Forecasting Malaria Infections in Rwanda Using Arima Model; Case Study of Eastern Province

Manishimwe Florent

Jomo Kenyatta University of Agriculture and Technology

Abstract: ***Malaria is a public problem for health in the Eastern Province of Rwanda. This study was carried out to develop a forecasting model of malaria infections with respect to climatic variables for the period 2008 to 2016 in Eastern Province in order to strengthen the prevention of the country and measures control. The monthly data on malaria infections in this region from January 2008 to December 2016 was obtained from Rwanda Biomedical Center whereas climatic data was obtained from Rwanda Meteorological Agency. The linear generalized models and SARIMA time series models were used in data analysis. These two models have been used in fitting monthly malaria infections as a function of monthly mean temperature, relative humidity and mean rainfall. SARIMA time series models provided a best fit for malaria infections as indicated by residual plots. Pearson' correlation test indicated a positive association between relative humidity and mean rainfall to malaria infections. High malaria infections were observed in July. This study is an important an important tool for policy makers and implementers in order to put in place effectively and efficiently malaria measure controls, because it provides a useful information for forecasting malaria infections and developing a warning system for the future.***

Keywords: Autoregressive processes, Malaria infections, Moving Average processes, Time series

1. Introduction

Malaria is a significant public health problem in Rwanda. Approximately 90% of Rwandans are at risk of malaria. Malaria is the leading cause of morbidity and mortality and it is responsible for up to 50% of all outpatient visits. In 2006, it was responsible for 37% of outpatient consultations and 41% of hospital deaths, with 42% which are children under five years [1].
Malaria transmission can be tested with non-climatic or climatic variables. Their impact on malaria transmissions still remains controversial in different regions. The aim of this this study is about modeling malaria infections in Eastern Province of Rwanda with respect to climatic variables.

Rwanda has two district malaria epidemiological strata: in one part of the country (Central and Eastern regions), malaria is characterized by seasonal peaks of transmission and in the other part of the country (Western region); malaria transmission is quasi stable around the whole year.

Malaria is still being one of the leading health problems not only in Rwanda, but also in other developing world. Rwanda took the world vision of eliminating malaria where the implementation was planned to be reached in 2025. Key malaria control measures to the country have been implemented progressively; specifically, distribution of insecticide bednets, utilization of artemisinin-based combination therapy to treat uncomplicated malaria, indoor residual spraying of insecticides and provision of intermitted preventive therapy for pregnant women and for children under five years. Unfortunately, based on malaria statistical data of a few years ago, this free malaria target plan has not been achieved.

Despite impressive increases in malaria intervention coverage, there was still an increase in malaria infections each year from 2013, contrary to the years before. Figures from Rwanda Biomedical Centre show that the morbidity rate in the country is 9% while the mortality rate is 4%.

The relationship between climatic variables and malaria transmission has been reported in many countries [2]. The impact of environment, weather or climatic changes, are the basic factors on undercurrents of malaria infections and attention in recent years has attracted considerably, but hesitations around future trends of malaria continue.

A poor or ineffective forecasting model building can lead to a weak or incomplete plan and bad management of malaria, therefore based on past experience from other researchers, an updated forecasting model would help in a good implementation of future periods about the disease. Conversely, the relationship between climatic variables and malaria infections in Eastern Province has not been studied.

In this study, association of malaria infections and climatic factors was modeled using linear regression and time series models respectively. This is essential for the expansion of malaria cautionary structure, and permit operative malaria regulator measures in Eastern Province and it will be prolonged to the whole country.

2. Methodology

2.1. Study area

The study was directed to all hospitals from the Eastern Province of Rwanda which is a supreme dominated area of malaria. The Eastern Province is the largest (9, 813 km^2around37.26% of the country), the most populated (2, 600, 812) and the least dense (275 people per km^2) of Rwanda's five province. This Province was created in January 2006 as a government program for decentralization that held for the country's local administrative structures. It has seven districts where we can in each find at least one referral hospital. Agriculture is the most dominating activity of the population in this province. The study area is located at a long plane altitude of 216 meters above the sea level and has a climate with variation between winter and long summer seasons with a high temperature and low precipitation in the country along the whole year.

