# Times Series Analysis Of Malaria Cases In Ejisu-Juaben Municipality

Takyi Appiah, S., Otoo, H., Nabubie, I.B.

**Abstract:** The number of malaria cases in the Ejisu-Juaben Municipality were modelled statistically to find the best model for forecasting the disease for a two year period. The Box-Jenkins approach was applied to Secondary data from the municipality to determine the best model fit. From the model obtained, the forecast was found to have an oscillatory trend for some period and then remain constant for the period of two years from 2014 and 2016.

**Keywords:** Autoregressive, Cyclic, Hypothesis, Trend, Time Series, Stationary

## 1 INTRODUCTION

Malaria has been one of the oldest prominent and ancient diseases which has been profiled and studied in tropical regions. It is a mosquito-borne infectious disease in humans caused by parasitic of the genus *Plasmodium and* is transmitted by means of a bite from an infected anopheles mosquito which introduces the *protozoans*. It remains the leading cause of death in children under five years in Africa [15]. Malaria is one of the killer diseases in tropical and subtropical countries. It therefore poses a serious health problem to these countries including Ghana [3]. This disease can be considered a disease of the poor because its prevalent rate is very high in poor tropical countries [14]. According to the World Health Organization (WHO) in 2010, there were 219 million cases of malaria. It is been estimated that economic growth per year of countries with intensive malaria was 1.3% lower than that of countries without malaria [10]. In Ghana, the peak period of malaria transmission occurs during the rainy season and often coincides with the peak period of agricultural activities such as planting and harvesting, thus, hindering the economic development of the affected people [3. According to National Malaria Control Program (NMCP), Ghana in 2013 recorded about 11.3 million malaria cases at Outpatients Department (OPD). On average 30,300 of such cases were attended to each day in the country's health facilities. Malaria burden is not felt only in the health sector, but in every aspect of our social and economic life [9]. Among pregnant women, malaria accounts for 14% of outpatient visits, 11% of admissions and 6% of deaths. Malaria accounts for 10.6% of disability adjusted life years which 6% of the annual gross domestic product (GDP) [6]. It has been the leading cause of morbidity and mortality in the Ejisu-Juaben Municipality since 2003 with malaria in pregnant women increasing annually [8]. There are several research on trend of malaria in Ghana.


- *Appiah Takyi Sampson: Department of Mathematics, University of Mines and Technology, Tarkwa, Ghana, E-mail: stappiah@umat.edu.gh*
- *Otoo Henry: Department of Mathematics, University of Mines and Technology, Tarkwa, Ghana, E-mail: hotoo@umat.edu.gh*
- *Nabubie I. B: Department of Mathematics University of Mines and Technology, Tarkwa, Ghana. boyuou@yahoo.com*


Asamoah *et al*. (2008) in their work used SARIMA (1,1,0)(1,0,0) for Total OPD reported cases, ARIMA (1,1,0) for Admission reported cases, SARIMA (1,0,0)(1,0,0) for female OPD reported cases and SARIMA (1,1,0)(1,0,0) for OPD pregnant cases in malaria reported cases. This paper therefore seeks to investigate and find the best model fit for the malaria cases and forecast for the period of two years.

## 2 STUDY AREA

The Ejisu-Juaben Municipal is one of the 30 administrative and political Districts in the Ashanti Region of Ghana. The Municipality stretches over an area of 637.2 km$^2$ constituting about 10% of the entire Ashanti Region. Agriculture is the main source of livelihood of the indigenous people and employs up to 69.57% of the residents [1]. The Municipal is known for its vibrant farming activities and high incidence of malaria cases. Thus, there is the need to determine the trend and forecast malaria cases in order to make evidence based suggestions to improve the health and economic status of the residents [7].

