Jenkins. We adoptted the Box-Jenkins methodology to build ARIMA model for malaria incidences during pregnancy for a period of 10 years spanning from January 2006 to December 2016. Among the candidate models considered, ARIMA (3,1,1) was identified to be the most robust based on some model performance measures. The model was further improved upon by incorporating Fourier residual modification on the fitted ARIMA model. The Fourier Residual Autoregressive Moving Average (FARIMA) model obtained yielded improved result. Besides, model evaluation criterion such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Bias Error (MBE), Mean Bias Error (MBE), Mean Absolute Scaled Error (MASE), were used to access the models. FARIMA Model out performed ARIMA Model. Several time series plots and tests like augmented dickey fuller test, correlogram, Ljung-Box test for serial correlation of the residuals, etc were carried out in this study to test for stationarity, identify the order of ARIMA model and serial correlation residual respectively.

Keywords: ARIMA, Modeling, Forecasting, FARIMA, Model Performance Measures

1. Introduction

These days, malaria has been considered as one of the most predominant sicknesses among children and pregnant women. Thus, no country is completely devoid of malaria infection. The control of the infection is essential as it helps in reducing mortality rate. Due to its relevance and the resulting impacts on human lives, studies concerning the phenomenon cannot be over-emphasized. Malaria parasite is transmitted from one person to another through the bite of an Anopheles Mosquito which require blood to groom her eggs. The mosquito borne disease is caused by a parasite called plasmodium. When Malaria parasites finds its way to the blood stream of a person, they infect and destroy the red blood cells. The destruction of these essential cells leads to fever and other symptoms such as headache, muscle aches, tiredness, nausea, vomiting and diarrhea [4]. Malaria causes at least 300 million cases of acute illness yearly, costing Africa more than US$12 million annually and slows economic growth by 1.3% a year, thus trapping malaria vulnerable countries into poverty [9]. One factor contributing to this problem is the known global climate change and this is considered as a big challenge in fight against the scourge of malaria [6]. The report released by the federal ministry of health in Nigeria, indicated that 11% of maternal deaths are

attributed to malaria [5]. Malaria infection is more frequent and severe in primigravidae both during pregnancy and at the time of delivery [3]. Thus, pregnant women, who are known to be one of the groups at high risk of the effects of malaria infection, deserve special protective measures to ensure their survival and improve birth outcome. The case of malaria during pregnancy in Nigeria, has been a regular occurence. The cases are relatively high and call for a robust health programme to forestall the outbreak. The protection of pregnant women living in malaria-endemic countries has been of particular interest to many National Malaria Control Programs because of their reduced immunity. Many pregnant women are concerned about their chances of being infected (or killed) by malaria, and many more about the risk it poses on their unborn children. The nation generally is concerned as that contributes to an increase in mortality rate. There is, therefore, an established concern about malaria infection among pregnant women and this concern is not entirely misplaced especially by health bodies.

The cases of malaria in seven endemic districts of Bhutan in Thailand was studies by [12]. Their interest was majorly to come up with a forecasting ARIMA model for the system. With monthly malaria cases from 1994 to 2008 in the seven endemic districts of Bhutan in Thailand, SARIMA (2,1,1) (0,1,1) was established as the best model with S = 12 as the length of the seasonal period. The identified model was used to run forecasts into the future so as to come up with a robust malaria prevention and control programme in the districts. In addition, they proposed a prediction model incorporating climatic factors such as temperature, humidity and rainfall; which are important in the development of malaria parasites and vector bionomics. The relationship between rainfall excesses and malaria incidence in certain ecoepidemiologic setting were examined by [11]. This study was prompted by the knowledge that excessive rainfall has the tendency to incubate the bacteria that cause malaria especially in children and pregnant women. He analyzed the data on malaria incidences and climate data over a period of time in many parts of the world and it was observed that rainfall excess correlatespositively with changes in malaria incidence in certain eco epidemiologic settings. The study of malaria mortality rate in Imo state Nigeria using SARIMA models was conducted by [4]. Box-Jenkins Seasonal Autoregressive Integrated Moving Average (SARIMA) was employed to analyze monthly malaria mortality rate in Imo State from January 1996 to December 2013. The study intended mainly to forecast the monthly malaria mortality rate for the following period of January, 2014 to December 2014. They developed many tentative models to forecast monthly malaria mortality rate, but based on minimum AIC and BIC values and after the estimation of parameters and series of diagnostic test were performed, ARIMA (1,1,1) (0,0,1) model was chosen to be the best model for forecasting having satisfied the model assumptions. One of the major problem in curbing an outbreak (system) is the knowledge of the future state of the system. To come up with a better programme to arrest the situation, a reliable knowledge of future occurence is necessary. The need for exact decision making against the future occurrences of malaria infection among pregnant women is the major reason behind time series modeling of the incidences. This study seeks to examine the variability on the incidences of malaria in pregnant women at Bishop Shanahan Hospital, Nsukka. This will enhance our knowledge on the likelihood of occurence in the hospital and consequently help the hospital management in coming up with a reliable malaria prevention and control programme for pregnant women in the hospital. The idea is to model the incidences using $ARIMA$ models where the incidences are regressed on lagged observations and random errors. In extension, the residual from the fitted model are modified with fourier series and incoporated into the model to improve the model accuracy.

