Employing seasonal autoregressive integrated moving average forecasting model to predict the number of dengue cases in Mumbai

Sujata Shirish Pol[1], Shekahr S Rajderkar[2], Seema S Bansode Gokhe[1], Kamlesh J Suthar[3]

[1]Department of Community Medicine, Lokmanya Tilak Municipal General Hospital and Medical College, Mumbai, Maharashtra, India, [2]Department of Community Medicine, Government Medical College, Miraj, Maharashtra, India, [3]Department of Mathematics, Institute of Chemical Technology, Mumbai, Maharashtra, India

Correspondence to: Sujata Shirish Pol, E-mail: sspol2004@gmail.com

Received: October 03, 2019; **Accepted:** November 04, 2019

ABSTRACT

Background: Dengue control suffers challenges such as the absence of specific treatment and lack of vaccine. Forecasting of dengue would facilitate allocation of resources that are needed for such activities. The seasonal autoregressive integrated moving average (SARIMA) is the mathematical model, which provides the estimated monthly figures for the given period. **Objectives:** The objectives of the study were to select the best prediction model for dengue fever (DF) by time series data over the past 13 years in the Mumbai city and to forecast monthly dengue incidence for 2019. **Materials and Methods:** Retrospective study design was employed at epidemiology cell, Mumbai – Integrated Disease Surveillance Project (IDSP). The reported DF/dengue hemorrhagic fever cases during January 2006–December 2018 were mobilized from epidemiology cell of the city. Data were recorded on Excel sheet. The SARIMA model was applied to the data with R software. **Results:** The cases showed a form with a seasonal difference. SARIMA (1, 1, 2)(0, 1, 1) model had the highest Akaike information criteria (AIC) of 1433.18 and mean absolute percentage error of 43.75 and performed to be the best model. Adequacy of the model was established through the Ljung–Box test, which showed no substantial correlation among residuals at different lag times. Seasonal peak is estimated in the month of September of 2019 with 323 (95% confidence interval: [196.56, 449.92]) cases followed by 222 (95% confidence interval: [95.58, 348.93]) in October. **Conclusions:** The function of the SARIMA model may be useful for a forecast of cases and impending outbreaks of dengue.

KEY WORDS: Dengue; Forecasting; Prediction Model; Seasonal Autoregressive Integrated Moving Averages Model

INTRODUCTION

Dengue control suffers challenges such as the absence of specific treatment and lack of vaccine. Forecasting of dengue would facilitate allocation of resources that are needed for such activities. The seasonal autoregressive integrated moving average (SARIMA) is the mathematical model, which provides the estimated monthly figures for the given period.[1] Dengue is a vector-borne disease that causes a substantial public health burden within its expanding range.[2] Accurate prediction of dengue outbreaks may lead to public health interventions that mitigate the effect of the disease. Predicting infectious disease outbreaks are a challenging task.[3] Knowledge of the geographical distribution and burden of dengue is essential for understanding its contribution to global morbidity and mortality burdens, in determining how to allocate optimally the limited resources available for dengue control and in evaluating the impact of such activities internationally. In addition, estimates of both apparent and inapparent infection distributions form a key requirement for assessing clinical surveillance and for scoping reliably future vaccine demand and delivery strategies.[2]

Access this article online	
Website: http://www.ijmsph.com	Quick Response code
DOI: 10.5455/ijmsph.2020.1028504112019	

Accurate prediction of dengue incidence levels weeks in advance of an outbreak may reduce the morbidity and mortality associated with this neglected disease. Several researchers have developed different models to predict high and low dengue incidence to provide timely forewarnings in different parts of the world.[4-17]

Few of those researchers tried climate-based models to predict dengue cases with the help of meteorological data.[6-9,14,15] The feasibility and utility of different models that have been used for forecasting dengue incidence were compared with methodical literature review. Various recent national and international studies indicated that the Seasonal Autoregressive Integrated Moving Average (SARIMA) model is a useful tool for monitoring and forecasting the dengue incidence. In India, during monsoon there is a threat of outbreaks of more than one communicable disease such as dengue, malaria, and chikungunya. At such times, forecasting would certainly enable the public health authorities for suitable allocation of resources for controlling communicable diseases. Thus, the present study was planned to choose the best SARIMA prediction model on the incidence rate of dengue fever (DF) in the city of Mumbai and find out the seasonal pattern of dengue for the next year (2019) with selected model.