## 3 TIME SERIES

Time series is a sequence of data point, measured typically at successive point spaced at uniform time intervals. This can be measured daily, weekly, monthly, quarterly and sales etc. at a time $t_1, \ t_2, t_3, t_4 \ldots$ thus Y is a function of t symbolically

$$Y = f(\mathrm{t}).$$

Time series analysis comprises methods for analyzing time series in order to construct a meaningful statistic and other characteristics of data [5]. Time series forecasting is the use of model to forecast for future event based on known data event to predict data points before they are measured. Time series data have a natural temporal ordering. This makes time series analysis very distinct from other common data analysis tools [2]. Mathematically, time series is defined by the values $\mathrm{y}_1, \mathrm{y}_2, \mathrm{y}_3, \mathrm{y}_4, \ldots$ of a variable. Time series analysis is also distinct from spatial data analysis where the observations typically relate to geographical location. Time series gives an understanding of the underlying forces and structure that produced the observed data. It also aims to find a model and proceed to forecast and monitor the process. . A time series model can be expressed as some combination of these four components. Any time series can contain some or all the following components:

1. Secular Trend (T)
2. Seasonal (S)
3. Cyclical (C)

4. Random (Irregular) (R)

It is usually assumed that the models are additive or multiplicative. Time series model is basically expressed as a multiplicative model where the value of time series at time t is specified as:

$$Y_t = T_t \times S_t \times C_t \times R_t \quad (1)$$

Or additive model where the value at time series at time t is specified as:

$$Y_t = T_t + S_t + C_t + R_t \quad (2)$$

### 3.1 IDENTIFICATION

The purpose of identification is to determine the differencing required to produce stationarity and also the order of seasonality and non-seasonality of Autoregressive (AR) and Moving Average (MA) operators for the series [11]. Ordinarily, model identification is an explanatory process and analysis done is based upon previous result. Identification consists of specifying the appropriate structure Autoregressive Integrated Moving Average (ARIMA) and the order of the model. Identification is sometimes done by looking at the autocorrelation function (ACF) and partial autocorrelation function (PACF) to determine whether the observations are stationary or not. Once stationarity is achieved, the second ARIMA parameter d, is simply the number of time the series is differenced to achieve stationarity. Next is the identification of the order of AR and MA, pure AR and MA processes have characteristics signature in the ACF and PACF. The steps use to identify AR and MA and their orders are simplified in the table.

***Table 1.1*** *Identification of AR and MA models and their orders*

| | **AR(q)** | **MA(p)** |
|---|---|---|
| ACF | Models have exponentially decaying and declining values of ACF (with alternative positive and negative value) or The ACF decline steadily and follow a dumped cycle. | Have precisely p spike in the value of the ACF or The ACF spike on the first p lag or ACF cut steadily after p lags. |
| PACF | Have precisely p spike in the value of the PACF<br>or<br>PACF spikes on the first p lag<br>PACF cut steadily after p lags | Models have exponentially decaying and declining values of ACF (with alternative positive and negative value)<br>or<br>The ACF decline steadily after a q lag. |

### 3.2 ESTIMATION

The second step is to estimate the co-efficient of the model. The estimation of the co-efficient is done by using Statistical Package Social Sciences (SPSS 16.0) software.

### 3.3 Model Adequacy

The third step is to check the adequacy of the model. This step is also called diagnostic checking or verification [12]. Diagnostic checking consists of evaluating the adequacy of the estimated model. It is important to ensure that the estimated parameters are statistically significant. Usually the model fitting process is guided by the principle of parsimony by which the best mode is the simplest possible model- the model with a fewer parameters- that adequately describe the data. An adequate model satisfies these four conditions:

1. The estimates of all the parameters must differ significantly from zero
2. All AR parameter estimates must be within the "bounds of stationary". This guarantees that the model is stationary about its mean. For example AR model, the requirement of the bound of stationary are:

For AR (1)

$$|\phi_1| < 1 \quad (3)$$

For AR (2)

$$|\phi_1| < 1 \Rightarrow -1 < \phi_1 > 1 \quad (4)$$

$$|\phi_2| < 1 \quad (5)$$

$$\left.\begin{array}{l} \phi_1 - \phi_2 < 1 \\ \phi_1 + \phi_2 < 1 \end{array}\right\} \quad (6)$$

