TIME SERIES MODEL FOR CARBON MONOXIDE (CO) AT SEVERAL INDUSTRIAL SITES IN PENINSULAR MALAYSIA

[1]Norshahida Shaadan and [2]Muhammad Soffi Rusdi [3]Nik Noorul Syakirin Nik Mohd Azmi [4]Shahira Fazira Talib [5]Wan Athirah Wan Azmi
Center for Statistical and Decision Science Studies, Faculty of Computer and Mathematical Sciences, UiTM, 40450, Shah Alam

e-mail: [1]shahida@tmsk.uitm.edu.my, [2]soffirusdi123@gmail.com, [3]noorul_syakirin@yahoo.com, [4]shahirafazira9@gmail.com, [5]wanathirah@yahoo.com

Abstract *Malaysia is reported to experience explosive rise in the demand of transport vehicles in recent years due to rapid economic development and population growth. As a result, air pollution is expected to increase in conjunction with the increase in the number of the vehicles. In particular, Carbon Monoxide (CO) has been identified as the main component of the emission sources from vehicles other than Nitrogen Oxide (NOx), hydrocarbon lead and particulate matter of size less than 10 micron (PM_{10}). This provides the reason why CO concentration is often used to reflect traffic density in an area. CO has both short-term and long-term effect on human's health. Thus, knowledge on CO behaviour and the future levels at an area is important to help decision makers in managing air pollution due to vehicles emission in the country. This study was conducted to describe CO data and to determine a suitable time series model to enable the prediction of CO levels at two industrial sites; Perai and Pasir Gudang, Malaysia. The model obtained could help management to mitigate CO pollution at the sites. The analysis was conducted using daily maximum data which was obtained from the Department of Environment Malaysia from 2010 to 2014. The performance of the best model was determined using several performance measures such as MAE, RMSE and MAPE. The study has found that the most appropriate time series model for Perai is ARIMA (3,1,1) and for Pasir Gudang is SARIMA (2, 1, 8) (1, 1, 2)$_7$.*

Keywords Carbon Monoxide; Time series; Prediction Model, Air Pollution

1   Introduction

In Malaysia, power stations, motor vehicles and open burning has been identified as the main contributors to the country's air pollution emissions (Afroz *et al*., 2003). In relation to this phenomenon, it is expected that there will be an increase in the amount of primary pollutant emitted by vehicles and industries such as carbon monoxide in the future years. In the context of air pollution from vehicles emission, Carbon Monoxide (CO) has been identified as the largest contributor followed by NO_2, SO_2 and PM_{10} (Azedah *et al*. 2015). Other than the mentioned pollutant, non-methane volatile organic compounds and ammonia are also omitted by industries (Fore and Mbohwa, 2010). Thus, this problem should have become a driven and motivating factor for the responsible body and the decision makers for the planning and designing of policy to mitigate and sustain the quality of atmospheric environment.

It has been reported that the number of transport vehicles registered in all types of categories has increased year by year in Malaysia (Noresah *et al*., 2012). This increment is in conjunction with the rapid economy growth, increase in the population and urban development. The private own vehicles registered in Malaysia had also show an increasing trend. These increasing trends can be traced back. In 1980 only 80,400 units of new vehicles registered. It had increased to 106,400 units (1990), 224,900 units (1995), 282,103 units (2000) and 416,600 units (2005) respectively (MAA, 2016). Malaysia Automotive Association (MAA), had also forecasted a further positive growth in number of new vehicle registration from 2017 to 2020 as shown in Table 1.

Table 1: Total Industry Volume (TIV) Forecast for 2017 to 2020

Year	Forecast			
	2017	**2018**	**2019**	**2020**
Passenger Vehicles	593,500	608,200	622,000	635,444
Commercial Vehicles	76,000	78,000	80,000	82,000
Total Industry Volume	669,500	686,200	702,000	717,444

Source: Malaysia Automotive Association: Market Review for 2015 and Outlook for 2016.

Exposure to air pollution is the main environmental threat to human's health in many busy and high population areas including industrial areas, cities and urban towns (Pasca *et al*., 2013). Long time exposure to high pollutant level in the air also contributes to wide range of chronic respiratory diseases, aggravates heart diseases and other types of particulates pollution (Miller *et al*., 2016). In particular, due to increase of human activities, emissions of particulate matters and gaseous matters including CO have been rising. Expansion of industries and transport systems has made this situation more critical. Other than human's health, such industrial activities also have great impact on the ecology and agriculture as well as safety effects (Xu *et al*., 2014).

CO is an odorless gas. It is a product of incomplete combustion and occurs when carbon in the fuel is partially oxidized rather than fully oxidized to Carbon Dioxide (CO_2) (Shuhaili *et al.,* 2016). CO also being released when engines burn fossil fuel which is a gas that comes out from the burning fossil fuels, mostly in motor vehicles. Emissions of CO are higher when engines are not tuned properly, and when fuel is not completely burned. Cars emitted lot of the CO found outdoors. CO has been identified as the main components of the emission sources from vehicles other than NOx, hydrocarbon lead and $PM_{10.}$ Vehicles sources have contributed to at least 70 % to 75% of the total air pollution in the Malaysia environment (Afroz *et al*., 2003; Yahya *et al*., 2006). CO is a significantly toxic gas that can lead to significant toxicity of the central nervous system and heart, thus affect human's health, in fact can cause fatality (Chen *et al.*, 2011).

