American International Journal of Research in Science, Technology, Engineering & Mathematics

Available online at http://www.iasir.net

ISSN (Print): 2328-3491, ISSN (Online): 2328-3580, ISSN (CD-ROM): 2328-3629

AIJRSTEM is a refereed, indexed, peer-reviewed, multidisciplinary and open access journal published by International Association of Scientific Innovation and Research (IASIR), USA
(An Association Unifying the Sciences, Engineering, and Applied Research)

Box-Jenkins Approach to Forecast Air Pollution at Fort Railway Station, Sri Lanka

N.A.M.R.Senaviratna
Department of Mathematics
The Open University of Sri Lanka
Nawala, Sri Lanka

***Abstract**: Air pollution has been identified as one of the major environmental problems faced by most countries across the world in the last few decades. The objective of this study is to develop models to forecast air pollutants SO_2 and NO_2 at Fort Railway station, which is one of the most congested area in Sri Lanka. In this study, monthly concentration of SO_2 and NO_2 ($\mu g/m^3$) were used from October 2012 to December 2016. Augmented Dickey-Fuller test is used to check the stationary condition of series. Diagnostic tests are carried out using Jarque-Bera test, Breusch-Godfrey LM test and White's test. Akaike information criterion (AIC), Schwartz's Bayesian Criterion (SBC), Mean Square Error (MSE) and coefficient of determination (R^2) are used as the model selection criteria. Mean Absolute Percentage Error (MAPE) value was used to validate the model. ARIMA(0,1,2) and ARIMA(0,0,2) model were found to be the optimal model to forecast the NO_2 and SO_2 concentration respectively as this model comply with all assumptions of ARIMA model.*

***Keywords**: Air pollution, forecast, ARIMA, MAPE, SO_2, NO_2*

I. INTRODUCTION

Air pollution has been identified as one of the major environmental problems faced by most countries across the world in the last few decades. Air pollution in Sri Lanka is also increasing day by day due to the rapid rise in emissions of noxious gases from vehicular traffic, industrial emissions particularly from thermal power generation plants and rapid urbanization. The number of motor vehicles almost tripled during the 1990s which also led to an increase in traffic jams. Major air pollutants in Sri Lanka are oxides of nitrogen, oxides of sulfur, oxides of carbon and particulates. These pollutants have negative impact on people's health as they can cause respiratory illnesses, asthma or even death. Dust falls are also an issue in areas with a high traffic density [6]. Therefore, it is important to regularly monitor forecast of air quality to protect our health.

The objective of the present study is to observe the trend of different air pollutant parameters and to forecast the concentration of the parameters. Autoregressive integrated moving average (ARIMA) model is developed for monthly forecast of the parameters.

II. MATERIALS AND METHODS

This study was carried out using secondary data which was collected from National Building Research Organization(NBRO), Sri Lanka. Passive Air Quality Monitoring techniques were used to collect data. Fort railway station is selected for this study. It is one of the most congested area in Sri Lanka which is situated in Colombo city. The parameters considered in this study are Sulphur Dioxide (SO_2) and Nitrogen Dioxide (NO_2). Monthly data of NO_2 and SO_2 were used from October 2012 to December 2016 and EVIEWS software is used to analyze the data.

ARIMA Model

A dependent time series that is modeled as a linear combination of its own past values and past values of an error series is known as a ARIMA model. Non-seasonal ARIMA models are generally denoted ARIMA(p, d, q) where parameters p, d, and q are non-negative integers, p is the order of the autoregressive model, d is the degree of differencing, and q is the order of the moving-average model.

Given a dependent time series $\{Y_t : 1 \le t \le n\}$ mathematically the ARIMA model is written as

$$(1-B)^d Y_t = \mu + \frac{\theta(B)}{\phi(B)} e_t \quad \text{where} \qquad \phi(B) = 1 - \phi_1 B - \ldots - \phi_p B^p$$

$$\theta(B) = 1 - \theta_1 B - \ldots - \theta_q B^q$$

μ-mean term, B- backshift operator, $\phi(B)$- autoregressive operator, $\theta(B)$- moving average operator, e_t- random error.

