Global Journal of Pure and Applied Mathematics.
ISSN 0973-1768 Volume 11, Number 4 (2015), pp. 2187-2190
© Research India Publications
http://www.ripublication.com

Predicting PM_{10} with Multiple Linear Regression Equaiton in Rayong

J. Mekparyup[1], K. Saithanu[2] and J. Thongchue[3]

[1,2,3]Department of Mathematics, Faculty of Science, Burapha University
169 Muang, Chonburi, Thailand
[1]jatupat@buu.ac.th, [2]ksaithan@buu.ac.th, [3]missjaruwan_thong@hotmail.com

Abstract

The purpose of this project was to study relationship among PM_{10}, SO_2 and NO_2. This study was divided into two periods, January-June and July-December, since 2005-2012. The project result found that the appropriate multiple linear regression equation estimating PM_{10} for the first period was $\widehat{PM}'_{10}(\text{Jan-Jun}) = 5.84+0.35SO_2(\text{Jan-Jun})-7.80NO'_2(\text{Jan-Jun})$ with standard error of estimation 1.066 and the adjusted coefficient of determination 36.60 and $\widehat{PM}'_{10}(\text{Jul-Dec}) = 0.140+0.0414SO'_2(\text{Jul-Dec})+1.41080NO'_2(\text{Jul-Dec})$ for the second period with standard error of estimation 0.028 and the adjusted coefficient of determination 41.30.

Mathematics Subject Classification: 62J05

Keywords: MLR method, best subset method

INTRODUCTION

The eastern seaboard of Thailand is an important part of the economy. This area is suitable for development of the industry. Eastern Seaboard Development Program (ESB) had occurred in 1981 under the fifth National Social and Economic Development Plan which was the plan developed by the government to encourage private sector investment by facilitating infrastructure in order to compete with international. The target areas were in Map Ta Phut District, Rayong province being developed into a major industrial and commercial city such as Map Ta Phut Industrial Estate, Hemaraj Eastern Industrial Estate (Map Ta Phut), etc. According to numerous industrial factories, many air pollutants, Sulfur dioxide (SO_2), Nitrogen dioxide (NO_2), Carbon monoxide (CO), Ozone (O_3), Particulate matter smaller than 10

microns (PM_{10}) and Volatile Organic Compound (VOCs), are emitted into the air caused pollution problems [1][2][3]. Rayong province has set up a unit to monitor the situation and resolve these problems continually. Once any air pollutant concentration has exceeded the standard level, it will result in human health problems. PM_{10} is one of the pollutants that cause environmental problems in Rayong so the present project was to predict PM_{10} for defense planning in order to reduce the effect to the community and environment in Rayong province.

MATERIALS AND METHODS

Average air pollutant concentrations, sulfur dioxide (SO_2), nitrogen dioxide (NO_2) and particulate matter smaller than 10 microns (PM_{10}), were collected from Air Quality and Noise Management Bureau, Pollution Control Department, Thailand since 2005 to 2013.

1. CORRELATION COEFFICIENT

Correlation coefficient (R) is used to monitor relationship among air pollutant concentrations.

2. THE MULTIPLE LINEAR REGRESSION MODEL

In the present project, multiple linear regression (MLR) method was used to analysis multivariate variables which consisted of one dependent variable, PM_{10}, and 2 independent variables, SO_2 and NO_2. These variables were used to generate the MLR equation following Equation 1 for the first period and Equation 2 for the second period.

$$\mathrm{PM}_{10}(\text{Jan-Jun}) = \beta_{01} + \beta_{11}\mathrm{SO}_2(\text{Jan-Jun}) + \beta_{21}\mathrm{NO}_2(\text{Jan-Jun}) + \varepsilon_{\mathrm{PM}_{10}(\text{Jan-Jun})} \quad (1)$$

$$\mathrm{PM}_{10}(\text{Jul-Dec}) = \beta_{01} + \beta_{11}\mathrm{SO}_2(\text{Jul-Dec}) + \beta_{21}\mathrm{NO}_2(\text{Jul-Dec}) + \varepsilon_{\mathrm{PM}_{10}(\text{Jul-Dec})} \quad (2)$$

3. THE MULTIPLE LINEAR REGRESSION EQUATION

For choosing the MLR equation, Mallows' C_p [4], standard error of estimation (S) and adjusted coefficient of determination (R^2_{adj}) were considered using the best subset method.

4. ASSUMPTIONS OF THE MULTIPLE LINEAR REGRESSION MODEL

There were four assumptions of the MLR model. (I) Normal distribution of the error term was tested by Anderson-Darling statistic (AD) [5], (II) Independence of the errors was tested by Durbin-Watson statistic (DW) [6], (III) Homoscedasticity of the errors was tested by Breusch-Pagan statistic (BP) [7] and (IV) Multicollinearity

among independent variables was tested by Variance Inflation Factor (VIF) [8]. If any of these four assumptions was violated, two ways will be used for solving these problems. First, the MLR equation would be rectified such as using the procedure of Box-Cox [9] or Johnson [10] to transform the dependent variable, etc. Secondly, others MLR equation would be chosen.