Having knowing the health effect of CO emitted by vehicles to human and the growing trends of vehicles in Malaysia, there is a need to investigate the temporal pattern of the pollutant as well as to predict its future concentration levels. So that the study results will be able to assist government agencies and authorities to provide inputs in solving air pollution problems and issues due to vehicles usage. This requires an appropriate time series prediction model in providing the input and the insights. Even though, for Malaysia data set, several modelling techniques have been used to predict CO including the application of Artificial Neural Network (ANN) and Imperial Competitive Algorithm (Mahmoudzadeh *et al*., 2013; Rathi et. al. 2018), a

hybrid approach of Radial Basis Function network (Mamat *et al*., 2003), Multi-Layer Perceptron Techniques (MLP) and the hybrid (Raji *et al*., 2005), multiple regression (Afsaneh *et al*., 2014; Sembiring et. al., 2010), only a few had applied ARIMA time series modelling. The studies focused on CO level at the chosen locations such as east coast region (Kelantan, Terengganu and Pahang) and Bachang, Melaka (Ibrahim et al., 2009; Hamid *et al*., 2017).

Unlike the previous studies conducted for Malaysia data, this research is conducted with the purpose to investigate and predict the reading of daily maximum CO focuses at industrial sites; Perai in Penang and Pasir Gudang in Johor states of Malaysia. The SARIMA model application on CO data which was never being applied for these two popular industrial sites data set in Malaysia, can be considered as the novelty of this paper. The methodology is further discussed in the following sections.

2   Methodology

2.1   Data and the Study Locations

Secondary data of daily maximum CO levels in (ppm) within the period from year 2010 until 2014 were obtained from Air Quality Division of Department of Environment, Malaysia. The analysis on CO data was conducted to investigate CO variation and to obtain the appropriate time series model to predict CO level at two selected industrial sites, Pasir Gudang in Johor and Perai in Penang, Malaysia. The reasons for choosing these sites is because the sites are among the busiest industrial location in Peninsular Malaysia, equipped with air quality monitoring stations and have high number of registered vehicles recorded. The instrument that used to collect CO is called Gas Filter Correlation CO Analyzer (Model T300). The software used in the analysis are R and MINITAB.

2.2   Research Framework

Figure 1 describes the process flow on how this study is conducted to achieve the objectives. First and foremost, after the data have been collected, a quality checking process was conducted. Missing values were replaced using median imputation method. Exploratory analysis was then conducted to describe the variation and pattern of CO as well as to detect data stationary and seasonal influence. Based on the exploratory analysis the process of building the time series model was conducted followed by performance evaluation and model validation.

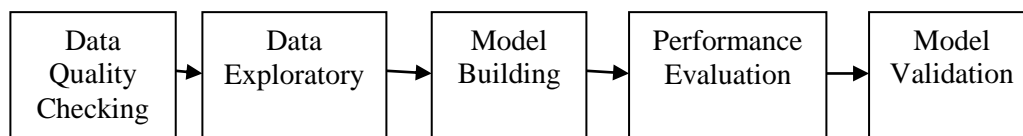| Data Quality Checking | → | Data Exploratory | → | Model Building | → | Performance Evaluation | → | Model Validation |

Figure 1: Flow of research process

The whole data set from 2010 until 2014 was used in exploratory analysis. The results of exploratory analysis will provide the basis on the building and development of CO time series

model. The data set is divided into two parts, daily recorded data from year 2010 to 2013 were used in model building and data of 365 days in year 2014 were used for model validation.

2.3   Time Series Models

Within the framework of the Box-Jenkins methodology, it consists of four basic models including the Autoregressive (AR), Moving Average (MA), Mixed Autoregressive and Moving Average (ARMA) and Mixed Autoregressive Integrated Moving Average (ARIMA) (Lazim, 2007; Mah et. al. 2018).

In the AR model, the current value of the variable is defined as a function of its previous values plus an error term. In other words, the dependent variable, y_t is taken as the function of the time-lagged values of itself. While in the MA model, the moving average is the average of the actual observation, y_t whereas in this case it is a function of the error terms. The MA model links the current values of the time series to random errors that have occurred in the previous periods rather that the values of the actual series themselves. Alternatively, there are series that fit better by combining both AR and MA models under the assumption of data stationary. Such model is known as the Mixed Autoregressive Moving Average (ARMA) model. However, when data is found not stationary, an improved model should be used involving an integrated (I) component. It is known as the Mixed Autoregressive Integrated Moving Average (ARIMA) model. This model is formulated when the stationary assumption of the variable is not met. A degree of data differencing is conducted to overcome stationarity problem. Stationary means that the properties of the series don't depend on the time when it is captured.

The general term for this model is ARIMA (p,d,q) where p refers to the AR order, d is degree of differencing and q is the order of MA. Generally, estimating parameter of ARIMA model is describing dependent variable Y by combining a pth order AR model, d^{th} degree of differencing and a qth order MA model. A non-seasonal ARIMA forecasting equation can be written like this. Let Y denote the original series; y denotes the differenced series. It is defined that:

No difference, when d=0: $y_t = Y_t$ (1)
First difference, when d=1: $y_t = Y_t - Y_{t-1}$ (2)
Second differencing, when d=2: $y_t = Y_t - 2Y_{t-1} + Y_{t-2}$ (3)
And so on.