3. All MA parameters estimate must lie within the bounds of "invertibility". This is the MA along to stationary to AR model. Where the model is re-express as infinite series as AR terms, inevitability guarantees that this series converges. For a simple MA models, the requirement of the bounds inevitability are

For MA (1) model

$$|\phi_1| < 1 \quad (7)$$

For MA (2) model

$$\left.\begin{array}{l} |\phi_2| < 1 \\ \phi_2 - \phi_1 \end{array}\right\} \quad (8)$$

For models of the same orders, that is AR (i) and MA (j), the bounds of invertibility place limit on $\emptyset_i$ that are identical to those on $\emptyset_i$ by the bounds of invertibility.

4. Residual must not differ significantly from a series of purely random error (White noise) with mean zero. For White noise the theoretical ACF and PACF are both zero at all lags. For residual, the calculated standard error tends to over-estimate the true standard error (Monserud, 1986).The simplest way of checking the best model is to use goodness of fit

statistics such as the real adjusted R-square mean absolute error, sum of square of error normalize BIC (Bayesian Information Criteria) and the residual plot of ACF and PACF. In summary, the best model is the one with relatively small of BIC, relatively small of mean absolute error, relatively small of sum of squares error, relatively high adjust R-square and Random pattern of the plot of the ACF and PACF

### 3.4 Diagnostic Check

Ljung-Box portmanteau Q statistics, Q is the test of null hypothesis which specifies that the ACF does not differ from zero up to lag $k$. It is evaluated as chi-square with $k-m$ degree of freedom, where k is the number of lags examined and m is the number of parameters estimated. The second test is to examine the ACF and the PACF plot of the first difference of the residual.

### 3.5 Model Selection

The selection of a forecasting method is a difficult task that must be based on the port of knowledge concerning the quantity being forecast. We can however, point out some simple characteristics of the methods that have been described. With forecasting procedure we are generally trying to recognize a change in the underlying processes of a time series while remaining insensitive to variation caused by purely random effects. The goal of planning is to responds to fundamental changes not to spurious effects. With a method based purely on historical data, it is impossible to filter out all the noise. The problem is to set parameters that find an acceptable trade-off between the fundamental processes and the noise. If the process is changing very slowly, both the moving average and the regression approach should be used with a long stream of data. For exponential smoothing method, the value should be small to the emphasis the most recent observation. Stochastic variation will be almost entirely felted out. If the process is changing rapidly with a linear trend, the moving average and the exponential smoothing methods are at disadvantage, because they are not designed to recognize trend. Because of the rapid changes, the time range of the moving average method must be set small and the parameter of the exponential smoothing method must be set to a large value, so that the forecast will respond to the changes. Nevertheless, these two methods will always fall behind a linear trend. The forecast will never converge to a trend line, even if there is no random variation. Of course, will the adjustment parameter to allow a response to a process change, the forecasts become more sensitive to a random effect? The exponential smoothing method with a trend adjustment and the regression methods are both designed to respond to linear trend and will eventually converge to a trend line. Thus in the absence of change in trend, the time range of the regression data can be large and the values of the exponential smoothing method can be small, thus reducing the random effects. If the process is changing rapidly with a rapid change in the trend, each of the methods discussed in the above will have troubles, because it is difficult to separate changes in the process from the random changes. The time ranges must be set small for moving average method and the regression method, resulting to sensitivity to random effects. Similarly the parameters of the exponential smoothing must be to a larger value with a corresponding increase in the sensitivity to randomness. Both the moving average and the regression methods have disadvantages that they are most accurate with respect to forecast in the middle of the time range. Unfortunately, the interesting forecasts are in the future, outside the range of the data. With all the methods, though, the accuracy of the results decrease with the distance into the future one wishes to forecast.