2.2. Data description

Data on monthly Malaria infections (**Fig.1**) from January 2008 to December 2016 were gotten from Rwanda Biomedical Center, climatic data (**Fig. 2;3;4**) were obtained from Rwanda Meteorological Agency, the well-known nearest neighbor and cross validation method was used to fill the missing data(~5%).

2.3. Data collection

Secondary data on blood examinations (tested on malaria and / or non-malaria cases) was collected from the laboratory archives of hospitals in Eastern Province region. This was carried out in two stages, specifically data on malaria infections and data on climatic factors. The complete number of monthly malaria infections slide positives between January 2008 and December 2016 was taken from the inventory of RBC. The data thus obtained comprises of slide positive rates (SPR) values as recorded by Eastern hospitals. The data for climatic factors (mean temperature, relative humidity and mean rainfall) for the equivalent months was taken from Rwanda Meteorological Agency. Ethical clearance was not sought since these data were collected from official registers.

2.4. Statistical analysis

The forecasting approaches included statistical modeling; mathematical modeling and machine-learning methods (STATA software) were used for fitting the best appropriate model for the time series data. The stationarity of the data was tested by autocorrelation function (ACF) and partial auto-correlation function (PACF). The Ljung-Box test was used to check whether the model is properly specified. To address the confounding factors, forecasts of monthly malaria infections was done together with climatic predictors using the best fitting model.

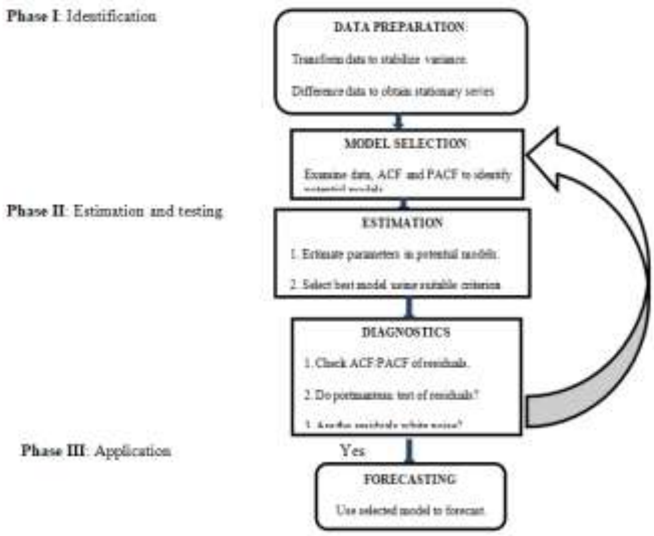
2.5. Study design

Box-Jenkins is a relatively accurate technique and powerful forecasting tool. **Table 1** shows a step by step process required for identifying the appropriate model, estimating parameters and checking that the model is adequate.

3. Results

The number of monthly laboratory confirmed cases for malaria infections showed a decreasing tendency from 2008 to 2012 **(Table 2)**. But it increased during the years 2013 to 2016. Investigation of the monthly malaria infections, the mean precipitation, mean temperature and relative humidity from 2008 to 2016 demonstrates no strong tendency and recommends a seasonal dependency in the series.

Table 3.1: Schematic Representation of the Box-Jenkins for time series modeling



In our case, monthly data on malaria infections and climatic factors are of interest. By observation, **Figure 1** show that, malaria infections series is not stationary because, it exhibits a long term pattern and the mean is not zero. The wildest fluctuations in **Figure 5** occur around January, 2013 where abruptly rose and dropped before settling towards more quasi usual levels. In addition, when we fit a first order autoregressive model for the raw data of malaria infections, we see that the coefficient of AR(1), $\phi_1 = 0.73$ which is less than one.

