Global Journal of Pure and Applied Mathematics.
ISSN 0973-1768 Volume 11, Number 4 (2015), pp. 1851-1856

http://www.ripublication.com

Performance of Artificial Neural Network Models based on Principal Components for prediction of PM_{10} Concentration at the East of Thailand

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

To predict the PM_{10} concentration at the east of Thailand for this study, the Feed-forward Artificial Neural Network (FANN) models with two different architectures were performed. One was the simple, Multi-Layer Perceptron (MLP), and the other was the advanced, Radial Basis Function (RBF). Principal component method was also applied combining with the artificial neural network scheme called the Principal Component Feed-forward Artificial Neural Network (PCFANN) to increase efficiency in prediction of PM_{10} concentration. For measuring the performance of artificial neural network models, the root mean square error (RMSE) of the validation data set was considered. The results of study indicated the MLP-FANN16-5-1 was the best performance with 11.572650 of RMSE. In addition, the principal component technique was not only helpfully reduce the complexity for training artificial neural network and get rid of data collinearity but also beneficially enhance the performance of artificial neural network models in prediction of PM_{10} concentration as seeing with the rather small value of RMSE (12.260360) for MLP-PCFANN8-5-1.

Keywords: Feed-forward Artificial Neural Network (FANN), Multi-Layer Perceptron
(MLP), Radial Basis Function (RBF), Principal Component Feed-forward Artificial Neural Network (PCFANN)

Mathematics Subject Classification: 62-07, 62G35

INTRODUCTION

Most of significant industrial estates of Thailand located at the eastern areas like the

Maptaput industrial estate in Chonburi or the Laem Chabang industrial estate in Rayong so it has continually encountered air pollution problem at present particularly the PM_{10}. The standard of 24 hours PM_{10} level stipulated by the Thai Environmental Protection Department is 120 $\mu g/m^3$ [1]. The annual concentration report of Chonburi [2] as well the air quality report of Rayong [3] notified that the PM_{10} concentration in many days was often beyond the standard level. There were many researches in Thailand then worked on the prediction of PM_{10} concentration basing on both of statistical methods, regression analysis and multivariate tool, or even the mathematical gadget like the artificial neural network (ANN) model [4], [5], [6], [7], [8]. Therefore, this study was objective to demonstrate the performance of ANN models in prediction of PM_{10} concentration following the recommendation of [8] by applying principal component analysis to obtain the input nodes hypothesized the principal components (PCs) might reduce the complication for training ANN models.

MATERIALS AND METHODS

Daily concentrations of 10 air quality variables (CO, NO, NO_2, NO_X, SO_2, HC, CH_4, NMHC, O_3 and PM_{10}) and 7 meteorological variables (Pressure, Rain, Relative Humidity: RH, Temperature: Temp, Sun Radiation: SR, Wind Direction: WD and Wind Speed: WS) were collected during 2006-2010 from the two agents of monitoring stations at the east of Thailand (the General Education Centre, Mueang District, Chonburi and the Map Ta Phut Health Office, Mueang District, Rayong). There were following 4 steps for performing the performance of ANN models relating on PCs in prediction of PM_{10} concentration at the east of Thailand.

1. Preparing data for analysis by dividing data into 2 sets. The 70% of all data was randomly selected for model training called training data set while the rest of data known the validation data set was utilized for validating suitability of model.
2. Grouping variables in the weather which might correlate to each other by applying principal component method of factor extraction from factor analysis.
3. Modeling two ANN models with both MLP and RBF architectures. The first model was the FANN and the other one was the PCFANN. According to the suggestion of [9] and [10], one hidden layer is enough for the simple task so three layers of network architecture were designed: input, hidden and output. Because there was no definite rule for identifying the suitable number of hidden nodes in a hidden layer, one should only concentrate on undertraining or overtraining problem. This study then merely used 3 and 5 nodes in a hidden layer while the output layer contained only one node representing the daily concentration of PM_{10}. For the input layer, there were 2 patterns of input nodes for training the ANN models. The number of input nodes for the first form was 16 nodes equal to the number of all variables (9 air quality and 7 meteorological variables) in the air. The other one was trained with the input nodes equal to the number of PCs obtaining from step 2. The activation function used at the computing nodes was also one of the crucial factors to be

considered. There are numerous activation functions to be applied at the hidden and output nodes then Mekparyup and Saithanu recommended the appropriate activation function in [8] for prediction PM_{10} concentration. Therefore, logistic and exponential functions were respectively used at the hidden and output nodes for the MLP. Otherwise, Gaussian and identity functions were successively computed instead at the hidden and output nodes for the RBF.

