

Characteristic and Prediction of Carbon Monoxide Concentration using Time Series Analysis in Selected Urban Area in Malaysia

Hazrul Abdul Hamid^{1*}, *Ahmad Zia Ul-Saufie Mohd Japeri*², and *Hasfazilah Ahmat*²

¹Faculty of Civil and Environmental Engineering, Universiti Tun Hussein Onn Malaysia

²Faculty of Computer Sciences and Mathematics, Universiti Teknologi MARA

Abstract. Carbon monoxide (CO) is a poisonous, colorless, odourless and tasteless gas. The main source of carbon monoxide is from motor vehicles and carbon monoxide levels in residential areas closely reflect the traffic density. Prediction of carbon monoxide is important to give an early warning to sufferer of respiratory problems and also can help the related authorities to be more prepared to prevent and take suitable action to overcome the problem. This research was carried out using secondary data from Department of Environment Malaysia from 2013 to 2014. The main objectives of this research is to understand the characteristic of CO concentration and also to find the most suitable time series model to predict the CO concentration in Bachang, Melaka and Kuala Terengganu. Based on the lowest AIC value and several error measure, the results show that ARMA (1,1) is the most appropriate model to predict CO concentration level in Bachang, Melaka while ARMA (1,2) is the most suitable model with smallest error to predict the CO concentration level for residential area in Kuala Terengganu.

1 Introduction

Air pollution is one of the important issues to deal with. Air pollution that happened today is from the result of a natural cycle phenomenon in which a very large impact on health in particular and economic development [1]. In Malaysia, Department of Environment (DoE) is responsible in monitoring the level of air pollutants and there are 52 monitoring locations belong to Department of Environment. There are 5 main pollutant that are significant and concerned by Department of Environmental Malaysia which is ground level ozone (O₃), carbon monoxide (CO), nitrogen dioxide (NO₂), sulfur dioxide (SO₂), and fine dust with aerodynamic diameter less than 10 micrometer (PM₁₀).

Carbon monoxide is one of the main pollutants in residential area since the main source of carbon monoxide is from motor vehicles [2]. Generally, Carbon Monoxide (CO) is a by-product of incomplete combustion of Hydrocarbon Fuels. Two kinds of health effects are attributed to CO, very long term and short term. Long periods of exposure to high

* Corresponding author: hazrul@uthm.edu.my

concentration of CO can raise the level of carboxyhemoglobin in the blood. Due to the high level of carboxyhemoglobin in blood, blood loses the capability to absorb oxygen. Due to this the amount of oxygen which is carried by blood to different parts of body is reduced [3]. Short periods of exposure to very high concentrations of CO can precipitate or hasten the death of individuals with severely impaired heart or respiratory functions. This hazardous particular substances can harm us by disturb our healthy body system. Every part of our body needs oxygen. Air is made up of approximately 20.9% oxygen and 79% nitrogen. The hemoglobin within our blood carries oxygen to every cell in our body. The hippocampus is the part of the brain that is involved in memory forming, organizing and storing. It is a limbic system structure that is particularly important in forming new memories and connecting emotions and senses, such as smell and sound, to memories [3].

Dangerous of carbon monoxide worsening because it can causes health effects to patients suffering respiratory problems. This is due it prevents the body from using oxygen as well as potential cancer patients [4]. Therefore prediction of carbon monoxide is important, especially as initial steps by give an early warning sufferers of respiratory problems such as asthma, wheezing as a precaution as well as early warning for us or the related authorities to be more prepared to prevent and take suitable action to overcome the problem. The main focus in this research is to understand the characteristic of CO concentration and also to find the most suitable time series model to predict the CO concentration in Bachang, Melaka and Kuala Terengganu.

