International Journal of Community Medicine and Public Health
Manikandan M et al. Int J Community Med Public Health. 2016 Mar;3(3):615-618
http://www.ijcmph.com
pISSN 2394-6032 | eISSN 2394-6040

DOI: http://dx.doi.org/10.18203/2394-6040.ijcmph20160620

Research Article

Forecasting the trend in cases of Ebola virus disease in West African countries using auto regressive integrated moving average models

Manikandan M.[1]*, Velavan A.[1], Zile Singh[1], Anil J. Purty[1], Joy Bazroy[1], Senthamarai Kannan[2]

[1]Department of Community Medicine, Pondicherry Institute of Medical Sciences, Puducherry, India
[2]Department of Statistics, Manonamaim Sundaranar University, Tirunelveli, India

Received: 15 October 2015
Revised: 12 February 2016
Accepted: 16 February 2016

***Correspondence:**
Dr. Manikandan M,
E-mail: manikandanmsu@gmail.com

ABSTRACT

Background: Ebola Virus Disease (EVD), formerly known as Ebola haemorrhagic fever, is a severe, often fatal illness in humans. The current outbreak in West African countries like Guinea, Liberia and Sierra Leone is one of the largest in the history. Hence, to forecast the trend in a number of cases of EVD reported in these Countries a univariate time series model was used.
Methods: We adopted an Auto Regressive Integrated Moving Average (ARIMA) models on the data collected between March 2014 to December 2014 and verified it using the data available between Jan 2015 to June 2015. The same has been used to predict the number of cases till December 2016 without any additional intervention.
Results: The results also showed an increasing trend in the actual and forecasted numbers of EVD cases. The appropriate ARIMA (1, 1, 0) model was selected based on Bayesian Information Criteria (BIC) values.
Conclusions: Hence, to prevent the disease from getting established as an endemic in these countries, additional interventions with an increase in the intensity of existing interventions and support of the international community along with WHO is essential to stop the epidemic.

Keywords: Univariate time series, Ebola, ARIMA, BIC, Forecasting

INTRODUCTION

The Ebola virus causes an acute, serious illness which is often fatal if untreated. The current outbreak in West Africa, (first cases notified in March 2014), is the largest and most complex Ebola outbreak since the Ebola virus was first discovered in 1976. The most severely affected countries include Guinea, Liberia and Sierra Leone. As of 14th June 2015, 27,341 cases and 11,184 deaths were reported from these three countries.[1]

The ARIMA methodology is also called as Box-Jenkins methodology.[2] The Box-Jenkins procedure is concerned with fitting a mixed ARIMA model to a given set of data. The main objective in the fitting ARIMA model is to identify the stochastic process of the time series and predict the future values accurately. These methods have also been useful in many types of situations which involves the building of models for discrete time series and dynamic systems. However, this method is not good for lead times or for seasonal series with a large random component.[3]

In the present study, this univariate time series model was used to forecast the trend in number of Ebola cases reported in the West African Countries.

METHODS

Auto regressive integrated moving average (ARIMA) model was fitted with the data available during March 2014 to December 2014 and verified using the data available between January 2015 to June 2015. In 1968, George Box and Gwilym Jenkins have extensively

studied ARIMA models and their names have frequently been used synonymously with general ARIMA process applied to time series analysis, forecasting and control.[4]

Autoregressive (AR) models can be effectively coupled with Moving Average (MA) models to form a general and useful class of time series models called Autoregressive Moving Average ARMA (p, q) models. However, they can only be used when the data are stationary. When a time series is studied based on the dependence relationship between the time-lagged values of the forecast variance and the past error terms, an Autoregressive Integrated Moving Average (ARIMA) model is more appropriate and it can be used when the time series is non-stationary. The general form of the ARIMA (p, d, q) model is,

$$Y_t=b_0+\varphi_1 Y_{t-1}+\ldots+\varphi_p Y_{t-p}+\theta_1 e_{t-1}+\ldots+\theta_q e_{t-q}+e_t,$$

Where Y_t and e_t are actual and random error at time period t, respectively; p, d and q represent the order of autoregressive part, the degree of differencing involved and the order of the moving average part. Random error, e_t independently and identically distributed (i.i.d) random variable with mean zero and variance (σ^2).[5-7]

Generally, ARIMA models consist of four stages:

1. Identification of the model. This involves selecting the most appropriate lags for the AR and MA parts, as well as determining if the variable requires first-differencing to induce stationary. The Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF) are used to identify the best model.
2. Estimation. This usually involves the use of a least squares estimation process.
3. Diagnostic testing, which usually is the test for autocorrelation. If this part fails, then the process returns back to the identification section and begins again, usually by the addition of extra variables.
4. Forecasting. The ARIMA models are particularly useful for forecasting due to the use of lagged variables.