## 4 Discussions and Results

Secondary data was collected from the Ejisu-Juaben Hospital. The data collected has the duration from January, 2009 to December, 2013. The recorded figure of the disease is considered as dependent variable with time as the independent variable.

***Table 1.1*** *Total Number of Malaria Cases*

| | 2009 | 2010 | 2011 | 2012 | 2013 |
|---|---|---|---|---|---|
| JANUARY | 1427 | 1759 | 1658 | 863 | 1260 |
| FEBRUARY | 1471 | 1233 | 1743 | 1386 | 1156 |
| MARCH | 1267 | 1848 | 2118 | 1231 | 1293 |
| APRIL | 1288 | 1449 | 2051 | 1294 | 1293 |
| MAY | 1431 | 1536 | 1704 | 1131 | 1248 |
| JUNE | 1161 | 712 | 1607 | 1131 | 1777 |
| JULY | 1209 | 1977 | 2456 | 1865 | 2203 |
| AUGUST | 1608 | 1324 | 2402 | 1164 | 1422 |
| SEPTEMBER | 1444 | 1823 | 1107 | 1042 | 1045 |
| OCTOBER | 1130 | 1122 | 1775 | 1186 | 1568 |
| NOVEMBER | 1142 | 1056 | 1492 | 1396 | 1588 |
| DECEMBER | 1400 | 1356 | 1502 | 1393 | 1035 |
| TOTAL | 15978 | 17195 | 21615 | 13951 | 16888 |

*(Source: Ejisu-Juaben Hospital, 2014)*

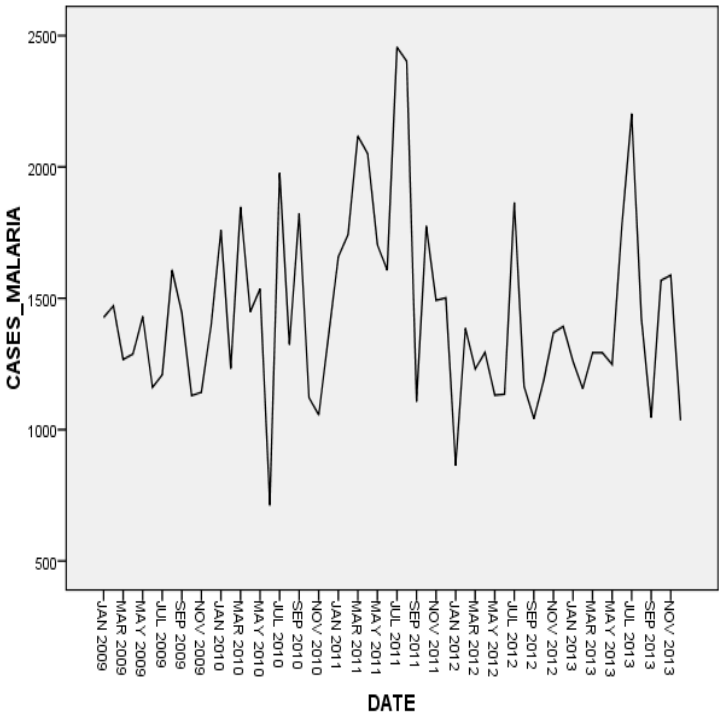


***Fig.1.1*** *Graph showing the original plotting of the Malaria cases*

### 4.1 TESTS FOR STATIONARITY

To make statistical inference about the structure of a time series, it is prerequisite to make some assumptions about the structure. The most important assumption is the stationarity. The basic idea of the stationarity is that, the probability laws that govern the behavior of the process does not change over time. This is to determine that the process is in a statistical equilibrium. By observation, the graph (Fig.1.1) is not stationary because the series display a long term pattern and the mean is not zero. It is therefore important to difference the graph in order to obtain stationarity. The stationarity is done in order to find an average pattern of the graph. The first differencing of the original data is represented graphically as shown in Figure 1.2 below.

