Available online www.jsaer.com

***Journal of Scientific and Engineering Research*, 2018, 5(9):35-42**

Research Article

ISSN: 2394-2630
CODEN(USA): JSERBR

Time Series Analysis and Forecasting Model for Monthly Malaria Infection by Box-Jenkins Techniques in Kass Zone, South Darfur State, Sudan

Hussien Eltom Hassan[1,2]*, Yong Bin[1]

[1]State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing 210098, China
[2]Nyala Technical College, Nyala, South Darfur, Sudan

Abstract Time series analysis has been extensively utilized in health fields and epidemic diseases. Malaria warning system and predicting the incidence of malaria infections is critical to sustain the health. The major goal of this study become to offer a malaria prediction model by means of the usage of Box-Jenkins statistics and historic malaria morbidity records for malaria-endemic areas in Kass zone. A total of 2002 locally malaria infections, which occurred in the Kass sector, South Darfur, Sudan over a period of 4 years from Jan 2005 to Dec 2008, were analyzed by seasonal ARIMA model. Twelve-monthly differencing of INFECT yields the series DINFECT which has an overall upward trend. By the Augmented Dickey-Fuller Unit Root Test both INFECT and DINFECT are adjudged non-stationary whereas DINFECT is adjudged stationary. The ARIMA forecast period is January 2008 to December 2008, there is deviation for month 1 and month 2. Prediction from month 9 to month 11 almost exact. Slight deviation in predicting in moth 6 to month 9, overall prediction is good. The ARIMA forecast model is a valuable tool that has the potential for malaria early warning and early detection in Nyala. It is able to provide dependable information to the applicable authority to act proactively, because the values of the malaria forecast from the best fit ARIMA model that has been chosen and fitted closely with the values of the mentioned malaria infections.

Keywords Forecasting, Malaria, Infection**,** ARIMA, forecast, Kass, South Darfur State

1. Introduction

The world population at malaria disease risk is estimated to be 3.2, billion, and the annual malaria cases were estimated to be 350–500 million, in addition to more than 1 million die from vector-borne diseases each year, more. The world health Organization (WHO) has mentioned that the severe and increasing threats of vector borne diseases are putting the health agenesis and organization in dire. Considering the vector-borne diseases, malaria is the most dangerous threat to the world's populations. 107 International locations had ongoing transmission of malaria in [1-3]. Regardless of diverse modern development in diagnostic and treatment modalities, malaria remains a public health problem in locations that grow round the world as the most critical undertaking in addressing the scourge of malaria and its vectors [2, 4]. Malaria is a very preventable and treatable infection because of parasites of Plasmodium species and transmitted absolutely via the bites of Anopheles mosquito. Although malaria can be prevented and treated, it causes significant fever and disease, particularly in areas far from prevention and treatment sources. In Africa Sub-Sahara, it considered the most affected region by malaria within the world; Sudan has the maximum range of infections with an estimated 17.6 million cases per year. In Africa, more than 80% of malaria mortality occurs [5]. Malaria-stricken families spend huge amounts of their income on malaria treatment. In addition to paying for prevention and suffering from income shortages [6, 7]. because of the extreme health effect of malaria, there is a growing need for

methods with a view to allow forecasting and early warning with timely case detection in zones of risky transmission, just so more manipulates measures can be implemented efficaciously [8]. Despite the fact that there has been a marked reduction within the number of malaria cases in Sudan under Roll back malaria-Sudan RBM, malaria nevertheless is the primary purpose of infectious diseases with the development of drug resistant Plasmodium species and insecticide resistant mosquitoes. Forecasting of malaria instances lets in for allocation of suitable assets to target prevention and treatment of malaria and moreover to offer for eventual elimination. Consequently, the existing study was designed with goals of developing a temporal model for forecasting malaria infections using Box-Jenkins statistics and historic malaria morbidity records for malaria-endemic areas in Kass zone, South Darfur, Sudan.