Table 3.2: Monthly malaria infections and climatic data

Year	Month	January	February	March	April	May	June	July	August	September	October	November	December
2008	**Malaria data**	1297	2182	2829	2403	2526	2968	2686	1848	1902	1820	1441	2347
	Temperature	24.6	25.5	25.8	26.2	25.9	26.6	27.1	27.3	27.8	26.3	25.2	25.8
	Rainfall	18.3	19.2	11.3	25.4	30.6	27.7	14	10.1	8.2	21.3	29.8	27.5
2009	**Malaria data**	2952	2455	3573	2057	5049	5372	4794	3580	3390	2826	4164	5755
	Temperature	27.1	26.9	26.5	25.9	25.7	26.2	26	26.5	26.7	26.4	26.6	25.4
	Rainfall	15.2	3.2	7.4	22	35.8	44.5	26	24.6	20	18	15.4	12.9
2010	**Malaria data**	3830	4112	4604	6297	11786	10575	4280	1214	883	990	988	1377
	Temperature	26.2	27.1	26.3	26.1	25.7	24.7	25.4	26.4	26.2	26.5	25.8	26.4
	Rainfall	14.7	23.2	12	9.9	8.5	5.5	5.3	3.8	2.8	8.1	8.1	6.9

2011	**Malaria data**	689	565	705	694	2157	4799	2272	1037	416	400	880	1367
	Temperature	27	26.4	26	26.3	25.4	25	25.3	25.4	25.5	25.4	26.3	24.8
	Rainfall	17.6	19.7	22.2	15.3	6.5	1.4	2.6	0.8	5.3	6.7	9	9.2
2012	**Malaria data**	1300	1038	579	659	1178	3821	3652	1752	1554	3455	37451	12061
	Temperature	25.8	26.6	25.9	25.6	26.4	23.6	26.2	25.9	26.6	25.8	26.7	26.5
	Rainfall	17	10.1	15.4	17	14.3	4.5	3.2	13.9	48.5	58.8	110.5	70.5
2013	**Malaria data**	11039	10386	7740	7992	12961	15908	5212	2685	3236	6861	12518	13799
	Temperature	24.9	26.2	25.5	25.2	25.8	24.9	26.7	27.2	26.9	26.2	25.2	25.3
	Rainfall	166.5	158	81	92	150.1	143.3	102	125.7	130.2	83.6	114	80.1
2014	**Malaria data**	18487	11985	8873	10428	16054	20450	8351	6012	11281	23160	38167	45114
	Temperature	25.3	25.7	24.7	24.2	25.5	26.9	28.7	27.6	27.2	27.5	25.7	26.5
	Rainfall	80.1	128.8	61.9	130.7	34	16.7	15	14.3	58.5	209.6	98.6	149.1
2015	**Malaria data**	36053	19066	12878	14652	17070	26265	20294	6200	7383	7273	7354	8740
	Temperature	25.7	26.2	25.8	25.3	25.9	27.6	27.5	27.5	26.7	27.3	27.2	26.8
	Rainfall	38	48.2	87.2	45.5	42.5	52.3	43.6	40.9	42.4	41.9	166.8	124.3
2016	**Malaria data**	25963	22465	13488	10324	10345	14987	15670	3024	4750	5765	5211	6066
	Temperature	24.8	25.1	25.6	25.7	26.6	27.4	27.9	27.2	26.9	27	26.8	26.5
	Rainfall	151.2	122	77.3	34.3	28.7	21.3	7.8	5.3	27.6	30.7	54.9	59.1

Source: Rwanda Biomedical Center and Rwanda Meteorological Agency

3.1. Impact of Climatic Factors on Malaria Infections

Malaria infections and relative humidity or mean rainfall all have positive correlations, as one might expect. Their correlation with mean temperature is negative: the warmer the air, the less malaria infections (or vice versa). At lag 0, **figure 6** shows that there is no correlation between malaria infections and mean temperature; **figure 7** shows that there is a positive direct correlation between mean rainfall and malaria infections and **figure 8** shows that there is again an immediate correlation between mean rainfall and malaria infections reaching their maximum point at lag 1(one month before). This means that an increase in humidity or in rainfall causes an abrupt upsurge in malaria infections.