The forecasting equation for y is denoted as follows:

$$\hat{y}_t = \mu + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} - \theta_1 e_{t-1} \dots - \theta_q e_{t-q} \tag{4}$$

Where μ - constant, ϕ - AR parameter and θ - MA parameter.

The construction of an ARIMA model must follows several stages as shown below:

Stage 1: Stationaries the series, if necessary, by differencing.
Stage 2: Plot ACF and PACF to study the pattern of autocorrelations and partial autocorrelations to determine if lags of the stationaries series and/or lags of the forecast errors should be included in the forecasting.
Stage 3: Identify the initial value for parameter p and q from ACF and PACF plot.
Stage 4: Fit the model that is suggested and check its residual diagnostics.

Often time series possess a seasonal component that repeats itself after every *S* observation (*S*=12 for monthly and *S*=4 for quarterly). In order to deal with seasonality, ARIMA processes have been generalized. ARIMA model for series with seasonal component present, an additional differencing is necessary to be performed in order to eliminate the seasonal effect. It is called '*seasonal differencing*'. This model is known as SARIMA. The non-seasonal & seasonal differences are combined to stationarize the series such that:

No difference, when d=0: $y_t = Y_t - Y_{t-s}$ (5)

First difference, when d=1: $y_t = Y_t - Y_{t-1} - (Y_{t-s} - Y_{t-s-1})$ (6)

And so on.

Seasonal ARIMA models rely on seasonal lags and differences to fit the seasonal pattern. The seasonal part of an ARIMA model is summarized by three additional numbers; P – the number of seasonal autoregressive terms, D – the number of seasonal differences and Q – the number of seasonal moving-average terms. The SARIMA term can be written as SARIMA (*p,d,q*)(*P,D,Q*)*s* where *s* is the number of period in a season. This is represented more formally by equation 7;

$$\lambda(B^s)\phi(B)(y_t - \mu) = \eta(B^s)\theta(B)\xi_t \quad (7)$$

Where *S* is the number of periods per season, y_t is the time series observation at time *t*, ξ_t is white noise, μ is the mean of the series, λ is seasonal AR parameter, ϕ is non-seasonal AR parameter, η is the seasonal MA parameter, θ is non-seasonal MA parameter and B is the back shift operator. The back shift operator can be simplified in AR and MA term as follows:

$$\phi(B)y_t = 1 - \phi_1 y_{t-1} - \dots - \phi_p y_{t-p} \quad (8)$$

and

$$\theta(B)\xi_t = 1 - \theta_1 \xi_{t-1} - \dots - \theta_q \xi_{t-q} \quad (9)$$

The non-seasonal components for AR and MA are;

$$\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p \quad (10)$$

and

$$\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q \quad (11)$$

The seasonal components for AR and MA are:

$$\lambda(B) = 1 - \lambda_1 B^s - \dots - \lambda_p B^{ps} \quad (12)$$

and

$$\eta(B) = 1 - \eta_1 B - \dots - \eta_q B^{qs} \quad (13)$$

It is noticed that, on the left hand side of equation (7), the seasonal and non-seasonal AR components multiply each other, and on the right hand side, the seasonal and non-seasonal MA components multiply each other.

2.3 Performance Indicator

Several performance indicators were used to identify the most appropriate Time Series model for the CO data at the study locations which include Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE) and Akaike Information Criterion (AIC) (Lazim, 2007). For n series of data points, with k numbers of parameter in the model, y_t is an actual data at time t, $\hat{y}_t$ is the predicted value, ε_t is the error defined by $\varepsilon_t = y_t - \hat{y}_t$ and $\hat{\sigma}_\varepsilon^2$ is the sum of sample squared residuals. The formula are as follows.

$$MAE = \frac{\sum_t^n |\varepsilon_t|}{n} \tag{13}$$

$$RMSE = \sqrt{\frac{\sum_t^n \varepsilon_t^2}{n}} \tag{14}$$

$$MAPE = \sum_t^n \frac{|(\varepsilon_t / y_t)*100|}{n} \tag{15}$$

$$AIC = n \ln\left(\frac{\hat{\sigma}_\varepsilon^2}{n}\right) + 2k \tag{16}$$

The most appropriate or optimal model is chosen based on the lowest value of MSE, RMSE, MAPE and AIC.

3 Results and Discussion

3.1 Exploratory Analysis

Figure 2 (a) shows the average maximum CO cycle over the seven days' period in a week and Figure 2 (b) exhibits the average monthly fluctuation. CO has increased during the weekdays that reach the peak on Friday at Perai and Thursday at Pasir Gudang. Meanwhile the concentration had drop during weekend. Over a year period, on the average there are two cycles of CO movement with the peak in May and October at Pasir Gudang, meanwhile no obvious cyclical pattern was observed at Perai. Month of May is the lowest while the highest is in December.
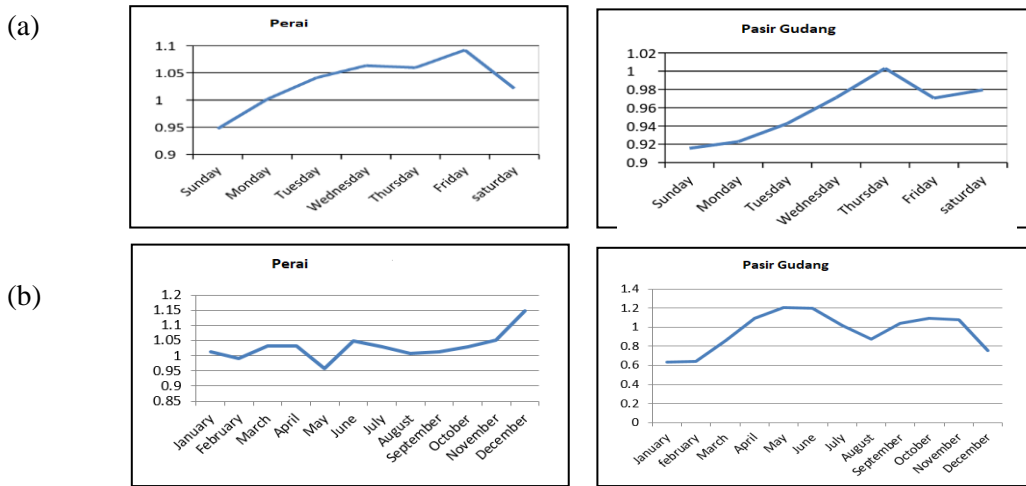
(a)