Hit implementation of RBM program in Sudan has brought about a large decrease in malaria instances in Sudan in current decades. Sudan is in the early stages in terms of malaria control programs, and the NMTD of the ministry of health has developed a programme to eliminate malaria in Sudan by the year 2025 [9, 10]. Many health and endemic time series are known to be seasonal as well as fluctuate. For instance, malaria incidence, annual blood examination rate, the annual parasite index, mortality and morbidity rate are a few such series. Seasonal time series can be modelled using seasonal autoregressive moving average (SARIMA) strategies. Malaria prevalence rates represent this kind of time series. The reason for this study is to model those charges via SARIMA strategies.

A few examples of such use of the SARIMA methodology to model epidemiology and health data contain the following. Krishnan Bhaskaran et al (2013) highlighted the environmental epidemiology using SARIMA models in terms of parsimony and the efficiency of data modelling, London. Ben Lopman et al. (2009) modelled Weather and Virological Factors Drive Norovirus Epidemiology by such methods England and Wels. Alberto Gomez-Elipe et al. (2007) Forecasting malaria incidence and environmental factors in Karuzi, Burundi.

2. Methodology

2.1. Study area

Kass is fall within longitude and latitude 23° 41' -24° 52' East and 11°08'-13°08' north, The altitude of Kass zone is fall within mountainous area ranges between 900 and 3,100 m (mean: 2,000 m) (figure 1). In addition the villages which are falling in the mountainous area can only be reached by foot or animals. Almost the targeted population is scattered all over the zone.

2.2. Data collection

Malaria infection data were collected in the neglected and Tropical Diseases Nyala office, South Darfur State, Sudan. The monthly data of examined blood for malaria has been achieved. Data were treated analyzed, developed and combined for the years 2005-2008. Population for the zone was obtained from Nyala Bureau of Statistics. It took as a right that every resident in Kass became in threat for malaria contamination.

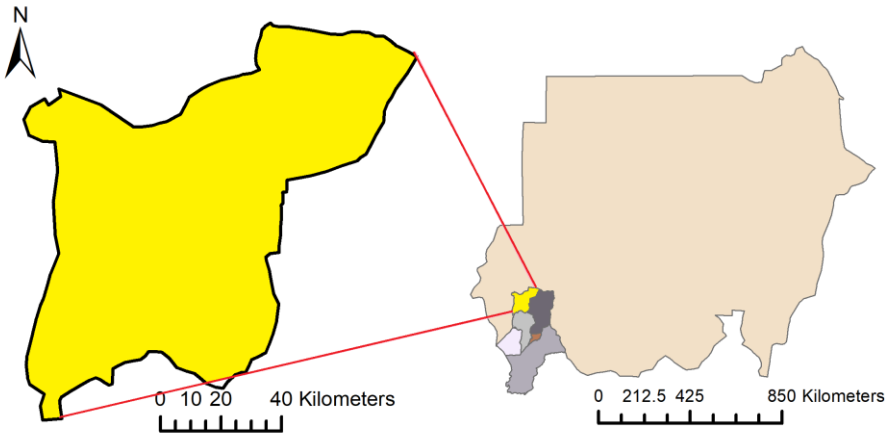

Figure 1: Location map of the study area

2.3. Models

The analytical technique to this study is limited through the box-Jenkins SARIMA model. The SARIMA model combines non-seasonal and seasonal components, and can be designated as SARIMA (p,d,q) × (P,D,Q) s, in which p, d and q talk over with the orders of the non-seasonal autoregressive (AR), non-seasonal differencing and non-seasonal moving average (MA) elements of the model. P, D and Q discuss with the orders of the seasonal AR, seasonal differencing and seasonal MA elements of the model, and s are the period of the seasonal length. AR procedure accounts for previously found values as much as a distinctive maximum lag, plus a blunders term. The technique of differencing is recognized as the combination part that accounts for stabilization of the data by means of removing seasonality or trend, whilst the MA procedure accounts for previous blunders terms, making forecasting less complicated. A stationary time series X_t is stated to comply with an *autoregressive moving average model of orders p and q*, specific ARMA (*p, q*), if it satisfies the following difference equation:

$$X_t - \alpha_1 X_{t-1} - \alpha_2 X_{2-2} - ... - \alpha_p X_{t-p} = \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + ... + \beta_q \varepsilon_{t-q} \quad (1)$$