An ARIMA model can be obtained at first by determining its parameters. The values of p and q can be determined from the patterns in the plotting of the values of ACF and PACF. The spikes falling above the time axis are used to estimate the value of p. The spikes falling below the time axis are used to estimate the value of q. For an AR (p) model, the spikes of ACF decay exponentially or there is a sine wave pattern and the spikes of PACF are close to zero beyond the time lag p. For a MA (q) model the spikes on the ACF end to zero beyond the time lag q whereas the spikes of PACF decay exponentially or there is a sine wave pattern.[8,9]

Once the model was identified and model parameter can be estimated, then the model is determined with a different set of parameters. It is basically checked with the assumption that the model about the random error e_t is satisfied. This can be identified as several diagnostic statistical measures and plots of the residuals can be used to examine the goodness of fit of different models to the historical data. The model selection can be made based on the values of certain criteria like Normalized Bayesian Information Criteria (BIC).

RESULTS

Data regarding the number of cases and deaths in the three severely affected counties (Guinea, Liberia and Sierra Leone) were collected from WHO situational reports.[1] This data were plotted on a graph to see the trend in individual country (Figure 1) and in total as well (Figure 2).

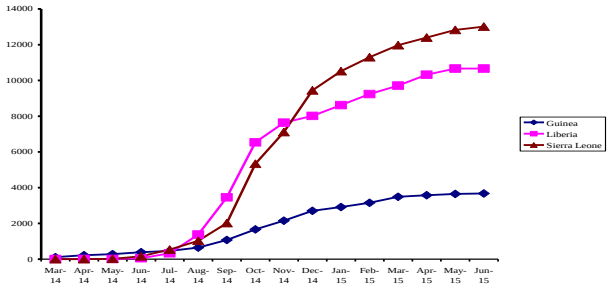

Figure 1: No. of cases in Guinea, Liberia, Sierra Leone.

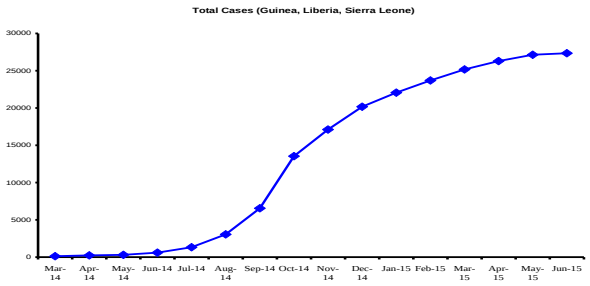

Figure 2: Combined total number of Ebola cases in three countries.

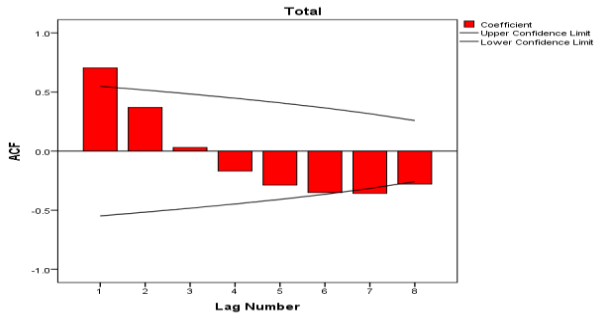

Figure 3: Auto correlation function.

