International Journal of Engineering & Technology, *7 (3.25) (2018) 420-422*

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Prediction PM_{10} Concentration Using VAR Time Series

Norazrin R.[1,2]*, Ahmad Shukri Yahaya[1], Hazrul Abdul Hamid[3]

[1]*School of Civil Engineering, Engineering Campus, Universiti Sains Malaysia, 14300 Nibong Tebal, Penang, Malaysia*
[2]*School of Environmental Engineering, Universiti Malaysia Perlis, Kompleks Pusat Pengajian Jejawi 3, 02600 Arau, Perlis, Malaysia*
[3]*School of Distance Education, Pusat Pengajian Pendidikan Jarak Jauh, Universiti Sains Malaysia, 11800 Gelugor, Penang, Malaysia*
**Corresponding author E-mail:norazrin.msia@gmail.com*

Abstract

This paper presents a case study from Kangar's monitoring station using a monthly average data (1999-2015). The objective of this study is to predict the PM_{10} concentration by using the VAR time series model. This model was adapted to quantify and understand the interaction of PM_{10} concentration and meteorological parameters for air quality control using (temperature, wind speed, and relative humidity) as independent parameters and particulate matter (PM_{10}) as a dependent parameter. The performance indicator results were (R^2 = 0.887), (IA = 0.954), (PA=0.966), and (NAE=0.087) respectively. This study indicates that the VAR time series model is a good model to predict PM_{10} concentration since the results obtained are close to the performance criteria.

Keywords: *Concentration; PM_{10}; Prediction; Time Series; VAR*

1. Introduction

Time series is a set of observations on the values that a variable takes at different times and widely used in statistics, econometrics [1-2], mathematical finance, weather forecasting, earthquake prediction and many other applications. Multivariate time series is a simultaneous study of several variables to analyze the interrelationships between the variables. It is more informative that univariate analysis. It is well known that VAR models [3] have become an increasingly powerful macroeconomic tool to gauge the dynamic response of a set of endogenous variables to exogenous shocks, and to identify the shocks that dominate the intrinsic volatility in a set of endogenous variables. The vector autoregressive (VAR) model is widely used in practice to test for the existence of a dynamic relationship between economic and financial time series [1, 2,4]. The ability of the VAR model approach to model and forecast is an advantage since only lagged variables are used on the right-hand side. Forecasts of the future values of the dependent variables can be calculated using only information from within the system and we could term these as unconditional forecasts since they are not constructed conditional on a set of assumed values. However, it may be useful to produce forecasts of future values of some variables conditional upon known values of other variables in the system [5].

2. Literature Review

Currently, there are many statistical analysis techniques to interpret environmental data using envirometric techniques. These techniques use multivariate analysis such as cluster, discriminant, and principal components analysis (PCA) [6-9]. Studies by Ul-Saufie et al. (2013) used a hybrid model which combines multiple models such as multiple linear regression with principal components analysis (MLR and PCA) and feedforward backpropagation with principal components analysis (FFFBP and PCA) for prediction [9]. In another study, the meteorological factors were determined to directly influence the air pollution quality at Klang, Perai, and Pasir Gudang as reported by [6]. The negative correlation was summarized between O_3 with NO, NO_2, CO, PM_{10}, and RH. However, the positive correlation for O3 was between SO_2, T, WS, and UV_B. The O_3 has an inverse relationship with rain and a positive relationship with temperature [10]. A study by Mohamed Noor et al. (2015) found that the positive correlation was indicated for PM_{10} with the temperature (r =0.241 to 0.421), while the negative correlation was PM_{10} with humidity (r = -0.118 to -0.406) [11].

Local air pollution issues which involve a high concentration of particulate matter with an aerodynamic diameter of less than 10 µm (PM_{10}) have become the most problematic issues in the cities. However, from 1990 to 2006 the PM_{10} concentrations had declined by 38 percent (38%). However, the annual average PM_{10} concentrations found in 230 cities that were monitored in 2008 was 4.5 times the World Health Organization (WHO) standard (12). Thus, this study is carried out to predict the PM_{10} concentration and VAR model that have been adapted into the air pollution modelling. In particular, the air pollution modelling that uses air quality data to quantify and understand the interaction of PM_{10} concentration with meteorological parameters such as relative humidity (RH), temperature (T), and wind speed (WS) for the purpose of air quality control.