Where the collection ε_t is a white noise method, $\alpha's$ and $\beta's$ are constants such that the model is each stationary and invertible. Assume the model (1) is positioned within the form:

$$A(L)X_t = B(L)\varepsilon_t \quad (2)$$

Where $A(L) = 1 - \beta_1 L - \alpha_2 L^2 - ... - \alpha_p L^p$ *and* $B(L) = 1 - \beta_1 L - \alpha_q L^q$

And *L* is the backward shift operator defined by means of. $L^k X_t = X_{t-k}$ *(AL)* is the autoregressive *(AR)* operator and *B(L)* is the moving average *(MA)*operator.

Assuming X_t is non-stationary. Box and Jenkins proposed that differencing of the series to the suitable degree *d* ought to render it stationary. Assume the *d*-th distinction, $\nabla^d X_t$ Of X_t is stationary and follows the ARMA*(p,q)* model. Then X_t is said to comply with an autoregressive moving average incorporated model of order*p, d, q*. designated ARMA*(p, d, q)*. Here $\nabla = 1 - L$.

If moreover X_t is seasonal of period *s*, Box and Jenkins [11] further proposed that it could be modelled by:

$$A(L)\Phi(L^s)\nabla^d \nabla^D{}_s X_t = B(L)\Theta(L^s)\varepsilon_t \quad (3)$$

Where $\Phi(L)$ and $\Theta(L)$ are the seasonal *AR* and *MA* operators, respectively, defined by:

$\Phi(L) = 1 + \Phi_1 L + \Phi_2 L^2 + ... + \Phi_P L^P$ *and* $1 + \theta_1 L + \theta_2 L^2 + ... + \theta_Q L^Q$, and the $\Phi's$ and the $\theta's$ contents such that the whole model (3) is both stationary and invertible. The symbol ∇_s is the seasonal difference operator defined by $\nabla_s = 1 - L^s$ The model is known as a multiplicative seasonal autoregressive integrated s moving average model of order *p, d, qP, D, Q,* and *s* and denoted by using SARIMA (*p, d, q P*)*(*P, D, Q*).

Suhartono proposed a subset SARIMA modelling algorithm thus: Fit the (1, 0, 0)*(1, 0, 0) SARIMA model:

$$X_t = \beta_1 \varepsilon_{t-1} + \beta_s \varepsilon_{t-s} + \beta_{s+1} \varepsilon_{t-s} - 1 \quad (4)$$

If $\beta_{s+1} = 0$, then the model is said to be additive but if $\beta_{s+1} \neq 0$ and $\beta_{s+1} = \beta_1 \beta_s$ the model is *multiplicative*. Otherwise, it is said to be *subset.*

2.4. Model Estimation

A SARIMA *(p,d,q)(P,D,Q)* model turned into built the use of monthly malaria case information from January 2005 to December 2007 and a forecast of malaria instances from January 2005 to December 2008, following the steps below:

2.4.1. Identification the model

The power transformation known as the Yeo-Johnson transformation was employed at the time series to stabilize the variance, while SARIMA non-seasonal and seasonal differencing had been performed to acquire

seasonality of the time series by way of eliminating the trend and seasonality. From the non-seasonal and seasonal difference information, the non-seasonal and seasonal components of the model had been formulated through examining their autocorrelation characteristic (ACF) and partial autocorrelation characteristic (PACF). The ACF and PACF have been applied to pick out the degree of differencing and appropriate autoregressive and moving average terms.

2.4.2. Estimating the Parameter

Parameters of the model in step 1 were expected to affirm that each one of the parameters within the achievable model have been significant.

