Asian Journal of Water, Environment and Pollution, Vol. 9, No. 4 (2012), pp. 63–69.

Time Series Analysis Model for Particulate Matter of Air Pollution Data in Dhaka City

Md. Arafat Rahman* and M. Amir Hossain[1]

ISRT, University of Dhaka, Dhaka 1000, Bangladesh
✉ arahman4@isrt.ac.bd

Received May 7, 2012; revised and accepted September 14, 2012

Abstract: Time series analysis and forecasting has become an important tool in many applications in the field of air pollution and environmental management. ARIMA (Autoregressive Integrated Moving Average) models form an important part of the Box-Jenkins approach to time series data modelling. In this study Box-Jenkins method was used to construct ARIMA model for monthly particulate matter of air pollution data with a total of 108 readings from Dhaka meteorological station for the period 2002-2010. An attempt has been made to construct an ARIMA (0, 0, 2) $(2, 1, 0)_{12}$ model in a systematic and scientific manner. Based on the fitted ARIMA model, monthly particulate matter of air pollution for further two years has been predicted. It will help to make better decision for controlling air pollution in Dhaka city.

Key words: Time series analysis, particulate matter, air pollution, forecasting, ARIMA model.

Introduction

An ordered sequence of values of a variable observed at equally space time intervals is referred to as time series (Makridakis et al., 1998). There are several types of data analysis method available for time series which are appropriate for different purposes. This study particularly deals with ARIMA (Autoregressive Integrated Moving Average) described by Box-Jenkins. In time series analysis, Box-Jenkins methodology, named after statistician George Box and Gwilym Jenkins, applies ARIMA model to find best fit of a time series data in order to make forecast.

Particle pollution, called particulate matter or PM, is a combination of tiny specks of soot, dust, and aerosols that are suspended in the air we breathe. "A mixture of mixtures" is how the U.S. Environmental Protection Agency (USEPA) describes them (e.g. http://www.epa.gov/ttn/naaqs/standards/pm/s_pm_cr_cd.html). PM can be solids, like dust, ash, or soot. PM can also be sulfate, nitrate, or carbonaceous aerosols. They are measured in microns. The largest of concern are 10 microns in diameter or smaller (PM_{10}). Burning fuel is a major source of the smallest types of particle pollution. The sources may be woodstoves, diesel trucks and buses, and coal-fired power plants. Larger particles also come from other sources including construction, agricultural practices, and mining. Short-term increases (over hours to days) in particle pollution have been linked to: death from respiratory and cardiovascular causes including strokes (Dominici et al., 2002; Hong et al., 2002; Franklin et al., 2007; D'Ippoliti et al., 2003; Ghio et al., 2000), increased numbers of heart attacks, especially among the elderly and in people with heart conditions (Dominci et al., 2006), inflammation of lung tissue in young, healthy adults (Tsai et al., 2003) hospitalization due to aggravated asthma attacks among children (Lin et al., 2002; Norris et al., 1999; Tolbert et al., 2000; Slaughter et al., 2003). Time series analysis techniques were used to investigate the association

*Corresponding Author

between outdoor air pollution and mortality effects by cause, age and socioeconomic status (Nelson and Tony et al., 2000).

Methodology

The basis of the Box-Jenkins approach for modelling time series data consists of three phases. Identification is the first step in this process which is broadly classified into two categories: data preparation and model selection. In data preparation part data is transformed to stabilize variance and differences of data are considered to obtain stationary series. In model selection, part data is examined by checking ACF (Auto Correlation Function) and PACF (Partial Autocorrelation Function) to identify potential model. ACF plot is frequently used to identify whether or not seasonality is present in a given time series data to identify appropriate model for specific situations and to determine if data are stationary. The general non-seasonal model is known as ARIMA (p, d, q) where p is order of autoregressive part, d is degree of first differencing involved, and q is order of moving average part. The ARIMA notation can be extended readily to handle seasonal aspects and general notation is ARIMA (p, d, q) (P, D and Q)$_{12}$ where P, D, Q are seasonal autoregressive, seasonal difference, and seasonal moving average part respectively. For illustrative purpose consider the following general ARIMA (1,1,1) (1,1,1)$_{12}$ model:

$$(1 - \phi_1 B)(1 - \alpha_1 B^{12})(1 - B)(1 - B^{12}) Y_t = (1 - \theta_1 B)(1 - \gamma_1 B^{12})\, e_t \qquad (1)$$

where $1 - \phi_1 B$ = non-seasonal autoregressive of order 1; $1 - \alpha_1 B^{12}$ = seasonal autoregressive of order 1; Y_t = current value of the time series examined; B = backward shift operator $BX_t = X_{t-1}$ and $B^{12}X_t = X_{t-12}$; $1 - B$ = 1st order non-seasonal difference; $1 - B^{12}$ = seasonal difference of order 1; $1 - \theta_1 B$ = non-seasonal moving average of order 1; and $1 - \gamma_1 B^{12}$ = seasonal moving average of order 1.

