Forecasting dengue incidence in Dhaka, Bangladesh: A time series analysis

M.A.H. Zamil Choudhury a,b #, Shahera Banu b, M. Amirul Islam c,d

a International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDR,B), Mohakhali, Dhaka-1212, Bangladesh
b Department of Microbiology and Hygiene, Bangladesh Agricultural University, Mymensingh 2202, Bangladesh
c Department of Agricultural Statistics, Bangladesh Agricultural University, Mymensingh 2202, Bangladesh
d Southampton Statistical Sciences Research Institute, University of Southampton, UK

Abstract

This article attempts to model the monthly number of dengue fever (DF) cases in Dhaka, Bangladesh, and forecast the dengue incidence using time series analysis. Seasonal Autoregressive Integrated Moving Average (SARIMA) models have been developed on the monthly data collected from January 2000 to October 2007 and validated using the data from September 2006 to October 2007. The results showed that the predicted values were consistent with the upturns and downturns of the observed series. The SARIMA (1,0,0)(1,1,1)_{12} model has been found as the most suitable model with least Normalized Bayesian Information Criteria (BIC) of 11.918 and Mean Absolute Percent Error (MAPE) of 595.346. The model was further validated by Ljung-Box test (Q18 = 15.266 and p > .10) with no significant autocorrelation between residuals at different lag times. Finally, a forecast for the period November 2007 to December 2008 was made, which showed a peak in the incidence of DF during July 2008, with estimated cases as 689.

Keywords: Dengue; Time series analysis; SARIMA; Disease prediction; Dhaka, Bangladesh.

Introduction

Dengue is one of the most important emerging viral diseases of major public health concern in Bangladesh. The disease is transmitted through the bite of the *Aedes aegypti* and *Ae. albopictus* mosquitoes[1]. It causes a broad spectrum of clinical manifestations in humans ranging from the acute febrile illness, dengue fever (DF), to the life-threatening dengue haemorrhagic fever/dengue shock syndrome (DHF/DSS)[2].

Dengue was first reported as “Dacca fever” in Bangladesh in 1964 by Aziz and his colleagues[3]. Subsequent reports suggested that...
Dengue fever may have been occurring sporadically in Bangladesh from 1964 to 1999\cite{3-9}. The first epidemic of dengue was reported in the capital city, Dhaka in the year 2000\cite{10,11}. Since then the disease has shown an annual occurrence in all major cities of the country. During January 2000–December 2007, Bangladesh recorded a total of 22,245 cases and 233 deaths (1.04%). Of these, Dhaka accounted for 20,115 cases and 181 deaths (0.9%).

In the absence of a vaccine and specific treatment available for dengue, vector control remains the only option. Early warning about the disease based on forecasting, therefore, becomes crucial for the prevention and control of dengue in Bangladesh. The time series analyses methodology has been increasingly used in the field of epidemiological research on infectious diseases, particularly in the assessment of health services\cite{12-16}. In health science research, Autoregressive Integrated Moving Average (ARIMA) models\cite{12-18} as well as Seasonal Autoregressive Integrated Moving Average (SARIMA)\cite{19,20} models are useful tools for analysing time series data containing ordinary or seasonal trends to develop a predictive forecasting model. There have been efforts in forecasting dengue incidence in different parts of the world using both ARIMA\cite{21,22} and SARIMA\cite{19} modelling. This study is aimed at developing univariate time series models to forecast the monthly dengue incidence in Dhaka based on reported monthly cases available from 2000–2007. This forecasting offers the potential for improved and consistent planning of public health interventions.

Materials and methods

Study area

Dhaka is the capital and principal city of Bangladesh located at 23° 42’ 0” N, 90° 22’ 30” E, covering an area of 815.85 km² (315 sq miles). According to the World Gazetteer (2006), the population in the Dhaka region was 11 million and the density was 14,608/km² (37,834.5/sq mile) making it the largest city in Bangladesh and the eleventh most populous city in the world\cite{23}. The Dhaka region was chosen as the study area because of its relatively high incidence of DF between 2000 and 2007 (average annual incidence: 2515.75 cases).

