Statistical Models for Predicting Cement Factory Emissions

Sudjit Karuchit* and JetiyaKwanma

School of Environmental Engineering, Suranaree University of Technology, Thailand
*Corresponding Author: sudjit@sut.ac.th, Tel: (662) 4224456, Fax: (662) 4224606

Abstract: This research studied the relationship among important factors of cement production – namely, raw materials, fuels, and manufacturing processes – and gaseous and particulate emissions. Two types of statistical prediction models, multiple regression (MR) and artificial neural network (ANN), were developed and compared. The recorded daily average data of raw materials, coal fuels, alternative (hazardous waste) fuels, production processes, and gaseous and particulate emissions in 2007 were used in the analysis. Results show that the MR and ANN models for predicting NO_2, SO_2, CO_2, HCl and TSP, have the Adjusted R^2 values in the range of 0.25-0.57 and 0.44-0.66, respectively. It is also found that the independent variables that have significant effect on the of models are quantity of clay, quantity of limestone, raw mill running time, alternative fuels used, kiln running time, and quantity of clinker. Overall, the ANN models perform slightly better than the MR models.

Keywords: Cement, artificial neural network, multiple regression, air pollution, model.

1. Introduction

The cement industry is a major source of air pollutants such as dust and gases that affect people's health and quality of life. Air pollutant emissions from the cement factories normally vary due to various factors. The variation in raw materials used and manufacturing processes can affect air pollutant emissions. In addition to conventional raw materials, some cement factories use hazardous wastes, such as used tires, lubricating oils, or solvents, as alternative fuels. Good management of such alternative fuels utilized can substantially reduce the emissions [1]. There are various types of pollutant emitted, but common control devices such as electrostatic precipitators (ESPs), when used alone in cement plants only control particulate matters, not gaseous pollutants.

The Siam City Cement in Saraburi Province, Thailand, was here chosen to study for the relationship among important factors of cement production and gaseous and particulate emissions. Saraburi Province is known as the center of cement production in Thailand, and Siam City Cement is a leading cement factory in the nation. The factory is among only a few cement factories that utilize hazardous wastes as alternative fuels in their production processes. The five gases and particulate matters studied are: nitrogen dioxide (NO_2), sulfur dioxide (SO_2), carbon dioxide (CO_2), hydrogen chloride (HCl), and total suspended particles (TSP). The concentration data of these gases and particulate matters were obtained from the factory's continuous emission monitoring system (CEMs). It is worth to mention that carbon monoxide (CO) was not monitored by the CEMs, and thus it was not included in the analysis of this study. Consequently, statistical models were developed for predicting the gaseous and particulate emissions, using raw material components and manufacturing processes data treated or assignedas predictors. Two statistical modeling techniques were employed in this study to formulate two types of models: the multiple regression (MR) modelsand the artificial neural network (ANN) models. The two techniques are commonly and widely used for developing air pollution prediction models [2-3]. The Siam City Cement consists of 3 main facilities: Facroty #1, #2, and #3. Each factory has 2 kilns in operation. This paper presents the results of Kiln #5, one of the two kilns in operation in Factory #3, due to its more complete process information. Each kiln has the capacity of about 10,000 ton clinkers/day.

2. Experimental

2.1 Process Study and Data Collection

The research started from studying the manufacturing processes, collecting the data, and building the database of the variables. After the processes and emission routes were investigated, important factors to be included in the models were identified. There were a total of 73 independent variables which can be divided into 3 groups: raw materials, fuels, and production processes. The dependent variables are the gas and particulate concentrations: NO_2, SO_2, CO_2, HCl, and TSP. Subsequently, the database of these variables were developed from their daily recorded average data in 2007. Descriptive analysis and outlier detection were performed on the collected variable data. Significant correlations among variables were identified. Variable transformation with natural logarithm was used in this study to investigate the best correlation among variables. As a result, there were 4 paths of model development: (1) independent variables vs. dependent variables, (2) logarithm of independent variables vs. dependent variables, (3) independentvariables vs. logarithm of dependent variables, and (4) logarithm of independent variables vs. logarithm of dependent variables.