(b)

Figure 2. The average Maximum CO by (a) days of week and (b) month
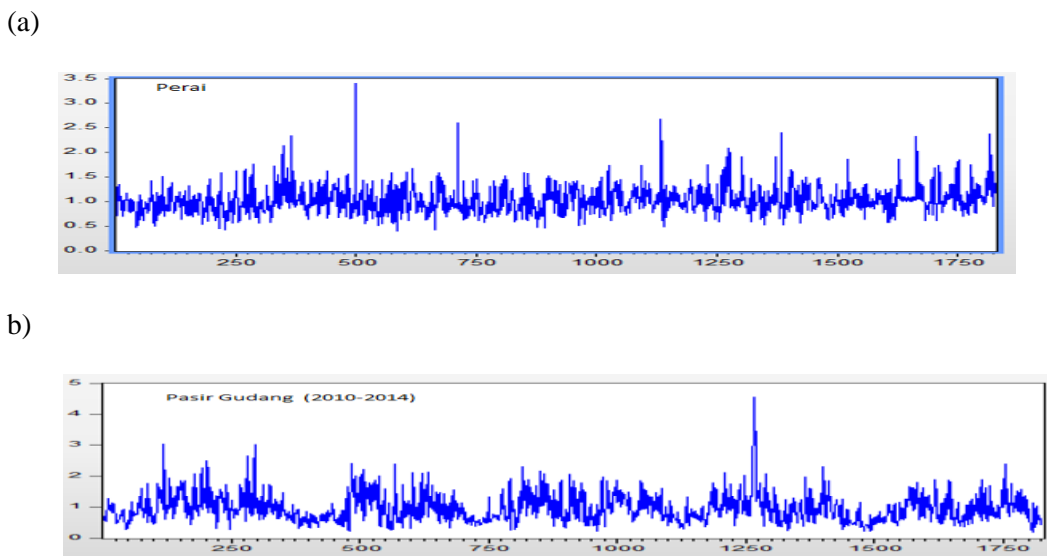at Perai and Pasir Gudang

(a)

b)

Figure 3. Time series plot of daily maximum CO for the period (2010-2014) at (a) Perai and (b)
Pasir Gudang

The time series plot for the data set in Figure 3 (a) indicates there exist an irregular pattern at Perai. For Pasir Gudang data set, Figure 3 (b) also depicts irregular and some cyclical pattern which indicates the present of seasonal behavior. With the null hypothesis that CO time series is stationary, the Augmented Dickey-Fuller (ADF) test results of p-value equals 0.01 has provided the evidence that the CO data are not stationary for both Perai and Pasir Gudang. This results suggest for data differencing to be conducted to improve this situation of non-stationarity. Other than that, based on Pettitt's test, the CO data is also proven not homogeneous for Perai and Pasir Gudang sites. The null hypothesis that a time series is homogeneous need to be rejected due to very small p-value (0.000) at both sites respectively.

3.2 Determining Time Series Model for Daily Maximum CO at Perai Site

In order to increase the quality of data stationary, first differencing was performed based on the results in Figure 4 (a) which is based on the decay pattern of ACF plot and one high spike followed by several smaller spike in PACF plot. Figure 4 (b) shows that the data is stationary.
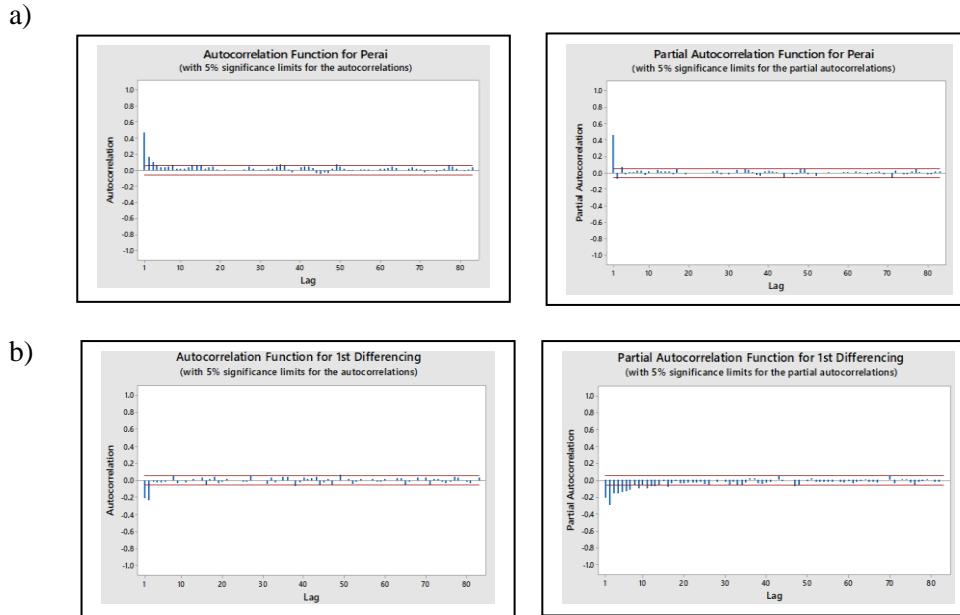
a)

b)