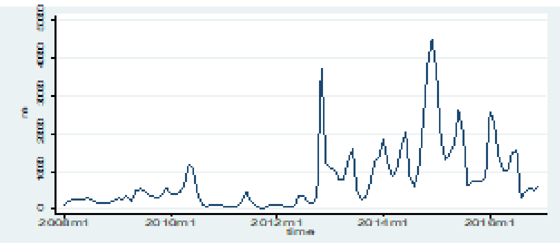


Figure 3.1: Malaria Infections against Time

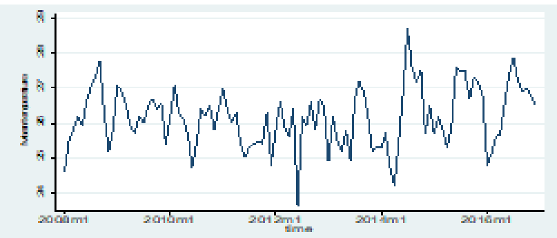


Figure 3.2: Mean temperature against time

Together, these three drivers (climatic factors) now explain 72.7% of the variance in monthly malaria infections where relative humidity and mean rainfall have by far the strongest effect, in a positive direction. Once we control either relative humidity or mean rainfall, the coefficient on mean temperature becomes positive as well.

$$\mathbf{mi_t = -40320.02 + 996.76mt_{t-1} + 99.77rh_{t-1} + 242.38mr_{t-1}} \quad (3.1)$$

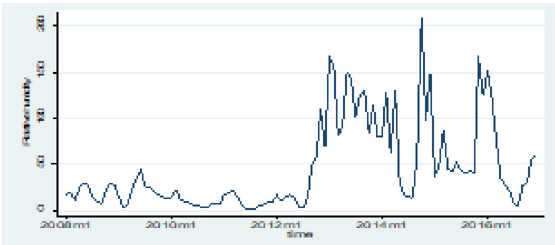


Figure 3.3: Relative Humidity against Time

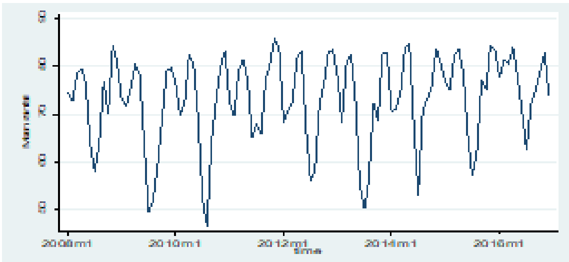


Figure 3.4: Mean Rainfall against Time

International Journal of Science and Research (IJSR)
ISSN: 2319-7064
Index Copernicus Value (2016): 79.57 | Impact Factor (2017): 7.296