4. Performing the performance of ANN models in prediction of PM_{10} concentration by 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 observation of PM_{10} concentration, $\hat{Y}_i$ be the ith predicted value of PM_{10} concentration from each of ANN model and n be the total observations in considered data set.

RESULTS

There were 8 factors in scree plot having eigen value greater than or closed to 1 so it roughly conducted the new eight variables would be later analyzed. Once varimax rotation of factor analysis was applied for assisting interpretation, these new eight factors were grouped into the new eight variables (NO_X, Temp&SR, CO&O_3, HC&CH_4, NMHC, Pressure, SO_2&WS and Rain) by regarding only the high factor loadings from each factor (F_i) as shown in Table 1.

Table 1: High factor loading of variables in each factor

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

Then, there were 8 networks for modeling the ANN in prediction of PM_{10} concentration. Four FANN networks were MLP-FANN16-3-1, RBF-FANN16-3-1, MLP-FANN16-5-1 and RBF-FANN16-5-1 illustrated in Figure 1 and Figure 2. The four remaining networks were MLP-PCFANN16-3-1, RBF-PCFANN16-3-1, MLP-PCFANN16-5-1 and RBF-PCFANN16-5-1 displayed in Figure 3 and Figure 4.

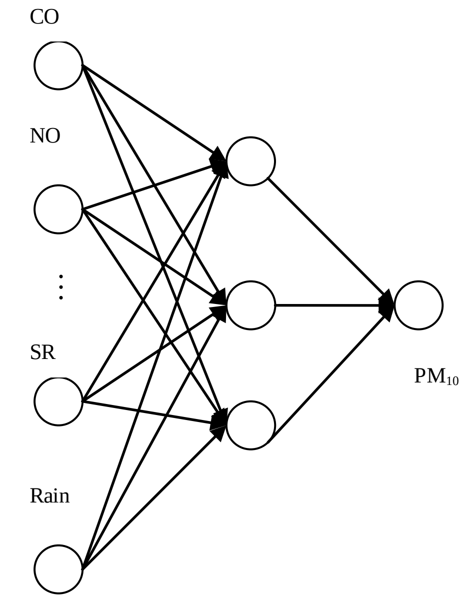

Figure 1 MLP-FANN16-3-1 and RBF-FANN16-3-1

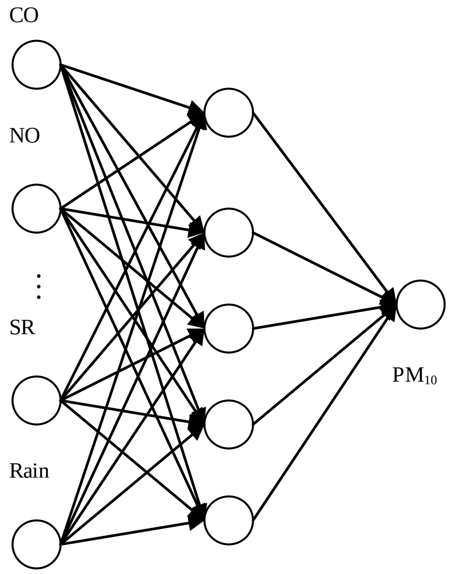

Figure 2 MLP-FANN16-5-1 and RBF-FANN16-5-1

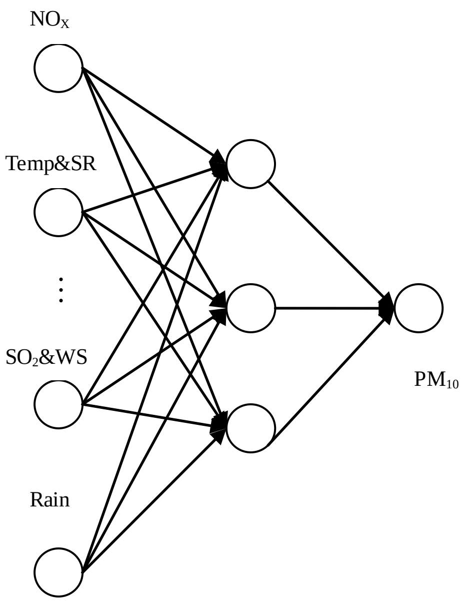

Figure 3 MLP-PCFANN8-3-1 and RBF-PCFANN8-3-1

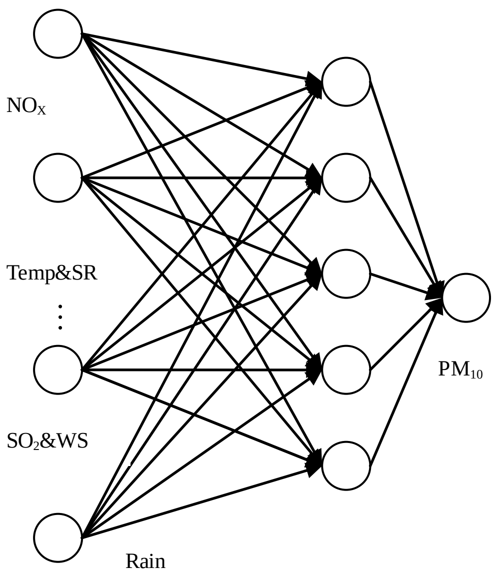

Figure 4 MLP-PCFANN8-5-1 and RBF-PCFANN8-5-1

Finally, the RMSE was measured and compared for the performance of ANN models in prediction of PM_{10} concentration as illustrated in Table 2**.**

Table 2: Comparison performance of ANN models with the RMSE

Architecture of ANN models	Activation function of		RMSE of	
	Hidden node	Output node	Training	Validation
MLP-FANN16-3-1	Logistic	Exponential	11.170529	12.402098
MLP-FANN16-5-1	Logistic	Exponential	**10.324785**	**11.572650**
RBF-FANN16-3-1	Gaussian	Identity	11.638943	19.945354
RBF-FANN16-5-1	Gaussian	Identity	16.078578	20.510113
MLP-PCFANN8-3-1	Logistic	Exponential	11.604290	12.464933
MLP-PCFANN8-5-1	Logistic	Exponential	11.284792	12.260360
RBF-PCFANN8-3-1	Gaussian	Identity	15.668867	19.573669
RBF-PCFANN8-5-1	Gaussian	Identity	11.666887	13.092489