Figure 4. ACF and PCF plot (a) before and (b) after first differencing for CO data at Perai

The next step is to identify the parameter *p, d* and *q* for ARIMA model. Parameter *d*=1 while based on Figure 4 (b), from the ACF plot, *q* is between 1 and 2 while PACF suggest that *p* is between 2 to 7. Thus we begin the initial model with the first *p*=2 significant spike for AR and the first *q*=1 significant spike for degree of MA shown by ACF and PACF respectively; ARIMA (2, 1, 1). Based on this model several other models will be proposed and the performance are compared to determine which one the best is given by the lowest MSE, RMSE, MAPE and AIC values.

Table 2: Performance comparison among several proposed ARIMA models

Model: ARIMA	Performance Indicator			
	MAE	RMSE	MAPE	AIC
(2,1,1)	0.1748	0.2417	17.910	9.31
(2,1,2)	0.1746	0.2412	17.8811	5.63
(3,1,1)	0.1745	0.2412	17.8687	5.49
(3,1,2)	0.1745	0.2142	17.8687	6.84
(4,1,1)	0.1745	0.2412	17.8704	8.81
(4,1,2)	0.1745	0.2412	17.8691	8.86
(5,1,1)	0.1745	0.2412	17.8700	8.81
(5,1,2)	0.1745	0.2412	17.8704	10.69

The best model is ARIMA (3,1,1) and the equation can be written as follows:

$$(1-\phi_1 B - \phi_2 B^2 - \phi_3 B^3)(1-B)y_t = (1-\theta_1 B)\xi_t,$$

By simplification,

$$(1 - B - \phi_1 B + \phi_2 B^3 - \phi_3 B^3 + \phi_3 B^4)y_t = (1-\theta_1 B)\xi_t,$$

Hence,

$$y_t - y_{t-1} - \phi_1 y_{t-1} + \phi_2 y_{t-3} - \phi_3 y_{t-3} + \phi_3 y_{t-4} = \xi_t - \theta_1 \xi_{t-1},$$

$$y_t = y_{t-1} + \phi_1 y_{t-1} - \phi_2 y_{t-3} + \phi_3 y_{t-3} - \phi_3 y_{t-4} + \xi_t - \theta_1 \xi_{t-1}$$

Where,

$\hat{\phi}_1 = AR(1) = 0.5014$, $\hat{\phi}_2 = AR(2) = -0.1119$, $\hat{\phi}_3 = AR(3) = 0.0640$ and $\hat{\theta}_1 = MA(1) = -0.9927$

Figure 5 illustrates the forecasting of the Daily Maximum CO for Prai sites for 30 days ahead. It is shows that CO maximum can be between 0.5 to 2.0 ppm in the next 30 days' period. The thick blue line represents the mean of the forecasted values.
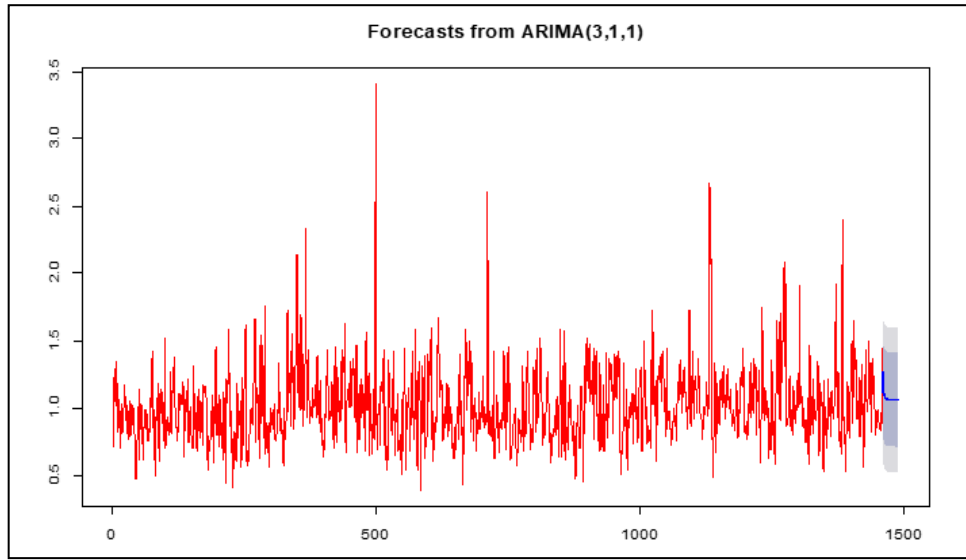
Figure 5. Forecasted value of daily maximum CO for 30 days ahead at Perai

The ARIMA (3,1,1) model is fitted to the validation data, the daily maximum CO for the period from January 2014 until December 2014. The result is shown in Table 3.