3. Methodology

3.1. Description of Data

In the study, the monthly average air quality monitoring data was used. The data that was used as a case study was obtained from the Department of Environment Malaysia (DOE) for the Kangar monitoring station in Perlis. Due to the fact that Perlis is a small state, only one device Continuous Air Quality Monitoring Station

(CAQMS) is used to measure the air pollution in the state which is located at the Institute Latihan Perindustrian (ILP), Kangar. The data used included the concentration of particulate matter less than 10 microns (PM_{10}), relative humidity (RH), wind speed (WS) and temperature (T). The data were analysed using a statistical software Eviews 9 Student Version with 80 percent (%) of the monitoring data used for VAR model and another 20 percent (%) of the data used for validation.

3.2. Vector Autoregression (VAR) Model

The VAR model is commonly used for forecasting system and analyzing the dynamic impact of random disturbances on the system (13). Unit root test is important and useful in the analysis to examine the stationarity data. Consider the equation as follows:

$$\Delta \, In \, Y_t = \alpha + \beta_t + \delta \, In \, Y_{t-1} + \sum_{i=1}^{p} \beta_i \Delta \, ln Y_{t-i} + \varepsilon_t \tag{1}$$

Where $\Delta \, In \, Y_t$ is the parameter of interest; α, β and δ are coefficients; *t* is the time trend, *p* is the number of lag length and ε is the residual term. In this study, the Augmented Dickey-Fuller (ADF) test is used to check stationarity data. The stationarity of the series can be strongly influenced by its behaviour and properties. The null hypothesis for this test is that there is a unit root.

The test for the unit root has the null hypothesis of H_0 is δ=0. If the coefficient is significantly different from zero, the hypothesis that δ contains a unit root is considered as rejected. If the test on the level series fails to reject, the ADF procedure is then applied to the first-differences of the series.

Rejection leads to the conclusion that the series are integrated of order one, I [1]. A limitation of the ADF test is its assumption that the errors are statistically independent and have constant variances [2,13]**.** The equation of the VAR model was carried out using regression analysis on the lag of the dependent parameter. The lag order selection criteria include Likelihood Ratio (LR) test, Akaike information criteria (AIC), and Schwarz information criterion (SC), Final prediction error (FPE), and Hannan-Quinn information criterion (HQ). The proper selection of lag is important since a long lag can reduce the autocorrelation of the error term, and may result in the inefficient model [4,13].

The Lagrange Multiple (LM) test or Breusch-Godfrey (BG) test is the common statistical test that is used for testing the residual autocorrelation in a VAR model. The null hypothesis of the LM test is that there is no serial correlation in residual up to the specified order. The p-value represents the probability of the test values. If the p-values are greater, these imply that there was no serial correlation in the specified order.

To ensure that the model is well-specified, the stability test was conducted. This was to ensure that the VAR had satisfied the stability condition and to test whether the estimated parameter had changed over time. The characteristic roots of the coefficient matrix were tested; if the roots are less than 1 and lie inside the unit circle the model is valid and stable [13]**.**

3.3. Performance Indicator

The model performance was evaluated by calculating the performance indicators. Performance indicator that used were coefficient of determination (R^2), index of agreement (IA), prediction accuracy (PA), normalized absolute error (NAE), and root mean square error (RMSE). The performance indicator equation used is illustrated in Table 1.

Table 1: Performance Indicators (14)

PI	Equation	Rang e
Coefficient of Determination (R^2)	$R^2 = \left(\frac{\sum_{i=1}^{n}(P_{i-} P)\,(O_i{}_-O)}{n.S_{pred.}S_{obs}}\right)$	[0,1]
Index of Agreement (IA)	$IA = 1 - \left(\frac{\sum_{i=1}^{n}(P_- O_i)^2}{\sum_{i=1}^{n}(\lvert P-\bar{O}\rvert + \lvert O_i-\bar{O}\rvert)^2}\right)$	
Prediction Accuracy (PA)	$PA = \left(\frac{\sum_{i=1}^{n}(P_{i\_}\bar{O})^2}{\sum_{i=1}^{n}(O_{i-}\bar{O})^2}\right)$	
Normalized Absolute Error (NAE)	$NAE = \frac{\sum_{i=1}^{n}\lvert P_{i-}O_i\rvert}{\sum_{i=1}^{n} O_i}$	≤ 0
Root Mean Square Error (RMSE)	$RMSE = \frac{1}{n-1}\sum_{i=1}^{n}(P_{i-}O_i)^2$	

4. Results and Findings

The data from 1999-2011 with four parameters PM_{10}, relative humidity (RH), wind speed (WS) and temperature (T) were used for modelling and the unit root test was conducted to measure the stationarity of all the parameters.

