Global Journal of Pure and Applied Mathematics.
ISSN 0973-1768 Volume 11, Number 4 (2015), pp. 2127-2130

http://www.ripublication.com

Application Of Neural Network Scheme Using PCR To Predict PM_{10} Concentration At The East Of Thailand

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

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

Abstract

To predict PM_{10} concentration at the east of Thailand, neural network scheme was applied by using the principal component regression (PCR) as the input nodes for training the network. The study results showed the MLP-PCR5-5-1 presented the best performance with small value of root mean square error (RMSE=0.144770) for the validation data set. That meant, the MLP trained with 5 input nodes obtained from the PCR method could efficiently predict the PM_{10} concentration because it helpfully reduced network dimensionality for training.

Keywords: PM_{10}, Neural network, PCR method

Mathematics Subject Classification: 62-07, 62G35

INTRODUCTION

PM_{10} is one of the main air pollutants which has negatively affected to human health [1]. Both of air quality and meteorological variables in the air are influential to PM_{10} concentration [2], [3], [4]. The air pollution problem of PM_{10} is also rapidly increasing in Thailand particularly in the eastern areas because it is the hub of manufacturer and agro-industry. Many researches, therefore, worked on different issues of PM_{10} concentration like prediction of PM_{10} concentration, for example; [5], [6], [7], [8], or AQI classification based on the standard level of PM_{10} notified by [9] such as [10]. There were different tools for PM_{10} analysis as mentioned in previous researches then this study purposed neural network scheme which contained input nodes combing principal component (PC) and regression analysis, called principal component regression (PCR), to test whether it could favorably reduce network complexity for training as well as led to precisely predict the PM_{10} concentration.

MATERIALS AND METHODS

The data for analysis consisted of 2 sets including 2,265 observations. One was the training contained 70% of all data. The remaining data was the validation. Both of them were figured in daily concentrations of PM_{10} (dependent variable), 9 air quality variables (CO, NO, NO_2, NO_X, SO_2, HC, CH_4, NMHC and O_3) and 7 meteorological variables (Pressure, Rain, Relative Humidity (RH), Temperature (Temp), Sun Radiation (SR), Wind Direction (WD) and Wind Speed (WS)) collected during 2006-2010 by the two delegates of eastern monitoring stations in Thailand, the General Education Centre, Mueang District, Chonburi and the Map Ta Phut Health Office, Mueang District, Rayong. To predict PM_{10} concentration at the east of Thailand with the neural network scheme, 4 steps were conducted as follows. First, the test of Pearson correlation coefficient was utilized to consider association of concentrations between PM_{10} and each of 16 independent variables. Secondly, the influential variables affecting to PM_{10} concentration were determined from the training data set by using factor analysis applied principal component method of factor extraction. Then, these new obtained variables were considered as the predictor variables for building regression model to predict PM_{10} concentration. This determined regression equation was consequently checked the assumptions of regression.

Third, both of simple (MLP) and advanced (RBF) architectures of neural network were trained with the number of input nodes equal to the number of predictor variables of regression equation obtained from previous step. The number of hidden nodes in a hidden layer was 3 and 5 following the suggestion of [11], [12], [13]. The activation function to train neural network at hidden and output layers for obtaining the potential PM_{10} prediction model was guided in [7]. Only one output node represented the dependent variable (PM_{10} concentration) which was also derived from regression equation of previous step. Finally, the root mean square error ($RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}$; 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 neural network model and n be the total observations in measured data set) was calculated to evaluate the network model performance.