Table 3: Model Validation Results for Perai

Model	MAE	RMSE	MAPE	AIC
Training	0.1745	0.2412	17.8687	5.49
Validation	0.1534	0.2153	14.6162	-71.61

Based on the results, it can be concluded that model ARIMA (3,1,1) is a valid model since the performance measures of the training data set has improved (i.e. shown by the reduction of MAE, RMSE, MAPE and AIC).

3.3 Determining Times Series Model for Daily Maximum CO at Pasir Gudang

Seasonality in a time series is a regular pattern of changes that repeats over S time periods, where S defines the number of time periods until the pattern repeats again. Figure 3 indicates that the daily maximum CO data contains seasonal component. However, the number of time periods of the seasonal cycles was not obvious. For this study, based on the results shown in Figure 2 (b), since the data is daily data, it is decided to let S=7 the period of a cyclical CO to repeat over seven days in a week. Because there is no obvious linear trend and no obvious seasonality based on exploratory data analysis results in section 3.1, both seasonal and non-seasonal differencing will be performed. The results of ACF and PACF after the process has been conducted are shown in Figure 6.

The ACF and PACF were plotted to collect conclusive evidence on stationarity condition and so as to guest for initial model parameter for SARIMA $(p,d,q)(P,D,Q)_7$. The results of Figure 4 (a) indicates, that it is still cannot be concluded that the data series in seasonal differencing is

not yet stationary, so in order to render the original series stationary, the first differencing for non-seasonal must be performed.
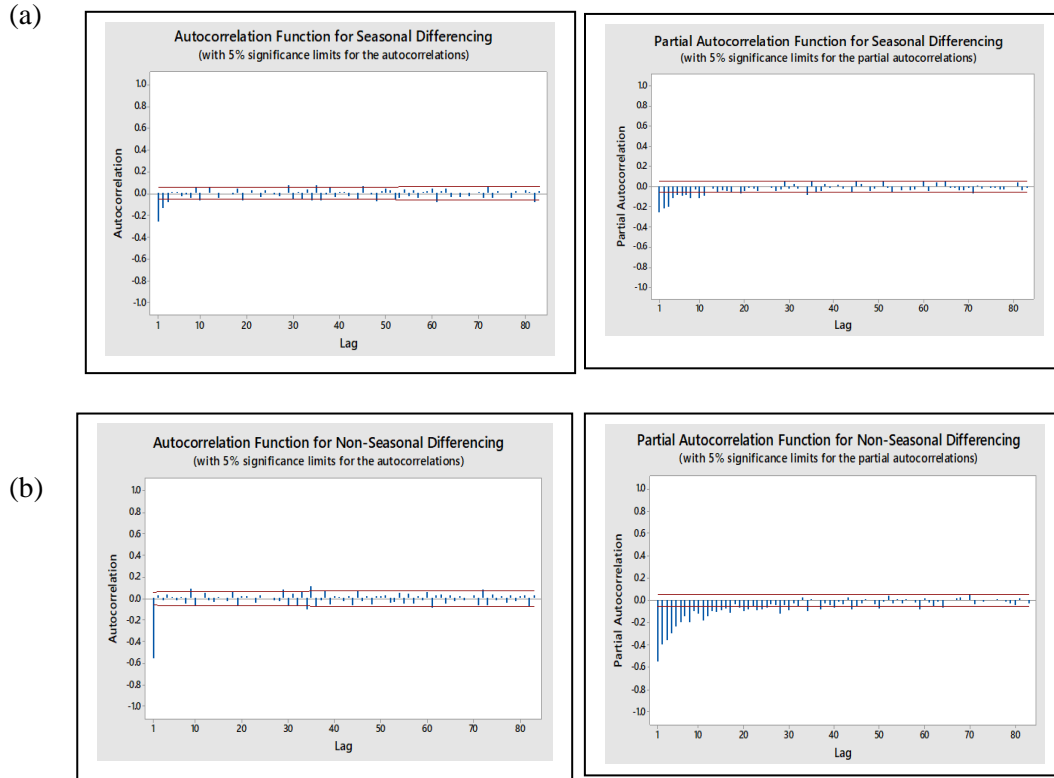
(a)

(b)

Figure 6. ACF and PCF plot for seasonal differencing (a) and for non-seasonal differencing (b)

When looking at the results after the first differencing for non-seasonal component, the ACF has shown a decay pattern, unfortunately not for PACF as shown in Figure 4 (b). Hence, aiming for improvement, a second differencing was conducted. However, it is found that the results were getting worse. Therefore, it was then decided to stop apply differencing and conclude that data series only be performed with first differencing. In order to determine the appropriate model formulations to be fitted to the data series one needs to observe for the behavior of significant spike(s) in ACF and PACF plots.

From ACF plot in Figure 4 (b), two significant spikes are observed which at point lag 1 and 34. This one spike can be used to specify the non-seasonal MA part of the model. Another significant spike is also observed at lag 35 to suggest the seasonal SMA part of the model. Similarly, to identify the AR part of the model one needs to observe the PACF plot. Eight significant spikes are observed, one at lag 1, 2, 3, 4, 5, 6, 8 and lag 11 to suggest the non-seasonal AR part of the model. There are also fairly significant spikes at lag 7, 14 and 28 to indicate the seasonal SAR part of the model. However, even if these observations are made it is still cannot be perfectly sure of the correct values of the respective *p, q*, P and Q that can be assigned to the model. Hence, to ensure that a well specified model is not missed out, several model formulations will be proposed and the performances are compared.   In this study, guided and supported by the

previous results, the parameter we let the *d*=D=1 and *p*=2, q is in the range [1, 8], P=1 and Q is in the range [1,3].