2 Data and study area

Bachang and Kuala Terengganu are categorised by Department of Environment Malaysia as an urban area located in Melaka and Terengganu, respectively. Previously, this two monitoring station was classified as a residential area [5]. Geographically, Melaka state is situated on the western coast of peninsular Malaysia with the coordinates is latitude $02^{\circ} 12.78'N$ of the equator and longitude $102^{\circ} 14.05'E$ of the prime meridian. Terengganu state is situated on the northwest coast of peninsular Malaysia. The coordinates is $05^{\circ} 20.23'N$ latitude and $103^{\circ} 09.45'E$ longitude. Hourly average of CO concentration in this two location for 2013 and 2014 was obtained from the Department of Environment Malaysia and was measured in part per million (ppm) using carbon monoxide gas detector at the continuous air quality monitorind stations. Monitoring data was computed by averaging direct measurement from the monitoring sites on a yearly basis and cross-reference with Malaysia Ambient Air Quality Guidelines as shown in Table 1.

Table 1. Malaysia Ambient Air Quality Guideline for Carbon Monoxide

Averaging Time	ppm	mg/g ³
1 Hour	30.0	35.0
8 Hours	9.0	10.0

3 Procedure of time series analysis

3.1 Stationarity

Unit root test which is Augmented Dickey-Fuller (ADF) test was used to test the stationarity of data since this is the most commonly used test in time series analysis to determine the stationarity of any data set. The ADF test uses the following Equation (1):

$$ADF = \alpha_0 + \rho_1 y_{t-1} + \sum_{j=2}^{p-1} \beta_j \nabla y_{t-j} + e_t \quad (1)$$

where α_0 is drift components and e_t is error term [6]. The hypothesis to determine the stationary of the data is as follows:

- H_0 : the time series data is non-stationary
 H_1 : the time series data is stationary

The null hypothesis will be rejected if the ADF value is greater than the critical value.

3.2 Model Identification and Estimation

Each time series model were identify part by part starting from Moving Average (MA), Autoregressive (AR), and Autoregressive Moving Average (ARMA). In this study Akaike Infomation Criterion was used as medium of lag length which is to identify the compatibility of error from those process come out either too small or large of error. In estimation period, each model is directly involved and process of its own equation. Their combination will create a new value that will be used in the next stage of analysis. Among the time series model used is :

- i. Moving Average (MA) Process
- ii. Autoregressive (AR) Process
- iii. Autoregressive Moving Average (ARMA) or Autoregressive Integrated Moving Average (ARIMA) Process

The moving average process of order q is denoted MA(q) and defined by

$$X_t = \sum_{s=0}^q \theta_s \epsilon_{t-s} \quad (2)$$

where $\theta_1, \dots, \theta_q$ are fixed constants, $\theta_0 = 1$, and $\{\epsilon_t\}$ is a sequence of independent (or uncorrelated) random variables with mean 0 and variance σ^2 [7].

The autoregressive process of order p is denoted AR(p), and defined by the following Equation [7].

$$X_t = \sum_{r=1}^p \phi_r X_{t-r} + \epsilon_t \quad (3)$$

where ϕ_1, \dots, ϕ_r are fixed constants and $\{\epsilon_t\}$ is a sequence of independent (or uncorrelated) random variables with mean 0 and variance σ^2 .

The autoregressive moving average process, ARMA(p, q), is combination of AR(p) model and MA(q) model which is can defined by

$$X_t = \sum_{r=1}^p \phi_r X_{t-r} + \sum_{s=0}^q \theta_s \epsilon_{t-s} \quad (4)$$

where again $\{\in t\}$ is white noise. This process is stationary for appropriate ϕ, θ .