2.2 Model Development

For the MR model development, the stepwise selection techniques were employed for selecting appropriate variables into the models. If X_i is the value of the input variable i and Y is the actual gas or particulate concentration, then the constant b_0 and the regression coefficients b_i are computed by the ordinary least-squares equation:

$$Y = b_0 + \sum_{i=1}^{n} b_i X_i + \varepsilon_i \quad (1)$$

The available data for MR procedure was randomly separated into 2 sets: regression analysis data set (90%) and validation data set (10%). Residual analysis was performed on the possible models obtained from the regression analysis. Model validation wascarried out using the validation data set which was separated from the data used in the model development.

For the ANN model development, the multi-layer feed forward approach (MLFF) and the error-back propagation algorithm (BP) with sigmoid function were used with the selected predictor variables. The sigmoid function according to the equation:

$$f(x)=\frac{1}{1+e^{-x}} \quad (2)$$

The model development algorithm is shown in Fig. 1 [4]. The available data for ANN procedure was randomly separated into 3 sets: training data (60%), test data (20%) and validation

data (20%). The model structure used was 1 hidden layer with the number of hidden nodes calculated from the following equation:

$$\text{Number of Hidden Nodes} = 0.5(I + O) + (P)^{0.5} \quad (3)$$

where I, O, and P are the number of input variables, output variables, and training patterns, respectively. Values of the 3 parameters used were varied in order to obtain the combinations which yield better model performance: learning rate (0.05, 0.1, 0.2), weight (0.3, 0.5, 0.7), and momentum (0.1, 0.5, 0.9). In each training cycle, the training data is used to train the model to predict the outputs. The prediction error is then calculated and used to modify the weighting coefficients in the model. Consequently, the prediction is compared to the test data and the corresponding error is calculated. If the error is more than 0.005, the training cycle is repeated. If the error is less than 0.005, or the number of training cycles reaches 20,000 cycles, the training is considered to be completed.

The final model structure is validated with the validation data set. The best models are those with minimum value of statistical measurement of mean absolute percentage error (MAPE), according to the equation:

$$\text{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{\text{Predict}_i - \text{Actual}_i}{\text{Actual}_i}\right| \times 100 \quad (4)$$

where Actual and Predict represents actual measurement data and predicted data obtained from the model, respectively. The N is the number of data used in the calculation.

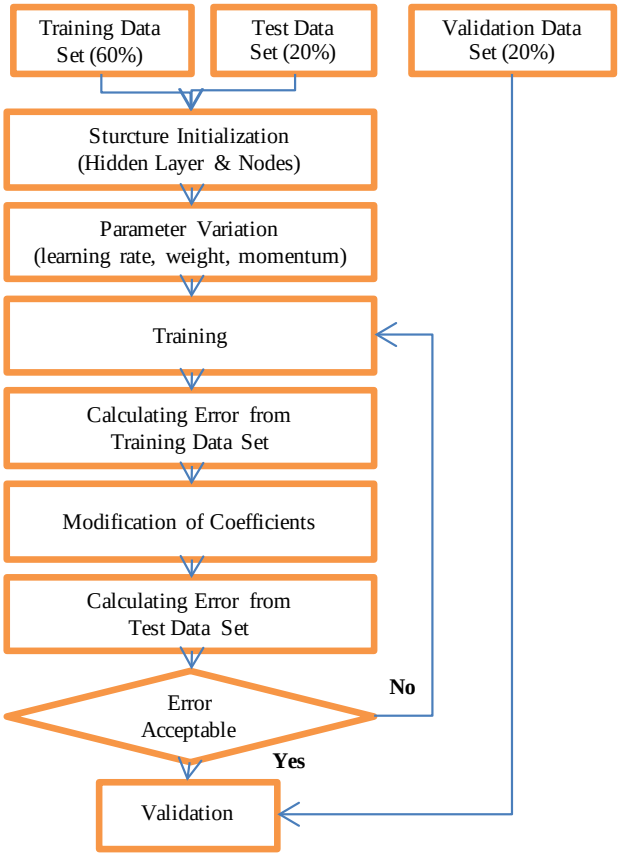

Figure 1. The ANN model development algorithm.