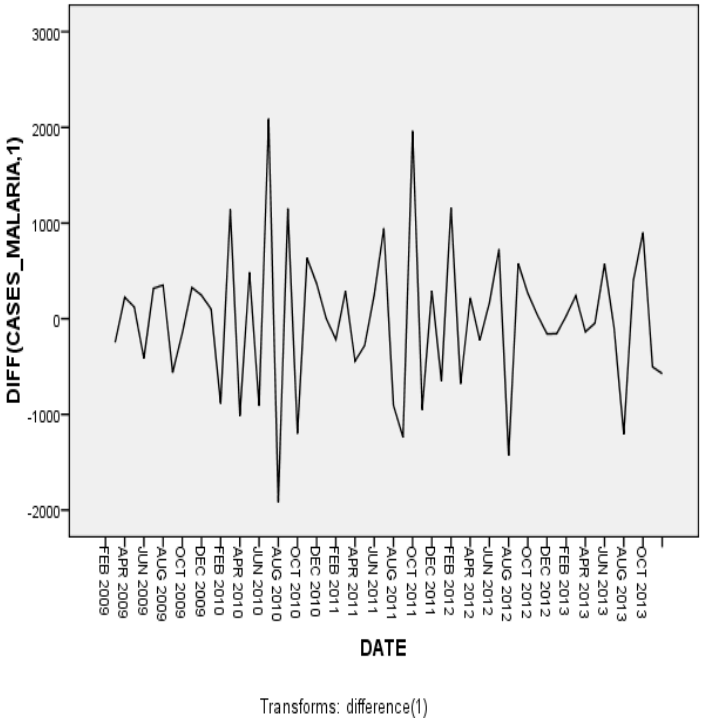


***Fig. 1.2*** *Graph showing the plot of first differencing of the Malaria cases.*

From the figure 1.2, it could be observed that the graph appears approximately stationary with a long term trend after differencing it for the first time. The trend has now been removed from the graph. There are changes of signs of pattern from one observation to another. This is justified by ACF and PACF plots. Thus as shown in Figure 1.3, it include two non-seasonal differencing, because the trend has been partially removed and the amount of autocorrelation remained is small. It appears as though the series may be is satisfactorily stationary.

### Model Diagnostics

Model diagnostics is concerned with testing the goodness of fit of a model and suggesting appropriate recommendations if found to be poor. From the Fig. 1.3 and Fig 1.4, the graph of autocorrelation and the partial autocorrelation show that all the points are random and hence one can conclude that there is a regular pattern which means that the model is fit.

### Graphical Representation of Autocorrelation

From the graphical representation of the autocorrelation function below (Fig 1.3), the value of the autocorrelation function is less than one. Also since the parameter of any given lag as shown in the figure 1.3 is less than one, it means that there is a stationarity and therefore any of the stationarity method can be used.