Data collection

We obtained computerized data sets of notifications of monthly DF cases in the Dhaka region for the period 1 January 2000 through 31 October 2007 from the Directorate-General of Health Services, Mohakhali, Dhaka-1212\cite{24}. It may be noted that the Directorate-General of Health Services collects information on dengue cases separately as DF, DHF and DSS but clubs this data as data for DF only.

Data analysis

A SARIMA (p,d,q)(P,D,Q)s model\cite{25} was fitted, where p is the order of autoregression, d is the order of integration, q is the order of moving average, P is the order of seasonal autoregression, D is the order of seasonal integration, Q is the order of seasonal moving average and s is the length of seasonal period. The analyses were performed using SPSS 17 software. The stationarity of the series was made by means of seasonal and non-seasonal differencing\cite{25}. Then the order of autoregression and moving average were identified using autocorrelation function (ACF) and partial autocorrelation function (PACF) of the differenced series. A model was fitted with a training set of data from January 2000 to October 2007 and the fitted model was used to predict values for a validation period (from
September 2006 to October 2007) to evaluate the time series model. Most suitable models were selected on the basis of their ability of reliable prediction. Two measures, namely, Normalised Bayesian Information Criteria (BIC)\(^{[26]}\) and Mean Absolute Percent Error (MAPE)\(^{[25]}\), were used. Lower values of Normalised BIC and MAPE were preferable. Furthermore, Ljung-Box test (portmanteau test) was performed to test if the residual ACF at different lag times was significantly different from zero, where not being different from zero was expected\(^{[27]}\). After the best model was identified, forecast for future values from November 2007 to December 2008 was made.

Results and discussion

The observed series of DF (January 2000 to October 2007) shows that the series is non-stationary and there are seasonal fluctuations in the dataset (Figure 1). ACF and PACF of one seasonal differenced series (Figure 2a, Figure 2b) as well as of one seasonal with one non-seasonal differenced series (Figure 2c, Figure 2d) instigated to explore a set of models based on the training set of data (January 2000 to October 2007), which are listed in Table 1. Among these models, SARIMA\((1,0,0)(1,1,1)_{12}\) has both lowest normalised BIC (11.918) and MAPE (595.346) values and appeared to be the best model. Moreover, the Ljung-Box test suggested that the ACF of residuals for the model at different lag times was not significantly different from zero \(Q_{18}=15.266\) and \(p>.10\). All the coefficients of SARIMA \((1,0,0)(1,1,1)_{12}\) model were significant (Table 2). The model has been used to predict values from September 2006 to October 2007 (Figures 3 and 4) for validation. It appeared that the predicted values could follow the upturn and downturn of the observed series reasonably well. Finally, Figure 5 represents the forecast values for the period from November 2007 to December 2008, which indicates a seasonal pick in July 2008, with the estimated number of patients as 689 and a sharp decrease from September 2008. The predicted values as well as the forecast values show some negative values, which is a common case with a series with too many zeros as observed values in the series.

**Figure 1:** Observed dengue fever from January 2000 to October 2007
Forecasting dengue incidence in Dhaka, Bangladesh: A time series analysis

Table 1: MAPE and normalized BIC of Time Series Models

<table>
<thead>
<tr>
<th>Models</th>
<th>MAPE</th>
<th>Normalized BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>SARIMA(2,1,1)(1,1,0)_{12}</td>
<td>1026.050</td>
<td>12.072</td>
</tr>
<tr>
<td>SARIMA(2,1,0)(1,1,0)_{12}</td>
<td>766.310</td>
<td>12.215</td>
</tr>
<tr>
<td>SARIMA(1,1,1)(1,1,0)_{12}</td>
<td>945.640</td>
<td>12.052</td>
</tr>
<tr>
<td>SARIMA(0,1,0)(1,1,0)_{12}</td>
<td>805.376</td>
<td>12.236</td>
</tr>
<tr>
<td>SARIMA(1,1,0)(1,1,0)_{12}</td>
<td>791.408</td>
<td>12.269</td>
</tr>
<tr>
<td>SARIMA(1,1,1)(1,1,1)_{12}</td>
<td>947.663</td>
<td>12.059</td>
</tr>
<tr>
<td>SARIMA(1,1,1)(2,1,0)_{12}</td>
<td>975.253</td>
<td>12.071</td>
</tr>
<tr>
<td>SARIMA(1,0,1)(1,1,1)_{12}</td>
<td>600.730</td>
<td>11.952</td>
</tr>
<tr>
<td>SARIMA(1,0,1)(1,1,0)_{12}</td>
<td>717.614</td>
<td>11.933</td>
</tr>
<tr>
<td>SARIMA(1,0,0)(1,1,1)_{12}</td>
<td>595.346</td>
<td>11.918</td>
</tr>
</tbody>
</table>