2.3 Model Performance Comparison

Once the best models from the 2 approaches were obtained, the validation data set from the MR procedure was employed again to compare their predicting ability. The model performance was evaluated using three criteria: correlation coefficient (R), index of agreement (I.A.), and root mean square error (RMSE). The I.A. and the RMSE are calculated using the following equations:

$$\text{I.A.} = 1 - \frac{\sum_{i=1}^{N}(\text{Predict}_i - \text{Actual}_i)^2}{\sum_{i=1}^{N}\left(\left|\text{Predict}_i - \overline{\text{Actual}}\right| + \left|\text{Actual}_i - \overline{\text{Actual}}\right|\right)^2} \quad (5)$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\text{Predict}_i - \text{Actual}_i)^2} \quad (6)$$

A value of I.A. close to unity indicates agreement of the model prediction values and the actual values. In contrast, the more RMSE value, the more error would be obtained from the prediction of the model. The RMSE represents the model error in the unit of the dependent variable.

3. Results and Discussion

3.1 Process, Emission Routes, and Predictor Variables

Raw materials are conveyed from silos to raw material mills, preheated in cyclone preheaters, and then mixed with additional materials such as alternative fuel and coal in the calciners. Subsequently, the kiln transforms the mixed materials in to cement pellets or “clinkers”, which are cooled down rapidly in the clinker cooling unit and transported to silos as finished products. Three types of coal: bituminous, lignite, and anthracite are processed in coal mills and preheated before feeding to the kiln as main heating fuels. Industrial wastes are occasionally added to the calciners as alternative fuels and for the purpose of waste treatment. There are as much as 46 types of waste listed and used, including solvents, used oil, sludge, agricultural waste, spent resin, rubber, and plastic.

Fig. 2 shows the process emission routes. Hot gas from the clinker cooling unit goes through successive heat recovery steps starting from the calciners, then the cyclone preheaters, and finally the raw mills. The cyclone and the main electrostatic precipitator (EP) then clean the gas before releasing it to the atmosphere. Other two routes are the gas from the kiln, which is released through EP #3 and EP#4 in a similar fashion. Lastly, a fraction of the clinker cooling unit gas is released directly through EP #1. The main EP emission route is selected for the model development in this study, due to available gas and particulate concentration data from the continuous emission monitoring system (CEMs). The CEMs was operating normally throughout the period of the data collection. Other routes, on the other hand, have only routine stack sampling results.

Once the process and its emission routes were thoroughly understood, pertinent factors of the emissions were determined as predictor variables and their data were collected. The raw material variables are: limestone (Lime), shale with silica (ShaleS), shale with alumina (ShaleA), laterite typ B (LB), clay (Clay), Klangdong soil (KD), alternative raw materials (AR), and other materials (Others). The fuel variables are: high quality bituminous (Coal_B), low quality bituminous (Coal_C), lignite (Lig), antracite (Ant), and 46 types of alternative fuel (W_01 to W_46). These two groups of variables represent the amount of materials used in the process per day. The process variables are: percent raw material residual that passed 90 micron screen (R_90), percent raw material that passed 200 micron screen (R_200), heating value of fuels in kiln (HVA), heating value of fuels in calciner B (HVB), heating value of fuels in calciner C (HVC), kiln temperature (T), kiln torque (KT), kiln excess oxygen (O2), raw material moisture (Moist_RM),

fuel moisture (Moist_F), non-reacted calcium oxide in kiln (CaO), kiln operating time (KR), raw mill operating time (RM), coal mill operating time (CM), and clinker produced (CK). The percent raw material residual that passed 90 and 200 micron screen represent coarse and fine characteristic of the raw materials.

