Journal of Sustainability Science and Management 2010 Volume 5 (1): 57-65

TIME-SERIES ANALYSIS OF POLLUTANTS IN EAST COAST PENINSULAR MALAYSIA

M.Z. IBRAHIM*[1], R. ZAILAN[1], M. ISMAIL[1], M.S. LOLA[2]

[1]Faculty of Science and Technology, Department of Engineering Science, [2]Faculty of Science and Technology, Department of Mathematic, Universiti Malaysia Terengganu, 21030 Kuala Terengganu, Terengganu, Malaysia.

**Corresponding author: zam@umt.edu.my*

Abstract: In keeping abreast with the country's rapid economic development and to meet with the nation's aspiration for an improved quality of life, clean-air legislation limiting industrial and automobile emissions was adopted in 1978. Yet, until today, air pollution from both sources still imposes a dilemma for the nation. In order to predict the status of future air quality in Malaysia, a Box-Jenkins ARIMA approach was applied to modelling the time series of monthly maximum 1-hour carbon monoxide and nitrogen dioxide concentrations in the East Coast states of Peninsular Malaysia, i.e. Terengganu, Pahang and Kelantan, respectively. In all the states, both carbon monoxide (CO) and nitrogen oxide (NO_x) concentrations have shown a fairly consistent upward trend since 1996. Nevertheless, the forecasting value till 2016 didn't exceed the permissible values of both NAAQS and DOE Malaysia which are 35 and 30 ppm at 1-hour average for CO and 0.053 and 0.17 ppm for NOx.

KEYWORDS: ARIMA Forecasting, time series, Carbon monoxide, Nitrogen dioxide, East Coast Peninsular Malaysia

---

Introduction

The time-series forecasting approach is of useful to predict the future air-quality status from various aspects of development in each country. The forecasting method analyses the sequence of historical data in a period of time to establish the forecasting model. The ARIMA method has been extensively studied and used in previous research proven to be effective in forecasting field study. Forecasting method applying the ARIMA time-series method for pollution field has been expounded in many previous publications. The hybrid model is applied to the time series of NO2 concentration observed at a site in Delhi. The prediction performance results show that the ARIMA modelling can be an effective tool to forecast the air-pollutant concentrations (Chelania and Devotta, 2006). A hybrid ARIMA and Artificial Neural Networks as well have been practiced for PM10 simulation or forecast (Diaz-Robles *et al.*, 2007). Prybutok *et al,* 2000 employed the similar method to forecast daily maximum ozone levels in Houston area.

Air-pollution data is obtained from Air Quality Division of Alam Sekitar Malaysia Sdn. Bhd. (ASMA) which was awarded a concession by the government of Malaysia to set up a systematic and comprehensive monitoring network for air quality for the nation and to establish the National Environmental Data Centre since 1995. Till now, there are 52 monitoring stations of continuous ambient air and 20 manual air-quality monitoring stations operated, managed and maintained by ASMA throughout Malaysia. The monitoring system employs the state-of the art instrumentation to continuously monitor the major pollutant gases in the air as well as providing precise and accurate monitoring data (ASMA, 2008).

Two criteria pollutants, carbon monoxide and nitrogen dioxide, are considered because each of the data sets covers at least 10 years with no missing data in between, and shows fairly apparently either trend or seasonality, or both. Scientific research has proven that these two gases have a lot of negative health effects, including some deadly diseases (Raub, 1999; Chaloulakou *et al.*, 2008). Carbon monoxide is a significantly toxic gas which can lead to significant toxicity of the central nervous system and heart. CO molecule is the agent that enters a person's nasal cavity and binds with the hemoglobin (Hb) in the blood to form carboxyhemoglobin (COHb) (Flachbart, 1999). The presence of COHb in the blood causes hypoxia which reduces the liberation of oxygen to body tissues (Raub, 1999).

