Global Journal of Pure and Applied Mathematics.
ISSN 0973-1768 Volume 11, Number 4 (2015), pp. 2177-2181

http://www.ripublication.com

Prediction Of PM_{10} Concentration In The Eastern Industrial Areas Of Thailand With Principal Component Regression Method

J. Mekparyup[1] and K. Saithanu[2*]

[1,2]Department of Mathematics, Faculty of Science, Burapha University
169 Muang, Chonburi, Thailand
[1]jatupat@buu.ac.th, [2]corresponding author: ksaithan@buu.ac.th,*

Abstract

There are various variables in the atmosphere affected to the PM_{10} concentration. All these variables may correlated to each others so the principal component extraction of factor analysis is helpful to eliminate multicollinearity problem then the method of regression analysis can be used later to predict the PM_{10} concentration. The principal component regression method was then developed for predicting the PM_{10} concentration in the eastern industrial areas of Thailand. The model performance index of PM_{10} concentration prediction was evaluated with the root mean square error (RMSE). The results of study indicated the new five influential variables (NO_X, $CO\&O_3$, $HC\&CH_4$, Pressure and $SO_2\&WS$) from factor analysis were selected to predict the PM_{10} concentration. The estimated multiple regression equation was then displayed as

$$Y' = 1.27 + 0.0405 NO_X + 0.129 CO\&O_3 + 0.0590 HC\&CH_4 + 0.0585\, Pressure + 0.0114\, SO_2\&WS$$

where $Y' = \log PM_{10}$. This equation was rectified in conformity with regression assumptions as well as the standard error of estimation valued 0.213454. Furthermore, the performance index of model provided the RMSE values equal to 0.157776 and 8.77367 for the training and validation data set, respectively.

Keywords: PM_{10}, factor analysis, principal component regression method

Mathematics Subject Classification: 62-07, 62G35

INTRODUCTION

The particulate matter with median aerodynamic of diameter below 10 μm is widely known as PM_{10}. It is one of the five main air quality variables considered to measure the air quality index (AQI). Since the pollution problem of PM_{10} impacted on human health has been increasing at present, an authorized department of every country is extremely worried so it is necessary to find the particular means for solving this significant problem. The PM_{10} concentration depends on various meteorological variables such as temperature, humidity, precipitation and ventilation [1], [2], [3] therefore the PM_{10} characteristics are varied across distinct geographical areas. The industrial areas at the east of Thailand, Chonburi and Rayong provinces, have also seriously confronted the air pollution problem of PM_{10} as mention in the annual concentration report of Chonburi [4] and the air quality report of Rayong [5] that the PM_{10} concentration was more frequently above the standard level in many days. Some researches utilized the complicated model like neural network, [6], [7], [8], or the simple one like regression, [9], to predict the PM_{10} concentration in Thailand. Both studies of [10] and [11] presented use of principal components obtained from factor analysis being the independent variables potentially assisted in prediction problem of O_3 which is also one of the important materials for measuring AQI. This study then exemplified the principal component regression model for prediction of PM_{10} concentration in the eastern industrial areas of Thailand.

MATERIALS AND METHODS

The General Education Centre, Mueang District, Chonburi and the Map Ta Phut Health Office, Mueang District, Rayong were two delegates of eastern monitoring stations in the industrial areas of Thailand. The daily concentration of PM_{10}, dependent variable, and sixteen independent variables (nine air quality variables: CO, NO, NO_2, NO_X, SO_2, HC, CH_4, NMHC and O_3 and seven meteorological variables: Pressure, Rain, Relative Humidity (RH), Temperature (Temp), Sun Radiation (SR), Wind Direction (WD) and Wind Speed (WS)) were measured 2,265 cases during 2006-2010. The procedure of prediction PM_{10} concentration in the eastern industrial areas of Thailand was following these five steps.