The second step of Box-Jenkins approach is modelling time series data by estimating parameters and testing the significance. This step involves the estimation of the parameters to fit potential models and to select a suitable criterion which can be used to determine the best model among plausible models. This criterion is well known as Akaike's Information Criterion (AIC) developed in 1974 defined as:

$$\text{AIC} = (-2 \log L + 2\, m) \qquad (2)$$

where L is the maximized likelihood function and m is the number of parameters estimated in the model. The model which attains minimum AIC value is considered as the best model. After choosing the most appropriate model, parameters are estimated by using the method of maximum likelihood. The method of maximum likelihood finds the values of the parameters, which maximizes the likelihood function. Computer programs for fitting ARIMA models will automatically find appropriate initial estimates of the parameters and then successfully refine them until the optimum values of the parameters are found. A test of significance of the estimated parameters is done from the parameter estimate and its standard error. Some of the estimated parameters may have been found insignificant (p-value larger than 0.05). If so, a revised model is considered dropping the insignificant explanatory variables. In diagnostic checking step check autocorrelation function (ACF) of residual. This is usually done by correlation analysis through the residual ACF plots and the goodness-of-fit test by means of Chi-square (χ^2) statistic. If the residuals are correlated, the model should be re-considered as discussed in step one. Otherwise, the model is adequate to represent the time series data.

The third step of Box-Jenkins approach is to use the best fit model (as adjudged in step 2) for forecasting purposes.

Result

Statistical software R version 2.13.0 has been used for developing the ARIMA (Autoregressive Integrated Moving Average) model. The descriptive statistics for our data is shown in Table 1. First, the data was checked for stationarity by using control chart which is a useful graphical device for detecting the lag of stationarity in a time series data. Figure 1 represents monthly particulate matter ($\mu g/m^3$) of air pollution data for Dhaka station; from the figure it can be reasonably concluded that there is a seasonal cycle of the data set indicating non-stationarity of the given time series data. To check non-stationarity of the data set, the ACF and the PACF were calculated and plotted. Figure 2 (a-d) represents (a): Autocorrelation function (ACF), (b) Partial Autocorrelation function (PACF) for original data set (particulate matter of air pollution data), (c) Autocorrelation function (ACF), and (d) Partial autocorrelation function (PACF) for first order seasonal differences (de-seasonalized) data set. It is clearly evident from Figure 2 (a-b) that the given time series data (monthly particulate matter) is not stationary. To fit ARIMA model to the data set it must be stationary. In order to make data stationary, the seasonal differences

Table 1: Basic statistics for Dhaka station monthly particulate matter data

No of obs	*Mean*	*St.Dev*	*Variance*	*Min*	*Max*
108	127.70	70.06362	4908.91	15.80	271.00

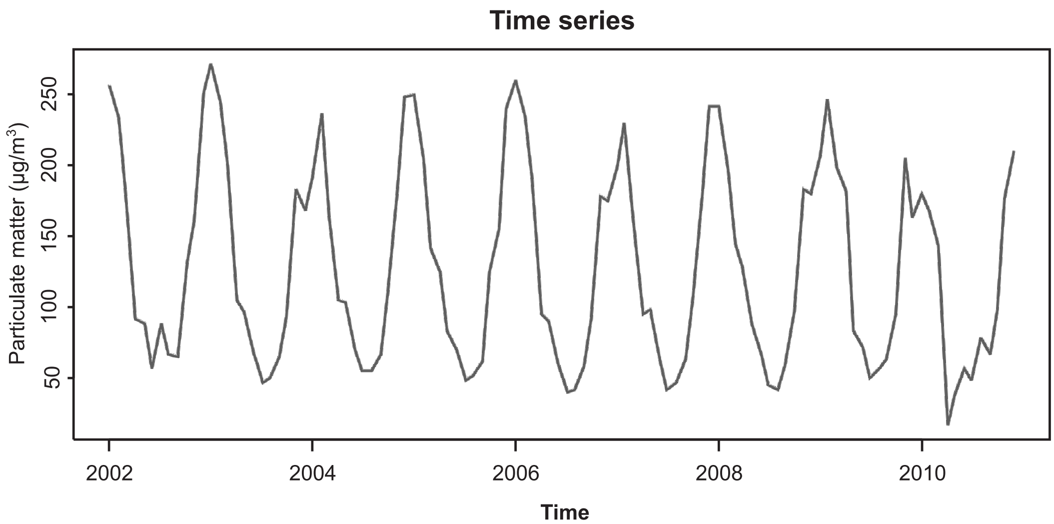

Figure 1: Monthly particulate matter (μg/m^3) of air pollution data for Dhaka station.