Table 2: Unit Root Test

	ADF Test	
	Level	**p-value**
PM_{10}	-5.521526	Less than 0.001
Temperature	-7.406011	Less than 0.001
RH	-18.09612	Less than 0.001
WS	-6.566101	Less than 0.001

Based on Table 2, the unit root test indicates that the data is significant with the level of p-value less than 0.001. Thus, the null hypothesis is rejected. The time series is stationary; the series can proceed to estimate the VAR model. The lag criteria are chosen based on the summary of the criterion analysis. Table 3 indicates that the results of the lag length criteria for LR (86.07), FPE (3829.3), AIC (24.20), SC (24.91), and HQ (24.49) are at lag 2. The optimal lag for a VAR model is chosen at lag 2 VAR (2).

Table 3: Lag Length Criteria

Lag	**LR**	**FPE**	**AIC**	**SC**	**HQ**
0	NA	2576	28.41	28.49	28.44
1	609.17	5599.4	24.58	24.97	24.74
2	86.07*	3829.3*	24.20*	24.91*	24.49*
3	15.54	4221.5	24.30	25.31	24.71
4	16.61	4607.0	24.38	25.71	24.92
5	23.25	4775.8	24.42	26.06	25.08
6	23.29	4930.7	24.44	26.40	25.24
7	25.52	4982.6	24.45	26.72	25.37
8	18.32	5314.0	24.50	27.08	25.55
9	24.92	5349.7	24.50	27.39	25.67
10	16.71	5757.4	24.56	27.77	25.86
11	13.01	6393.7	24.65	28.17	26.08
12	23.76	6416.2	24.63	28.46	26.19

* indicates lag order selected by the criterion

The VAR (2) Model

$$PM_{10} = 0.641237^* \; PM_{10,\; t\text{-}1} + 0.011468^* \; PM_{10,\; t\text{-}2} - 0.150857^*RH_{,\; t\text{-}1} + 0.162642^* \; RH_{,\; t\text{-}2} + 0.204975^* \; TEMP_{,\; t\text{-}1} - 0.071029^* \; TEMP_{,\; t\text{-}2} + 0.816580^* \; WS_{,\; t\text{-}1} - 0.541649^* \; WS_{,\; t\text{-}2} + 8.176422 \tag{2}$$

The Lagrange Multiple (LM) test was carried out to determine that there is no serial correlation in residual up to the specified order. Table 4 indicates the result for residual test and the results for lag 1 (*p*-value=0.2105) and lag 2 (*p*-value=0.3139) which are not significant; this implies that there is no serial correlation in the specified order.

Table 4: Residual Test

Lags	**LM-Stat**	***P*-value**
1	20.21984	0.2105
2	18.17246	0.3139

The stability test is conducted to ensure that the VAR (2) model satisfies the stability condition. Based on Figure 1, there is no root lies outside the unit circle. Thus, the VAR (2) model has satisfied the stability condition.

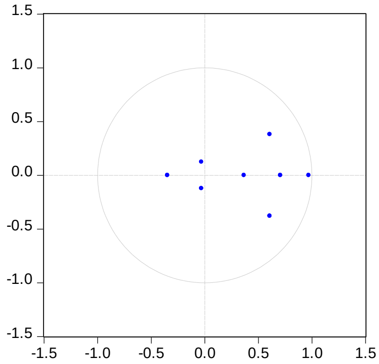

Figure 1:. Stability test

The VAR (2) model is adapted in the monitoring data to predict the PM_{10} concentration, and the performance of the model is evaluated based on the performance indicator results. The actual and predicted of PM_{10} concentration is shown in Figure 2. Based on the plots, the VAR time series model fits the data which indicates that the VAR (2) model is good.

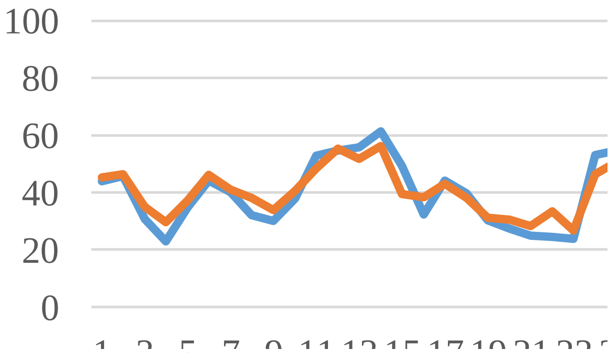

Figure 2:. Plots of Actual and Predicted PM_{10} concentration

The VAR (2) model performance indicators are shown in Table 5.

