

# ARIMA Modelling for Malaria Infection in Kachin State

Theingi Aye

## Abstract

This study attempts to model and to forecast malaria infection of Kachin State which had been affected malaria highest risk areas at 2004 in Myanmar. This study utilized monthly time series data from January, 2011 to December, 2016 and employed the well-known Box-Jenkins Seasonal ARIMA Modeling procedures. The objectives of this study are to study Malaria incidence in Kachin State, to examine the best fitted ARIMA model and to forecast the incidence of Malaria infection in Kachin State based on the best fitted model. Following the Box and Jenkins methodology, the time series modeling involves transformation of the data to achieve stationary followed by identification of appropriate models, estimation of model parameters, diagnostic checking of the assumption model and finally forecasting of future data values. SARIMA (1,0,0) x (1,1, 0)<sub>12</sub> was defined the best model to predict the future Malaria infection in Kachin state and forecasted the future values using the fitted model. The results of this paper indicate that over 50% of malaria incidence in Kachin State is decreased in 2017. That is why malaria incidence in Kachin State may be eliminated in 2020 although the Kachin State is not included in 2020 targeted areas for malaria elimination in Myanmar. There is also observed that the SARIMA model is capable of representing with relative precision the number of malaria infection in the next year.

**Keywords:** Malaria, Seasonal ARIMA, Time Series, ARIMA Modeling

## 1. Introduction

Malaria is an infectious disease caused by a parasite; it is spread by the bite of an infected mosquito. Every year, 300 to 700 million people get infected. Malaria kills 1 million to 2 million people every year.

Globally, more countries are moving towards elimination. In 2016, WHO identified 21 countries with the potential to eliminate malaria by the year 2020.

Asia ranks second to Africa in terms of malaria burden. In 2010, WHO estimated around 34.8 million cases and 45,600 deaths due to malaria in Asia. Sri

Lanka was declared malaria-free in 2016, becoming only the second country in Southeast Asia, after the Maldives, to successfully eliminate malaria. Apart from India, Indonesia, Myanmar, and Thailand, malaria-endemic countries have reported reductions in malaria incidence of more than 75% since 2000.

In Myanmar, Malaria was one of the major public health problems about 1976. In year 1978, Peoples Health Plan was initiated and malaria control programme was integrated with Basic Health Services. Compared to 2002 data, cases and deaths were lower in 2003 and 2004. Year 2004 is the lowest recorded number of malaria cases in outpatient and inpatient department, malaria deaths as well as malaria morbidity rate and mortality rate in Myanmar during the last two and half decade period.

In term of year 2004 statistics, the areas of high malaria morbidity rate (per 1000 population) are Rakhine State (62.43), Chin State (46.41), Kayah State (28.92), Kachin State (24.34) and Tanintharyi Division (21.46). High malaria mortality rate (per 100,000 population) were seen in Kayah State (14.00), Kachin State (8.80), Chin State (8.58), Shan State (7.55) and Tanintharyi Division (7.44). Myanmar has made impressive progress in malaria control during the past 5 years: 80 % reduction in the number of confirmed malaria cases has been registered from 2011 to 2016 (from 567,452 to 110,146 respectively) and 96 % reduction in the number of deaths attributable to malaria has been reported in the same period (from 581 to 21 respectively).

In term of world malaria report (2017), effective surveillance of malaria cases and deaths is essential for identifying the areas or population groups that are most affected by malaria. Malaria related cases had dropped significantly and there was zero malaria death since 2015. Although significant progress has been made in recent years, the malaria burden in Myanmar remains the highest among the six countries of the GMS (Greater Mekong sub region).

The ministry is trying to stop transmission of malaria in five regions; Yangon, Ayeyarwaddy, Bago, Mandalay and Magwe by 2020. These five regions are excluded the regions which has been the lowest recorded number of malaria cases, death, morbidity rate and mortality rate in Myanmar, 2004. Those lowest recorded areas of 2004 were Rakhine State, Chin State, Kayah State, Kachin State and Tanintharyi Division. Kachin state was the highest infectious area among the lowest recorded areas of 2004. This study is mainly focused on Kachin State in order to figure out why this area is not excluded in targeted areas of malaria elimination in 2020.

The objectives of the study are

- i) to study Malaria incidence in Kachin State
- ii) to examine the best fitted ARIMA model of Malaria infection in Kachin State
- iii) to forecast the incidence of Malaria infection in Kachin State based on the best fitted model.

## **2. Data and Methods**

Monthly time series secondary malaria data of Kachin State from January 2011 to December 2016 are used for this study. The required data are obtained from National Malaria Control Program (NMCP), Nay Pyi Taw, Myanmar.

Box and Jenkins Method (SARIMA) model is applied for forecasting the incidence of Malaria in Kachin State, Myanmar. Model identification is made based on autocorrelation (ACF) and partial autocorrelation function (PACF). The parameters are estimated by using the Least Square Method depending on the model. The adequacy of the models was verified by plots of the correlograms and ACF and PACF of the residuals and Ljung-Box test, which is a test for hypotheses of no correlation across a specified number of time lags. ACF of the residuals and Ljung-Box statistics are useful for testing the randomness of the residuals.