3.3 Validation

To judge the developed model, three error measures were used. The root mean square error method is the most common indicator. For a good estimator, the RMSE value must approach zero. Therefore, a smaller RMSE value means that the model is more appropriate by [8]. It can be defined by :

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (P_i - O_i)^2} \quad (5)$$

The normalized absolute error is a more sensitive measure of residual error than RMSE [8]. It is defined as

$$NAE = \frac{\sum_{i=1}^N |P_i - O_i|}{\sum_{i=1}^N O_i} \quad (6)$$

Mean absolute percentage error is a relative error measure that uses absolute values to keep the positive and negative errors from canceling one another out and uses relative errors to enable you to compare forecast accuracy between time-series models. The formula for calculating the MAPE is as follows :

$$MAPE = \frac{\sum_{i=1}^N \left| \frac{(O_i - P_i)}{O_i} \right|}{N} \times 100 \quad (9)$$

where N is the number of monitoring records, O_i is the observed monitoring records and P_i is the predicted monitoring records [9].

4 Results and discussions

This section will discuss the results that have been obtained from the analysis of CO concentrations in Bachang, Melaka and Kuala Terengganu. Time series plot and descriptive statistics are used to get a general overview and compare the level of CO concentration at both monitoring stations before carried out the procedure for prediction.

4.1 Time Series Plot and Descriptive Statistics

Acquisition data for CO concentration level in Bachang, Melaka and Kuala Terengganu show respectively the time series plotting. Fig.1 and Fig.2 show the time series plot for Bachang and Kuala Terengganu respectively. Missing values were replaced using the mean top bottom method. Descriptive statistics for both monitoring site are shown in Table 2. Basically, the mean and mode of CO concentration at both monitoring sites are almost the

same and both set of data are positively skewed, meaning that most of recorded data are lower than median.

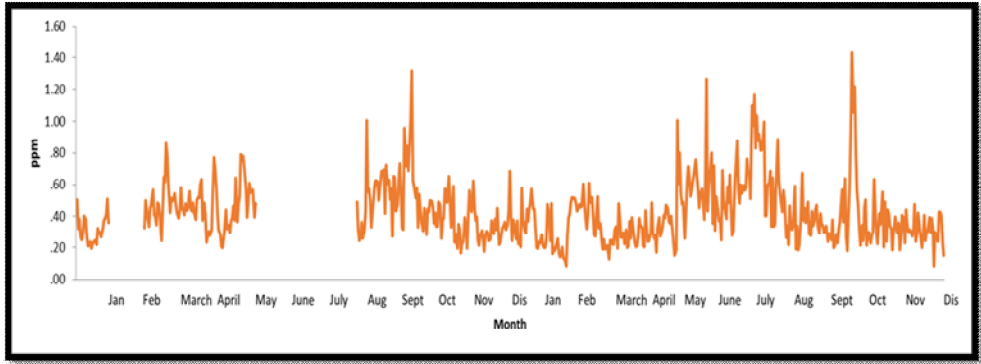


Fig. 1. Time Series plotting of CO concentration in Bachang, Melaka with missing value

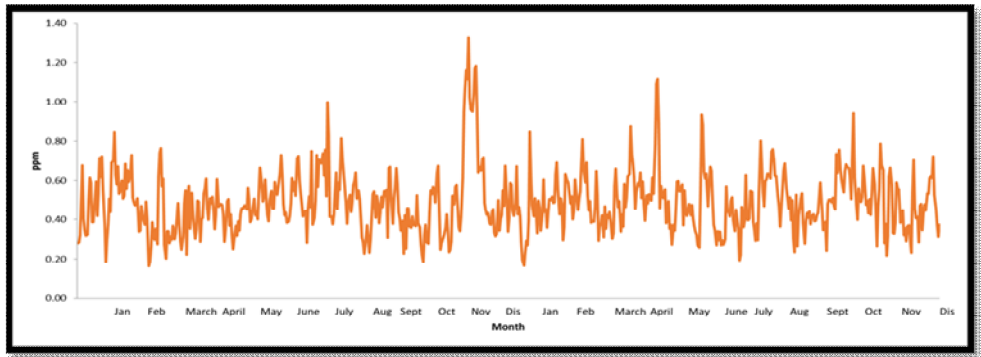


Fig. 2. Time Series plotting of CO concentration in Kuala Terengganu.