Nitrogen dioxide is also toxic to humans by its ability to form nitric acid with water in the eye, lung, mucous membrane and skin. Persons exposed to high concentrations of NOx can suffer lung irritation and potential lung damage. It can increase susceptibility to respiratory infection and airway constriction in those with asthma. In the urban area of Athens, it was shown that both mortality due to respiratory problems and hospital admissions for cardiac and respiratory causes are related with observed NOx level (Chaloulakou *et al.*, 2008).

In this paper, pollutant's data of selected monitoring station of Pahang, Terengganu and Kelantan from 1996 to 2006 were analysed to establish the forecasting model of these parameters as well as to observe the upcoming trend of these pollutants. Subsequently, the root of pollution problem in this study area will be deliberated.

Methodology

Box -Jenkins ARIMA Modelling.

Monthly data covering the periods of 1997 to 2006 were acquired from the Air-Quality Division of Alam Sekitar Malaysia Sdn. Bhd. (ASMA). The Box-Jenkins ARIMA model was used to model the time series behaviour to generate the forecasting trend. It stands for Autoregressive Integrated Moving Average with each term representing steps taken in the model construction until only random noise remains. The methodology consisting of a four-step iterative procedure was used in this study.

The first step is tentative identification where the historical data are used to tentatively identify an appropriate Box-Jenkins model. If the data is not stationary, differencing process should be performed until the obvious pattern, such as trend or seasonality, in the data fade away. It is followed by estimation of the parameters of the tentatively-identified model by the plot of autocorrelation function (ACF) and partial autocorrelation function (PACF) of the stationary data. After that, the diagnostic checking step must be executed to check the adequacy of the identified model in order to choose the best model. In this study, we used the Ljung-Box statistic in order to verify the adequacy of the model. The Ljung-Box statistic is

$$Q^* = T\,(T+ 2) \sum_{j=1}^{k} \tau j^{\,2} / T\text{-}j \tag{1}$$

The Q-statistic at lag k is a test statistic for the null hypothesis that there is no autocorrelation up to order k, where τ is the j-th autocorrelation and T is the number of observations. If the calculated value of Q^* is larger than the $x^2_{[\alpha]}$ value for k-p-q degrees of freedom, estimated residual series do not appear to be white noise, and the model should be considered inadequate (Hong Wu, 1997). Frequently, α is chosen to equal to .05 but this choice is not sacred. Usually, α between 0.01 and 0.05 is acceptable (Bowerman & O'Connel, 1993).

Finally, the best model will be used to establish the time-series forecasting value (Hyndman, 2001). The accuracy of the model is checked with the mean-square error (MS) to compare fits of different ARIMA models. The lower MS value shows the better-fitting model.

Autoregressive models (AR), moving average model (MA) and autoregressive integrated moving average models (ARIMA).

An autoregressive model of order *p*, AR (*p*) has the form of

$$z_t = \rho_1 z_{t-1} + \rho_2 z_{t-2} + \ldots + \rho_p z_{t-p} + \varepsilon_t \quad (2)$$

The term 'autoregressive' refers to the fact that this model expresses the current time-series values z_t as a function of past time-series values z_{t-1}, z_{t-2}, ..., $z_{t-p.}$ The $\rho_{1,}$ $\rho_{2,,}$ ..., ρ_3 are unknown parameters relating z_t to z_{t-1}, z_{t-2}, ..., $z_{t-p.}$

A moving-average forecasting model uses lagged values of the forecast error to improve the current forecast. A first-order moving-average term uses the most recent forecast error, a second-order term uses the forecast error from the two most recent periods and so on. An MA(q) and has the form of

$$zt = \varepsilon_t - \theta_1 \varepsilon_{t-1} + -\theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q} \quad (3)$$

Here, ε_{t-1}, ε_{t-2}, ..., ε_{t-p} are the past random shocks and θ_1, θ_2, ..., θ_q are unknown parameters relating z_t to ε_{t-1}, ε_{t-2}, ..., ε_{t-p}.