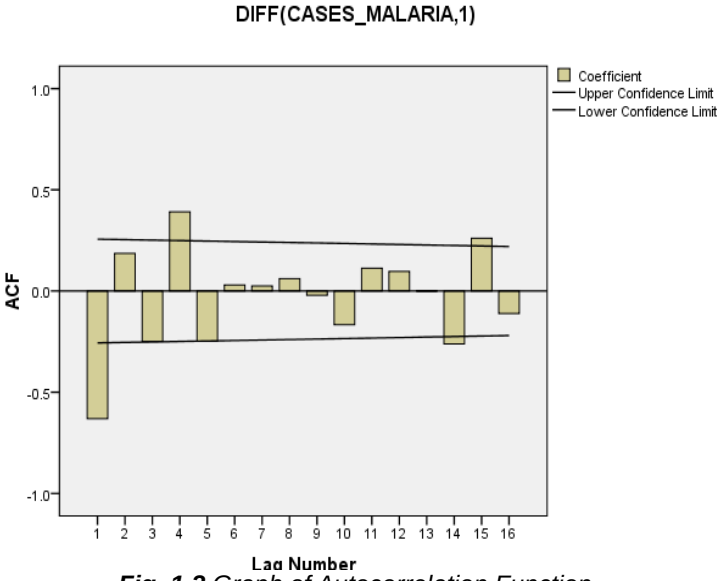


***Fig. 1.3*** *Graph of Autocorrelation Function*

The ACF plot has a negative spike at lag 1 and there are no changes of sign from observation of the next indicating that the series is not over difference. From the ACF plot, it can be observed that, the plots decay fairly to zero from both above and below the mean. From figure 1.3 above, it can be concluded that the graph is lag 1 and hence moving average of order 1.

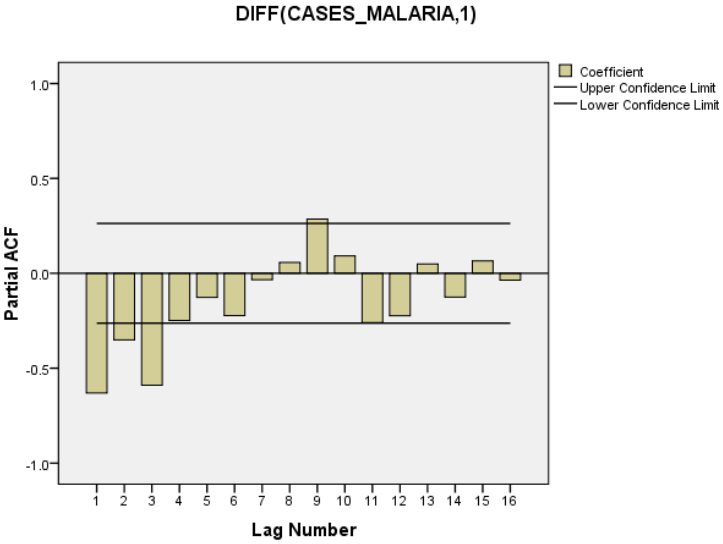


***Fig. 1.4*** *Graph of Partial Autocorrelation Function*

From the diagram representing ACF and PACF, it could be noticed that the ACF shows shaper cut off than the PACF. It can be seen clearly that the ACF has only one significant spike whilst the PACF has two significant spikes. From Fig 1.4, conclusion can be drawn that the graph is of order 2. Hence the autoregressive process is of order 2.

**Table 1.2** Suggested Models

| Model | Normalized BIC | Stationary R-Square | Absolute Mean Error |
|---|---|---|---|
| ARIMA (1,1,0) | 12.942 | 0.401 | 443.484 |
| ARIMA (1,1,1) | 12.337 | 0.701 | 321.233 |
| ARIMA (2,1,2) | 12.511 | 0.701 | 321.387 |
| ARIMA (0,1,1) | 12.479 | 0.623 | 358.462 |
| ARIMA (2,1,1) | 12.334 | 0.727 | 310.404 |

From the above table 1.2, ARIMA (2, 1, 1) is the best model. ARIMA (2, 1, 1) means autoregressive process of order 2, differencing of order 1 and moving average of order 1.

***Table 1.3*** *Model Statistics*

| Model | Number of Predictors | Model Fit statistics | | | | Ljung-Box Q | | | Number of Outliers |
|---|---|---|---|---|---|---|---|---|---|
| | | Stationary R-squared | RMSE | MAE | Normalized BIC | Statistics | DF | Sig. | |
| DIFF(MALARIACASES,1)-Model_1 | 0 | .727 | 414.442 | 310.404 | 12.334 | 41.660 | 15 | .000 | 0 |
| | | | | | | | | | |