3.2 Prediction Models

Results show that the MR models for predicting NO_2, SO_2, CO_2, HCl and TSP, have the Adjusted R^2 values in the range of 0.25-0.57 (Table 1).The NO_2 model has the highest R^2,it can explain 57% of the variation in the emission. The TSP model has the lowest R^2 and can explain only 25% of the emission variation. The reason could be that the TSP concentrations were controlled by the EP and thus were not varied significantly with other variables. The independent variables frequently appear as predictors in the MR models are clay, limestone, raw mill operating time, alternative fuels, kiln operating time, and clinker produced. More discussion on the MR models can be found in the related paper published earlier [5]. Note, however, that some results in the current paper are different from the referred one due to improved analysis. The ANN modelshave the R^2 values inthe range of 0.44-0.66 (Table 2).Their MAPE are in the range of 15.20-71.30. The CO_2 model has the highest R^2 values and low MAPE, with 27 independent variables in use. The SO_2 model does not perform well relative to the rest of the models.

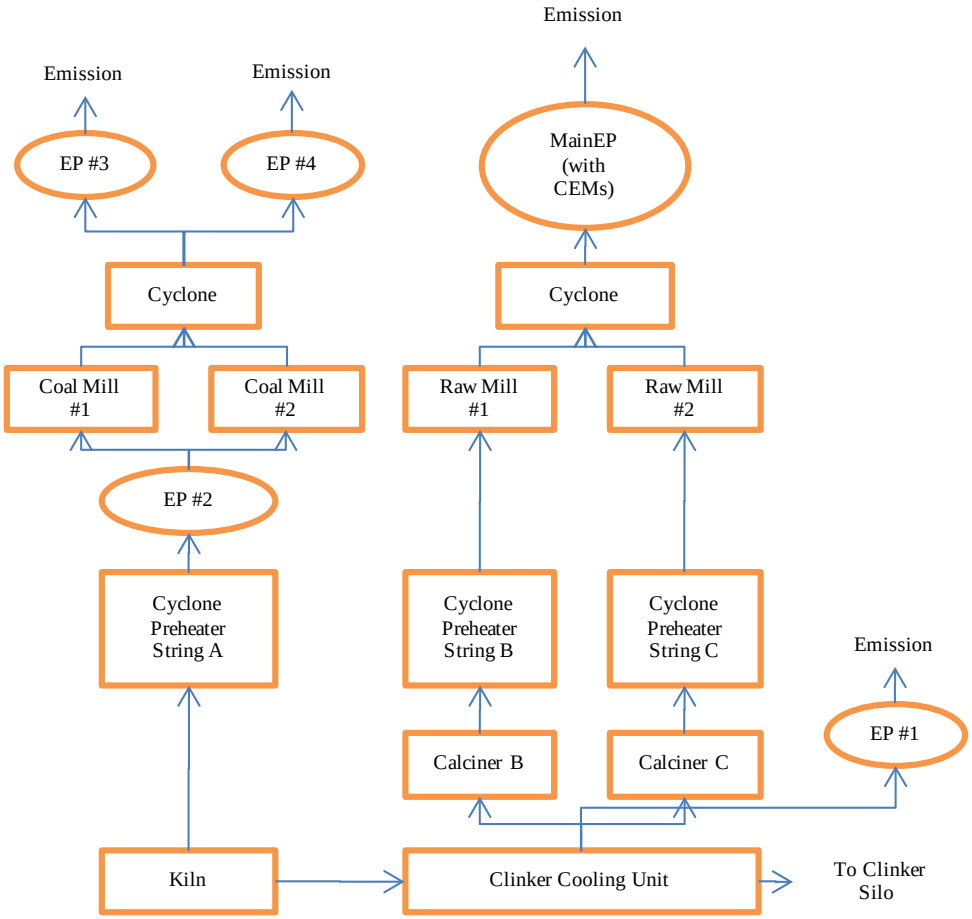

Figure 2. Emissions from Cement Kiln Operation.

Table 1. Optimum MR models for prediction of gas and particulate concentrations.