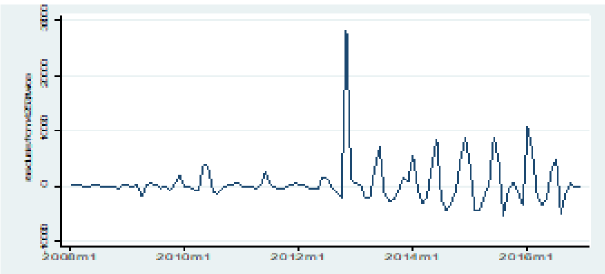


Figure 3.5: Malaria Infection Residuals plot

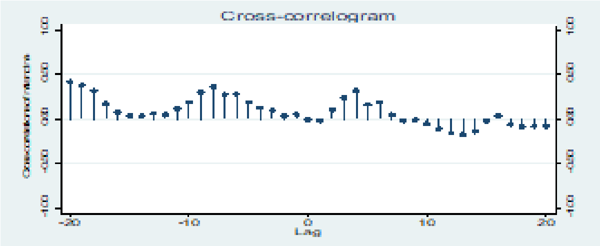


Figure 3.6: cross correlation of Malaria infections and Mean Temperature

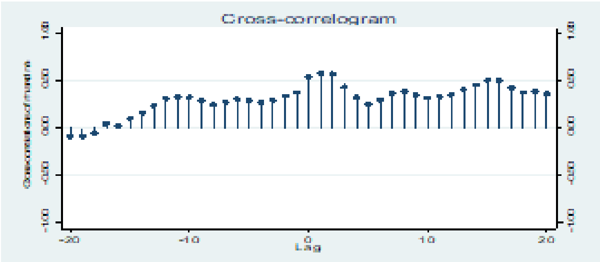


Figure 3.7: cross correlation of Malaria Infections and Relative Humidity

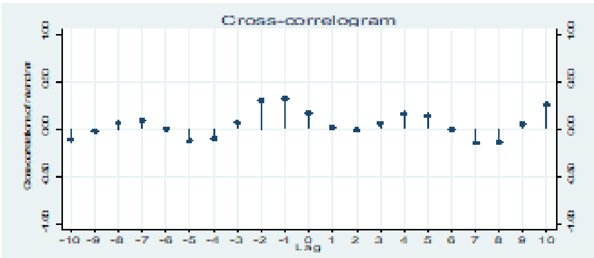


Figure 3.8: Cross Correlation of Malaria Infections against Mean Rainfall

Test for stationarity of malaria infections

The dfgls reports tests of the nonstationary null hypothesis, that malaria infections series represents a random walk, or has a unit root for lags from 1 to 12 months. MAIC recommend 5 lags. The DF-GLS statistic for 5 lags is –2.185, not greater than the 10% critical value of –2.661, not greater than the 5% critical value of –2.946 and not greater than the 1% critical value of –3.570. If the test statistic does not exceed the 5% critical value in absolute terms, we cannot reject non-stationarity. In each case here we fail to reject the null hypothesis and conclude that malaria infections series has a unit root.

In addition, Box-Pierce'Q statistics tests the null hypothesis that all correlations up to lag k are equal to zero. The series of malaria infections show significant autocorrelation as shown in **figure 9** that for any k vakue are less than 0.05; consequently, we reject the null hypothesis that all lags are not autocorrelated. We can see from **figure 9** and **figure 10** that the sample ACF dies out very slowly, while the sample PACF is only significant at the first lag. Also note that the PACF value at the first lag is very close to one.

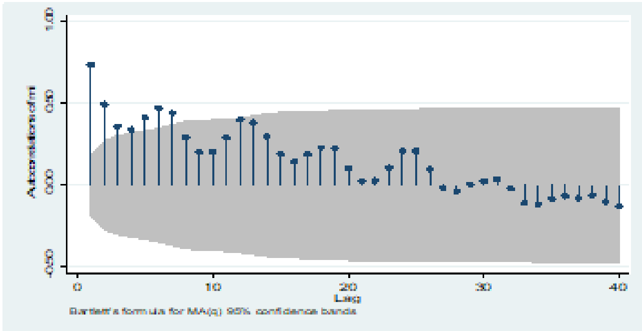


Figure 3.9: malaria infections ac

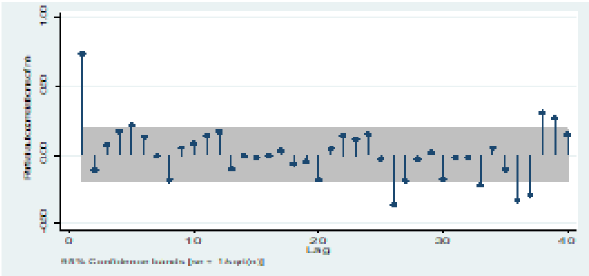


Figure 3.10: malaria infections pac

Expert modeler of the STATA statistical program ver. 11.2 was used to find the best fit model for forecasting malaria infections. Based on characteristics of a good model (the one with a few number of parameters), malaria infections could be modeled as $ARIMA(1,1,1) \times (0,1,1)_{12}$ process of the form

$$\boldsymbol{\phi(B)\varphi(B)(1-B_{12})y_t = \theta(B)\psi(B)(1-B_{12})\mu_t}(3.2)$$

Where B denotes the backward shift operator; $\phi(B), \varphi(B), \theta(B)$ *and* $\psi(B)$ are polynomials of order p=1, P=0, q=1, and Q=1 respectively and $\mu_t$is the purely random process with mean zero and constant variance σ_ε^2.