The best model is the one with the smallest Normalized BIC, smallest absolute mean error and biggest Stationary R-Square. Considering the models above, ARIMA (2, 1, 1) is the best model. The high value of the R-Square thus 0.727(72.7%) indicates that 72.7% of the variation in the malaria cases can be explained by the data. The ARIMA (2, 1, 1) equation is

$$Y_t - Y_{t-1} = \mu + \phi_1\left(Y_{t-1} - Y_{t-2}\right) + \phi_2\left(Y_{t-2} - Y_{t-3}\right) - \theta_1\varepsilon_{t-1} + \varepsilon_t \qquad (9)$$

Hence considering the second order Autoregressive process

$$Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t \qquad (10)$$

Where $\phi$ is the value of the parameter, for stationary to exist, then $|\phi| < 1$ but $\rho_k = \phi^k$, Where $\rho$ the autocorrelation of the first order of autoregressive processes and k is the number of lags. From table 1.4, it is found that the value of $|\phi|$ is always less than one from the relationship shown above. Since the time series has been found to be stationary, then any of the stationary methods could be used. By using the second order Autoregressive method, a model was fitted. Residual plot of the autocorrelation and partial autocorrelation shows a random pattern. This indicates that the ARIMA (2, 1, 1) model is the best fit for the observations.

***Table 1.4*** *ARIMA Model Parameters*

| | | | | | Estimate | SE | T | Sig. |
|---|---|---|---|---|---|---|---|---|
| DIFF(MALARIACASES,1)-Model_1 | DIFF(MALARIACASES,1) | No Transformation | | Constant | -.358 | 1.778 | -.202 | .841 |
| | | | AR | Lag 1 | -.532 | .142 | -3.750 | .000 |
| | | | | Lag 2 | -.261 | .141 | -1.851 | .070 |
| | | | | Difference | 1 | | | |
| | | | MA | Lag 1 | .997 | 1.973 | .505 | .615 |

The AR (2) model is

$$Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} \qquad (11)$$

Where
$$\left.\begin{aligned}\phi_1 &= -0.532\\ \phi_2 &= -0.261\\ \mu &= -0.358\end{aligned}\right\} \qquad (12)$$

The MA (1)

$$Y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} \quad (13)$$

Where

$$\theta_1 = 0.997 \quad (14)$$

$$Y_t = \varepsilon_t - 0.997 \varepsilon_{t-1} \quad (15)$$

The ARIMA (2, 1, 1) model is $Y_t = -0.358 - 0.532 Y_{t-1} - 0.261 Y_{t-2} + \varepsilon_t - 0.997 \varepsilon_{t-1}$ (16)

The above forecast is represented graphically in Fig. 1.5 below.

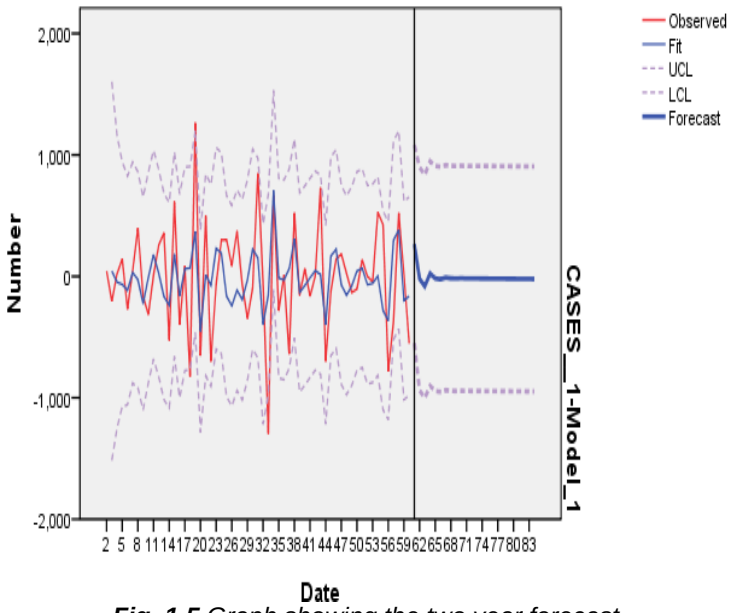


***Fig. 1.5*** *Graph showing the two year forecast*

Fig 1.5 shows the forecast for Malaria reported cases that would be recorded in the next two years. It can be seen clearly that, the forecast shows a downward trend for some time and then remain constant for the period of two years.