2.4.3. Diagnostic checking

To test for the adequacy of the chosen SARIMA model, we used the residuals of the fitted model to locate the (*ACF)* plot of the residuals and the box-Ljung test. The Q-Q plot and Shapiro-Wilk test have been utilized to test take a look at for normality of the residuals. If all of the diagnostic test outcomes are within desirable limits, the SARIMA model in step 2 is appropriate.

2.4.4. Forecasting

The chosen SARIMA model in step 3 turned into used to forecast malaria cases from January 2007 to December 2008. Mentioned malaria cases for 2014 have been utilized to validate the forecast.

In an effort to fit the model (3) the orders *p, d, q, P, D, Q* and *s* must first be determined. Frequently seasonality isn't evident from the time plot. After differencing a spike within the autocorrelation characteristic *(ACF)* and regularly inside the partial autocorrelation function *(PACF)* shows seasonality of period equal to the corresponding lags. So as to avoid undue model complexity it has been suggested that *(d+D<3)*. Most usually, it is allowed that *(d= D=1)*. The non-seasonal autoregressive order *p* is predicted by the non-seasonal cutoff lag of the *(PACF)* and *q* by the non-seasonal cut-off lag of the autocorrelation function *(ACF)*. The seasonal orders *P* and *Q* are respectively decided by way of the seasonal cut-off lags of the *(PACF)* and *(ACF)*. The parameters of the model are continuously forecast by using nonlinear optimization techniques due to the involvement of items of the white noise process inside the model. An initial estimate of the parameters is normally made and on its basis similarly estimates are acquired iteratively, each estimate predicted to be a development on its predecessor until optimality is attained depending on the degree of accuracy required. The optimality criterion employed is commonly the least squares criterion, the maximum likelihood criterion, the maximum entropy criterion, and so forth. After model fitting evaluation of the residuals is typically done which will ascertain the adequacy of the model. That is known as diagnostic checking.

In this study, all analytical work is accomplished the use of the Eviews 9.5 software program which uses the least squares technique to the estimation. Besides Eviews 9.5 displays values of three data criteria on the basis of which model evaluation can be done. The facts criteria are the Akaike information criterion (AIC) (Akaike, 1977), Schwarz information criterion (Schwarz, 1978) and HannanQuinn data criterion (Hannan and Quinn, 1979). A minimal criterion value is a sign of relative model adequacy.

3. Results

For the sake of this article, the time series information cover 48 months, from January 2005 to December 2008, realization analyzed is referred to as INFECT. And depict notable seasonality and a downward trend of malaria infection, as showed in figure 2A. Twelve-monthly differencing of malaria infection yields the series DINFECT, which reveals similar patterns as INFECT as showed in figure 2B.

3.1. Model identification

In a preliminary attempt to treatment the non-stationary of the time series, and eliminate the trend and seasonality indicated within the ACF plot, non-seasonal differencing became employed. Also in figure3A presents the output of the monthly malaria cases after transformation and non-seasonal differencing. The ACF plot shows that seasonality remains evident (lags 3, 6 and 12). We consequently employed seasonal differencing to remove the impact of seasonality in our model and to look for a better model fit. The non-seasonal aspect of our model becomes recognized by analyzing the ACF and PACF plots of the transformed non-seasonal differences malaria instances. ACF values in figure3B declines gradually after one lags and the PACF decay

exponentially in a sine-wave style. This suggests a moving average of order 1, leading to an autoregressive moving average (0, 1, 1) model (i.e. p = 0, d = 1 and q = 1).