The autoregressive and moving-average specifications can be combined to form an ARMA (p,q) specification:

$$z_t = \rho_1 z_{t-1} + \rho_2 z_{t-2} + \ldots + \rho_p z_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q} \quad (4)$$

The point estimate for each parameter in a Box-Jenkins model is associated with its standard error and *t*-value. The parameters are tested whether it is zero (null hypothesis, H_o) or different from zero (alternative hypothesis, H_a) (Ediker & Akar, 2007). If the $t > 1.96$, we can reject H_o: $\theta_1 = 0$ in favour of H_a: $\theta_1 \neq 0$ by setting α equal to 0.05.

Seasonal ARIMA model (SARIMA)

Seasonality is defined as a pattern that repeats itself over a fixed interval of time. In general, seasonality can be found by identifying a large autocorrelation coefficient or large partial autocorrelation coefficient at seasonal lag. For the seasonal model, we used the Akaike Information Criteria (AIC) as the criteria for model selection. AIC is a combination of two conflicting factors which are the mean square error and the number of estimated parameters of a model. Generally, the model with smallest value of AIC is chosen as the best model (Hong Wu, 1997).

Results and Discussion

Plot of raw datas

The application chosen for this study is the concentration, ppm of pollutants (CO and NOx) for the East Coast of Peninsular Malaysia. The data used were monthly data from 1997 to 2006. The raw data for each parameter of every state are included in Figure 1 and Figure 2.

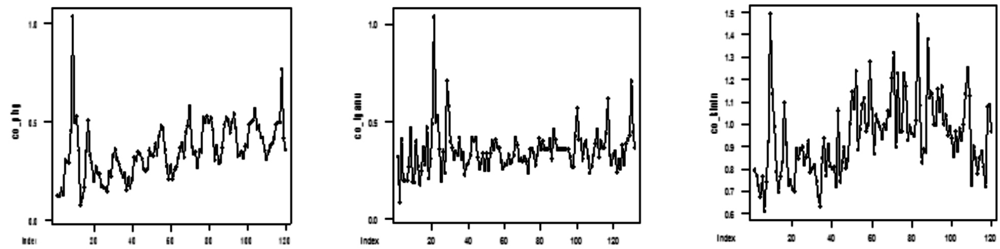

Figure 1. Raw data of CO (Pahang, Terengganu and Kelantan)

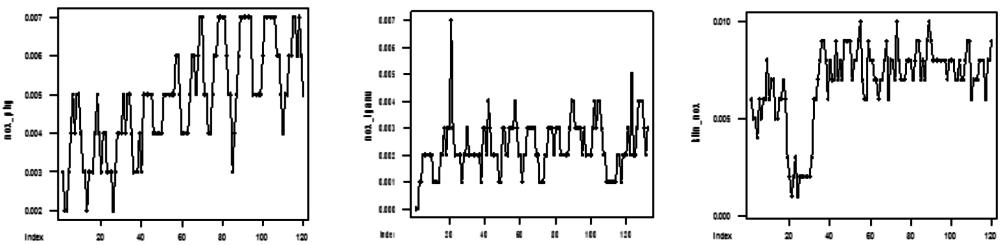

Figure 2. Raw data of NOx (Pahang, Terengganu and Kelantan)

The model-development process began by studying the original acf of the raw data. If the nonstationary condition emerges, the differentiation process will be executed to obtain the stationary time series. The number of lags to display the ACF is 30 lags. Then, the ACF and PACF of difference data in Appendix A were examined to determine the combination of ARIMA model for each time series.

After estimating all the possible models, the best-fitted models were selected through the diagnostic-checking procedure. The ACF plot then was tested with Ljung-Box Test or AIC test. The accepted values are included in the table provided for each model. Figure 3 and Figure 4 in the next section show the forecasting graph for both pollutants.

Carbon monoxide forecasting model for Pahang, Terengganu and Kelantan.