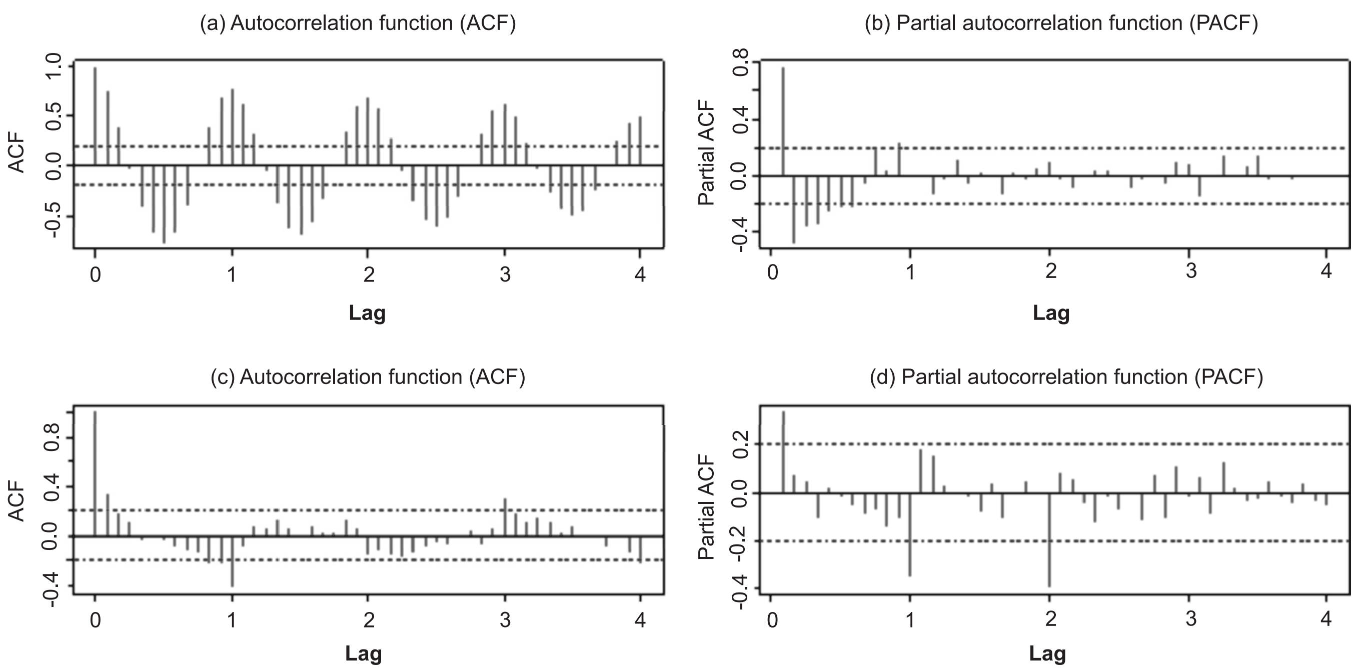

Figure 2: (a): Autocorrelation function (ACF), (b) Partial autocorrelation function (PACF) for original data set (particulate matter of air pollution data), (c) Autocorrelation function (ACF), and (d) Partial autocorrelation function (PACF) for first order seasonal differences (de-seasonalized) data set.

of original data was considered. Thus considering $D = 1$ the differences of the original data were calculated and again the ACF and PACF were calculated for the differences and presented in Figure 2(c-d). Figure 2 (c-d) suggests that differences (de-seasonalized) time series data is stationary.

Suppose the ARIMA (p, 0, q) (P, 1, Q)$_{12}$ model is considered for the given time series data (particulate matter in air). After considering the ARIMA model, the problem is to determine the values of the parameters *p*, *q*, *P* and *Q* of the considered model. Considering different combinations of the values of the parameters *p*, *q*, *P* and *Q*, a number of ARIMA model has been defined. To identify the best fit model, the AIC corresponding to each identified ARIMA model has been calculated. From the set of identified ARIMA models it was found that the ARIMA $(0, 0, 2)(2, 1, 0)_{12}$ model attains the lowest AIC value 884.18. Thus, the ARIMA $(0, 0, 2)(2, 1, 0)_{12}$ model is tentatively the best fit model for the given time series data.

Considering the ARIMA $(0, 0, 2)(2, 1, 0)_{12}$ model, the seasonal autoregressive parameters α_1 and α_2 were estimated to be 0.9239 (*p*-value = 0.00) and 0.6936 (*p*-value = 0.00) respectively and the moving average parameters θ_1 and θ_2 were estimated to be 0.2788 (*p*-value = 0.00) and 0.2440 (*p*-value = 0.01), respectively as shown in Table 2. Small *p*-value indicates that the coefficients of the selected model are significant.