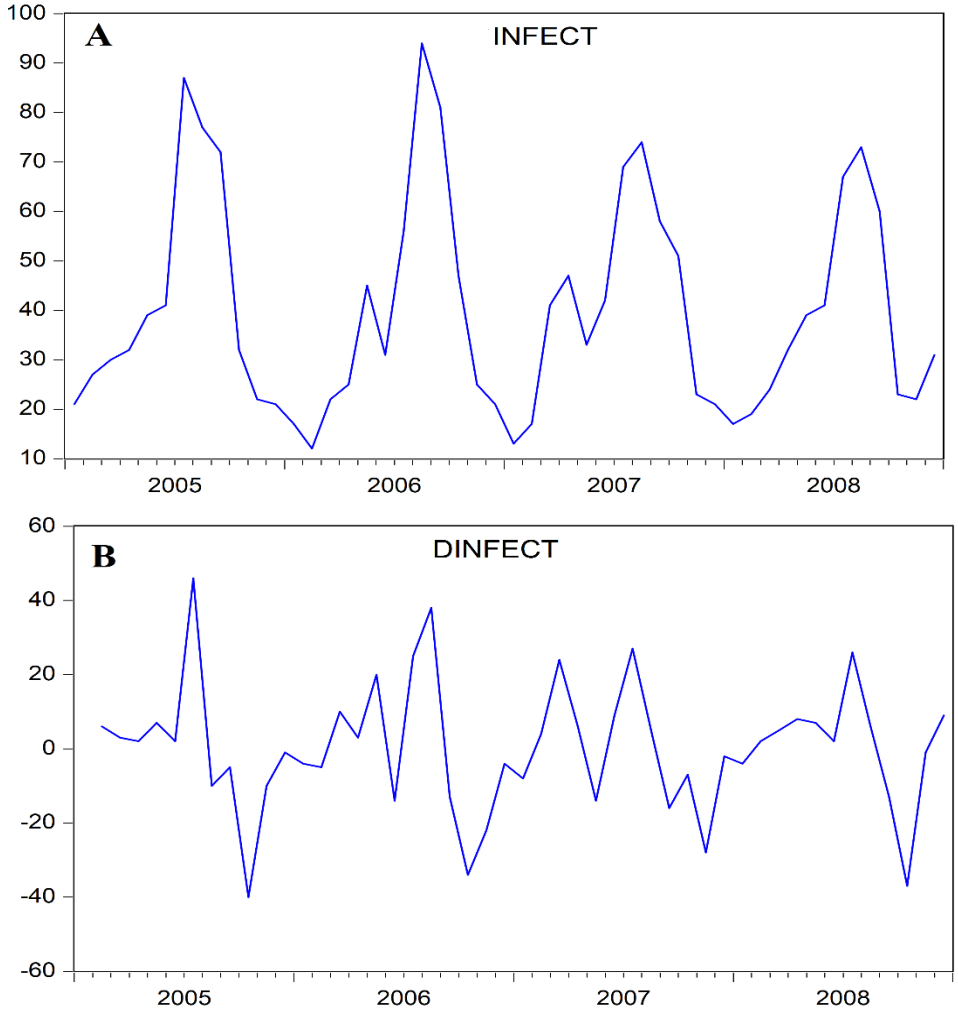

Figure 2: The original and differenced malaria time series plots in Kass (A) time series plots of monthly malaria cases INFECT (B) Twelve-monthly differencing of malaria infection yields the series DINFECT

3.2. Parameters estimation

The goodness-of-fit statistics have been using the Akaike information criterion (AIC), the Bayesian information criterion (BIC), log likelihood and the standard errors. The model with the lowest BIC value and a p-value <0.05 became selected as the best model fit. The BIC values are built on the likelihood function and the AIC The ARIMA (6, 1, 6) model has the smallest AIC, BIC, and volatility (table 1).

3.3 Diagnostic checks

This was done by verifying the correlogram of the model and examine the ACF and PACF we find that the correlogram is not flat since lag 12 is significant and there are some information must be captured (figure 3). So the selected model must be re-estimated by adding ar(12) and ma(12). By testing AIC, BIC, log likelihood and the standard errors value and with a p-value <0.05 became selected as the best model fit. Then the best adjusted ARIMA model was selected to be ar(6) ma(6) ma(12) as showed in table 2. In addition, a correlogram of residuals squared for the adjusted model and the Q-Q plot (Fig 4) depicts some out lines on the tails, suggesting that the normality of the residuals is not rejected. We therefore proceeded to use the SARIMA (0, 1, 1) $(0, 1, 1)_{12}$ models for forecasting, since it provides a reasonable fit to the highly seasonal and non-seasonal time series data.