The prediction trend of carbon monoxide concentration is shown in Figure 3 for each state from 1997 up to 2016. The selected models are ARIMA models with first differentiation, but the Kelantan data excluded the stage of differentiation as the ACF of original data is assumed stationary. The best models for each state for CO can be seen in Table 1.

Table 1. Models for Carbon Monoxide (CO)

Description	Model	T-test, t, > 1.96	Ljung-Box Test, Q^* /AIC	Model equation
Pahang	ARIMA (1,1,1)	AR(1): 7.4, MA(1): 44.62	18.3<23.2 at df=10	$z_t = 7.9x10^{-4} + 0.57z_{t-1} + \varepsilon_t - 0.99\varepsilon_{t-1}$
Terengganu	ARIMA (4,1,1)	AR(1): |-5.97|, AR (2): 3.22, AR (3) :|-3.86|, AR (4): |-2.86|, MA(1): 474.85	9.3<14.06 at df=7	$z_t = -3.38x10^{-5} + -0.52z_{t-1} - 0.31z_{t-2} - 0.37\ z_{t-3} - 0.25\ z_{t-2} + \varepsilon_t - 1.0\varepsilon_{t-1}$
Kelantan	ARMA (1,2)	AR(1): 33.62, MA (1): 5.98, MA (2):3	14.9<16.91 at df=9	$z_t = 0.02 + 0.98z_{t-1} + \varepsilon_t - 0.57\varepsilon_{t-1} - 0.28\ \varepsilon_{t-2}$

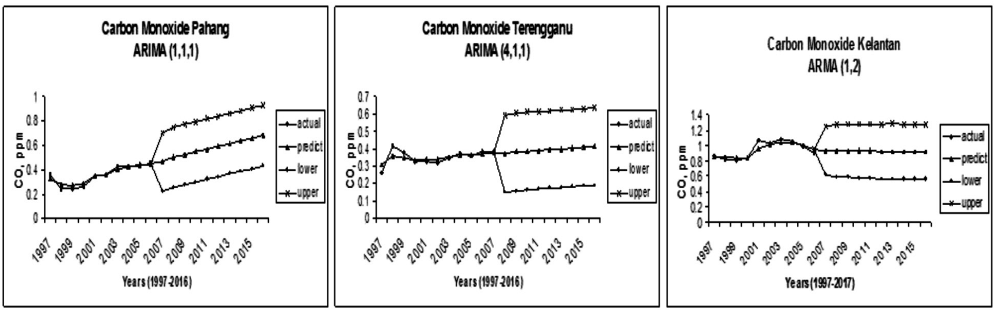

Figure 3. Prediction model for Carbon Monoxide
(Pahang, Terengganu and Kelantan, 1997-2016)

From the forecasting graph above, it can be seen that the CO concentration for Pahang state increases steadily from actual value of 0.36 to 0.68 ppm in 2016. Meanwhile, for Terengganu state, the concentration of CO shows a slight increasing to 0.42 in 2016 from an actual value of 0.26 in 1997. As for Kelantan, the value of forecast concentration lies in the range of 0.9 ppm compared to the actual value for initial year of 0.86 ppm. The increasing value of CO for Pahang and Terengganu states is faster than in Kelantan. So far, the predicted value of CO for the states is still under the limit of DOE which is 30 ppm (Talib *et al.*, 2002) and 35 ppm (EPA, 2008) for average 1-hour CO.

Nitrogen Dioxides forecasting model for Pahang, Terengganu and Kelantan.

In this section, the prediction trends of nitrogen dioxides concentration are displayed in Figure 4, while the descriptions for each model are available in Table 2. The selected models are SARIMA model with differentiation at seasonal level, ARIMA model without differentiation for Terengganu and ARIMA with first differentiation for Kelantan.