2. Materials and Methods

Autoregressive Integrated Moving Average (ARIMA) models is a well known forecasting model for a time series which can be made stationary by differencing or logging [7]. Consider a stochastic process $(Z_t, t = 1, 2, \dots, N)$ indexed with time. This process if linear, can be modeled using autoregressive integrated moving average ($ARIMA$) model propounded by [2]. This is a mathematical model designed to forecast data based on past observations and random shocks. The model alters the time series to make it stationary by using the differences between data points. The model type is generally referred to as $ARIMA\,(p,d,q)$, with the components (p,d,q) referring to the autoregressive, integrated and moving average parts of the model, respectively. The values of p and q are the number of autoregressive (AR) and moving average (MA) components in $ARIMA\,(p,d,q)$ model. These two simple components are incorporated in representing the behavior of the process. The AR is used to describe a time series in which the current observation depends on its preceding values, whereas the moving average (MA) is used to describe a time series process as a linear function of current and previous random errors. The $ARIMA\,(p,d,q)$ model therefore expresses a process as a linear function of AR and MA.

Let the time series under consideration be Z_t. For Z_t to be stationary, we have $\bar{Z}_t = \nabla^d Z_t$ where d is a non-negative integer order of differencing. The $ARMA$ model for the stationary process $\bar{Z}_t$ is therefore expressed as

$$\bar{Z}_t = \emptyset^{-1}(B)\theta(B)e_t \qquad (1)$$

$\phi(B) and\ \theta(B)$ are polynomials of degree p and q respectively. The process $\{Zt\}$ is said to be stationary if $d = 0$, in which case the $ARIMA\,(p,d,q)$ reduces to an $ARMA\,(p,q)$. The $ARIMA$ model is commonly used in analyzing data with a correlation among neighbouring observations.

Fourier Residual Modification of ARIMA Model

Making use of fourier series to modify the residual of ARIMA model can significantly improve the prediction ability of the model by optimizing the model performance measures. Forecasting models have been proved to be significantly improved after their residual series are modified with Fourier series [10]. This methodology when

incorporated into the $ARIMA$ model has the tendency of improving the model's predictability. The procedure for Fourier residual modification of $ARIMA$ model is as follows:

We generate a residual series $e_t = \hat{z}_t - z_t$ based on the fitted $ARIMA$ model.

Using a fourier series model on the generated residual, we express the residuals as follows:

$$e_t = a_0 + \sum_{i=1}^{f} a_i cos(iwt) + b_i sin(iwt) \quad (2)$$

The least square estimates for the parameters $a_0, a_1 b_1 a_2, b_2, \ldots\ldots\ldots\ldots, a_f b_f$ are obtained as

$$a_0 = \frac{\sum_{t=1}^{N} e_t}{N};\ a_i = 2\frac{\sum_{t=1}^{N} e_t cos(iwt)}{N};\ b_i = 2\frac{\sum_{t=1}^{N} e_t sin(iwt)}{N}$$