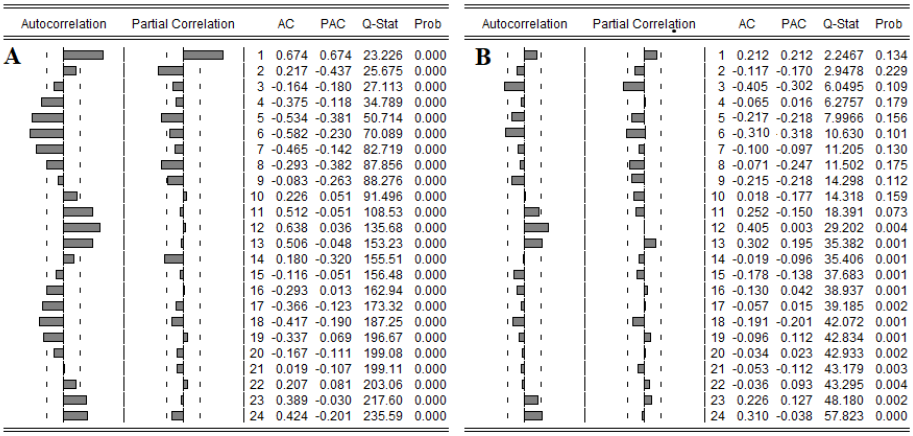

Figure 3: Correlogram of malaria time series in Kass (A) Correlogram of INFECT (B) Correlogram of DINFECT

3.4. Fitting and overcasting

The selected ARIMA model was used to forecast monthly malaria infections from January 2008 to December 2008 (Fig. 5). 12 The plot of the observed monthly malaria infections and predicted cases for 2014 (Fig. 6) suggests that the values for monthly expected infections generally tend to observe the stated values quite intently except in

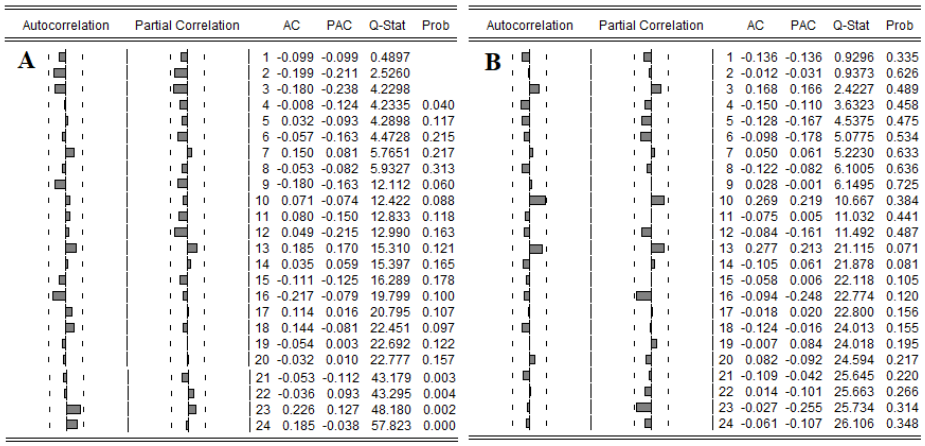

Figure 4: Correlogram of adjusted ARIMA modelQ-Q plot (A) Correlogram of residuals (B) Correlogram of residuals squared