Using autocorrelation function (ACF), partial autocorrelation function (PACF) and degree of differencing, appropriate autoregressive (AR) and moving average (MA) terms are determined. Then several ARIMA models were tried. Augmented Dickey- Fuller (ADF) test is used to check whether the series has a unit root. Lagrange's Multiplier (LM) test, White's test, Jarque-Bera (J-B) test are carried out to test adequacy of the fitted models. LM test is used to test the serial correlation among residuals. White's test is used in order to check heteroscedasticity. The normality assumption is checked by using Jarque-Bera test, which is a goodness of fit measure of departure from normality, based on the sample kurtosis and skewness. Akaike information criterion (AIC), Schwartz's Bayesian Criterion (SBC), Mean Square Error (MSE) and coefficient of determination (R^2) are used as the model selection criteria and select the optimal model among the identified models. Lower AIC, SBC, MSE values and higher R^2 indicate optimal model.

Mean Absolute Percentage Error (MAPE)

MAPE is the most common measure of forecast error. It is a measure of prediction accuracy of a forecasting method in statistics. It usually expresses accuracy as a percentage, and is defined by the formula:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{X_t - \hat{X}_t}{X_t}\right| \times 100$$; where X_t - Actual value, $\hat{X}_t$ - Fitted value, n - Sample size

III. RESULTS AND DISCUSSION

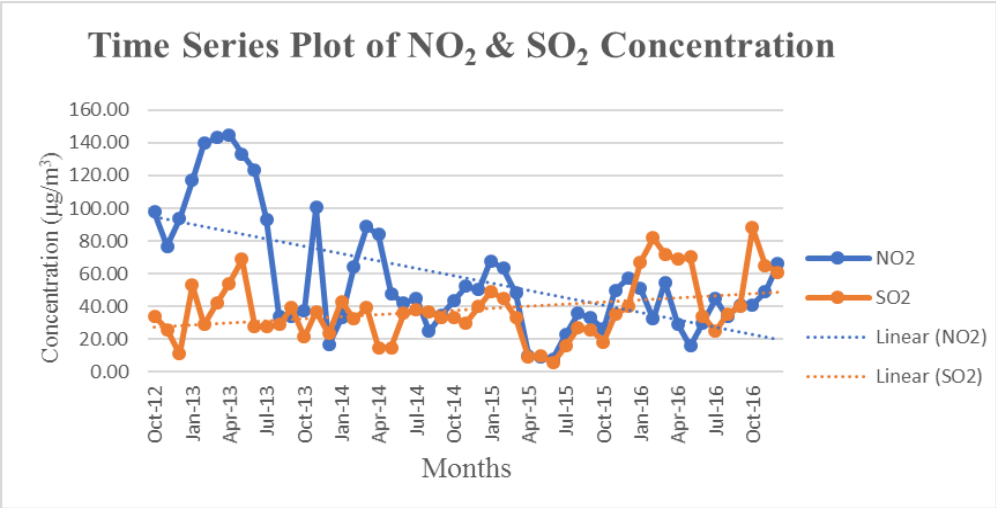

Figure 1: Time Series Plot of NO_2 and SO_2 Concentration

According to Figure 1 it can be seen that the NO_2 concentration does not vary in a fixed level which indicates that the series is non-stationary in both mean and variance, as well as exhibits seasonal trend. It also illustrates a downward trend throughout the period considered. Maximum permissible limit of NO_2 in Sri Lanka is and 100 μg/m^3. It can be seen that this maximum limit exceeds only in 2013. SO_2 series shows an upward trend. However, its fluctuation is not much high. It seems series is stationary. However, the maximum permissible limit which is 80 μg/m^3 for Sri Lanka exceeds February and October in 2016.