## **3. Results and Discussion**

Monthly malaria data kindly supported from National Malaria Control program (NMCP) Myanmar are analyzed by using Box-Jenkins Method. There are four steps in this method which are model identification, parameters estimation, diagnostic checking and forecasting. The data of malaria infection cases was used as a platform for creating the ARIMA models.

In term of descriptive statistics, over the six-year period, the table indicates that the average number of malaria infected people in 2011 and 2012 are the peak and its of 2016 is the least. It shows that the malaria infection in Kachin State are declined. Comparing every month of a year, the average incidence of malaria is peak on June and July.

The data series has a seasonal pattern with peaks and valley in the same month of the year. There are regular patterns with different magnitudes. Therefore, seasonality is testing by using ANOVA table. According to the decision rule, the

result is lead to reject  $H_0$  which is no monthly affect in data. Therefore, there is seasonality in Malaria Infection of Kachin State data series. As per plot of monthly malaria infection in Kachin State, both mean and variance are not stationary as well. The mean level depends on time and that variance decreases as the mean level decreases.

#### a.) Model Identification

To remove non stationarity, logarithms transformation is performed to stabilize the variance of a time series. The ACF of  $\ln(Z_t)$  series show damped sine wave and there is significant spike of PACF at Lag 1 and 2. Therefore, the log transformed data series seems to fit an AR (2) Model. The log transformed series is not stationary in the mean. To remove non-stationary, the log transformed data series is needed to compute as seasonal differencing because of existing seasonality. It means that it leads to remove changes in the level of the time series.

#### b.) Parameter Estimation and Diagnostic Checking

As a result of this, the sample ACF decays exponentially and sample PACF is significant spike at only Lag 1 which suggest SARIMA (1,0,0) x (0,1,0)<sub>12</sub> may be for this data series. The estimated parameter of  $\phi_1$  is 0.908 which is less than one, supporting the required stationary and invertibility condition. Since its p-value is zero, there is no evidence to reject the null hypothesis. It is statistically significant at 0.01 significance level.

SARIMA (1,0,0) x (0,1,0)<sub>12</sub> model is

$$(1 - \phi_1 B) (1 - B^{12}) \ln(Z_t) = \theta_0 + a_t$$

the estimated model is

$$(1 - 0.908) (1 - B^{12}) \ln(Z_t) = -0.547 + a_t$$

With respect to the estimated results of malaria infected people in Kachin State, the feasible model is SARIMA (1,0,0) x (0,1,0)<sub>12</sub> model for the data series of Malaria Incidence in Kachin State.

The residuals values of the ACF and PACF for the Malaria Incidence are all small and fall within the two limits, lower confidence limit (LCL) and upper confidence limit (UCL) as well as exhibit no pattern, it can be said that the residual series are white noise process. It means SARIMA (1,0,0) x (0,1,0)<sub>12</sub> model is adequate

to represent the seasonally log transformed data series of Malaria Incidence in Kachin State.

On the other hand, the autocorrelation among residuals are checked by using the test statistic (Q).

$$H_0 : \rho_1 = \rho_2 = \dots = \rho_k = 0 \text{ (There is no autocorrelation among residuals.)}$$

The observed value of Q is 15.268 and it is not significant since p-value is 0.576 which is greater than 0.05. It means there is no autocorrelation among residuals. Thus, the model SARIMA (1,0,0) x (0,1,0)<sub>12</sub> is adequate.

In addition, another possible model, SARIMA (1,0,0) x (1,1,0)<sub>12</sub> Model, to represent seasonal log transformed Malaria Incidence in Kachin State data series is also fitted as follow.

The results are providing that the seasonal autoregressive order one,  $\phi_1$  was estimated to be 0.927 (SE= 0.061) and seasonal  $\Phi_1$  was estimated to be -0.300 (SE= 0.179) and statistically significant at 0.1 significance level. The estimated parameters of  $\phi_1$  and  $\Phi_1$  are less than one, supporting the required stationary and invertibility conditions.

SARIMA (1,0,0) x (1,1,0)<sub>12</sub> model is

$$(1 - \Phi_1 B^{12}) (1 - \phi_1 B) (1 - B^{12}) \text{Ln}(Z_t) = \theta_0 + a_t$$

the estimated model is

$$(1 + 0.3B^{12}) (1 - 0.927) (1 - B^{12}) \text{Ln}(Z_t) = -0.536 + a_t$$

With respect to the estimated results of malaria infected people in Kachin State, SARIMA (1,0,0) x (1,1,0)<sub>12</sub> model is a tentative model for data series of Malaria Incidence in Kachin State.

The residuals values of ACF and PACF for the Malaria incidence are all small and lie inside the confidence limits, lower confidence limit (LCL) and upper confidence limit (UCL), as well as exhibit no pattern. This suggested that the residuals are white noise. SARIMA (1,0,0) x (1,1,0)<sub>12</sub> model is adequate to perform the seasonally log transformed data series of Malaria Incidence in Kachin State.