MATERIALS AND METHODS

Study Setting

The present study was conducted at Mumbai district, Maharashtra, India. There is an epidemiology cell also known as integrated disease surveillance project (IDSP). There are 24 wards in the city and Medical Officer of Health is the incharge of health for that particular ward. There is on-going surveillance for all monsoon-related diseases (MRDs) such as dengue, malaria, leptospirosis, and H1N1. All the health care centers including tertiary hospitals, peripheral hospitals, and dispensaries submit the details of MRDs (line list) to IDSP on daily basis. The data are collected and compiled at IDSP, Mumbai.

Study Design

The retrospective study design was planned at epidemiology cell, Mumbai – IDSP. Records were checked for details of all dengue cases reported to IDSP during 2006–2018. Official permissions were taken for data collection from records. Ethical approval was taken from IEC of the institute. (Ref. no- IEC/343118; dated 10/03/18).

Data Analysis

The data on dengue cases at Mumbai during 2006–2018 (13 years) were entered in Excel sheet. Tables and graphs were prepared wherever appropriate. The SARIMA model was applied to the data with R software version 3.5.2. A total of three different mathematical models were calculated for dengue forecasting at Mumbai city. The model determined the mean absolute percentage error (MAPE) and the Akaike information criterion (AIC). Adequacy of the model was recognized through Ljung–Box test. The statistical model was built based on the data during 2006–2017 and the data for 2018 were used for model validation. When the best model was known, the forecast for monthly dengue rate of the year 2019 was determined.

RESULTS

The time series plot of the reported dengue cases demonstrated seasonal fluctuations and therefore regarded non-stationary. Large autocorrelations were recorded for lags 0 and 1 with values 1.0 and 0.686, respectively. The sharp fall in autocorrelation values after lag 1 specified no evidence of a long-term trend [Figure 1]. In juxtapose, large autocorrelation values were conveyed at annual lags (and its multiplies), which indicated the need to comprise a 12-month difference term in the models (S = 12, D = 1).

The ACF and PACF plots of the differenced series provided further provision for these inferences [Figures 1 and 2].

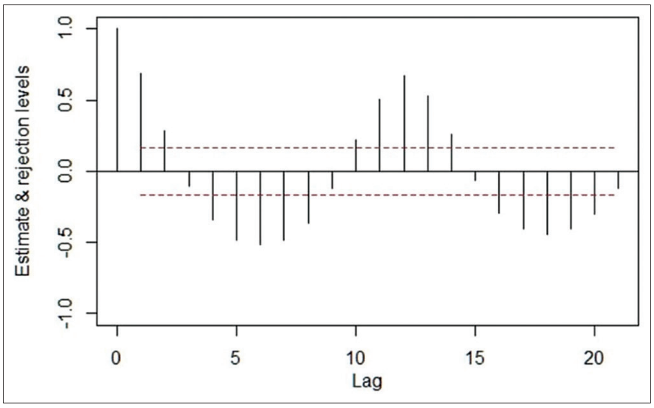

Figure 1: Autocorrelation function plot (without any differences)

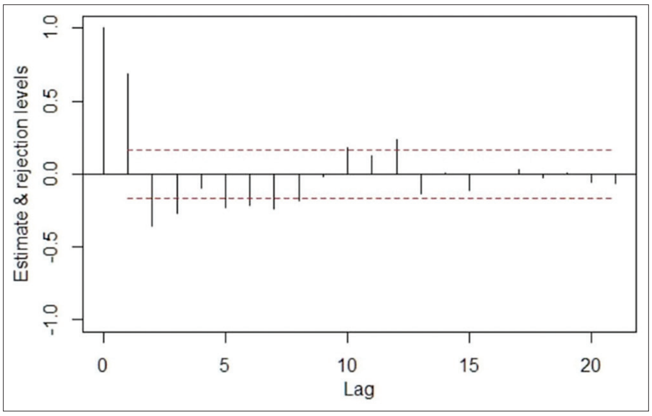

Figure 2: Partial autocorrelation function plot (without any differences)

Therefore, a SARIMA (p, 1, q)(P, 1, Q) was selected as the basic structure of the applicant model.