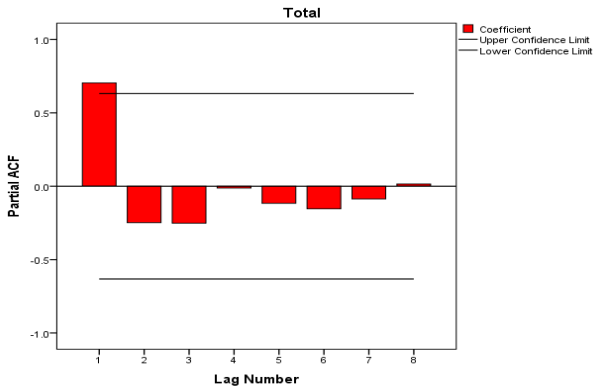

Figure 4: Partial auto correlation function.

Auto Correlation Function (Figure 3) and Partial Auto Correlation Function (Figure 4) showed that an irregular increasing pattern in the number of cases of EVD.

Table 1: Model selection.

ARIMA (p, d, q)	BIC	R-Squared
1,0,0	17.327	0.684
1,1,0	15.663	0.945
1,1,1	16.058	0.931
0,1,1	15.769	0.930
0,1,0	15.718	0.915

Henceforth, ARIMA models (p, d, q), apt for such scenario was applied. The best suitable model was selected based on minimal Bayesian Information Criteria (BIC) value. In this study, the least BIC value is 15.663 (Table 1) and the corresponding model is ARIMA (1, 1, 0) with goodness of fit (R^2=94%).

The model verification is done by checking the residuals of the model (Figure 4). This is done through examining the autocorrelation and correlation of the residuals of various orders.

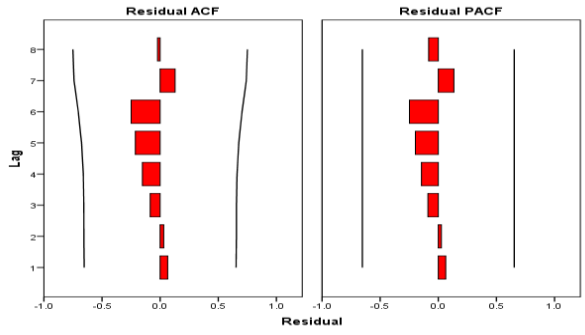

Figure 4: Residual ACF and PACF.

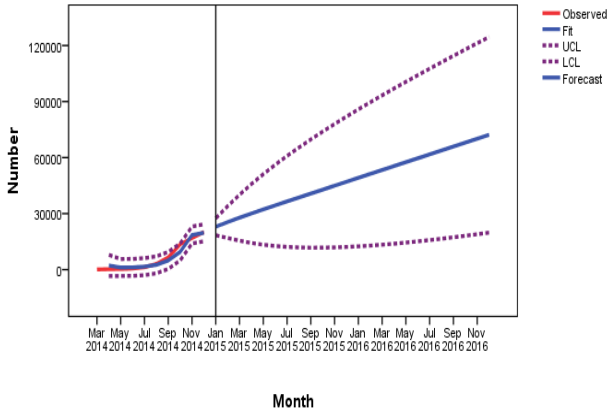

Figure 5: Actual and forecasted cases.

Table 2: Actual no. of cases and forecasted cases ARIMA (1, 1, 0).

Month	Actual no. of cases	Forecas-ted no. of cases	Lower confide-nce limit	Upper confide-nce limit
Mar-14	120	-		
Apr-14	234	2216	-3448	7879
May-14	309	1147	-3398	5693
Jun-14	599	1199	-3346	5744
Jul-14	1322	1617	-2928	6163
Aug-14	3052	2599	-1947	7144
Sep-14	6553	4930	384	9475
Oct-14	13540	9487	4942	14032
Nov-14	16899	18554	14009	23099
Dec-14	20171	19748	15203	24294
Jan-15	22057	22968	18423	27514
Feb-15	23694	25483	16920	34046
Mar-15	25178	27828	15496	40161
Apr-15	26298	30073	14296	45850
May-15	27145	32258	13353	51162
Jun-15	27341	34406	12654	56158
Jul-15	-	36534	12175	60893
Aug-15	-	38648	11885	65411
Sep-15	-	40755	11759	69752
Oct-15	-	42858	11772	73943
Nov-15	-	44957	11906	78009
Dec-15	-	47055	12144	81967
Jan-16	-	49153	12472	85833
Feb-16	-	51249	12880	89619
Mar-16	-	53345	13357	93334
Apr-16	-	55441	13897	96986
May-16	-	57537	14492	100582
Jun-16	-	59633	15138	104128
Jul-16	-	61729	15830	107628
Aug-16	-	63825	16563	111086
Sep-16	-	65920	17334	114506
Oct-16	-	68016	18141	117891
Nov-16	-	70112	18980	121244
Dec-16	-	72207	19849	124566

The forecasted values for January 2015 to June 2015 were in proximity to the actual values (Figure 5 and Table 2) and there by the validity of the model was ensured.