RESULTS

The P-values of Pearson Correlation coefficient test between the concentrations of PM_{10} and each of 13 variables (CO, NO, NO_2, NO_X, SO_2, HC, CH_4, O_3, Pressure, RH, Temp, SR and WD) were rather small values closed to zero. When factor analysis with the principal component method of factor extraction was applied, there were 8 high factor loadings. That meant, the 8 influential variables affecting to PM_{10} concentration were F_1 (NO_X: grouped from NO, NO_2 and NO_X), F_2 (Temp&SR:

grouped from Temp and SR), F_3 (CO&O_3: grouped from CO and O_3), F_4 (HC&CH_4: grouped from HC and CH_4), F_5 (NMHC), F_6 (Pressure), F_7 (SO&WS: grouped from SO and WS) and F_8 (Rain). Once these eight new factors counted as the predictor variables were regressed on the PM_{10} concentration, the appropriate model selected with the best subsets and stepwise technique remained 5 variables (F_1, F_3, F_4, F_6 and F_7) in the model. However, normality assumption of this model was not satisfied. Box-Cox procedure was then required to transform the PM_{10} concentration in the pattern of logarithm so the principal component regression (PCR) complied with regression assumptions were the dependent variable ($\log(PM_{10})$) and the 5 predictor variables (F_1, F_3, F_4, F_6 and F_7).

Then, 4 architectures (MLP-PCR5-3-1, MLP-PCR5-5-1, RBF-PCR5-3-1 and RBF-PCR5-5-1) for training neural network represented with 5 input nodes derived from PCR technique, activation function at hidden nodes was applied with logistic and exponential functions for MLP as well Gaussian and identity functions for RBF and only one output node denoted $\log PM_{10}$ obtained from regression analysis. The performance of all 4 network models for prediction PM_{10} concentration was finally compared with the RMSE as shown in Table 1.

Table 1: Comparison of neural network model performance

Architecture of Neural Network	Activation Function at		RMSE of	
	Hidden Node	Output Node	Training Set	Validation Set
MLP-PCR5-3-1	Logistic	Exponential	0.207734	0.150602
MLP-PCR5-5-1	Logistic	Exponential	**0.200900**	**0.144770**
RBF-PCR5-3-1	Gaussian	Identity	0.207896	0.149691
RBF-PCR5-5-1	Gaussian	Identity	0.203408	0.146928

DISCUSSION

The MLP-PCR5-5-1 performed the best performance as seeing of the smallest RMSE of training set (0.2009) and validation set (0.144770). As of such a result, it implied using PCR as the input nodes usefully relieved the complexity of network model for training also then gained the desirable performance. Finally, one could say if the same networks (MLP or RBF) were trained with more hidden nodes (5 nodes), it would give the better performance than with fewer hidden nodes (3 nodes), regardless of data set.

ACKNOWLEDGEMENT

The authors wish to thank the Air Quality and Noise Management Bureau, Pollution Control Department for kindly supporting all data.

REFERENCES

[1] Biggeri, A., Bellini, P., & Terracini, B., 2004, "Meta-analysis of the Italian Studies on Short-term Effects of Air Pollution--MISA 1996-2002," Epidemiologia & Prevenzione, 28(4-5 Suppl), 4-100.

[2] 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).

[3] 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).

[4] 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).

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

[6] 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 7th National Conference on Computing and Information Technology.

[7] Mekparyup, J., & Saithanu, K., 2013, "Development of Neural Network Technique for Prediction of PM10 Concentration in the Industrial Area, at the East of Thailand," Applied Mathematical Sciences. 7(93), 4631-4638.

[8] 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.

[9] Office of Natural Resources and Environmental Policy and Planning, Notice of the National Environment Committee NO.28 (B.E.2550) on Air Quality Standard. Retrieved October 25, 2011, from http://www.legalbase.pti.org/Law.aspx?lid=3596.

[10] Saithanu, K., & Mekparyup, J., 2014, "Modeling of Air Quality Index in the Eastern Urban Areas of Thailand using Neural Network Method," International Journal of Applied Environmental Sciences, 9(4), 1885-1891.

[11] Cybenko, G., 1989, "Approximation by superpositions of a sigmoidal function," Mathematics of Control, Signals, and Systerms, 2(4), 303-314.

[12] Hornik, K., Stinchcombe, M., & White, H., 1989, "Multilayer feedforward networks are universal approximators," Neural Networks, 2(5), 359-366.

[13] Guo, Y., & Dooley, K. J., 1992, "Identification of Change Structure in Statistical Process Control," The International Journal of Production Research, 30(7), 1655-1669.