IOP Conf. Series: Earth and Environmental Science **18** (2014) 012100 doi:10.1088/1755-1315/18/1/012100

PM_{10} Pollution: Its Prediction and Meteorological Influence in PasirGudang, Johor

A Afzali[1], M Rashid[2,*], B Sabariah[3], M Ramli[1]

1- Department of Chemical Engineering, Faculty of Chemical Engineering, UniversitiTeknologi Malaysia, 81310 Skudai, Johor, Malaysia
2- Air Resources Research Laboratory, Malaysia-Japan International Institute of Technology, UTM Kuala Lumpur, 54100 Kuala Lumpur, Malaysia
3- Malaysia-Japan International Institute of Technology, UTM Kuala Lumpur, 54100 Kuala Lumpur, Malaysia

*drrashid@ic.utm.my

Abstract. Ambient PM_{10} (i.e particulate diameter less than 10um in size) pollution has negative impacts on human health and it is influenced by meteorological conditions. Although the correlation between meteorological parameters and PM_{10} concentrations is significant in most cases, the linear relationship between them implies that the fraction of the variance, R2 rarely exceeds 25%. However, considering the previous day's concentration of pollutants to the multi-linear regression enhances the model performance and increases the value of R2. Alternatively, artificial neural networks (ANN) are used to capture the complex relationships among many factors considered which present a better prediction. Thus, this study presents the results of predicting ambient PM_{10} concentration and the influence of meteorological parameters based on the data sampled from 2008 – 2010 in an industrial area of PasirGudang, Johor.

1. Introduction

Ambient PM_{10} participates in the chemical and physical processes in the atmosphere resulting in its characteristics diversity over time. In addition, meteorological conditions play a major influence on the formation of PM_{10}, and elevated PM_{10} concentration is expected as the result of unfavorable meteorological conditions [1,2].

Statistical techniques have been used to explain the relatiosnhip between PM_{10} concentrations meteorological parameters. Kartal and Ozer (1998) applied multiple regression technique considering meteorological parameters and previous day's pollutant concentrations to predict SO2 and smoke in Kayseri-Turkey [3]. Kovač-Andrić et al., (2009) used a multiple linear regression (MLR) to explain the variations in ozone concentrations in Slavonia, eastern part of Croatia [4]. The authors reported that the variations in ozone concentrations are associated with the variation of few meteorological parameters with the R2=0.7 and R2=0.8 in the spring and summer, respectively.

Artificial Neural Network (ANN) has also been applied widely in different environmental issues. The ANN was used to predict PM concentration prediction in association to the meteorological

* To whom any correspondence should be addressed

variables [5,6]. Grivas and Chaloulakou (2006) predicted PM_{10} hourly concentrations in four locations within the Greater Athens Area by developing various neural network models [1]. The authors applied a genetic algorithm optimization procedure for selecting the meteorological and time-scale input variables for the period of 2001–2002 and found that the ANN models were rather satisfactory than the multiple linear regression models.

In this study the relationship between PM_{10} concentrations and meteorological parameters such as wind speed, relative humidity, solar radiation and temperature has been analysed statistically for the case study of PasirGudang industrial area. In the study, the previous day's PM_{10} concentration was considered in the prediction of PM_{10} concentration of the area. In addition, results of the statistical model were compared with that of artificial neural networks. Two different Artificial Neural Networks were developed to predict the one day ahead PM_{10} concentration and the results are presented in this paper.

2. Materials and Methods

2.1. *PM_{10} and Meteorological Data*

The PM_{10} concentration and meteorological parameters were obtained form an air quality monitoring station sited in the area of PasirGudang, Johor. The sample collected from 2008 to 2010 was used in the study. Table 1 presents the summary of the basic statistics of the PM_{10} and meteorological parameters sampled and used in the analysis.

Table 1. Mean, standard deviation and ranges of pollutantsand meteorological parameters at PasirGudang (2008-2010).