In the process of developing the prediction model, a total of three models were applied to the data. Of three models, one was chosen as best based on the criteria such as AIC, MAPE, and *P* value with Ljung–Box test. Among the statistical models, SARIMA (1,1,2)(0,1,1) was the one with lowermost normalized AIC of 1433.18 and a MAPE of 43.75. To decide the best SARIMA model, the (1,1,2)(0,1,1) model that showed $P = 0.5$ and lowest AIC value and highest MAPE value was chosen [Table 1]. We anticipated the parameters of the SARIMA model by full likelihood. The model parameters were significant ($P < 0.001$). Ljung–Box (Q statistics 1.855 $\times 10^{`6}$ and $P = 0.9989$) implied that there were no significant autocorrelations between residual lags at diverse times and residuals were white noise.

Plotting residual ACF and histogram of this particular model indicated its normality [Figures 3 and 4].

Having tested its validity, the prediction model [SARIMA (1,1,2)(0,1,1)] was developed to forecast the incidence of dengue cases in Mumbai for the year 2019. Figure 5 shows the trend of dengue cases for the past 13 years. The graph shows the peak of dengue cases in the year 2011 and 2017. Figure 6 illustrates the predicted figures of dengue cases for the year 2019. A total of 1037 cases of dengue was predicted for the coming year. As per model, seasonal peak is estimated in the month of September of 2019 with 323 (95% confidence interval: [196.56, 449.92]) cases followed by 222 (95% confidence interval: [95.58, 348.93]) in October.

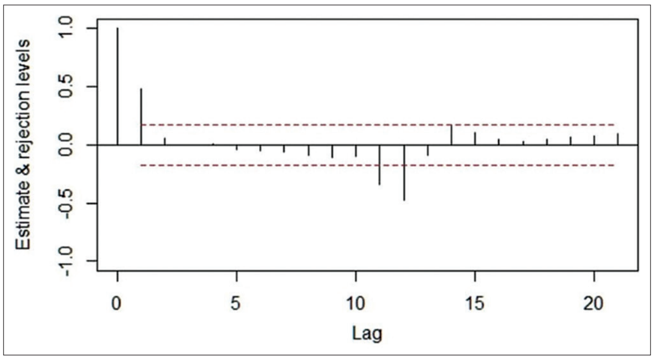

Figure 3: Autocorrelation function plot (with one seasonal difference [D = 1] and one non seasonal difference [d = 1])

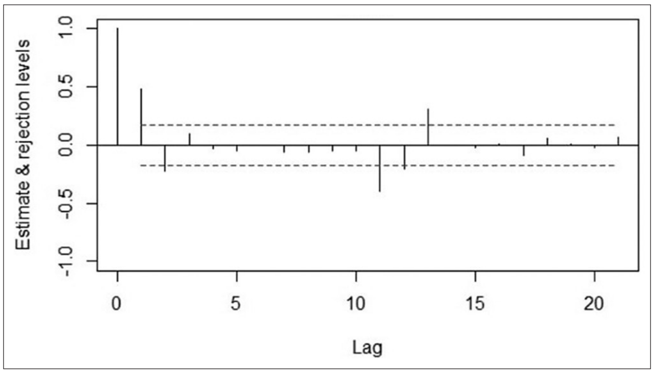

Figure 4: Partial autocorrection function plot (with one seasonal difference [D = 1] and one non seasonal difference [d = 1])