Table 2. Models for Nitrogen Dioxides (Nox)

Description	Model	T-test, t, > 1.96	Ljung-Box Test, Q^* /AIC	Model equation
Pahang	SARIMA (0,0,0)(0,1,2)12.	-	-1071.73708	$z_t = 1.375\varepsilon_{t-1} - 0.379\varepsilon_{t-12} + \varepsilon_t$
Terengganu	ARMA (1,3)	AR(1):13.55, MA (1): 17.86, MA (3) :3.05,	9.4<15.5 at df=8	$z_t = 2.78 \times 10^{-4} + 0.98_1 z_{t-1} + \varepsilon_t - 0.56\varepsilon_{t-1} - 0.16\varepsilon_{t-2} - 0.29\varepsilon_{t-3}$
Kelantan	ARIMA (4,1,0)	AR(1): |-11.21|, AR (2): |-6.46|, AR (3):|-4.84|, AR (4):|-4.4|	18.4<20.09 at df=8	$z_t = 2.1 \times 10^{-5} - 0.97z_{t-1} - 0.74 z_{t-2} - 0.56z_{t-3} - 0.39 z_{t-4}$

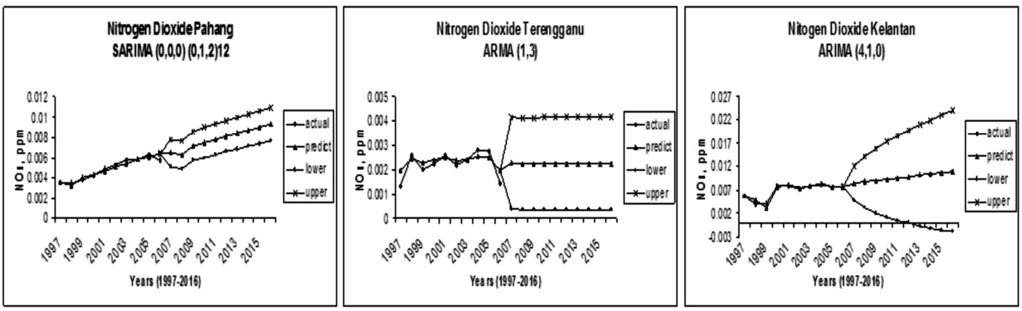

Figure 4. Prediction model for Nitrogen Dioxide (Pahang, Terengganu and Kelantan, 1997-2016)

The predicted values of NOx for Pahang and Kelantan increase evenly, respectively. The NOx Pahang rise from 0.0035 to 0.009 at 2016. Whereas, NOx Kelantan rise from 0.005 at initial actual value up to 0.011 ppm by year 2016. The CO concentration for Terengganu increases from 0.0013 ppm and varies steadily between 0.002 for the forecasted years. After all, Pahang still stands to be the highest-polluted state as it is the most developed state among the three states. But, the value is less than the standard of DOE and NAAQS which are 0.17 ppm and 0.053 ppm.

From the study, we can see that Pahang shows the increment trend for both parameters instead of Terengganu and Kelantan. The increase value of the pollutants can be related to the development of the states. The construction of the East Coast Highway which connects Kuala Lumpur to Kuantan, Pahang and continues to Kuala Terengganu has given a big impact to the escalating concentration of the pollutants. With the new linkage, many investors from other states will be interested to initiate business here. The industrialisation sector also will further develop and increase the amount of transportation for both states simultaneously. Recently, the government also has given more attention to the East Coast area by holding big events here. It is mainly to expand the economy and tourism sector here, as the East Coast of Peninsular Malaysia is a well known place for tourism.

However, the actual and forecast value for all the states are considered harmless as it is under the permissible value of DOE and NAAQS. It is different from the study at the West Coast area which shows the higher value of pollutants. The highest concentration of CO was recorded in Nilai Industrial Area with concentration 4.35 ± 0.80 ppm respectively, while the highest concentration of NO_2 was recorded in Sepanggar Industrial Area (0.057 ± 0.027 ppm) (Talib *et al.*, 2002). These two values are higher than the actual and predicted value of both parameters in this study. A report from Malaysia Meteorological Department also highlighted that, generally, the rainfall from the West Coast of Peninsular Malaysia is more acidic than the East Coast of Peninsular Malaysia. This situation supported the less-polluted condition in the East Coast area.