Table 1: Descriptive Statistics of Parameters

Statistic	NO_2	SO_2
Mean	57.18	38.01
Median	48	35.07
Standard Deviation	36.18	19.15
Skewness	1.01	0.71
Kurtosis	3.15	3.04
Jarque-Bera test statistic	8.67	4.24
Probability	0.01	0.12

Table 1 shows that summary of descriptive statistics of NO_2 and SO_2 series. It can be seen that the average values of NO_2 and SO_2 series do not exceed the maximum permissible limit. P value of Jarque-Bera test of NO_2 series indicates that the series is not normally distributed. However, p value of Jarque-Bera test of SO_2 series indicates that the series is normally distributed. Then stationarity of both series is tested using ADF test and results are displayed in Table 2.

Table 2: Results of Augmented Dickey-Fuller unit root test

Series	NO_2	SO_2
Original Series	0.1195	0.02
First differenced series	0.0000	-

According to Table 2, results of ADF test depict that NO_2 series is non-stationary. However, ADF test confirmed that first difference of NO_2 series is stationary. Furthermore, it is confirmed that SO_2 series is stationary.

Model Development for NO_2 series

Then several models of ARIMA including combination of AR(1), AR(2), MA(1), MA(2) terms based on observing the Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF) were tried and model assumptions were tested using model diagnostic tests. However only one model satisfied the model assumptions and summarized in following Table 3.

Table3: Results of ARIMA (0,1,2) model

Parameter	Coefficient	Standard Error	Probability
C	57.43	8.08	0.0000
MA(1)	0.87	0.11	0.0000
MA(2)	0.63	0.11	0.0000

Diagnostic Tests for NO_2 series

Table 4: Diagnostic test results for NO_2 series

	White's Test	BG LM Test	J-B Test
Model (MA(1), MA(2))	0.9681	0.4370	0.3138

Table 4 illustrates results for diagnostic tests. Probability of observed R-squared of White's test (0.9681) is greater than 0.05. It reveals that there is no Heteroskedasticity in this model at 5% level of significance. BG LM test exhibits that residuals are not serially correlated at 5% level of significance since probability of observed R-Squared(0.4370) is greater than 0.05.

Probability of J-B test (0.3138) is greater than 0.05. Thus it can be concluded that residuals are normally distributed with 5% level of significance. After the model estimation and residual analysis, it can be concluded that identified model is suitable for further analysis and forecasting.

The comparison of forecast values and observed values for the period from October 2016 to December 2016 is shown in Table 5.

Table 5: Forecasting performance of the model

Date	Observed value	Forecasted value
Oct-2016	41.01	48.83
Nov-2016	49.04	53.94
Dec-2016	66.03	56.29
MAPE		14.63%

Based on the MAPE value shown in Table 5, it is suggested that the ARIMA (0,1,2) the optimal model for NO_2 series as MAPE value gives 14.63%. By using this model, one can forecast the NO_2 concentration in Fort railway station for the near future and can take actions accordingly.

Accordingly, the fitted model for NO_2 series is shown as follows.

$$Y_t = 57.43 + 0.87e_{t-1} + 0.63e_{t-2} + e_t$$

Model Development for SO_2 series

Several models of ARMA including combination of AR(1), AR(2), MA(1), MA(2) terms based on observing the Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF) were tried and model assumptions were tested using model diagnostic tests. Models which satisfied conditions of diagnostic tests are only considered and summarized in following Table 6.

Table 6: Diagnostic results of identified models

Model	White's Test	BG LM Test	J-B Test
AR(1)	0.85	0.39	0.14
MA(1), MA(2)	0.89	0.56	0.13

Table 6 illustrates results for diagnostic tests for SO_2 series. Probability of observed R-squared of White's test confirms that there is no Heteroskedasticity in both models. BG LM test exhibits that residuals are not serially correlated. Probability of J-B test confirms that residuals are normally distributed in both models.