Having estimated the parameters, the predicted series for the residual $\hat{e}_t$ is then achieved based on the fitted Fourier series model:

$$\hat{e}_t = a_0 + \sum_{i=1}^{f} a_i cos(iwt) + b_i sin(iwt) \quad (3)$$

Based on the predicted series $\hat{z}_t$ obtained from $ARIMA$ model and the predicted series of the residuals $\hat{e}_t$ obtained from the Fourier series model, the predicted value using the modified model is thus, expressed as

$$\hat{z}_t^{(m)} = \hat{z}_t + \hat{e}_t \quad (4)$$

More explicitly, equation (4) transforms to

$$\hat{z}_t^{(m)} = a_0 + \sum_{i=1}^{f} a_i cos(iwt) + b_i sin(iwt) + \emptyset^{-1}(B)\theta(B)\, e_t + \varepsilon_t \quad (5)$$

In order to examine the accuracy of the forecast model, the residual error for the modified model is expressed as

$$\varepsilon_t^{(m)} = \hat{z}_t^{(m)} - z_t \quad (6)$$

The Fourier series residual modified $ARIMA$ model improves on the accuracy of the conventional *ARIMA* model. It does this by inculcating the Fourier series model for the residuals into the $ARIMA$ model and boosts significantly, the accuracy. The methodology was adopted first by [10] and subsequently by [1] to study the inbound tourism demand in New Zealand and climatic variables respectively.

In assesing the model accuracy, the following model performance measures were considered.

Root Mean Square Error (RMSE) is the square root of the mean square error. This gives an extensive interpretation of the average of the sum of square error. The smaller the measure, the better the model. It is calculated using

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(O_t - F_t)^2}$$

Mean Absolute Error (MAE) measures the average of the absolute deviation of the forecast value from the observed values. It measures how close forecasts or predictions are to the eventual observations. It is expressed as

$$MAE = 1/T\frac{\sum_{t=1}^{T}|O_t - F_t|}{Ot}$$

Mean Bias Error (MBE) measures the average deviation of the forecast values from the observed values. *MBE* provides information on the long term performance of the developed model. A positive value of MBE suggests the rate of over-prediction while a negative value of it suggests the rate of under-prediction. It is calculated using the mathematical expression

$$MBE = \frac{1}{T}\sum_{t=1}^{T}(O_t - F_t).$$

Mean Absolute Percentage Error (MAPE) is a measure of prediction accuracy of a forecasting method in statistics. It usually expresses accuracy as a percentage, and is defined by the formula:

$$MAPE = \frac{\sum_{t=1}^{T}|O_t - F_t|}{|Ot|} \times \frac{100}{T}$$

Mean Absolute Scaled Error (MASE) is a measure of the accuracy of forecasts. It is a generally applicable measurement of forecast accuracy without the problems seen in the other measurements. It is calculated as

$$MASE = \frac{\sum_{t=1}^{T}|e_t|}{\frac{T}{T-1}\sum_{t=2}^{T}|Z_t - z_{t-1}|}$$

3. Results

The best way to start with any time series analysis is to examine the sequence plot of the series as this will help to know whether the series has different characteristics at different intervals of time. A sequence plot is a graph of the series values, usually on the vertical axis, against time usually on the horizontal axis. The purpose of the sequence plot is to give the analyst a visual expression of the behavior (nature) of the series. This visual impression suggests to the analyst whether there are certain behavioral patterns present within the time series or not. The presence/absence of such components can help the analyst in selecting the model with the potential to produce the best forecasts [8]. To give a graphic expression of the changes in the incidences of malaria during pregnancy within the period under study, the sequence plot of the cases are given below:

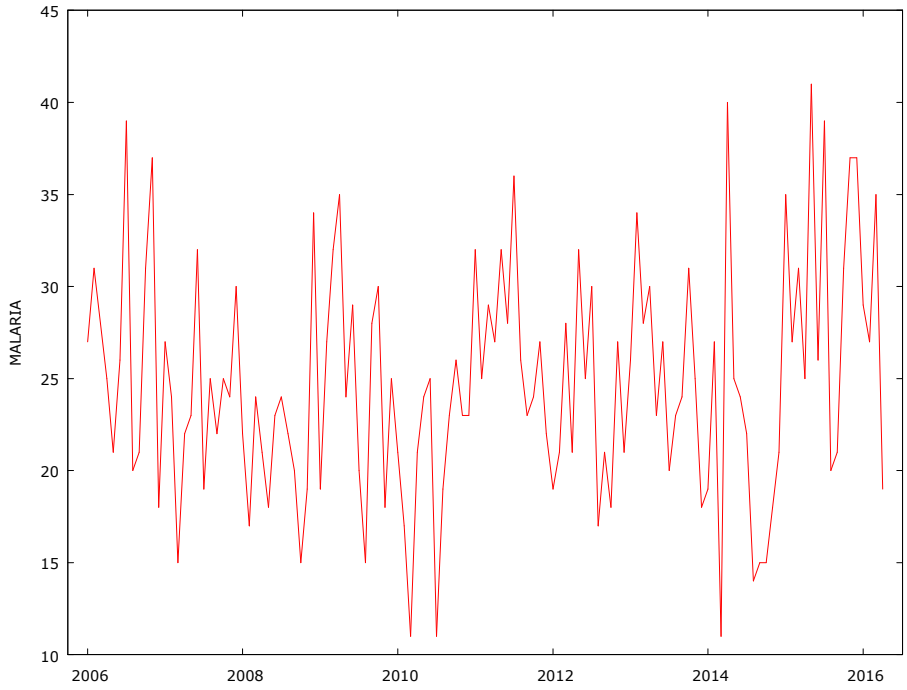

Figure 1. *Time series plot of malaria incidence during pregnancy.*

From the plot above, it is clear that the series are not stationary. Added to this fact is the unit root test (test for stationarity) conducted below, using augmented dickey fuller test. It is clear from a look at the test statistic value and p-value of the original series as contained in table 1, that the series are not stationary. However, first differencing (differencing of order 1) made the series stationary as the p-values of the tests on the differenced series is so small (less than 0.05) that it suggests the rejection of non-stationarity. Thus, the malaria incidence series were differenced once. The plot and correlogram of the differenced series is plresented below.

Table 1. *The Result of the Augmented Dickey-Fuller Test.*

Malaria Incidences		
	Original series	**Differenced series at order 1**
ADF	-0.5068	-14.7695
P-Value	0.4972	0.0000

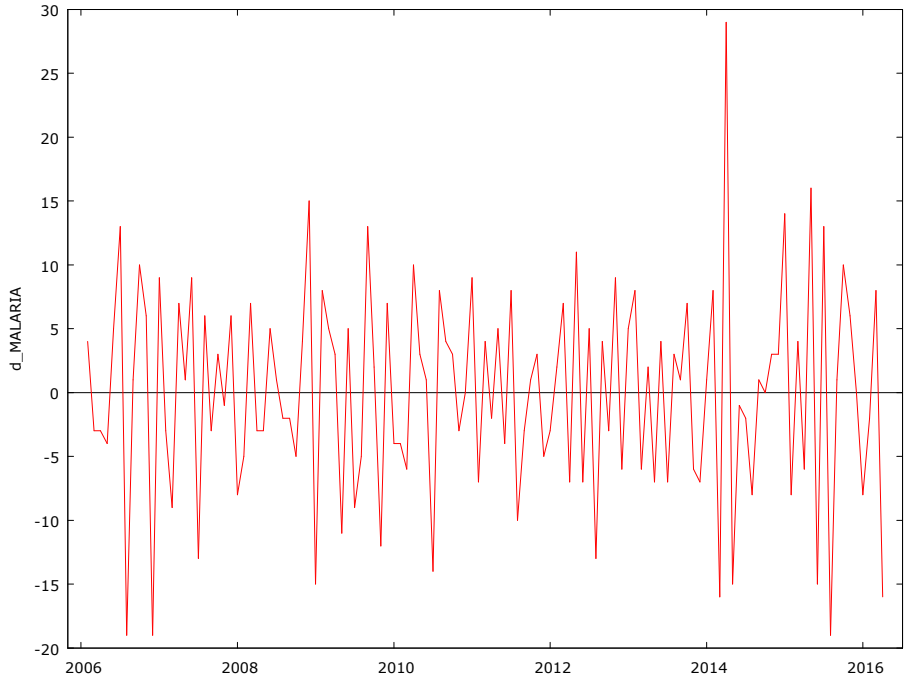

Figure 2. *The sequence plot of the differenced Malaria Incidences during pregnancy.*