Table 4: Performance comparison among several proposed SARIMA models

Model: SARIMA	Performance Indicator			
	MAE	RMSE	MAPE	AIC
$(2,1,8)(1,1,3)_7$	0.25705	0.34634	30.6909	1092.89
$(2,1,8)(1,1,2)_7$	0.25674	0.34697	30.6800	1090.87
$(2,1,8)(1,1,1)_7$	0.25762	0.34839	30.6910	1091.26
$(2,1,6)(1,1,1)_7$	0.25815	0.34898	30.6800	1104.59
$(2,1,4)(1,1,1)_7$	0.25857	0.3484	30.73682	1101.39
$(2,1,3)(1,1,1)_7$	0.25818	0.34897	30.73341	1098.54
$(2,1,2)(1,1,3)_7$	0.25685	0.34774	30.7658	1094.03
$(2,1.1)(1,1,2)_7$	0.25713	0.34785	30.8019	1096.12

The results and the performance of the several proposed models are given in Table 4. It is found that the model that gave the best performance is SARIMA $(2, 1, 8) (1, 1, 2)_7$ when fitted to the training data set. Backward shift operator for SARIMA is written as:

$$(1-\phi_1 B - \phi_2 B^2)(1-B)(1-\phi_1 B^7)(1-B^7)y_t$$
$$= (1-\theta_1 B - \theta_2 B^2 - \theta_3 B^3 - \theta_4 B^4 - \theta_5 B^5 - \theta_6 B^6 - \theta_7 B^7 - \theta_8 B^8)(1-\theta_1 B^7 - \theta_2 B^{14})\xi_t$$

Where,

$\hat{\phi}_1 = AR(1) = -1.0781,, \quad \hat{\phi}_2 = AR(2) = -0.7917, \quad \hat{\theta}_1 = MA(1) = 0.4666, \quad \hat{\theta}_2 = MA(2) = 0.0945,$

$\hat{\theta}_3 = MA(3) = -0.8681, \quad \hat{\theta}_4 = MA(4) = -0.276, \quad \hat{\theta}_5 = MA(5) = -0.1576, \quad \hat{\theta}_6 = MA(6) = -0.0561,$

$\hat{\theta}_7 = MA(7) = -0.1063,, \quad \hat{\theta}_8 = MA(8) = -0.0971, \quad \hat{\phi}_1 B^7 = SAR(1) = -0.5273, \quad \hat{\theta}_1 B^7 = SMA(1) = -0.3072$

and

$\hat{\theta}_2 B^{14} = SMA(2) = -0.5796$

Figure 7 illustrates the forecasting of the Daily Maximum CO for Pasir Gudang site for 30 days ahead. It is shows that CO maximum can be between 0.01 to 1.5 ppm in the next 30 days' period. The thick blue line represents the mean of the forecasted values.
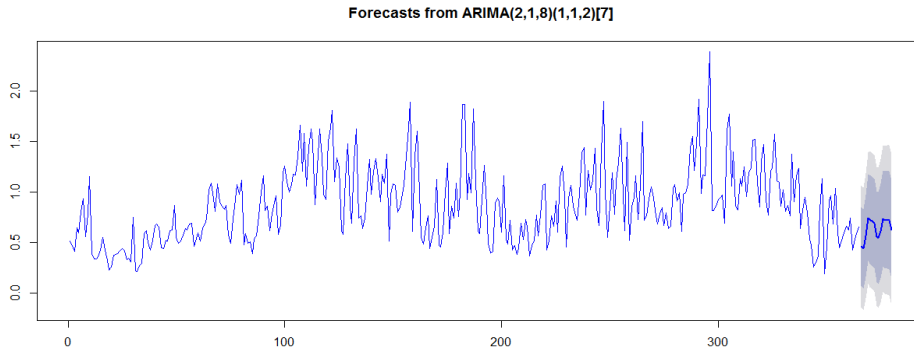
Forecasts from ARIMA(2,1,8)(1,1,2)[7]

Figure 7: The forecast value for Pasir Gudang site

The validity of the obtained model is verified, using validation data set (i.e. daily maximum CO for year 2014) and the results is shown in Table 4. The reduction in the value of MAE, RMSE, MAPE and AIC provided the evidence that the model is valid.

Table 5: Model Validation Results for Pasir Gudang

Model	MAE	RMSE	MAPE	AIC
Training	0.25674	0.34697	30.6800	1090.87
Validation	0.22834	0.2228	29.30966	1082.56

4   Conclusion

In summary, this study has shown that Perai and Pasir Gudang sites have different temporal behavior of CO levels. By day of week basis, the results indicate that Perai site has the highest traffic level on Thursday. While at Pasir Gudang the day with highest maximum CO is Friday. However, both sites experiencing lowest CO during weekend (Sunday) and less prominent at the beginning month of the year. Pasir Gudang has clear pattern of bimodal peak in May and October while Perai has the lowest maximum CO in May with no clear pattern of monthly cyles. Daily maximum CO time series at Perai shows an irregular pattern at Perai while for Pasir Gudang data, the series depicts irregular and some cyclical pattern which indicates the present of seasonal behavior. Based on several performance measures such as MAE, RMSE, MAPE and AIC, the study has found that the best appropriate prediction model for Perai is ARIMA(3,1,1) and SARIMA $(2,1,8)(1,1,2)_7$ for Pasir Gudang.