Table 3: SARIMA model output

ARIMA regression

Sample: 2009m2 - 2016m12 | Number of obs = 95 | Wald chi2(3) = 151.36 | Log likelihood = -964.9904 | Prob > chi2 = 0.0000

DS12.mi	Coef.	OPG Std. Err.	z	P>|z|	[95% Conf.	Interval]
mi						
_cons	-45.37505	94.83082	-0.48	0.632	-231.24	140.4899
ARMA						
ar L1.	.5935283	.1396207	4.25	0.000	.3198767	.8671799
ma L1.	-.8895468	.1111172	-8.01	0.000	-1.107333	-.6717611
ARMA12						
ma L1.	-.755662	.1108186	-6.82	0.000	-.9728625	-.5384615
/sigma	5889.48	257.9461	22.83	0.000	5383.915	6395.045

From **table 3**, the model describes the first difference or month-to-month change in the number of malaria infections as a function of present and one month lagged random noise.

$$(1-0.59B)(1-B_{12})(1-B)y_t = (1+0.88B)(1+0.75B_{12})\mu_t (3.3)$$

Where y_t represents malaria infections number at time t. Parameter estimates are $\phi_1 = 0.57$, $\theta_1 = -0.88$ and $\psi_1 = -0.75$. The terms ar(1) ma(1) and sma(1) are all statistically significant (p 0.000), and the model's residuals are indistinguishable from white noise. It was again found that the mean rainfall (*P* value = 0.002) and relative humidity (*P* value = 0.000) as climatic factors were significant predictors of malaria infections in the study area both lagged at one month while the mean temperature has no significant proofs in predicting malaria infection. Graphically, predicted values from this model appear almost indistinguishable from the observed first regular and seasonal differences of malaria infections **(Figure 11).**

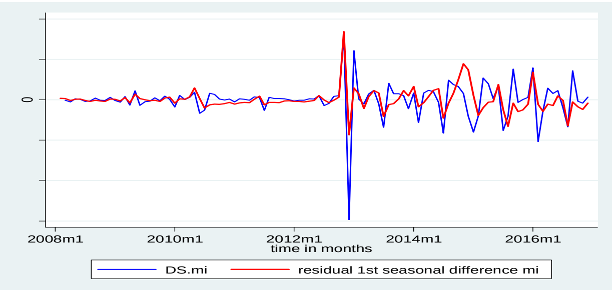


Figure 3.11: Predicted 1st regular and seasonal differences

The forecasting model proposed seasonal ARIMA model, provides a comprehensive set of tools for univariate time series model identification, parameter estimation and forecasting, and it provides a great flexibility in analysis.

After getting the final model SARIMA (1, 1, 1)(0, 1, 1)$_{12}$ of the monthly malaria infections in Eastern Province, that has been expressed above which can be expressed in equation (2), the researcher used it for forecasting future malaria infections number. We forecasted the monthly malaria infections in 2017 for 6 months with the last 4 actual values not included in the original series in order to compare them with the forecasted values of the series.

Figure 12 shows the result of predictions and follows the same behavior of the original series of monthly malaria infections in Eastern Province and the results of forecasts for the year 2017 are all between the upper and lower boundaries of the 95% confidence intervals. This confirms that the forecasting is very efficient.