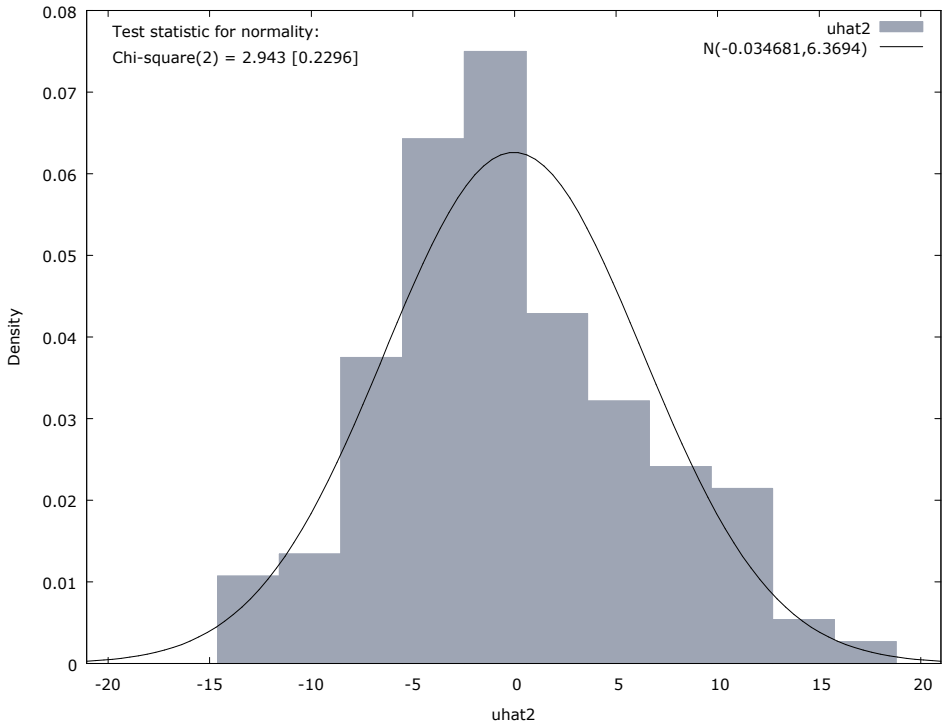

Figure 3. *Correlogram for the differenced Malaria Incidences during pregnancy.*

Based on the correlogram above, the appropriate order of the *ARIMA* model was identified as *ARIMA (3, 1, 1)* since there is a significant cut-off at lag three of the partial autocorrelation function *(PACF)* and a significant cut-off at lag one of the autocorrelation function *(ACF)* after the first differencing. The parameter values for the identified model are presented in the table below:

Table 2. *The model coefficients.*

Coefficient	Values
phi_1	0.135575
phi_2	0.148892
phi_3	0.011308
theta_1	-1.000000