1. Considering relationship between the concentrations of PM_{10} and each of all sixteen independent variables with test of Pearson correlation coefficient.
2. Separating data into two sets. The first set, training data set, consisted 1,586 cases randomly chosen 70% of all data was utilized for model training. The remaining data, validation data set, was considered for validating suitability of model.
3. Determining the significant variables influenced to the PM_{10} concentration from the training data set by factor analysis applied principal component method of factor extraction.
4. Predicting the PM_{10} concentration with regression analysis by using the new influential variables obtained from step 3.
5. Evaluating the model performance index of PM_{10} concentration prediction with the root mean square error, RMSE, from both of data sets as

$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2}$ where Y_i be the ith observed value of PM_{10} concentration, $\hat{Y}_i$ be the ith predicted value of PM_{10} concentration from the model and n be the total observations in considered data set.

RESULTS

The results of this study were as follows:

1. There were eight air quality variables (CO, NO, NO_2, NO_X, SO_2, HC, CH_4 and O_3) and five meteorological variables (Pressure, RH, Temp, SR and WD) which were associated to the PM_{10} concentration as considering from a really small P-value closed to zero of Pearson Correlation coefficient test.
2. Once the principal component method of factor extraction from the training data set was applied, there were eight factors (NO_X, Temp&SR, CO&O_3, HC&CH_4, NMHC, Pressure, SO&WS and Rain) shown as the crucial variables influencing on the PM_{10} concentration as seeing of the high factor loadings from each factor (F_i) in Table 1.

Table 1: High factor loading of variables in each factor

Factor	Variable	Loading
F_1	NO	0.637
	NO_2	0.816
	NO_X	0.966
	Grouping to NO_X	
F_2	Temp	-0.773
	SR	-0.783
F_3	CO	0.703
	O_3	0.813
F_4	HC	0.719
	CH_4	0.898
F_5	NMHC	0.969
F_6	Pressure	0.822
F_7	SO_2	0.729
	WS	0.696
F_8	Rain	-0.932

3. After regressing the PM_{10} concentration on the new eight influential variables obtained from the previous result, called the principal component regression method, with best subsets method, two models with the smallest values of Mallow C_p and standard error of estimation (S) as well as the biggest R^2_{adj} were chosen. First model was composed of six independent variables (NO_X,

CO&O_3, HC&CH_4, Pressure, SO&WS and Rain) with Mallow $C_p = 7.5$, $S = 12.745$ and $R^2_{\text{adj}} = 33.6$ while the second model consisted seven independent variables (NO_X, Temp&SR, CO&O_3, HC&CH_4, Pressure, SO&WS and Rain) with Mallow $C_p = 7.2$, $S = 12.740$ and $R^2_{\text{adj}} = 33.6$. However, both of these two models indicated the large P-value of regression coefficient test comparing to the defined significant level on Temp&SR and Rain. Thus, the five variables remained to regress on the PM_{10} concentration were NO_X, CO&O_3, HC&CH_4, Pressure and SO&WS). When this model was checked for regression assumptions, it revealed the normality of error term was violated with the P-value of Normal Probability Plot less than 0.005. Box-Cox transformation was then suggested to transform the PM_{10} concentration to the logarithm of PM_{10} concentration. Therefore, the estimated multiple regression equation with $S = 0.213454$ and $R^2_{\text{adj}} = 35.2$ was illustrated as

$$Y' = 1.27 + 0.0405 NO_X + 0.129 CO\&O_3 + 0.0590 HC\&CH_4 \\ + 0.0585\, Pressure + 0.0114\, SO_2\&WS$$

where $Y' = \log\ PM_{10}$. Then, all regression assumptions of this model were retested. It appeared all assumptions of error terms were satisfied as normality (P-value of normality test valued 0.145), independence (No pattern displayed in scatter plot between the residuals and fitted values or time series plot of error term) and constant variance (Chi-square Test of Breusch-Pagan: $\chi^2_{BP} = 5.78\times 10^{-7} < \chi^2_{0.95;\,5} = 11.0705$).