DISCUSSION

Dengue is the most common arbovirus infection globally, but its burden is poorly quantified.[18] Many popular dengue forecasting techniques have been used by several researchers to extrapolate dengue incidence rates, including the K-H model, support vector machines, and artificial neural networks. The time series analysis methodology, particularly ARIMA and SARIMA, has been increasingly applied to the field of epidemiological research for DF, dengue hemorrhagic fever (DHF), and other infectious diseases.[7] However, such analysis has been minimal undertaken in an Indian situation. A few years back Bhatnagar *et al*. attempted dengue forecasting at Rajasthan.[11] In the present study, SARIMA (1,1,2)(0,1,1) among all models was the fittest predictive model, which showed normalized AIC and MAPE values. Further, it revealed a typical seasonal variation of dengue cases with peak in September month. Data analysis estimated seasonal peak of dengue in the month of September for the year 2019.

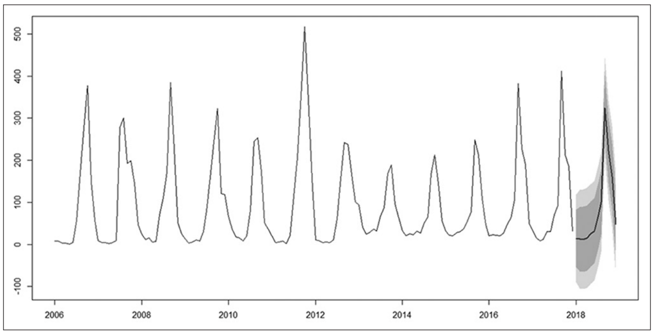

Figure 5: Forecast from autoregressive integrated moving average (1,1,2)(0,1,1)[12]

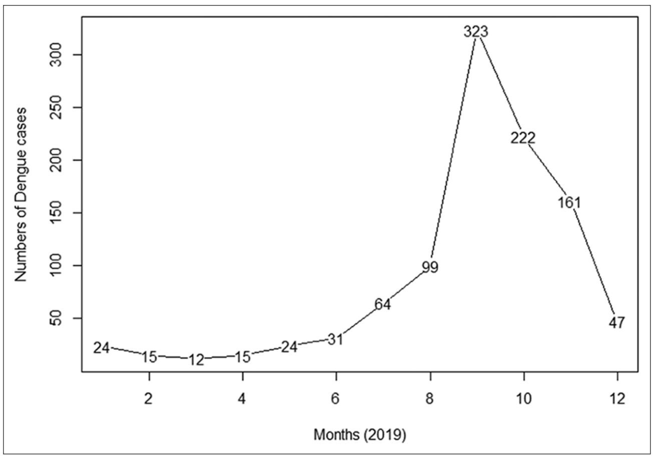

Figure 6: Forecast incident cases of dengue using seasonal autoregressive integrated moving average (1,1,2)(0,1,1)[12]

Table 1: Values of different SARIMA models

SARIMA model	Box–Ljung (*P*-value)	X-squared-error	AIC	ME	RMSE	MAE	MPE	Mean absolute percentage error
(1,1,2)(0,1,1)	0.99	1.85E-06	1433.18	−1.291	50.32	29.41	−16.21	43.75
(1,0,0)(0,1,1)	0.62	0.23	1434.25	−0.668	50.27	29.06	−11.97	39.29
(0,1,2)(0,1,1)	0.37	0.79	1434.95	−1.546	51.03	30.61	−18.25	46.84

SARIMA: Seasonal autoregressive integrated moving average, AIC: Akaike information criteria

SARIMA (1, 1, 2)(0, 1, 1) model with the highest AIC and MAPE seemed to be the best model. A similar study in Rajasthan found SARIMA $(0,0,1)(0,1,1)_{12}$ model was best suitable for dengue occurrence data.[11] Whereas in Thailand, SARIMA (2,1,0)(0,1,1) and (1,0,1)(0,1,1) were found as best fit.[5,19] Mertinez *et al.* and Luz *et al.* reported SARIMA (2,0,0)(1,0,0) and (2,1,1)(1,1,1) as the best fit for forecasting the dengue incidence in Brazil.[1,20] A Cruz *et al.* found (1,0,1)(0,1,1) best suitable model in Philippines[17] and Choudhary *et al.* deemed (1,0,0)(1,1,1) as best model at Dhaka.[21] The model was also tried to forecast the DHF in Indonesia by Siregar *et al*. who found (1,0,0)(0,1,1) best fit.[4]