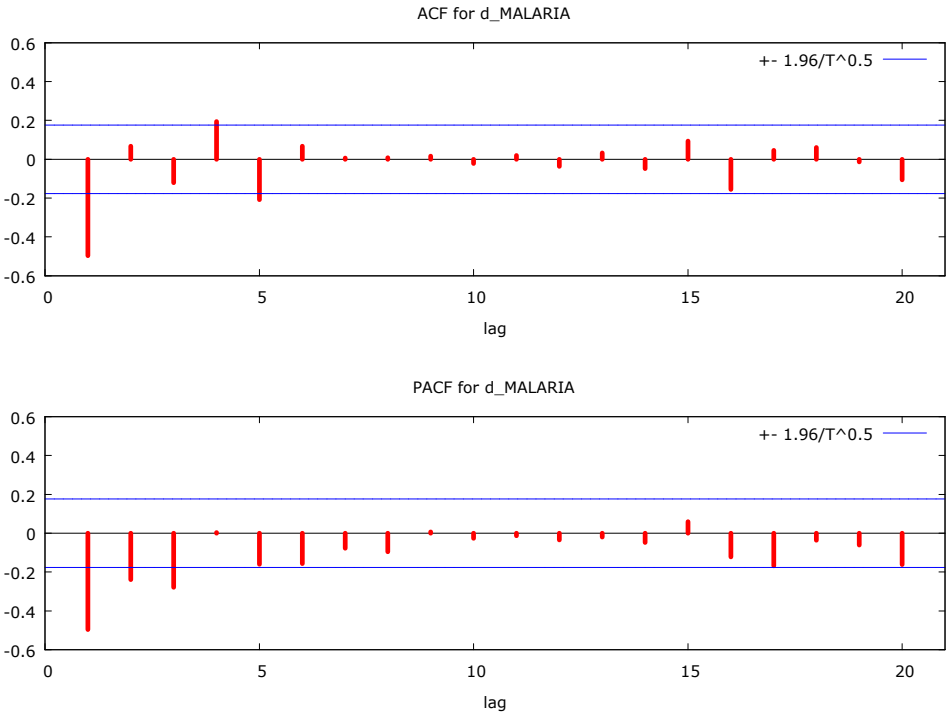

Figure 4. *Normality curve for the residuals.*

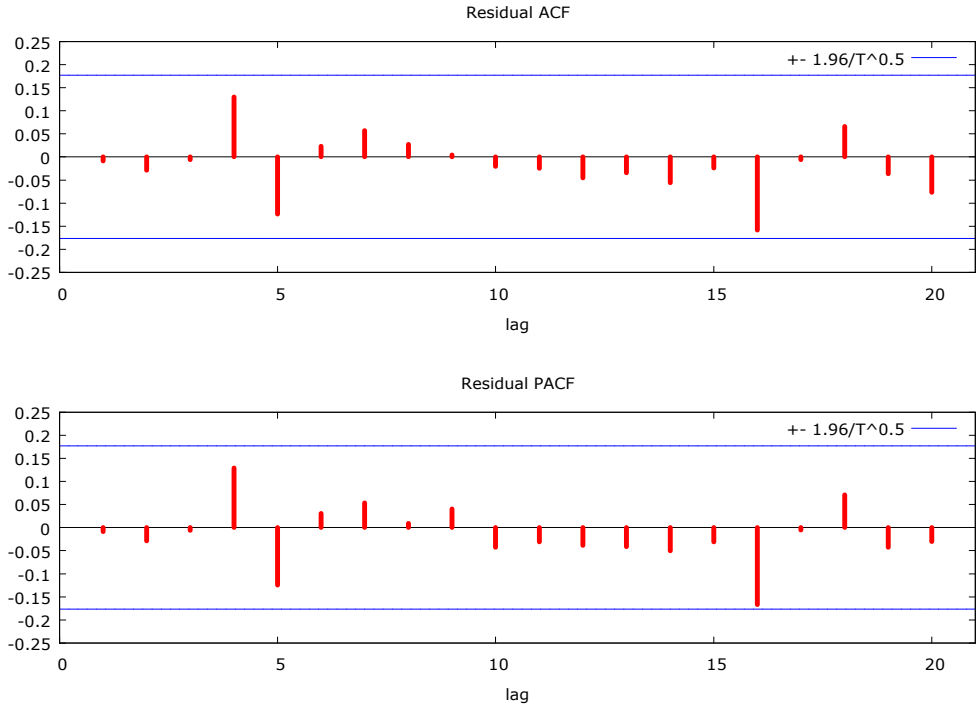

Figure 5. *The correlogram of the residuals.*

3.1. Residual Analysis

Here we try to analyze the residual of the fitted model to see whether the identified model fits the series appropriately. We first conduct Ljung-Box test for serial correlation of the residuals. The test statistic gives a value of Q = 5.32911 with p-value = P (Chi-square (8) >5.32911) = 0.7219. The P-value is relatively high, implying that there is a serial correlation among the residuals of the fitted model. As can be seen in figure 5, the autocorrelation and partial autocorrelation of the residuals lie within ± 1.96 of the standard error, implying that the model is a good one. Also, the chi-square test for normality of the residuals was conducted and it was observed that the residuals are normally distributed with mean (-0.034681) and variance (6.3694).

3.2. Fourier Residual Modification of the ARIMA Model

Based on the methodology above, a fourier series model is to be fitted on the residuals. The best order of the fourier model was identified based on their model accuracy measures and it was identified that a fourier series model of order 4 is the best fit for the residuals with the coefficients presented in table 3. Inculcating the fourier residual into the *ARIMA* model, we have the modified model to be

$$Z_t^m = 0.136z_{t-1} + 0.149z_{t-2} + 0.011z_{t-3} - e_t + e_{t-1} - 0.17 + \sum_{i=1}^{4} a_i \cos(iwt) + b_i \sin(iwt) + \varepsilon_t \quad (7)$$

Table 3. *Fourier coefficients.*

Fourier coefficients	Values
α_0	-0.17
α_1	0.826
b_1	-0.07
α_2	-2.27
b_2	-0.45
α_3	-1.62
b_3	-1.05
α_4	1.16
b_4	-1.49

To reflect the improvement in the modified model, the performance measures of the two models based on the out of sample test were examined. The last 20 data points in the entire data set were used to validate the models. For each of them, the performance measures are tabulated below.