Model Selection for SO_2 series

Table 7: Selection Criteria for Identified Models

Model	AIC	SBC	MSE	R^2
AR(1)	8.36	8.44	15.53	36.78
MA(1), MA(2)	8.35	8.42	15.24	38.31

According to the results in Table 7, second model has lower AIC, SBC and MSE values and higher R^2 value. Therefore the second model is selected as the optimal model. The comparison of forecast values and observed values for the period from October 2016 to December 2016 is shown in Table 8.

Table 8: Forecasting performance of the model

Date	Observed value	Forecasted value
Oct-2016	58.03	41.74
Nov-2016	65.02	65.43
Dec-2016	61.04	56.88
MAPE		11.82%

Based on the MAPE value shown in Table 8, it is suggested that the ARIMA(0,0,2) model is optimal model for SO_2 series as MAPE value gives 11.82%. By using this model, one can forecast the SO_2 concentration in Fort railway station for the near future and can take actions accordingly.

Accordingly, the fitted model for SO_2 series is shown as follows.

$$Y_t = 38.09 + 0.61e_{t-1} + 0.41e_{t-2} + e_t$$

IV. CONCLUSION

This study was undertaken to forecast air pollutants NO_2 and SO_2 at Fort railway station, which is one of the most congested area in Sri Lanka using Box-Jenkins ARIMA approach. ARIMA(0,1,2) model was found to be the optimal model to forecast the NO_2 concentration as this model comply with all assumptions of ARIMA model and MAPE value is 14.63%. ARIMA(0,0,2) model was found to be the optimal model to forecast the SO_2 concentration as this model comply with all assumptions of ARIMA model and MAPE value is 11.82%.

References

[1] Klemm, O., & Lange, H. (1999). Trends of Air Pollution in the Fichtelgebirge mountains, NE Bavaria. *Environmental Science and Pollution Research*.

[2] LEE, J., & LIST, J. (2004, October 28). Examining Trends of Criteria Air Pollutants: Are the Effects of Governmental Intervention Transitory? *Environmental & Resource Economics*, pp. 21-37.

[3] Chaudhuri, S., & Dutta, D. (2014, August). Mann–Kendall trend of pollutants, temperature and humidity over an urban station of India with forecast verification using different ARIMA models. *Environmental Monitoring and Assessment*, pp. 4719-4742.

[4] (2006). *Country Synthesis Report on Urban Air Quality Management.* Sri Lanka: Asian Development Bank.

[5] Dı´az-Robles, A., C, O., Fu, S., Reed, D., Chow, C., Watson, J., & Moncada-Herrera, A. (2008, July 18). A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: The case of Temuco, Chile. *Atmospheric Environment*, pp. 8331-8340.

[6] Ileperuma, O. (2000). Environmental pollution in Sri Lanka: a review. *Journal of the National Science Foundation of Sri Lanka* , 301-325.

[7] Ileperuma, O., & Abeyratne, V. (2002). Monitoring Air Pollution Levels in Kandy using Passive and Active Gas Sampling Techniques. *Ceylon Journal of Science: Physical Sciences*, 54-61.

[8] Matharaarachchi, S., Manawadu, L., & Gunatilake, J. (2016). Evaluation of Urban Air Pollution Distribution in the Colombo Municipal Council Area, Sri Lanka. In *Geostatistical and Geospatial Approaches for the Characterization of Natural Resources in the Environment* (pp. 405-414). Springer International Publishing.

[9] Perera, M., Premasiri, H., Basnayake, G., & Fernando, A. (2004). Air Pollution Trends in the Largest Industrial Area in Sri Lanka. *Air Resource Management.* Sri Lanka.

[10] Salcedo, R., Ferraza, M., Alves, C., & Martinsa, F. (1999, July 1). Time-series analysis of air pollution data. pp. 2361-2372.