After estimating parameters for this model the adequacy of the model was checked by their residuals. Figure 3 represents the diagnostic checking values of the residuals. From Figure 3 we can conclude that

Table 2: ARIMA $(0, 0, 2)(2, 1, 0)_{12}$ model characteristics parameters and *p*-values

Coefficients	*Parameters*	*Standard error*	*p-value*
ma1(θ_1)	0.2788	.0995	0.00
ma2(θ_2)	0.2440	.1037	0.01
sar1(θ_1)	0.9239	.0872	0.00
sar2(θ_2)	0.6936	.0753	0.00

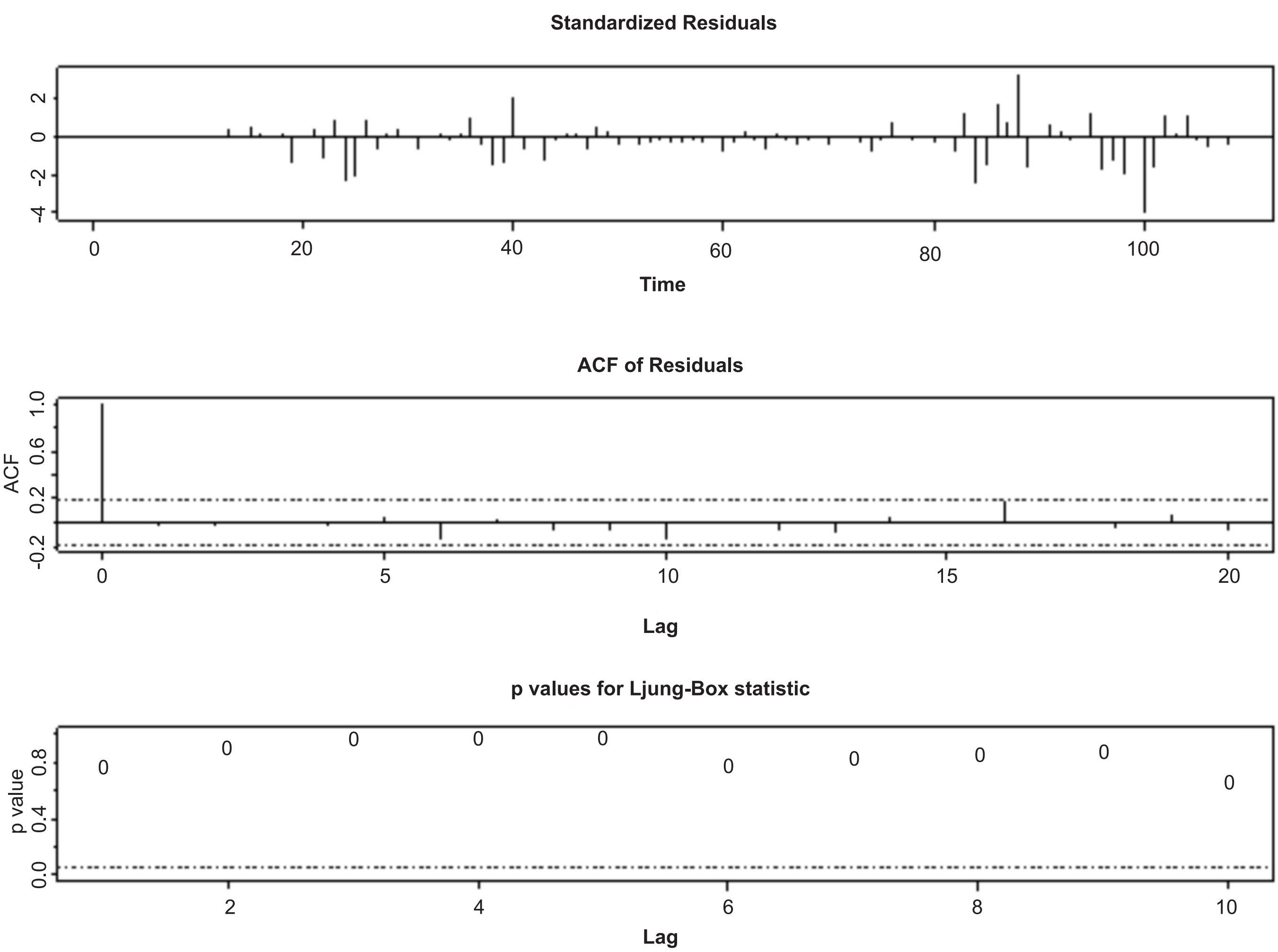

Figure 3: The diagnostic checking of ARIMA $(0, 0, 2)(2, 1, 0)_{12}$ model.

residuals are independently and identically distributed sequence with mean zero and constant variances. ACF of residual suggests that residual did not deviate significantly from zero mean white noise process. Ljung-Box shows high p-value associated with the statistic.

Figure 4 shows that the histogram of standardized residuals for the selected model follows approximately normal distribution which is also shown in the normal Q-Q plot. The goodness-of-fit of the model was also tested by chi-square statistic ($\chi^2 = 17.19$) and observed that the residuals are white noise.