Gas/Particulate	MR Model	Regression		Validation
		Adj. R^2	RMSE	R
NO_2	NO_2 = - 51.686 + 0.020Clay + 1.445W_46 + 0.062T + 1.093W_23 + 20.584Moist_RM + 41.917R_200 - 0.007ShaleA - 2.907R_90 - 2.101O2 - 0.111KT	0.57	7.05	0.63
SO_2	Log_ SO_2 = - 9.160 + 0.445Log_Lig - 0.316Log_Ant - 0.060Log_Coal_C + 8.546Log_R_200 + 0.067Log_W_42 + 0.406Log_Clay - 0.115Log_LB	0.39	29.99	0.74
CO_2	CO_2 = 19.308 + 1.172Log_W_38 - 1.141Log_RM + 1.277Log_W_23 - 0.618Log_KD + 0.113Log_W_03 - 0.082Log_Coal_C	0.52	1.27	0.44
HCl	HCl = 55.418 - 1.966Log_RM + 0.687Log_W_38 - 5.281Log_HVA - 0.168Log_Coal_C + 0.118Log_W_03 + 0.189Log_W_42 + 0.471Log_W_34 - 0.290Log_Ant + 0.953Log_W_36	0.40	1.63	0.60
TSP	Log_TSP = 3.253 + 7.12E-5 ShaleS + 0.006AR + 1.53E-4 Clay - 0.002RM - 0.007KD	0.25	8.41	0.34

Table 2. Optimum ANN models for prediction of gas and particulate emission.

Gas/ Particulate	Variable Transformation		Architecture[1]	Parameter[2]			Statistics	
	Independent	Dependent		η	w	α	R^2	MAPE
NO_2	-	-	27-29-1	0.2	0.5	0.9	0.44	21.23
SO_2	-	Natural Logarithm	18-24-1	0.1	0.7	0.1	0.37	71.30
CO_2	-	-	13-22-1	0.2	0.5	0.5	0.66	15.34
HCl	-	Natural Logarithm	19-25-1	0.05	0.3	0.9	0.65	25.27
TSP	-	Natural Logarithm	7-19-1	0.2	0.3	0.9	0.61	15.20