4. The RMSE was later measured for the model performance index of PM_{10} concentration prediction. It provided the RMSE values equal to 0.157776 and 8.77367 for the training and validation data set, respectively.

DISCUSSION

Using of principal components determined from factor analysis being the independent variables instead of all independent variables was more efficient in prediction of the PM_{10} concentration by regression analysis as seeing of the smaller values of Mallow C_p and S as well as the bigger R^2_{adj}. In addition, the RMSE of this principal component regression model revealed smaller values for the validation data sets as comparing with two network models, MLP 12-5-1 and ORBF 12-5-1 [8]. Above all, one who authorized could use this finding for controlling the PM_{10} concentration not beyond the standard level.

ACKNOWLEDGEMENT

This work was financially supported by the Research Grant of Burapha University through National Research Council of Thailand (Grant no. 71/2558). The authors were also grateful to the Air Quality and Noise Management Bureau, Pollution Control Department for kindly providing all data.

REFERENCES

[1] Mickley, L.J., Jacob, D.J., Field, B.D., & Rind, D., 2004, "Effects of future climate change on regional air pollution episodes in the United States," Geophysical Research Letters, 31(24).

[2] Liao, H., Chen, W.T., & Seinfeld, J.H., 2006, "Role of Climate Change in Global Predictions of Future Tropospheric Ozone and Aerosols," Journal of Geophysical Research: Atmospheres (1984-2012), 111(D12).

[3] Heald, C.L., Henze, D.K., Horowitz, L.W., Feddema, J., Lamarque, J.F., Guenther, A., Hess, P.G., Seinfeld, J.H., Goldstein, A.H., & Fung, I., 2008, "Predicted change in global secondary organic aerosol concentrations in response to future climate, emissions, and land use change," Journal of Geophysical Research: Atmospheres (1984-2012), 113(D5).

[4] Khaenamkaew, P., Iamraksa, P., Raksawong, S., Wongsontam, K., Angwanisakul, C., & Khuntong, S., 2011, "Annual Concentration Report and Emission Sources Analysis of the Air Pollutants Measured by the AQM Station," Research Exhibition "Research in Kasetsart University 2011" in National Agricultural Fair.

[5] Thai Universities for Healthy Public Policies, 2011, "Air quality around Map Ta Phut in Rayong, Thailand," Retrieved November 13, 2012, from http://www.ehwm.chula.ac.th/maptaphut/air-maptaphut.pdf

[6] Arampongsanuwat, S., & Meesad, P., 2010, "Development of a Prediction Model of PM_{10} in Bangkok Using Artificial Neural Networks," The 6^{th} National Conference on Computing and Information Technology.

[7] Arampongsanuwat, S., & Meesad, P., 2011, "Development of a Prediction Model of PM_{10} in Bangkok Using Support Vector Regression and Radial Basis Function Network," The 7^{th} National Conference on Computing and Information Technology.

[8] Mekparyup, J., & Saithanu, K., 2013, "Development of Neural Network Technique for Prediction of PM_{10} Concentration in the Industrial Area, at the East of Thailand," Applied Mathematical Sciences, 7(93), 4631-4638.

[9] Saithanu, K., & Mekparyup, J., 2014, "Using Multiple Linear Regression To Predict PM10 Concentration In Chonburi, Thailand," Global Journal of Pure and Applied Mathematics, 10(6), 835-839.

[10] Phansiri, S., & Saithanu. K., 2013, "Prediction of Daily Maximum Concentration of Ground Level Ozone with Principal Component Regression Model," Burapha Sci. J., 18(1), 144-152.

[11] Mekparyup, J., & Saithanu, K., 2014, "Application of Artificial Neural Network Models to Predict the Ozone Concentration at the East of Thailand," International Journal of Applied Environmental Sciences, 9(4), 1291-1296.