Finally, based on the above results, the ARIMA (0, 0, 2) $(2, 1, 0)_{12}$ model was found adequate to represent the considered time series data on particulate matter in air for Dhaka meteorological station and used for forecasting purpose.

Figure 5 shows the predicted value and the observed value of the particulate matter in air. Close proximity of the original and the predicted values indicate that the considered model provide acceptable fit to predict particulate matter of air in Dhaka city.

Hence, the fitted ARIMA (0, 0, 2) $(2, 1, 0)_{12}$ model could be written as:

$$(1 - \alpha_1 B^{12} - \alpha_2 B^{24})(1 - B^{12}) X_t = (1 - \theta_1 B - \theta_2 B^2)\, e_t \tag{3}$$

This can be re-written as:

$$(X_t = (1 + \alpha_1)\, X_{t-12} + (\alpha_2 \alpha_1) X_{t-24} - \alpha_2 X_{t-36} - \theta_1 e_{t-1} - \theta_2 e_{t-2} + e_t \tag{4}$$

Finally the model in equation (4) is used for the forecasting purpose.

Figure 6 shows a comparison between the observed data and fitted data that resulted from the developed ARIMA model for the period between 2002 and 2010. The observed data is relatively close to the fitted data. Besides, it is clear that particulate matter of air pollution

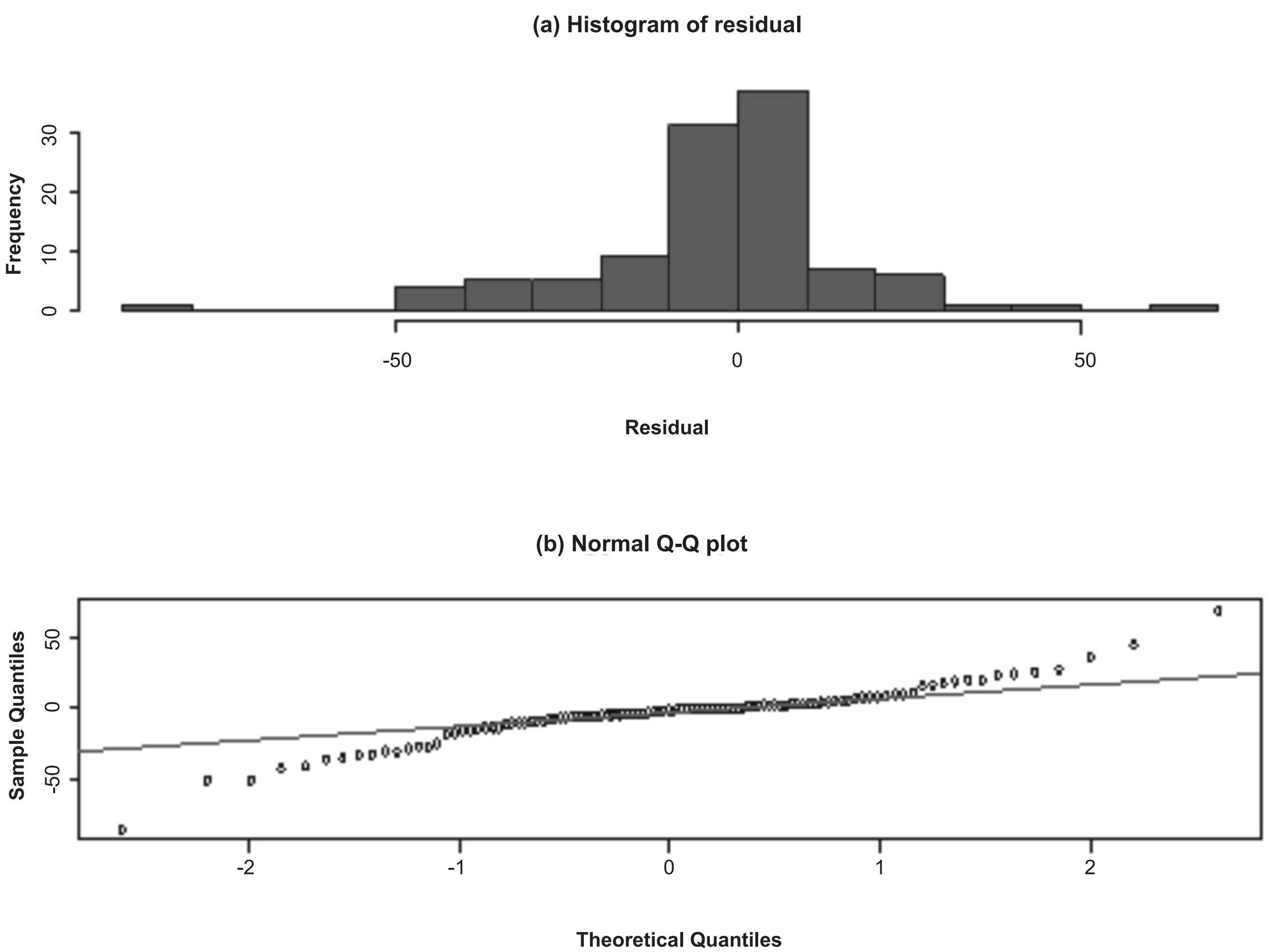

Figure 4: Testing normality and identifying outliers of residual.