Conclusion

For the conclusion, Pahang appears as the most polluted state in the East Coast as it shows the increment value of both pollutants in future. However, the forecasting value of each concentration parameters are still in the protected condition as it does not exceed the limits of both NAAQS and DOE Malaysia. This condition appears for the reason that the cities in the East Coast of Peninsular Malaysia are still not as developed as the West Coast area.

Acknowledgements

The authors would like to thank DOE Malaysia for providing pollutants data from 1997 to 2006 and the Ministry of Higher Education (MOHE) for allocation for research grant to accomplish this study.

References

Bowerman, B.L., R.T. O'Connel. (1993). *Forecasting and time series. An applied approach*, Belmont, California. Pp: Duxbury Press.

Chelani, A.B., S. Devotta. (2006). Air quality forecasting using a hybrid autoregressive and nonlinear model. *J. Atm. Env*. 40:1774–1780.

Chaloulakou, A., Mavroidis, I., Gavriil, I. (2008). Compliance with the annual NO2 air quality standard in Athens. Required NOx levels and expected health implications. *J. Atm. Env*. 42: 454–465.

Diaz-Robles, L.A., J.C. Ortega-Bravo, M. Ramirez, J.S Fu, G.D. Reed. (2008). A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: the case of Temuco, Chile. *J. Atm. Env*. 42:8331–834.

Ediger, V.S., S. Akar. (2007). ARIMA forecasting of primary energy demand by fuel in Turkey. *J. Energy Pol*. 35:1701-1708.

Environment Protection Agency. (2008). National ambient air quality standards (NAAQS). <http.www.epa.gov> (26th December 2008).

Flachsbart, P.G. (1999). Human exposure to carbon monoxide from mobile sources. *J. Chem: Global Change Sci*. 1:301-329.

Hong Wu. (1997). *A time series analysis of United States carrot exports to Canada*.

Horacio, B., R. Martin Negri, S.R Enrique. (1996). Continuous measurement of gaseous pollutants in Buenos Aires city. *J. Atm. Env*. 33:2587-2598.

Hyndman,R.J.(2001). Box Jenkins modeling. http://www.itl.nist.gov/div898/handbook/pmc/section4/pm445.htm.

Prybutok Victor, R., Junsub Yi, David Mitchell. (2000). Comparison of neural network models with ARIMA and regression models for prediction of Houston's daily maximum ozone concentrations. *European Journal of Operational Research* 122:31-40.

Raub, J.A. (1999). Health effects of exposure to ambient carbon monoxide. *Chemosphere: Global Change Science*. 1:331-351.

Talib, M.L., M.O. Rozali, S. Norela, M.N. Ahmad Daud, N.J Permata. (2002). Air quality in several industrial areas in Malaysia. *Proceedings of the Regional Symposium on Environment and Natural Resources.* 1: 703-71.