[1]Input–Hidden layer–Output;[2]η: learning rate, w: initial weight,α: momentum

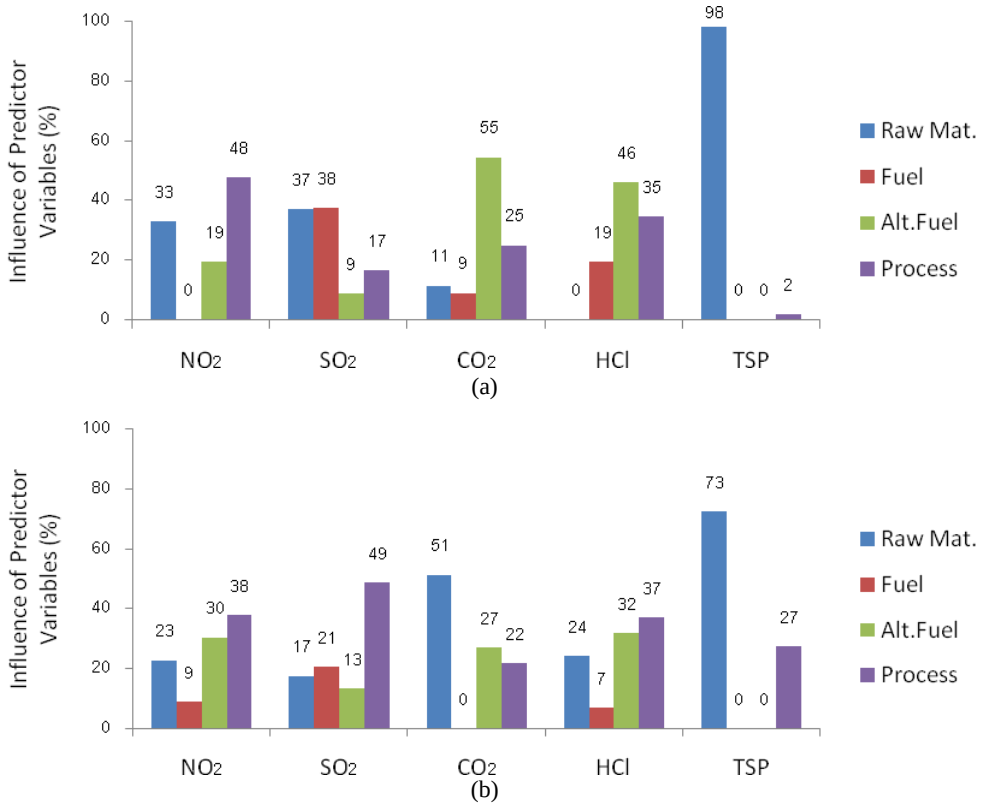

Figure 3. Influence of predictor variables to gaseous and particulate emission: (a) MR models and (b) ANN models.

In each MR model, the influence of predictor variables to the dependent variable can be evaluated using standardized regression coefficients, or beta. The more beta value, the more influence of that predictor to the gas or particulate concentration emitted. Based on the beta value of the independent variables, the most influential predictor of NO_2, SO_2, CO_2, HCl and TSP are clay, clay, alternative fuel – other waste, raw mill operating time, and alternative raw material, respectively. Fig. 3(a) shows the influence of 4 groups of predictors – raw material, fuel, alternative fuel, and process – on the concentrations of gas and particulate matter in the emission. The raw material group is clearly the main influence of TSP concentration, while the alternative fuel group highly influences CO_2 and HCl concentration values. SO_2 is influenced almost equally by the fuel and raw material groups. Concentration of NO_2, on the other hand, is influenced most by the process group.

In the similar analysis, the influence of predictor variables to the dependent variable in each ANN model was evaluated using their contribution factors – values given to each independent variable after the model development is completed. The influence of the 4 groups of predictors is shown in Fig. 3(b). The results are comparable to those of the MR models, only the influence is less pronounced. It is worth mentioned that, while the MR approach eliminates less significant variables from the final model, the ANN approach contains all independent variables assigned to the model. Therefore, the influence of the predictors in the ANN model spreads out more.

3.3 Model Performance Comparison

The results of the final model validation process are presented in Fig. 4, which shows the scatter plots of predicted values (X-Axis) versus actual values (Y-Axis). The R values range from 0.34-0.74 and 0.52-0.78 for the best MR and ANN models, respectively. The prediction values from ANN models correlate with actual data better in all 5 cases. The SO_2 model has the highest R value in both types of model, due partly to the wider range of values. Despite the fact that the TSP concentrations were controlled by the EP, as mentioned earlier, the ANN model still managed to achieve moderate prediction ability – R value equals 0.59.

Conclusion could be drawn in the same direction when considering the I.A. and RMSE results (Fig. 5). The I.A. values of both types of model are rather close together, ranging from 0.64-0.94 and 0.63-0.81 for the MR and ANN models, respectively. The RMSE values range from 1.7-11.2 and 1.6-22.7, respectively. With the exception of SO_2 models, both types of statistic suggested that the ANN models perform marginally better than the MR models. This conclusion is common with relevant study [6-7].