Table 5: Performance Indicators for VAR (2) Model

Performance Indicators	**Results**	**Range**
R^2	0.877	[0,1]
IA	0.954	
PA	0.966	
NAE	0.087	≤ 0
RMSE	4.886	

The results show that the VAR (2) model has fulfilled the PI criteria. The results obtained for R^2 (0.877), IA (0.954), PA (0.966) and NAE (0.087) are closer to the criteria. Based on the results obtained, the VAR (2) model is a good model to predict the PM_{10} concentration.

5. Conclusion

The findings of this study indicate that the VAR time series model could be adapted for prediction in the air quality studies. This study will serve as a base for future studies that allows for the VAR model to be explored in various research works and studies concerned with prediction and forecasting. Further studies might explore different parameters or variables of the PM_{10} concentration.

Acknowledgement

This study was supported by Ministry of Higher Education Malaysia (MOHE) and Universiti Malaysia Perlis (UniMAP) under SLAB Scheme. We would like to thank the Department of Environment Malaysia (DOE) for providing the air monitoring data.

References

[1] Shaari MS, Hussain NE, Ismail MS. Relationship between Energy Consumption and Economic Growth : Empirical Evidence for Malaysia. Bus Syst Rev. 2012;2(1).

[2] Shaari MS, Rahim HA, Rashid IMA. Relationship Among Population , Energy Consumption and Economic Growth in Malaysia. Int J Sci. 2013;13(1):39–45.

[3] Sims CA. Macroeconomics and Reality'. Sims Source Econom [Internet]. 1980;48(1):1–48. Available from: http://www.jstor.org/stable/1912017%0Ahttp://about.jstor.org/terms

[4] Zhang C, Zhou K, Yang S, Shao Z. Exploring the transformation and upgrading of China's economy using electricity consumption data: A VAR–VEC based model. Phys A Stat Mech its Appl [Internet]. Elsevier B.V.; 2017;473:144–55. Available from: http://dx.doi.org/10.1016/j.physa.2017.01.004

[5] Brooks C. Introductory Econometrics for Finance. Second. Cambridge University Press; 2008. 674 p.

[6] Awang NR, Elbayoumi M, Ramli NA, Yahaya AS. Diurnal variations of ground-level ozone in three port cities in Malaysia. Air Qual Atmos Heal. 2016;9(1).

[7] Dominick D, Talib M, Zain SM, Zaharin A. Spatial assessment of air quality patterns in Malaysia using multivariate analysis. 2012;60:172–81.

[8] Latif MT, Dominick D, Ahamad F, Khan MF, Juneng L, Hamzah FM, et al. Long term assessment of air quality from a background station on the Malaysian Peninsula. Sci Total Environ. Elsevier B.V.; 2014;482–483(2):336–48.

[9] Ul-Saufie AZ, Yahaya AS, Ramli NA, Rosaida N, Hamid HA. Future daily PM10 concentrations prediction by combining regression models and feedforward backpropagation models with principle component analysis (PCA). Atmos Environ. 2013;77:621–30.

[10] Rajab JM, MatJafri MZ, Lim HS. Air Surface Temperature Correlation with Greenhouse Gases by Using Airs Data Over Peninsular Malaysia. Pure Appl Geophys [Internet]. 2013;171(8):1993–2011. Available from: http://link.springer.com/10.1007/s00024-013-0762-y

[11] Mohamed Noor N, Yahaya AS, Abdullah M, Bakri. Variation of air pollutant (particulate matter - PM10) in peninsular Malaysia : Study in the southwest coast of peninsular Malaysia Variation of Air Pollutant (Particulate Matter - PM 10) in Peninsular Malaysia Study in the southwest coast of peninsul. 2015;(August 2016).

[12] Nations U, Programme E, Bank W, Indicators WD, Prices IF, Islands M, et al. Air Pollution and Air Climate Change. In: Statistical Yearbook for Asia and the Pasific. 2011. p. 79–84.

[13] Talbi B. CO2 emissions reduction in road transport sector in Tunisia. Renew Sustain Energy Rev [Internet]. Elsevier; 2017;69(June 2015):232–8. Available from: http://dx.doi.org/10.1016/j.rser.2016.11.208

[14] Ul-Saufie AZ, Yahaya AS, Ramli N, Hamid HA. Performance of Multiple Linear Regression Model for Long-term PM10 Concentration Prediction Based on Gaseous and Meteorological Parameters. J Appl Sci [Internet]. 2012 Dec 1;12(14):1488–94. Available from: http://www.scialert.net/abstract/?doi=jas.2012.1488.1494