Table 2. Characteristics of CO concentration level in Bacang and Kuala Terengganu

	Bachang, Melaka	Kuala Terengganu
Mean	0.428	0.488
Median	0.410	0.470
Mode	1.430	1.330
Standard deviation	0.178	0.159
Skewness	1.463	1.202

4.2 Time Series Analysis

Results from Augmented Dickey Fuller test shows that the monitoring record of CO concentration from 2013 to 2014 are stationary for both monitoring stations with ADF statistic for Bachang, Melaka is -4.562 and ADF statistic for Kuala Terengganu is -4.106 . Since both p -value are less than 0.05 , the time series data can be concluded as stationary.

Autoregressive model (AR model), Moving Average model (MA model) and Autoregressive Moving Average model (ARMA model) were considered to find the most

appropriate model for these set of data by comparing the Akaike Info Criterion (AIC) statistic values. The small AIC statistic values indicate the most appropriate model with smallest error. For Bachang, Melaka, the most suitable model to make the prediction with the smallest AIC statistic value is ARMA (1, 1) where the AIC statistic value is -1.2462 while for Kuala Terengganu, ARMA (1,2) is the best choice based on the smallest AIC statistic value, -1.5151 compared to others.

Time series equation based on the best time series model for Bachang, Melaka is as follows:

$$X_t = 0.429 - 0.836X_{t-1} - 0.314e_{t-1} + e_t \tag{10}$$

For Kuala Terengganu, the time series equation based on the best time series model is

$$X_t = 0.489 + 0.747X_{t-1} - 0.023e_{t-1} - 0.077e_{t-2} + e_t \tag{11}$$

For the validation, 14 days ahead was predicted and the performance indicators for the time series model are summarized in Table 3. For the purpose of seeing more clearly, the 14 days observed and predicted of CO concentration using the most appropriate time series model were plotted in Fig.3 and Fig.4 for CO concentration in Bachang and Kuala Terengganu respectively.

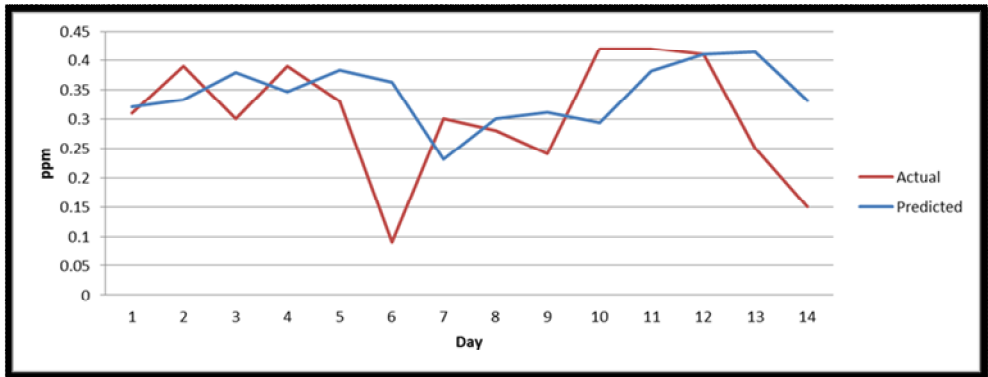


Fig. 3. Actual and predicted of CO concentration using ARMA (1, 1) for Bachang, Melaka.