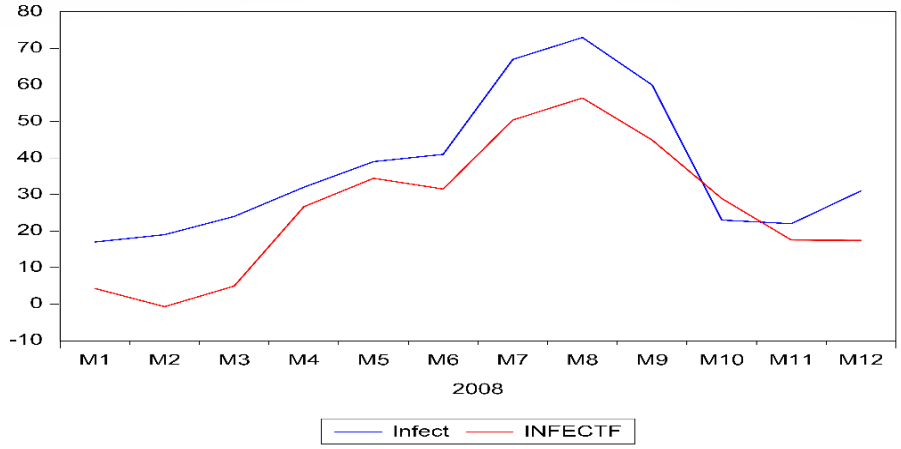

Figure 5: Observed malaria cases from January 2008 to December 2008 and predicted malaria cases from January 2008 to December 2008.

Table 1: Estimated ARIMA Model for Differenced INFECT

Differenced INFECT	ARIMA(3, 1, 3)	ARIMA(3, 1, 6)	ARIMA(6, 1, 6)	ARIMA(6, 1, 3)
Significant coefficient	0	2	2	1
Sigma^2 (Volatility)	185.2378	190.3531	168.5762	216.5609
Adjusted R^2	0.377769	0.360586	0.433737	0.272552
AIC	8.374595	8.453614	8.386996	8.434011
BIC	8.532055	8.611103	8.544455	8.591471

Table 2: Adjusted ARIMA Model for Differenced INFECT

Differenced INFECT	ARIMA(6, 1, 6)	AR(6)AR(12)MA(6)	AR(6)MA(6)MA(12)
Significant coefficient	2	2	3
Sigma^2 (Volatility)	168.5762	138.9875	118.7390
Adjusted R^2	0.433737	0.533128	0.601145
AIC	8.386996	8.340591	8.330681
BIC	8.544455	8.537415	8.527505

4. Discussion

However, an inspection of the data reveals that yearly minimums malaria infection happens in January, February, November, and December respectively. In other hands, the maximum malaria infection happens in May, June, July, and August respectively. That means a demonstration of 12-monthly seasonality. This warrants a SARIMA modelling technique. A twelve-month differencing of malaria infection series yields the series (DINFECT) which has a horizontal trend and a correlogram with the autocorrelation characteristic function (ACF) displaying seasonality and confirming the hypothesized seasonality of 12 months periodicity (figure 3 &4).

Time series predictions are generated by models primarily based on modifications over time in formerly observed values or historic datasets [12]. The ARIMA forecast period is January 2008 to December 2008, there is deviation for month 1 and month 2. Prediction from month 9 to month 11 almost exact. Slight deviation in predicting in moth 6 to month 9, overall prediction is good (fig 6). The model can serve as a beneficial tool for public health employees and epidemiologists. It may be performed as a malaria early-warning system and, can provide vital statistics to allow the applicable authority to behave proactively. This study indicates how the ARIMA model became employed in modelling and predicting malaria cases in a relatively low malaria transmission region, wherein focused interventions are vital to strengthening malaria management and removal efforts.

The epidemiological potential and capability of the ARIMA time series was explored by using distinctive capacities. The authors ensured that the time series techniques attained seasonality within the homogenous sense and variance, which might be elementary situations of an ARIMA model. This becomes completed via carrying out the first differencing and the seasonal differencing, which leads to a stationary time series via removing trends and seasonal results. However, in time where the variance of a time series trends downwards or upwards as the extent of the series decreases or increases, the time series has to be transformed earlier than the analysis or differencing. Although malaria transmission in Kass is restricted because of effective malaria control measures, the NMTD Nyala Office, federal ministry of health nonetheless regards malaria as a significant disease due to its propensity to cause an epidemic.