DISCUSSION

As early mentioned above ARIMA model consist of four steps, the first step was the identification of the model. The model identification done by ACF and PACF (Fig 3 & 4), it revealed increasing pattern in the number of cases of EVD. Model parameters were estimated using SPSS ver. 20.0. In the forthcoming months (July 2015 to December 2016), the model predicted a steadily increased. The outbreak, which started with 120 cases in March 2014, is expected to increase 600 folds in December 2016. This implies that the existing interventions in these countries will not be adequate to control the on-going epidemic even in another 18 months. Based on the forecasting model and assuming the current conditions of the Ebola virus outbreak in West Africa remains as in the past, it is expected a quick and exponential increase in the number of cases.

High numbers of cases were occurred in Sierra Leone compared to other two countries. Liberia and Sierra Leone have the highest average rate of cases per week. In order to control the outbreak of Ebola cases, it should be take actions and interventions to avoid more cases and deaths.

CONCLUSION

In this study, forecasting methods were applied to predict the number of Ebola virus disease cases in three Western African countries based on monthly cases reported by WHO. Model prediction was done using ARIMA models in which the appropriate model was identified using minimum BIC value. The trend in forecasted values Jan 2015 to December 2016 reveals that there is a steady increase in the number of cases of EVD which is of serious concern. Hence, to prevent the disease from getting established as an endemic in these countries, additional interventions with an increase in the intensity of existing interventions and support of the international community along with WHO is essential to stop the epidemic.

Funding: No funding sources
Conflict of interest: None declared
Ethical approval: Not required

REFERENCES

1. Ebola situation report. Geneva. World Health Organization. Available from http://apps.who.int/ebola/ebola-situation-reports. Accessed on 15th November 2015.
2. Bhatnagar S, Lal V, Gupta SD, Gupta OP. Forecasting Incidence of Dengue in Rajasthan, Using Time Series Analyses. Indian Journal of Public Health. 2012;56(4):281-5.
3. Widerström M, Omberg M, Ferm M, Pettersson AK, Eriksson MR, Eckerdal I, et al. Autoregressive Integrated Moving Average (ARIMA) Modelling of Time Series of Local Telephone Triage Data for Syndromic Surveillance. Online Journal of Public Health Informatics. 2014;6(1):e18.
4. Wangdi K, Asivanon PS, Silawan T, Lawpoolsri S, White NJ, Kaewkungwal J. Development of temporal modelling for forecasting and prediction of malaria infection susing time-series and ARIMAX analyses: A casestudy in endemic districts of Bhutan.Malaria Journal. 2010;9:251.
5. Prabakaran C, Sivapragasam C. Forecasting Areas and Production of Rice in India using ARIMA model, International Journal of Farm Sciences. 2014;4(1):99-106.
6. Zhang PG. Time series forecasting using a hybrid ARIMA and neural network models. Neurocomputing. 2003;50:159-75.
7. Promprou S, Jaroensutasinee M, Jaroensutasinee K. Forecasting Dengue Hemorrhagic Fever Cases in Southern Thailand using ARIMA Models, Dengue Bulletin. 2006;30:99-106.
8. Box G, Pierce DA. Distribution of residual autocorrelations in autoregressive- integrated Moving average time-series models. J Am Stat Ass. 1970;65:1509-26.
9. Box GEP, Jenkins GM. Time Series Analysis: Forecasting and Control, 2nd ed. San Francisco: Holden-Day. 1976:575.

Cite this article as: Manikandan M, Velavan A, Singh Z, Purty AJ, Bazroy J, Kannan S. Forecasting the trend in cases of Ebola virus disease in West African countries using auto regressive integrated moving average models. Int J Community Med Public Health 2016;3:615-8.