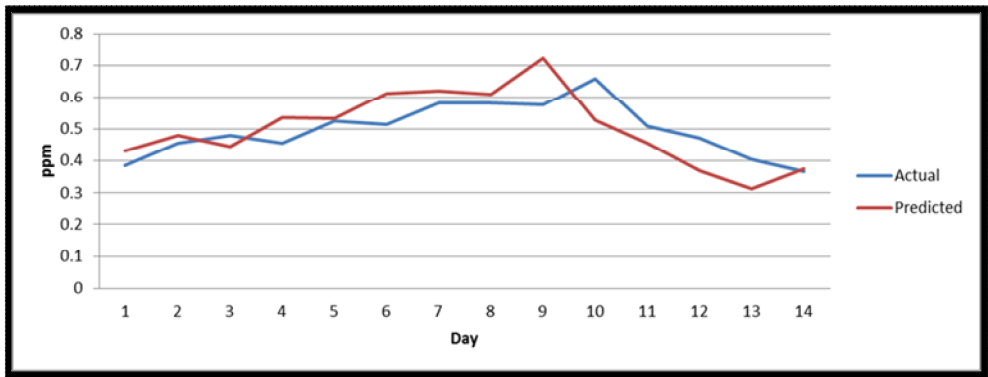


Fig. 4. Actual and predicted of CO concentration using ARMA (1, 2) for Kuala Terengganu

Table 3. Performance indicator

Performance Indicator (Error measure)			
	RMSE	NAE	MAPE
Bachang, Melaka	0.1119	0.0845	17.074
Kuala Terengganu	0.0773	0.0641	13.131

5 Conclusion

Set of secondary data from DOE were analyzed by varying the descriptive statistics and the result shows that mean value of CO concentration in Bachang, Melaka is slightly lower than Kuala Terengganu but for the skewness of CO concentration show, both urban areas have a positive values, indicated that data were skewed to the right which is means most of data are lower than median. To predict the CO concentration level using time series analysis by choosing the appropriate model for both monitoring station located in urban area, only autoregressive (AR), moving average (MA) and autoregressive moving average (ARMA) were considered since both sets of data are stationary. The quality and reliability of the developed models were validated using AIC and three performance indicators (NAE, RMSE, MAPE) and the result of this study show that ARMA (1, 1) is the most suitable model to predict the CO concentration for urban area in Bachang, Melaka. Similarly, another area which is for Kuala Terengganu, ARMA (1, 2) was best fitted to predict the CO concentration level.

References

- [1] Janet C., Neidelf M. and Schmieder J.F, Air Pollution and Infant Health: Lesson from New Jersey, *J. of Health Economic*, **28**, 688-703, (2009)
- [2] Kayode S.J. and Feyisayo K., Air Pollution by carbon Monoxide (CO) poisonous gas in Lagos area Southwestern Nigeria, *Atmospheric and Climate Science*, **3**, 510-514 (2013)
- [3] Akinyemi M.L and Usikali M.R., Investigation of carbon monoxide concentration from anthropogenic sources in Lagos, Nigeria. *Int. J. of Physical Sciences*, **8**(21), 1128-1132. (2013)
- [4] Peter A., Jack A. and Maureen A., *Healthy Home Issues : Carbon Monoxide. U.S Department of Housing and Urban Development*, Version 3, December 2005 (2015)
- [5] Department of Environment Malaysia, *Malaysia Environmental Quality Report 2004. Kuala Lumpur : Ministry of Science, Technology and Environment*, Malaysia (2005)
- [6] Abdel-Aziz A. and Frey, H.C., Development of hourly probabilistic utility NOx emission inventories using time series techniques : Part I – Univariate approach. *Atmospheric Environment*, **37**, 5379 – 5389A (2003)
- [7] Madsen H., *Time Series Analysis*, Chapman & Hall/CRC texts in statistical science series, London (1955)
- [8] Maxim V.S, Adriaan B., Nataliya L.S, Anton P.T., Timur A.J. and Valery A.K., A survey of forecast Error Measures, *World Applied Sciences J.*, **244**, 171-176 (2013)

- [9] Abdul Hamid H., Yahaya A.S., Ramli N.A., Ul-Saufie A.Z., Yasin M.N., Short-term prediction of PM₁₀ concentrations using seasonal time-series analysis. *MATEC Web of Conference* , **47**, 05001, (2016)