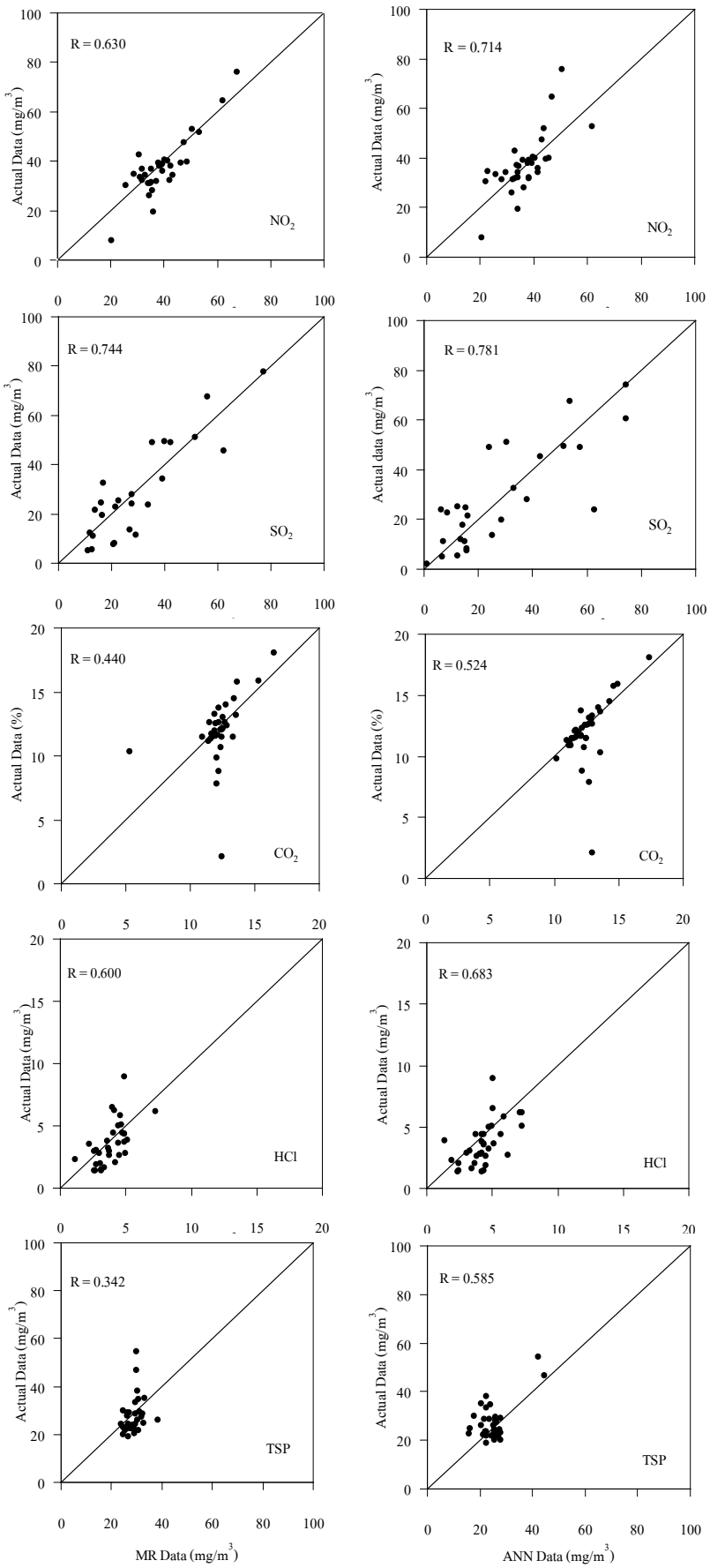

Figure 4. Scatter plots of predicted values versus actual values.