Table 4. *Forecast evaluation statistics.*

Performance Measures	ARIMA Model	F-ARIMA Model
Mean error	2.2086	2.09
Mean square error	77.888	70.95
Root mean square error	8.825	8.423
Mean absolute error	0.250	0.236
Mean absolute scale error	0.670	0.650

4. Conclusion

The case of malaria infection among pregnant women is a serious case that calls for immediate attention. The knowledge of the future occurence is also very necessary as that will help in planning health policy that will help control the outbreak. The Box-Jenkins approach used in the analysis revealed the data generating process which can be used to make a reliable forecast for the future occurence so that the institution concerned can plan for corrective measures. Based

on the available data, *ARIMA* (3, 1, 1) was identified as the generating process for the case of malaria during pregnancy. The robustness of the fitted *ARIMA* (3, 1, 1) was confirmed through analysis of residuals.

Generally, the problem of overfitting and underfitting in *ARIMA* models has become a common happening. Also, the inability of a standard *ARIMA* model to predict zero values has become a noticeable shortfall. The Fourier series residual modified *ARIMA* model accounts for the identified shortfalls by using the Fourier component of the model which allows the residual to take up both positive and negative values. The improvement in the model was shown in the model performance measures which was based on out of sample data set.

Acknowledgements

The researcher sincerely acknowledged and appreciated the management of Bishop Shanahan Hospital Nsukka for providing us with data set which were used in carrying this work.

References

[1] Aniefiok, I. I and Murphy, D. (2017). Mixed seasonal and subset fourier model with seasonal Harmonics. *Science Journal of applied mathematics and statistics. Vol. 5 (1): 1-9.*

[2] Box, G. E., and Jenkins, G. M. (1976). Time Series Analysis: Forcasting and Control. *San Francisco: Holden-Day*.

[3] Brabin, B. J. (1983). An analysis of malaria in pregnancy in Africa. *Bulletin of the World Health Organization, 61 (6): 1005-1016.*

[4] Ekezie, d. Opara, j., and Okenwe, I (2014). Modeling 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 (1): 31-41.*

[5] Federal Ministry of Health. Malaria situation analysis document. Nigeria: *Federal Ministry of Health; 2001. p. 14.*

[6] McMichael, A. J., Woodruff, R. E and Hales, S. (2006). Climate change and human health: present and future risks. *Lancet 2006, 367: 859-869*.

[7] Nguyen, T., Chen, P and Huang, Y. (2013). Forecasting with fourier residual modified ARIMA model – An emperical case of inbound tourism demand in New Zealand. *Recent researches in applied economics and management. Vol. 2.*

[8] Prajakta, S. K. (2004). Time Series Forecasting Using Holt-Winters Exponential Smoothing. *Kanwal Rekhi School of Information Technology, 4329008, 1-13.*

[9] Sachs, J and Malaney, P. (2002). The Economic and Social Burden of Malaria. *Macmillan Publisher. Ltd.*

[10] Shu, M., Hsu, B and Nguyen, T. (2013). Forecasting international tourism deman – An emperical case of Thaiwan. *Asian journal of emperical research. Vol. 3 (6), 711-724.*

[11] Thomson C. M, Mason J. S, Phindelia, T and Connor J. S. (2005). Use of Rainfall and Sea Surface Temperature Monitoring for Malaria Early Warning in Botswana. *American Journal Tropical Medicine and Hygiene 73 (1), pp. 214-221.*

[12] Wangdi, K., Pratap, S. Tassanee, S., Saranath, L., Nicholas, J and Jaranit, K. (2010). Development of temporal modeling for forecasting and prediction of malaria infections using time-series and ARIMAX analyses: A case study in endemic district of Bhutai. *Journal of Malaria. Vol. 9, pg. 251.*