Seasonal peak of dengue in the city was estimated in September while Siregar *et al*. and Bhatnagar *et al*. predicted it in October.[4,11] This is in contrast with the across study in Brazil that showed epidemics of dengue between January and May.[20] Wongkoon found peak of cases between May and June and Choudhary *et al*. found in month of July.[5,21] In India, July-August is the months of heavy rain. September, there is waning of rain and eventually many small water pools, which do not get washed away as it happens with heavy rains in July-August. Thus, many small pools may act as breeding sites and giving rise of dengue outbreak. One recent research on dengue trends in Mumbai also highlighted September as the month of peak in dengue cases.[22] This study denotes another effort after Bhatnagar *et al*. at Rajasthan, who developed an epidemic forecasting model using SARIMA model for predicting dengue transmission in India. The data analysis implies the lower dengue incidence in year 2019 as compared to 2018, which explains insignificant risk of outbreak if there are no environmental challenges in coming year. The outcomes of this study indicate that the key elements of the DF transmission include autoregression, moving average, and seasonal moving average. These variables may be used to support in anticipating outbreaks of dengue incidence rate in India. The cases of dengue typically vary throughout the year and assume a regular pattern, normally in association with changes of temperature and rainfall as described as seasonal by the WHO.[5] Our results specify that the forecast values could follow the upturn and downturn of the observed data equitably well.

The SARIMA modeling is a useful tool for inferring and employing surveillance data in disease control and prevention. Once a suitable model has been obtained, it can be used to predict the numbers of cases for a given time intervals. The weakness of the SARIMA method is that it does not reflect other climatic variables such as rainfall, humidity, and temperature that are correlated with the dependent variable. Such obstacles may subsequently downgrade the projecting influence of the outbreak. Thus, predictions may not be credible during the actual epidemic years. The model considers only reported dengue cases (hospital admissions); the real number of cases in a community may be missed out.

Recommendations

This early warning model structure may be useful to public health facilities. If done before the peak dengue season each year, it will probably control extensive dengue epidemic. Incorporating SARIMA model in the routine surveillance system at Integrated Disease Surveillance System of Mumbai city would enable suitable allocation of resources for other MRDs as well.

CONCLUSIONS

The results of the study suggested that the SARIMA (1,1,2)(0,1,1) was found as best-suited model. The time series developed with the selected model indicated the dengue epidemic in the year 2011 that was true as per available actual data. The second epidemic was indicated in the year 2017; however, as per actual data, the cases were close to normal expected number. The selected model in the present study was able to express the number of dengue cases in a succeeding year with relative precision.

REFERENCES

1. Martinez EZ. Article/Artigo A SARIMA forecasting model to predict the number of cases of dengue in campinas, state of São Paulo, Brazil Um modelo SARIMA para predição do número de casos de dengue em campinas. Rev Soc Bras Med Trop 2011;44:436-40.
2. Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, *et al.* The global distribution and burden of dengue. Nature 2013;496:504-7.
3. Buczak AL, Baugher B, Babin SM, Ramac-Thomas LC, Guven E, Elbert Y, *et al.* Prediction of high incidence of dengue in the philippines. PLoS Negl Trop Dis 2014;8:e2771.
4. Siregar, Fazidah A, Makmur Tri SS. Forecasting Dengue Hemorrhagic fever Cases Using ARIMA Model : A Case Study in Asahan District. In: IOP Conference Series: Material Science and Engineering; 2018.