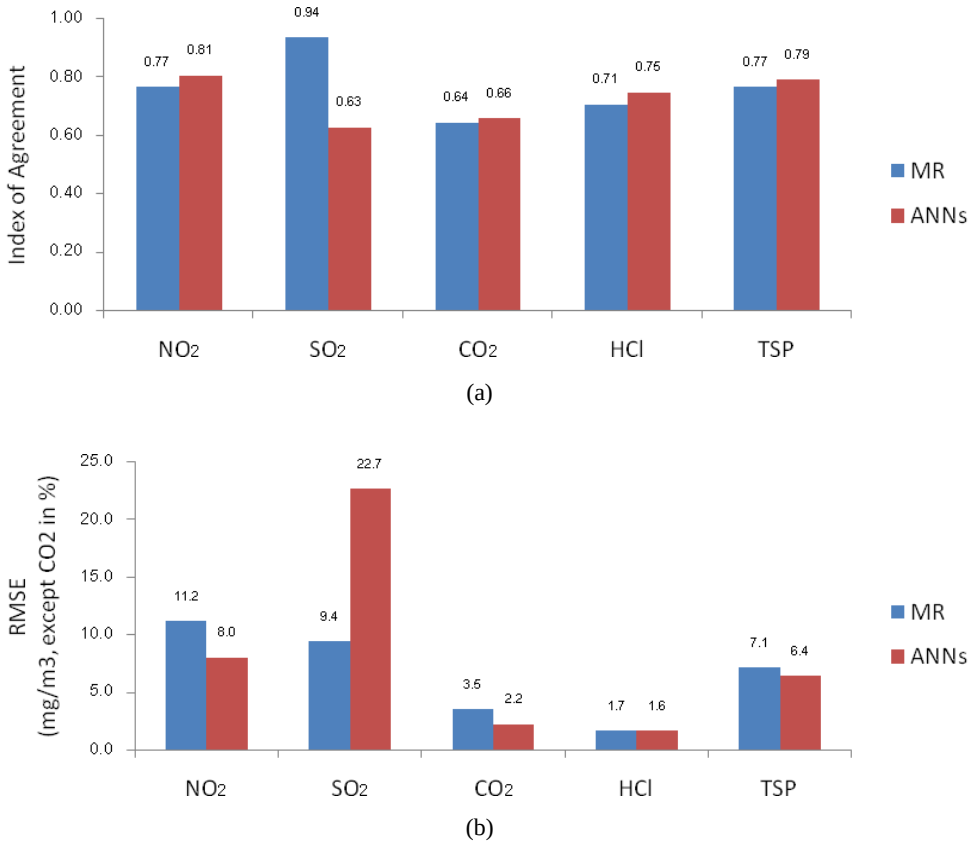

Figure 5. Model validation statistics: (a) I.A. and (b) RMSE.

4. Conclusions

The pertinent factors of cement production which affect the gas and particulate concentrations in the emission were investigated in this study. The statistical models obtained could predict NO_2, SO_2, CO_2, HCl and TSP with the R^2 values in the range of 0.25-0.66. Influential predictors in the models were identified. The outcomes of this study are beneficial tools for managing the plant emission, such as selecting production condition or raw material ratio to reduce pollution emission, or predicting future emission in different scenarios.

Acknowledgments

The authors gratefully acknowledge the contribution of the Thailand Research Fund in cooperation with Siam City Cement Public Company Limited.

References

[1] Gabel K, Tillman A, Simulating Operational Alternatives for Future Cement Production, *Journal of Cleaner Production* 13 (2005) 1246-1257.

[2] Slini T, Kaprara A, Karatzas K, Moussiopoulos N, PM10 Forcasting for Thessaloniki, Greece, *Environmental Modelling & Software* 21 (2006) 559-565.

[3] Fuller GA, Carslaw DC, Lodge HW, An Empirical Approach for the Prediction of Daily Mean PM10 Concentration, *Atmospheric Environment* 36 (2002) 1431-1441.

[4] Jiang D, Zhang Y, Hu X, Zeng Y, Tan J, Shao D, Progress in developing an ANN model for air pollution index forcast, *Atmospheric Environment* 40 (2004) 7055-7064.

[5] Kwanma J, Karuchit S, Regression Models for NO_2, SO_2, CO_2, HCl, and PM Emission from Cement Incineration Process,*Thai Environmental Engineering Journal* 24/2 (2010) 119-126.

[6] Pansripong S, Karuchit S, Kongjan T, Prediction of PM10 Levels 5 Days in Advance Using Neural Networks and Regression Analysis, *Thai Environmental Engineering Journal* 23/3 (2009) 63-72.

[7] Pansripong S, Karuchit S, Kongjan T, Prediction of PM10 Concentration 24h in Advance Using Neural Networks in Bangkok, Thailand, KKU Research Journal 13/9 (2008) 1049-1057.