On the other hand, the autocorrelation among residuals are checked by using the test statistic (Q).

$$H_0 : \rho_1 = \rho_2 = \dots = \rho_k = 0 \text{ (There is no autocorrelation among residuals.)}$$

As the result of above table, the observed value of Q is 12.641 and it is not significant since p-value is 0.699 which is greater than 0.05. It means that there is no autocorrelation among residuals.

By comparing these two feasible Models, SARIMA (1,0,0) x (0,1,0)<sub>12</sub> Model and SARIMA (1,0,0) x (1,1,0)<sub>12</sub> Model, the value of R-squared for SARIMA (1,0,0) x (1,1,0)<sub>12</sub> Model is slightly larger than the other one. On the other hand, the values of MAPE and Normalized BIC are smaller than other one as well. That is why, SARIMA (1,0,0) x (1,1,0)<sub>12</sub> Model is adequate to fit and used to forecast the malaria infection of Kachin State next year.

### **c.) Forecasting with SARIMA (1,0,0) x (1,1,0)<sub>12</sub> Model for Kachin Malaria**

#### **Infection Series**

The forecast values for 12 months period from January to December, 2017 indicate that the predicted values of SARIMA (1,0,0) x (1,1,0)<sub>12</sub> Model is not very close to the true value but all of them fall within lower confidence level (LCL) and upper confidence level (UCL) which proved the reliability of data series. According to these forecasting results, Malaria infection of Kachin state lead to eliminate in near future.

## **4. CONCLUSION**

This study focuses on the modeling and forecasting of Malaria Infection in Kachin State by modeling data from January, 2011 to December, 2016 using time series SAREMA model. SARIMA model can be obtained by using four iteratively Box-Jenkins steps and provide the prediction of the malaria infected number of people in Kachin State. Following Box and Jenkins methodology, the time series modeling involves for transformation of the data to achieve stationary, followed by the identification of appropriate models, estimation of model parameters, diagnostic checking of the assumption model and finally forecasting of the future data values.

In model identification process, test of seasonality is conducted by using ANOVA. Theoretical and estimated autocorrelation function (ACF) and partial autocorrelation function (PACF) play important role in the construction of SARIMA model. The estimated residuals are analyzed using the ACF and PACF to diagnose if the residuals are consistent with the hypothesis that the residuals are white noise.

Non-constant variance is removed by performing a natural log transformation. Then, trend in the series is removed by taking seasonal differencing. The results indicate that SARIMA (1,0,0) x (1,1,0)<sub>12</sub> Model is the fitted models. The model was also be able to represent the past data with MAPE, R<sup>2</sup> and normalize BIC are 14.083, 0.928 and 12.168 for malaria infection in Kachin state. As forecasting is essential for planning and operational control in a variety of areas, forecasting is made based on the best fitted SARIMA model.

In term of “Myanmar Times” publication on July 2017, The Public Health Department is aiming to make five regions; Yangon, Ayeyarwaddy, Bago, Mandalay and Magwe to free from malaria by 2020. Anti-malaria campaigns are especially being implemented in those regions according to the department.

Evaluation and forecasting the volume of malaria infection in Kachin State are significantly declined. The results of malaria infection in Kachin state is decreased over 50 % of infection (3517 to 1631) from 2016 to 2017. Although the Kachin State is not included in 2020 targeted areas for eliminating Malaria Infection in Myanmar, Malaria Incidence in Kachin State might be eliminated in 2020. It is because the finding of this study is shown that malaria incidence in Kachin State was declining. That is why it can be clearly realized how the implementation of anti-malaria activities were continuously conducted by the responsible people.

The SARIMA model was used the data of malaria infection of Kachin State from January, 2011 to December, 2016, which contained 72 observations. Next 12 months was forecasted by the SARIMA (1,0,0) x (1,1,0)<sub>12</sub> model well reflected the trend in the malaria infection of Kachin State. Results are indicated that SARIMA model was capable of representing the number of malaria infection in the following month with relative precision. That is why forecasting using time series models are useful of policy-making, supports for the planning and future analysis in Myanmar’s health industry.

## **ACKNOWLEDGEMENT**

I would like to thank to Professor Dr. Tin Win, Rector of Yangon University of Economics and Professor Dr. Nilar Myint Htoo, Pro-rector of Yangon University of Economics and Professor Dr. Maw Maw Khin, Professor, Head of Department of Statistics and Director of Master of Applied Statistics, Yangon University of

Economics who gave me a chance to participate in the Journal of Yangon University of Economics.

## **REFERENCES**

1. Box, G.E.P. and Jenkins, G.M. (1976) *Time Series Analysis for Forecasting and Control*. Holden-Day, San Francisco.
2. Terence, C. M. (2019). (n.d.). *Applied Time Series Analysis*. Loughborough University, United Kingdom: Candice Janco.
3. WHO. (2017). *WORLD MALARIA REPORT 2017*. World Health Organization.
4. WHO. (2010). *WORLD MALARIA REPORT 2010*. World Health Organization.
5. Wei, William W.S., W. (2006). *Time Series Analysis: Univariate and Multivariate Methods* (Second Edition), Pearson Education, Inc., U.S.A.