The FANN network performed both of the best and the worst performance. That meant, the MLP-FANN16-5-1 showed the minimum RMSE for both training (10.324785) and validation (11.572650) data set while the RBF-FANN16-5-1 gave the maximum RMSE with 16.078578 and 20.510113 for training and validation data set, respectively.

DISCUSSION

A discussion could be criticized as following points: (1) the principal component scheme might helpfully improve the ANN model performance in prediction of PM_{10} concentration as considering from the MLP-PCFANN8-5-1 really provided rather good performance with the small RMSE value of validation data set (12.260360). The difference of RMSE value between the training and validation data set of PCFANN was also smaller than of FANN. (2) Basing on the same ANN models, the model architecture with 5 hidden nodes almost provided the smaller RMSE value than with 3 hidden nodes except only one case between the RBF-FANN16-3-1 and the RBF-FANN16-5-1. (3) The study results could be applied in guidelines for developing ANN model performance to predict the concentration of any significant air quality variable in the air (not only for PM_{10}) by adjusting the number of hidden layers, the number of nodes in the hidden layer or even the type of activation function.

ACKNOWLEDGEMENT

The authors really appreciated to the Faculty of Science, Burapha University for supporting their research fund. Also, the authors were grateful to the Air Quality and Noise Management Bureau, Pollution Control Department for kindly providing all data.

REFERENCES

[1] Office of Natural Resources and Environmental Policy and Planning, 2011, “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.

[2] P. Khaenamkaew, P. Iamraksa, S. Raksawong, K. Wongsontam, C. Angwanisakul, & S. Khuntong, 2011, “Annual Concentration Report and Emission Sources Analysis of the Air Pollutants Measured by the AQM Station,” Research in Kasetsart University 201 in National Agricultural Fair (2011).

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

[4] 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.

[5] 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.

[6] 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.

[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 7th 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] Cybenko, G., 1989, “Approximation by superpositions of a sigmoidal function,” Mathematics of Control, Signals, and Systerms, 2(4), 303-314.

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