Table 2: Model parameters of SARIMA (1,0,0)(1,1,1)_{12}

<table>
<thead>
<tr>
<th>Variable</th>
<th>β</th>
<th>SE</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR (Lag 1)</td>
<td>0.385</td>
<td>0.106</td>
<td>0.000</td>
</tr>
<tr>
<td>AR, Seasonal (Lag 1)</td>
<td>-0.587</td>
<td>0.109</td>
<td>0.000</td>
</tr>
<tr>
<td>Seasonal Difference</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MA, Seasonal (Lag 1)</td>
<td>0.405</td>
<td>0.151</td>
<td>0.009</td>
</tr>
</tbody>
</table>

Figure 2a: ACF of transformed series with seasonal difference (1, period 12)
**Figure 2b:** PACF of transformed series with seasonal difference (1, period 12)

![PACF of transformed series with seasonal difference (1, period 12)](image)

**Figure 2c:** ACF of transformed series with non-seasonal difference (1) and seasonal difference (1, period 12)

![ACF of transformed series with non-seasonal difference (1) and seasonal difference (1, period 12)](image)
Figure 2d: PACF of transformed series with non-seasonal difference (1) and seasonal difference (1, period 12)

![Figure 2d: PACF of transformed series with non-seasonal difference (1) and seasonal difference (1, period 12)](image)

Figure 3: Dengue incidence: Observed values and predicted values of SARIMA (1,0,0)(1,1,1)_12

![Figure 3: Dengue incidence: Observed values and predicted values of SARIMA (1,0,0)(1,1,1)_12](image)
Figure 4: Dengue incidence – Observed values and predicted values of SARIMA (1,0,0)(1,1,1)_{12} for the period September 2006 to October 2007.

Figure 5: Forecast of dengue incidence from November 2007 to December 2008 by SARIMA (1,0,0)(1,1,1)_{12}.
Conclusion

The incidence of dengue fever every year in Bangladesh, especially in Dhaka city, is a constant threat to the population and a recurring problem for the health authorities. Furthermore, all environmental conditions that can trigger an outbreak are more or less present in the country. Forecasting a dengue outbreak can help the authorities to take effective measures to handle any unexpected situation. Such an effort is cost-effective considering the financial constraints of the health sector. The present study is the first of its kind in Bangladesh. The SARIMA results revealed that the number of dengue patients in 2008 will have a seasonal pick with the highest value as 689 in July, which is concordant with our previous experience. SARIMA models are well-practised tools in epidemiological research which may offer further accuracy in prediction if some relevant variables, for example, temperature, humidity, rainfall, are considered during the modelling process. Efforts should be made in the future to use such additional information, which was not possible in the current study due to lack of coordination between different sources as well as dissimilarity of area of coverage by different authorities. Separate modelling approaches for DF, DHF and DSS would provide better information to policy-makers and planners. Relevant data should be made available in a timely manner, possibly from one service point, with proper coordination between different data sources.

Acknowledgements

We thank the Disease Control Directorate, Directorate-General of Health Services, Dhaka, Bangladesh, for providing the data of notified dengue cases between 2000 to 2007. We also thank Professor Dr Md Alimul Islam for his suggestions and inspiration. Finally, we would like to thank the three anonymous reviewers for their valuable comments on the previous version of the manuscript.

References


Forecasting dengue incidence in Dhaka, Bangladesh: A time series analysis