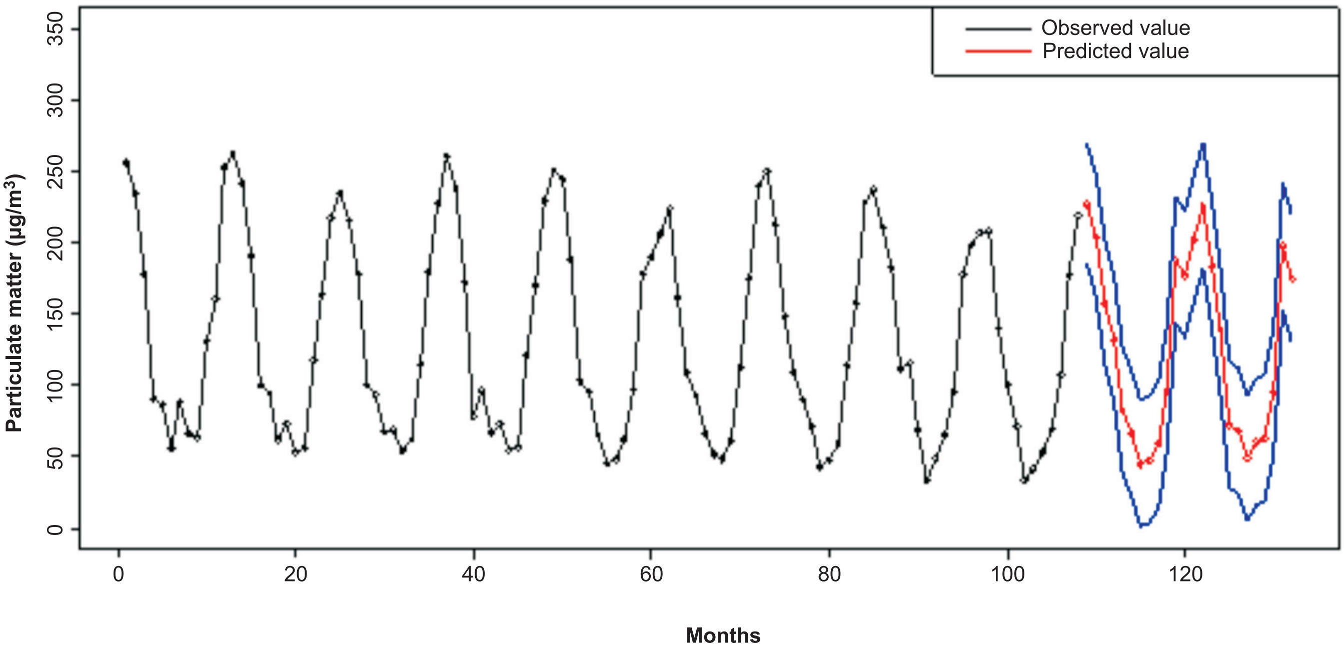

Figure 5: Actual and predicted values for 2002-2012 of overall time series by using ARIMA (0, 0, 2) $(2, 1, 0)_{12}$ model.