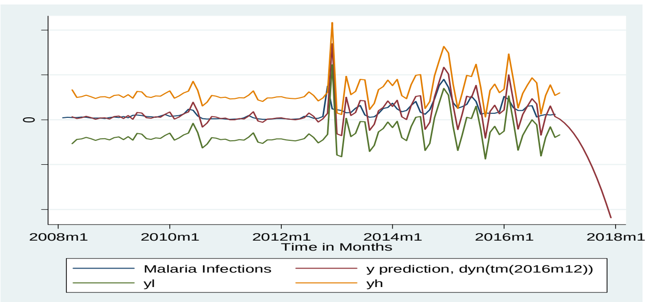


Figure 3.12: Plot of the forecasted data with 95% confidence interval

3.2. Conclusion

The forecasted values through 2016 to 2017 showed harmony with its counterparts in the original series values. Moreover, the forecasted values for the year 2017 are all between the upper and lower boundaries of the 95% confidence intervals. Thus, it provided a future image of the reality of monthly malaria infections in Eastern Province. It showed a slight increase in the monthly malaria infections number in the Eastern Province and there is a real problem that faces that area. Therefore, the officials and decision-makers can adopt the results of this study to face monthly malaria infections phenomena.

The increase monthly malaria infections in the upcoming years in Eastern Province may be due to increasing population, climatic changes or resistance to medicine of mosquito. This problem need for provision of alternative measures for malaria infections control.

3.3. Recommendations

Through the results that have been reached, we recommend the following:

To adopt the results of this research and the adopted formula of forecasting by the related agencies because it uses the suitable scientific style in forecasting as well as taking in account that there is a real problem facing Eastern Province and our country through the upcoming years which is eliminating malaria as one of millennium goals. Furthermore, it helps the officials and decision makers in finding solutions and quick alternatives to face this problem and putting the future plans of the monthly malaria infections to stop aggravating the problem.

To use this method in deducting the standard method and improving it, to forecast not only monthly malaria infections but also other fields of research that predictions be produced every year.

Generalizing this study to similar studies on other Provinces and comparing between them.

Malaria is a vital problem to socio- economic development and progress, thus it should be treated carefully to meet the millennium goals that to eliminate it in year 2030.

References

[1] Karema et al., (2012): Trends in malaria cases, hospital admissions and deaths following scale-up of anti-malarial interventions, 2000–2010, Rwanda. Malaria Journal.

[2] Aribodor, D.N., Ugwuanyi, I.K. and Aribodor, O.B. (2016) Challenges to Achieving Malaria Elimination in Nigeria. American Journal of Public Health Research, 4, 38-41

[3] Varun Kumar et al (2014), Forecasting Malaria Cases Using Climatic Factors in Delhi, India: A Time Series Analysis; Department of Community Medicine, Vardhman Mahavir Medical College and Safdarjung Hospital, New Delhi 110029, India

[4] Nkurunziza, H., Gebhardt, A. and Pilz, J. (2011) Geo-Additive Modeling of Malaria in Burundi. Malaria Journal, 10, 234. https://doi.org/10.1186/1475-2875-10-234
[5] Stella W. and Robert W.(2016), Forecasting paediatric malaria admissions on the Kenya coast using rainfall; Spatial Health Metrics Group, Kenya Medical Research Institute-Wellcome Trust Research Programme, Nairobi, Kenya.
[6] Lindsay, S.W., Bodker, R., Malima, R., Msangeni, H.A. and Kisinza, W. (2000) Effect of 1997-98 El-Nino on Highland Malaria in Tanzania. The Lancet, 355, 989-990. https://doi.org/10.1016/S0140-6736 (00)90022-9
[7] Martens, P., Kovats, R.S., Nijhof, S., De Vries, P., Livermore, M.T.J., Bradley, D.J., Cox, J. and McMichael, A.J. (1999) Climate Change and Future Populations at Risk of Malaria. Global Environmental Change, 9, S89-S107. https://doi.org/10.1016/S0959-3780(99)00020-5
[8] Abeku TA (2002); forecasting malaria incidence from historical morbidity patterns in epidemic-prone areas of Ethiopia-simple seasonal adjustment method performs best. Tropical Medicine and International Health 7(10): 851–857.

Author Profile

MANISHIMWE Florent, Student in Masters of Applied Statistics, Department of Statistics and Actuarial Sciences, School of Mathematical Sciences, Jomo Kenyata University of Agriculture and Technology; Nairobi; Kenya, Email: maniflores1@gmail.com, Tel: 250783529023