Acknowledgement

References

Afroz, R., Hassan, M.N. and Ibrahim, N.A. 2003. Review of air pollution and health impacts in Malaysia. *Environmental Research. 92*(2), 71-77.

Afsaneh, A. Trends and Prediction of Air Pollutants in Pasir Gudang Industrial Area, Johor. 2014. PhD thesis, UTM, Johor. http://dms.library.utm.my:8080/vital/access/manage...

Azedah, G., Abdullah, A.M., Mande, K.H., Nazanin, A., Raziyeh, S., Aramesh, A. and Tahoora, S.N. 2015. Review on CO2 Emission from Transport Sector in Malaysia. *Journal of Environmental Science, Toxicology and Food Technology 9*(5), 61-70.

Chen, R., Pan, G., Zhang, Y., Xu, Q., Zeng, G., Xu, X., Chen, B. and Kan, H. 2011. Ambient carbon monoxide and daily mortality in three Chinese cities: the China air pollution and health effects study (CAPES*). Science of Total Environment 409*(23), 4923–4928.

Fore, S. and Mbohwa, C.T. 2010. Cleaner Production for Environmental Conscious. Manufacturing in the Foundry Industry. *Journal of Engineering Design and Technology 8*(3), 314-333.

Hamid, A.H., Japeri, A.Z. Ul-Saufie. M. and Ahmat, H. 2017. Characteristic and Prediction of Carbon Moxide Concertration. MATEC Web of Conferences 103, 1-8.

Ibrahim, M.Z, Ismail, M., Zailan, R. and Lola, M.S. 2009. Forecasting and Time Series Analysis of Air pollutants in several area of Malaysia. *American Journal environmental science 5*(5), 625-632.

Lazim, M. A. 2007. *Introductory Business Forecasting a practical approach* (3rd ed.), Publisher UiTM.

MAA, Malaysia Automotive Association. 2016. Market Review for 2015 and Outlook for 2016, Press Conference, 21 January 2016.

Mah, P., Ihwal, N., & Azizan, N. 2018. Forecasting Fresh Water And Marine Fish Production In Malaysia Using Arima And Arfima Models. *Malaysian Journal of Computing, 3*(2), 81-92.

Mamat, M., Mashor, M.Y., Sadullah, A.F. and Saad, A.R. 2003. Carbon Monoxide Concentration Forecasting Using Hybrid Radial Basis Function Network. Proceeding of the International Conference on Robotices, Vision, Information and Signal Processing, Penang Malaysia.

Mahmoudzadeh, S., Othman, Z., Yazdani, A.M. and Bakar, A.A. 2013. Carbon Monoxide Prediction Using Artificial Newral Network and Imperialist Competitive Algorithm. *Australian Journal of Basic and Applied Science 7*(4), 735-744

Miller, K.A., Siscovick, D.S., Sheppard, L., Shephard, K., Sullivan, J.H., Anderson, G.L. and Kaufman, J.D. 2007. Long-Term Exposure to Air Pollution and Incidence of Cardiovascular Events in Woman. *The New England of Medicine* 356, 447-458.

Noresah, M.S. Private Vehicle Ownership and Transportation Planning in Malaysia. 2012. In proceedings of International Conference on Traffic and Transportation Engineering.

Pasca, M., Pascal, L., Bidondo, L.M, Cochet, A. Sarter, H. Stempfelet, M. and Wagner, V. A 2013. Review of the Epidemiological Methods Used to Investigate the Health Impacts of Air Pollution around Major Industrial Areas. *Journal of Environmental and Public Health. 2013,* 1-17

Raji, U., Mashor, M.Y. Adom, A.H., Ali, A.N. and Sadullah, A.F. 2005. Performance Comparison between HMLP, MLP and Recurrent Networks with Applications to Carbon Monoxide Concentrations Forecasting. *WSEAS Transaction on System. 4*(6), 812-820

Rathi, S., Cik Soh, S., & Akbarally, A. 2018. Coefficient Bounds For A Certain Subclass Of Close-To-Convex Functions Associated With The Koebe Function. *Malaysian Journal of Computing, 3*(2), 172-177.

Sembiring, R. W., Mohamad Zain, J., and Embong, A. 2010. "A Comparative Agglomerative Hierarchical Clustering Method to Cluster Implemented Course", Journal of Computing, Vol. 2, Issue 12 , December 2010, pp. 33-38.

Shuhaili, A.F. A., Ihsan, S. I., Faris, W. F. 2013. Air Pollution Study of Vehicles Emission In High Volume Traffic: Selangor, Malaysia As A Case Study. *WSEAS TRANSACTIONS on SYSTEMS 2*(12), *56-304*

Xu, M., Duan, X., Wang, Z. and Pau, B. 2014. Effects of Pollution on Aquatics Ecology and Water quality Bio Assessment. *Hydrobiology 729*(1), 1-13

Yahya, N., Ali, A., Ishak, F. 2006. Air Pollution Index (API) and the Effect on Human Health: Case Study in Terengganu City, Terengganu, Malaysia. Paper submitted to the International Association for People-Environment Studies (IAPS) Conference, September 2006, Alexandria, Egypt.