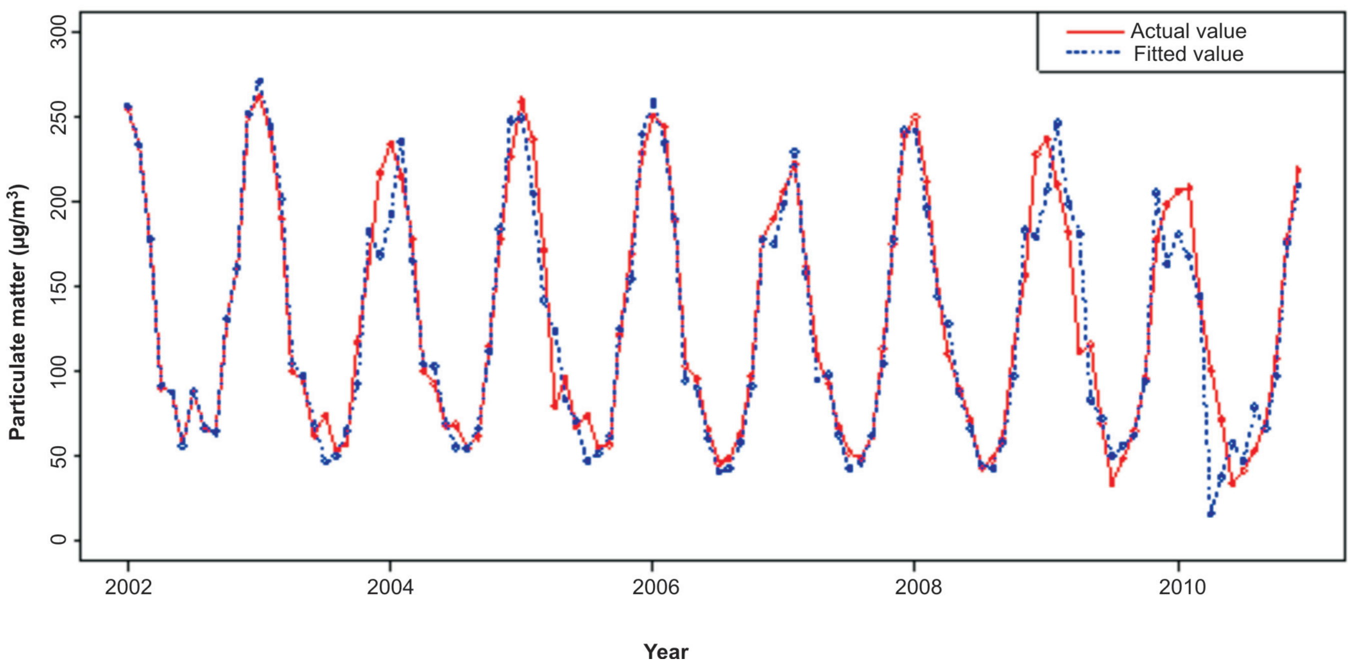

Figure 6: The comparison of actual and fitted value of particulate of air pollution.

pattern continues for the upcoming years and there is an indication of slight decreasing trend of the amount of particulate matter of air pollution with time.

Conclusion

Air pollution is a constant threat to the population of Dhaka city. In this study, ARIMA (0, 0, 2) $(2, 1, 0)_{12}$ model well reflected the trend in the particulate matter of air pollution in Dhaka city. It was observed that the ARIMA model is capable of representing air pollution in subsequent year with relative precision. This result also indicates that statistical time series models lead to better understanding to prevent air pollution from Dhaka city.

References

Choudhury, Z., Banu, S. and A. Islam (2008). Forecasting dengue incidence in Dhaka, Bangladesh: A time series analysis. *Dengue Bull*., **32:** 29-36.

D'Ippoliti, D., Forastiere, F., Ancona, C., Agabity, N., Fusco, D., Michelozzi, P. and C.A. Perucci (2003). Air Pollution and Myocardial Infarction in Rome: A case-crossover analysis. *Epidemiology*, **14:** 528-535.

Dominici, F., McDermott, A., Zeger, S.L. and J.M. Samet (2002). On the Use of Generalized Additive Models in Time-Series Studies of Air Pollution and Health. *Am. J. Epidemiol.*, **156:** 193-203.

Dominci, F., Peng, R.D., Bell, M.L., Pham, L., McDermott, A., Zeger, S.L. and J.M. Samet (2006). Fine Particulate Air Pollution and Hospital Admission for Cardiovascular and Respiratory Diseases. *JAMA*, **295:** 1127-1134.

Edson et al. (2011). Predicting the number of cases of dengue infection in Ribeirão Preto, São Paulo State, Brazil, using a SARIMA model. *Revista da Sociedade Brasileira de Medicina Tropical*, **44(4):** 436-440.

Franklin, M., Zeka, A. and J. Schwaratz (2007). Association between PM2.5 and All-Cause and Specific-Cause Mortality in 27 U.S. Communities. *Journal of Exposure Science and Environmental Epidemiology*, **17:** 279-287.

Ghio, A.J., Kim, C. and R.B. Devlin (2000). Concentrated Ambient Air Particles Induce Mild Pulmonary Inflammation in Healthy Human Volunteers. *Am. J. Respir. Crit. Care Med.*, **162:** 981-988.

Hong, Y.-C., Lee, J.-T., Kim, H., Ha, E.-H., Schwartz, J. and D.C. Christiani (2002). Effects of Air Pollutants on Acute Stroke Mortality. *Environ. Health Perspect*, **110:** 187-191.

Lin, M., Chen, Y., Burnett, R.T., Villeneuve, P.J. and D. Krewski (2002). The Influence of Ambient Coarse Particulate Matter on Asthma Hospitalization in Children: Case-crossover and time-series analyses. *Environ. Health Perspect.*, **110:** 575-581.

Makridakis, Spyros, Steven, C., Wheelwright and Rob J. Hyndman (1998). Forecasting: Methods and Applications. Third edition. John Wiley and Sons.

Momani, P.E.N.M. (2009) Time Series Analysis Model for Rainfall Data in Jordan: Case Study for Using Time Series Analysis. *American Journal of Environmental Sciences*, **5(5):** 599-600.