APPENDIX A

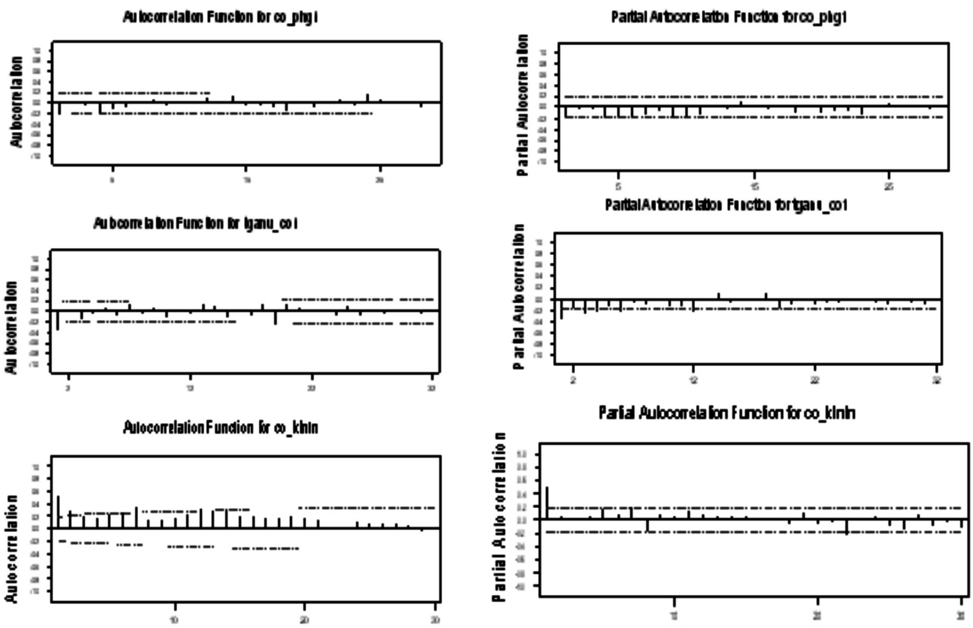

Figure 1. The *acf* and *pacf* of Carbon Monoxide (Pahang, Terengganu and Kelantan).

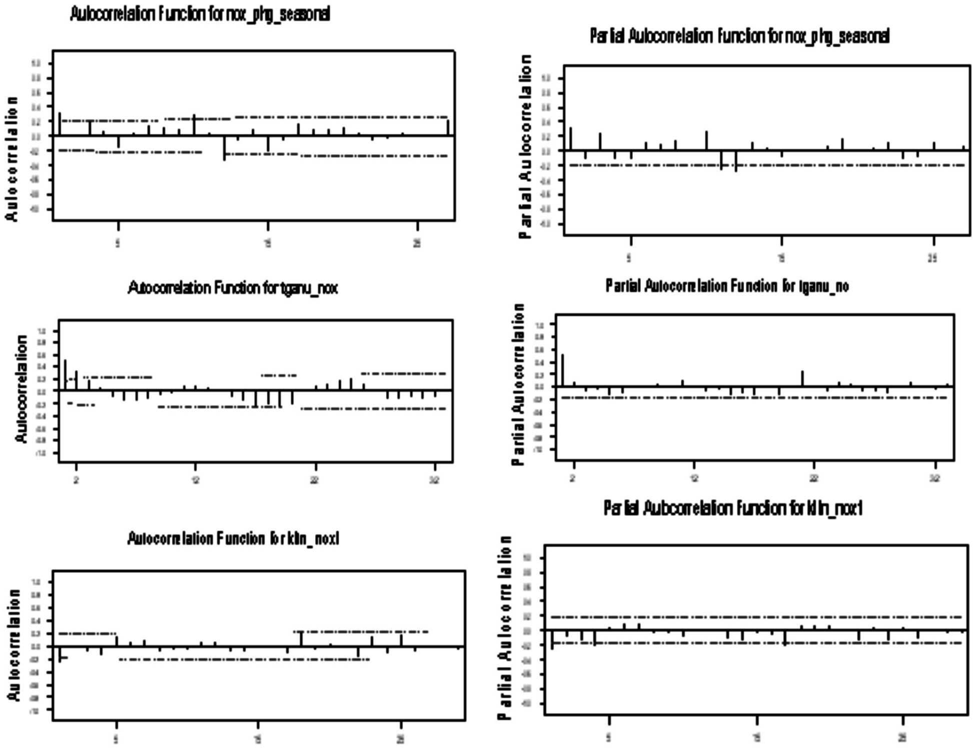

Figure 2. The *acf* and *pacf* of Nitrogen Dioxides (Pahang, Terengganu and Kelantan).