5. VALIDATION OF THE MULTIPLE LINEAR REGRESSION EQUATION

Once the MLR equation which was in agreement of all assumptions obtained, comparison between the observed data (OBS) and the predicted data (PRE) are validated using time series plot, scatter plot and the percentage of error (PE) calculated as of Equation 3.

$$PE = \frac{|OBS - PRE|}{OBS} X100\% \quad (3)$$

RESULTS AND DISCUSSION

For the first period (Jan-Jun), PM_{10} and NO_2 showed the high positive correlation coefficient with R=0.575 (P-value=0.000) and PM_{10} and SO_2 was the second highly one with R=0.505 (P-value=0.000). For the second period (Jul-Dec), the results were the same as the first period which correlation coefficient between PM_{10} and NO_2 was the highest one with R=0.450 (P-value=0.001) and PM_{10} and SO_2 was the second highly one with R=0.330 (P-value=0.015). All above results were the same period in agreement of the previous studies [1][2][3].

SO_2, and NO_2 were used to build the MLR equation using the best subset method and the MLR equations were $\hat{PM}'_{10}(\text{Jan-Jun}) = 5.84 + 0.350SO_2(\text{Jan-Jun}) - 7.80NO'_2(\text{Jan-Jun})$ where $\hat{PM}'_{10}(\text{Jan-Jun}) = \sqrt{PM_{10}(\text{Jan-Jun})}$ and $\hat{NO}'_2(\text{Jan-Jun}) = 1/NO_2(\text{Jan-Jun})$ with Mallows' C_p= 3.0, S=1.066 and R^2_{adj}=36.60 for the first period and $\hat{PM}'_{10}(\text{Jul-Dec}) = 0.140 + 0.0414SO'_2(\text{Jul-Dec}) + 1.41NO'_2(\text{Jul-Dec})$ where $\hat{PM}'_{10}(\text{Jul-Dec}) = \sqrt{PM_{10}(\text{Jul-Dec})}$, $\hat{SO}'_2(\text{Jul-Dec}) = 1/SO^2_2(\text{Jul-Dec})$ and $\hat{NO}'_2(\text{Jul-Dec}) = 1/NO^2_2(\text{Jul-Dec})$ with Mallows' C_p= 3.0, S=0.028 and R^2_{adj}=41.30 for the second period. Then, the assumptions of MLR analysis were monitored; (I) the error of the MLR distributed normally (AD=0.527 with P-value=0.170 for the first period and AD=0.377 with P-value=0.396 for the second period), (II) the errors were not significant independence (DW=1.304 with a critical value D_L=1.269 for the first period and DW=1.267 with a critical value D_L=1.245 for the second period), (III) the variance of errors was significant constant (BP=0.539 with a critical value 3.841 for the first period and BP=0.69 with a critical value 3.841 for the second period), (IV) there was no multicollinearity problem among the independent variables of both MLR equations with all VIF values less than 5 [11]. Finally, the PE was considered and the

results were displayed in Table 1. The maximum and minimum PE values were found with 57.99% in December and 5.49% in November consequently.

Table 1. The percentage of error (PE)

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
PRE	7.24	6.85	6.65	6.85	6.45	4.93	0.20	0.20	0.20	0.18	0.16	0.18
OBS	8.00	7.87	6.48	5.57	4.47	5.10	0.21	0.22	0.22	0.16	0.17	0.12
PE	9.51	13.01	5.64	23.02	44.12	3.28	3.29	8.41	7.21	12.26	5.49	57.99

ACKNOWLEDGEMENT

The authors wish to thank to the Air Quality and Noise Management Bureau, Pollution Control Department, Thailand for kind support collecting data.

REFERENCES

[1] Kuo, C. Y., Chen, P. T., Lin, Y. C., Lin, C. Y., Chen, H. H., & Shih, J. F., 2008, "Factors affecting the concentrations of PM_{10} in central Taiwan," Chemosphere, 70(7), 1273-1279.

[2] Vardoulakis, S., & Kassomenos, P., 2008, "Sources and factors affecting PM 10 levels in two European cities: implications for local air quality management," Atmospheric Environment, 42(17), 3949-3963.

[3] Kar, S., & Mukherjee, P., 2012, "Studies on Interrelations among SO_2, NO_2 and PM_{10} Concentrations and Their Predictions in Ambient Air in Kolkata," Open Journal of Air Pollution, 1(02), 42.

[4] Hocking, R.R., & Leslie, R.N., 1967, "Selection of the best subset in regression analysis," Technometrics, 9(4), 531-540.

[5] Lewis, P.A.W., 1961, "Distribution of the Anderson-Darling Statistic," The Annals of Mathematical Statistics, 32(4), 1118-1124.

[6] Durbin, J., & Watson, G.S., 1951, "Testing for Serial Correlation in Least Squares Regression II," Biometrika, 38(2), 159-177.

[7] Breusch T.S., & Pagan, A.R., "A Simple Test for heteroscedasticity and Random Coefficient Variation," Econometrica, 47(5), 1287-1294.

[8] O'brien, R. M., 2007, "A caution regarding rules of thumb for variance inflation factors," Quality & Quantity, 41(5), 673-690.

[9] Sakia, R.M., 1992, "The Box-Cox transformation technique: a review," The statistician, 169-178.

[10] Johnson, N.L., 1949, "Systems of frequency curves generated by methods of translation," Biometrika, 149-176.

[11] O'brien, R.M., 2007, "A caution regarding rules of thumb for variance inflation factors," Quality & Quantity, 41(5), 673-690.