Nelson, Gouveiaa and Tony Fletcherb (2000). Time series analysis of air pollution and mortality: Effects by cause, age and socioeconomic status**.** *J. Epidemiol. Community Health*, **54:** 750-755 doi:10.1136/jech.54.10.750.

Norris, G., YoungPong, S.N., Koenig, J.Q., Larson, T.V., Sheppard, L. and J.W. Stout (1999). An Association between Fine Particles and Asthma Emergency Department Visits for Children in Seattle. *Environ. Health Perspect.*, **107:** 489-493.

Promprou, S., Jaroensutasinee, M. and K. Jaroensutasinee (2005). Forecasting Dengue Haemorrhagic Fever Cases in Southern Thailand using ARIMA Models. Dengue Bulletin.

Slaughter, J.C., Lumley, T., Sheppard, L., Koenig, J.Q. and G.G. Shapiro (2003). Effects of Ambient Air Pollution on Symptom Severity and Medication Use in Children with Asthma. *Ann. Allergy Asthma Immunol.*, **91:** 346-353.

Tolbert, P.E., Mulholland, J.A., MacIntosh, D.D., Xu, F., Daniels, D., Devine, O.J., Carlin, B.P., Klein, M., Dorley, J., Butler, A.J., Nordenberg, D.F., Frumkin, H., Ryan, P.B. and M.C. White (2000). Air Quality and Pediatric Emergency Room Visits for Asthma in Atlanta, Georgia. *Am. J. Epidemiol.*, **151:** 798-810.

Tsai, S.S., Goggins, W.B., Chiu, H.F. and C.Y. Yang (2003). Evidence for an Association between Air Pollution and Daily Stroke Admissions in Kaohsiung, Taiwan. *Stroke*, **34:** 2612-2616.

U.S. Environmental Protection Agency (2004). Air Quality Criteria for Particulate Matter. At http://www.epa.gov/ttn/naaqs/standards/pm/s_pm_cr_cd.html

Calendar of Events

15th International Riversymposium
8th to 11th October 2012
Melbourne, Victoria, Australia
Website: http://www.riversymposium.com
Contact person: Carla Mathisen
Organized by: International Water Centre

3rd International Conference on Environmental Aspects of Bangladesh (ICEAB12)
13th to 14th October 2012
Fukuoka, Japan
Website: http://www.benjapan.org/iceab/
Contact person: Md. Atiqur Rahman Ahad
Organized by: BENJapan

Storm Warning: Water, Energy and Climate Security in a Changing World
15th to 19th October 2012
Banff, Canada
Website: http://www.stormwarning2012.ca
Contact person: Ann Clemensen
Organized by: Stormwater Industry Association

WATER 2012
5th to 6th November 2012
London, United Kingdom
Website: http://marketforce.eu.com/Conferences/water12/?utm_source=conferencealerts.com&utm_medium=event_calendar&utm_campaign=water12_conferencealerts.com
Contact person: Robert Champion
Organized by: Marketforce Business Media

2th International Conferences of Water Resources
5th to 9th November 2012
Langkawi, Malaysia
Website: http://seminar.utmspace.utm.my/icwr2012/
Contact person: School of Professional and Continuing Education
Organized by: Faculty of Civil Engineering Universiti Teknologi Malaysia & Department Irrigation & Drainage Malaysia(DID)

2nd Annual Effluent & Waste Water Management Conference
13th to 14th November 2012
Nairobi, Kenya
Website: http://aidembs.com/effluent_conference/
Contact person: Hellen
Organized by: Aidem Business Solutions

Water Crisis Management under Changing Climate
16th to 17th November 2012
Bhubaneswar, India
Website: http://www.gugly.org/National%20Conference%202012.htm
Contact person: Debahuti Acharya
Organized by: Gugly Centre for Biological Research

2012 4th Journal Conference on Environmental Science and Development (JCESD 2012 4th)
24th to 25th November 2012
Bangkok, Thailand
Website: http://www.ijesd.org/jcesd/4th/
Contact person: Secretary of JCESD
Organized by: CBEES

2012 3rd International Conference on Biology, Environment and Chemistry (ICBEC 2012)
24th to 25th November 2012
Bangkok, Thailand
Website: http://www.icbec.org/
Contact person: Mr. Lee
Organized by: CBEES

International Conference on Environment (ICENV 2012)
11th to 13th December 2012
Penang, Malaysia
Website: http://chemical.eng.usm.my/ICENV2012/
Contact person: Professor Dr. Azlina bt harun@kamaruddin
Organized by: Universiti Sains Malaysia

2012 International Conference on Environment, Chemistry and Biology - ICECB 2012
29th to 30th December 2012
Hong Kong, China
Website: http://www.icecb.org/
Contact person: Miss Yang
Organized by: CBEES