To prevent an epidemic, NMTD has in place an outbreak threshold of confirmed cases at districts and provinces endemic to malaria, and health facilities time series into the ARIMA model over a longer time frame may want to improve the model fit and the forecast if the exogenous elements liable for trend, seasonality and outliers are integrated into the model.

5. Conclusion

It may be concluded that Kass malaria incidence follows the multiplicative ARIMA model. Which means that a modern value of the time series depends on the past value of a month ago and that of a year ago of shock or its

error terms. This model has been shown to be adequate and may be used for forecasting of Kass malaria incidence. But efforts have to be made to discover the possibility of models that higher account for the variability in the time series. The ARIMA forecast model is a valuable tool that has the potential for malaria early warning and early detection in Nyala NMTD. It is able to provide dependable information to the applicable authority to act proactively, because the values of the malaria forecast from the best fit ARIMA model that has been chosen and fitted closely with the values of the mentioned malaria infections. However, the sensible application of the generated model is recommended. Moreover, studies that employ daily records and contain possible malaria transmission risk factors and confounding in multivariate time-series models are recommended.

Acknowledgments

The author profusely wishes to thank the Malaria and Neglected Tropical Diseases, Ministry of Health, South Darfur, Sudan. Dr. Taha Hussin Musa and Dr. Mohamed Mustafa for theirguidance, encouragement and unwavering support in this undertaking.

References

1. World Health Organization. Malaria Country Profile 2016. Geneva: WHO, 2017. http://www.who.int/malaria/publications/country-profiles/en/ (accessed 15 August 2018).
2. National Human Resources for Health Strategic Plan for Sudan, 2012-2016. Available on line: http://www.who.int/workforcealliance/countries/Sudan_HRHPlan_2012-16.pdf.(Accessed on 15 August 2018).
3. World Health Organization. Proposal for an Evidence Review Group on Border Malaria. 2017. Available online: http://www.who.int/malaria/mpac/mpac-oct2017-proposed-erg-border-malaria-session6.pdf?ua=1 (accessed on 15 August 2018).
4. Adam I, Elhassan EM, Haggaz AED, et al. A perspective of the epidemiology of malaria and anaemia and their impact on maternal and perinatal outcomes in Sudan. The Journal of Infection in Developing Countries 2010, 5(02):83-87.
5. National Human Resources for Health Strategic Plan for Sudan, 2012-2016. Available on line: http://www.who.int/workforcealliance/countries/Sudan_HRHPlan_2012-16.pdf. (Accessed on 15 August, 2018).
6. Malik, E. M.; Atta, H. Y.; Weis, M.; Lang, A.; Puta, C.; Lettenmaier, C.; Bell, A. Sudan roll back malaria consultative mission: essential actions to support the attainment of the abuja targets. 2004, 29.
7. Nmor JC, Sunahara T, Goto K, et al. Topographic models for predicting malaria vector breeding habitats: potential tools for vector control managers. Parasites & vectors 2013, 6(1):1.
8. Akande, T.; Musa, I. Review Article: Epidemiology of Malaria in Africa. African Journal of Clinical and Experimental Microbiology 2005, 6, doi:10.4314/ajcem.v6i2.7409.
9. Malaria and VHF Outbreak in Darfur, Sudan, Situation Report No 14: 18 January - 7 February 2016 Federal Ministry of Health, World Health Organization.
10. Malik EM, Khalafalla O. Malaria in Sudan: past, present and the future. Gezira journal of health sciences 2004, 1(1 Sup).
11. Brockwell, PJ, Davis RA. Time Series: Theory and Methods. New York: Springer Science & Business Media, 1991.
12. Momia Yashpal, Gupta Anil, Talati Shweta, Koushal Vipin, Bhagat Hemant, Kumar Ashok, Singh Ranjana. 2018.Time Series Forecasting of Postponement Rates in Operation Theaters of Advanced Eye Centre -A Robust Optimization Strategy, International Journal of Scientific & Engineering Research Volume 9, Issue 7, 12-17.