5. Wongkoon S, Jaroensutasinee M, Jaroensutasinee K. Assessing the temporal modelling for prediction of dengue infection in Northern and North-Eastern, Thailand. Trop Biomed 2012;29:339-48.
6. Gharbi M, Quenel P, Gustave J, Cassadou S, La Ruche G, Girdary L, *et al.* Time series analysis of dengue incidence in guadeloupe, french west indies: Forecasting models using climate variables as predictors. BMC Infect Dis 2011;11:166.
7. Siriyasatien P, Phumee A, Ongruk P, Jampachaisri K, Kesorn K. Analysis of significant factors for dengue fever incidence prediction. BMC Bioinformatics 2016;17:166.
8. Adde A, Roucou P, Mangeas M, Ardillon V, Desenclos JC, Rousset D, *et al.* Predicting dengue fever outbreaks in french guiana using climate indicators. PLoS Negl Trop Dis 2016;10:e0004681.
9. Ramachandran VG, Roy P, Das S, Mogha NS, Bansal AK. Empirical model for estimating dengue incidence using temperature, rainfall, and relative humidity: A 19-year retrospective analysis in east delhi. Epidemiol Health 2016;38:e2016052.
10. Eastin MD, Delmelle E, Casas I, Wexler J, Self C. Intra- and interseasonal autoregressive prediction of dengue outbreaks using local weather and regional climate for a tropical environment in colombia. Am J Trop Med Hyg 2014;91:598-610.
11. Bhatnagar S, Lal V, Gupta SD, Gupta OP. Forecasting incidence of dengue in Rajasthan, using time series analyses. Indian J Public Health 2012;56:281-5.
12. Sitepu MS, Kaewkungwal J, Luplerdlop N, Soonthornworasiri N, Silawan T, Poungsombat S, *et al.* Temporal patterns and a disease forecasting model of dengue hemorrhagic fever in Jakarta based on 10 years of surveillance data. Southeast Asian J Trop Med Public Health 2013;44:206-17.
13. Wiwanitkit V. SARIMA for predicting the cases numbers of dengue. Rev Soc Bras Med Trop 2012;45:140.
14. Hii YL, Rocklöv J, Wall S, Ng LC, Tang CS, Ng N, *et al.* Optimal lead time for dengue forecast. PLoS Negl Trop Dis 2012;6:e1848.
15. Descloux E, Mangeas M, Menkes CE, Lengaigne M, Leroy A, Tehei T, *et al.* Climate-based models for understanding and forecasting dengue epidemics. PLoS Negl Trop Dis 2012;6:e1470.
16. Martinez EZ, da Silva EA. Predicting the number of cases of dengue infection in Ribeirao Preto, Sao Paulo State, Brazil, using a SARIMA model. Cad Saude Publica 2011;27:1809-18.
17. Cruz AC, Lubrica JA, Punzalan BV, Martin MC. Forecasting dengue incidence in the national capital region, Philippines : Using time series analysis with climate variables as predictors. Acta Manila Ser A 2012;60:19-26.
18. Stanaway JD, Shepard DS, Undurraga EA, Halasa YA, Coffeng LE, Brady OJ, *et al.* The global burden of dengue: An analysis from the global burden of disease study 2013. Lancet Infect Dis 2016;16:712-23.
19. Silawan T, Singhasivanon P, Kaewkungwal J, Nimmanitya S, Suwonkerd W. Temporal patterns and forecast of dengue infection in northeastern thailand. Southeast Asian J Trop Med Public Health 2008;39:90-8.
20. Luz PM, Mendes BV, Codeço CT, Struchiner CJ, Galvani AP. Time series analysis of dengue incidence in Rio de Janeiro, Brazil. Am J Trop Med Hyg 2008;79:933-9.
21. Choudhary, Zamil MA, Shahera IM. Forecasting dengue incidence in dhaka, Bangladesh : A time series analysis. Dengue Bull 2008;32:29-37.
22. Pol SS, Rajderkar SS, Gokhe SS. Trends of dengue cases reported at tertiary care hospital of metropolitan city of Maharashtra : A record based study. Natl J Community Med 2017;8:411-5.

How to cite this article: Pol SS, Rajderkar SS, Gokhe SSB, Suthar KJ. Employing seasonal autoregressive integrated moving average forecasting model to predict the number of dengue cases in Mumbai. Int J Med Sci Public Health 2020;9(1):51-55.

Source of Support: Nil, **Conflicts of Interest:** None declared.