## 5 Conclusions

The data was found to non-stationary. However, the graph was found to be stationary after the first differencing. The forecast was to have an oscillatory trend for some time and then remain constant for the period of two years from 2014 and 2016. Moreover, the model was found to have a good fit hence appropriate for the study. Therefore hospitals in the municipality should expect a reduction in the number of malaria cases.

## 6 References

[1] Anon. 2013. Ejisu-Juaben [Online]. Ejisu-Juaben Ghana Ministry of Food and Agriculture. www.mofa.com/gov [Accessed 25-01-2015 2015].

[2] Anderson, T. W. 2011. The statistical analysis of time series, John Wiley & Sons.

[3] Binka, F. N., Morris, S. S., Ross, D. A., Arthur, P. & Aryeetey, M. 1994. Patterns of malaria morbidity and mortality in children in northern Ghana. Transactions of the Royal Society of Tropical Medicine and Hygiene, 88**,** 381-385.

[4] Brockwell, P. J. & Davis, R. A. 2002. Introduction to time series and forecasting, Taylor & Francis.

[5] Djangmah F. M & ESENA. R. K. 2013. Assessment Of Provider Training On The Use Of Amfm Acts In Private Drug Outlets In Kumasi Metropolis Of Ghana. International Journal of Scientific & Technology Research, 2**,** 45-55.

[6] Gemperli, A., Vounatsou, P., Kleinschmidt, I., Bagayoko, M., Lengeler, C. & Smith, T. 2004. Spatial patterns of infant mortality in Mali: the effect of malaria endemicity. American Journal of Epidemiology, 159**,** 64-72.

[7] Ghana News Agency. 2007. Reducing Malaria Deaths Among Children In Rural Ghana [Online]. Ejisu-Juaben Kumasi: Ghana News Agency.

http://www.ghanaweb.com/GhanaHomePage/health/artikel.php [Accessed 25-01-2015 2015].

[8] Ghana News Agency. 2014. Ghana records over 11 million cases of OPD malaria in 2013 [Online]. Ghana. [Accessed 16-01-2015 2015].

[9] Guinovart, C., Navia, M., Tanner, M. & Alonso, P. 2006. Malaria: burden of disease. Current molecular medicine, 6**,** 137-140.

[10] Hipel, K. W., Mcleod, A. I. & Lennox, W. C. 1977. Advances in Box- Jenkins modeling: 1. Model construction. Water Resources Research, 13**,** 567-575.

[11] Mcleod, A. I. & Li, W. K. 1983. Diagnostic checking ARMA time series models using squared- residual autocorrelations. Journal of Time Series Analysis, 4**,** 269-273.

[12] Monserud, R. A. 1986. Time-series analyses of tree-ring chronologies. Forest Science, 32**,** 349-372.

[13] Snow, R. W., Guerra, C. A., Noor, A. M., Myint, H. Y. & Hay, S. I. 2005. The global distribution of clinical episodes of Plasmodium falciparum malaria. Nature, 434**,** 214-217.

[14] Snow, R. W., Trape, J.-F. & Marsh, K. 2001. The past, present and future of childhood malaria mortality in Africa. Trends in Parasitology, 17**,** 593-597.

[15] Yantai, S., Minfang , Y., Oliver, Y., Jiankun, L. & Huifang, F. 2005. Wireless traffic modelling and prediction using seasonal ARIMA models. . The Institute of Electronics, Information and Communications Engineers (IEICE) Transaction on Communications, 88**,** 3992-3999.

[16] World Health Organisation,(2010) , World Malaria Report, www.who.int /malaria/world_malaria_report [Accessed :30/01/2015]