Parameter	Mean	Standard Deviation	Range
PM_{10} (ug/m^3)	53.4	19.8	20.3 - 157
SO_2 (ppm)	0.005	.0036	0-0.023
NO_2 (ppm)	0.026	0.0137	0.0034-0.08
CO (ppm)	0.62	0.23	0.121-1.769
O_3 (ppm)	0.014	0.0058	0.002-0.035
Temperature (°C)	27.3	1.28	23.3 - 31.0
Wind Speed (m/s)	6.07	1.44	3.32 - 13.6
Humidity Rate (%)	81.6	4.88	67.6 - 93.8
Solar Radiation (J/m^{2}hr)	195	69.2	33.3 - 497

2.2. *Statistical Analysis*

In this study, a statistical technique was applied to study the relationship of each individual meteorological parameter (as independent variable) on the PM_{10} concentration found at the site. Then, the relationship of meteorological parameters as a whole (i.e temperature, wind speed, relative humidity and solar radiation) on the PM_{10} was investigated through a multiple linear regression analysis. The MLR model was also developed by including the previous days of PM_{10} concentrations as the independent variable into the fitting. Similarly, the prediction of the PM_{10} concentrations using ANNs also took this approach into consideration.

Regression analysis can be applied to analyse the relationship among various variables and obtain a suitable prediction equation and model. Multiple linear regression is used when the number of independent variables are going to be more than one. For example considering four independent variables, a general regression equation can be expressed as:

$$y = a + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + e$$

where and are the constant and coefficient of regression, respectively and is the error. To minimize the error, the least squares method is used to determine the values of the constant and coefficients. The T and Z distribution are used to test the significance level of the constant and coefficients statistically. The coefficient of determination (R2) is defined as the measurement of goodness of fit of a linear model.

2.3. *Artificial Neural Network Approach*

Artificial neural network (ANN) is based on the neurons connected or functionally-related to each other, imitate the behaviour of human biological neurons. Neurons which are the basic components of the neural network are interconnected through different layers such as input, hidden and output layers. The degree of interconnection is defined by the weight which means the impact of neuron on neuron . Each neuron is activated by an activation function based on a threshold value. The transfer activation is processed by both input signals neighbour nodes belonging to different layers and output signal such as bias to get the output of neuron. The common transfer functions in the ANN are tangent sigmoid function and linear function.

The important factor on performance is the topology of ANN. The main categories of neural networks are Feed Forward and Recurrent Neural Networks [7]. In Feed Forward networks such as Multi-Layer Perceptron (MLP) and Feed Forward Network (FNN), data enter into input layer for processing layer by layer to achieve the output layer. Recurrent neural networks such as Elman and Hopfield have the recurrent path through feed backing from output layer to hidden layer for the detection and generation of time-varying patterns. ANN should be configured to produce the desired output by adjusting the weights of interconnections among all neuron pairs. This process is called as training. The data are considered in three different data sets including training, test and validation in artificial neural network.

3. Results and Discussion

3.1. *statistical Model*

Table 2 presents the correlation coefficient between PM_{10} concentration and meteorological variables which showed that there was a negative correlation between PM_{10} and wind speed, solar radiation and relative humidity. Increase in wind speed would cause PM_{10} to dilute by dispersion and hence decrease its concentrations in the atmosphere [8]. The reason for negative relationship between solar radiation and PM_{10} is that during the sun radiation, the surface is warmer and the exchange of heat in the air is more intense. So mixing of air and increases the size of eddies which helps reducing the concentration of pollutant through dispersion [9]. While, relative humidity is commonly affected by the number of rain occasions which through wash-out processes of the atmospheric aerosols, reduces the concentration of pollutant in the air [10,11].

On the contrary, there was a positive relationship between the PM_{10} and temperature in this study suggesting that ambient temperature affects the concentration of PM_{10}. Temperature increases the chemical reaction in the atmosphere resulting in the formation of finely divided particulate matter that naturally contributes to the concentration of PM_{10}. The fact that there is a positive correlation between PM_{10} and other gaseous pollutants like SO2 and NO2 suggests that these gases also contribute to the formation of fine sulfate and nitrate particles as part of the PM_{10}concentration in the atmosphere.

Table 2.Correlation coefficient between $PM_{10}$and meteorological variables for PasirGudang.

	SO_2	NO_2	PM_{10}	CO	O_3	T	SR	HR	WS
PM_{10}Pearson Correlation	$.360^{**}$	$.424^{**}$	1	$.185^{**}$	$.067^{*}$	$.321^{**}$	$-.110^{**}$	$-.242^{**}$	$-.153^{**}$
Sig. (2-tailed)	.000	.000		.000	.026	.000	.000	.000	.000

N	1096	1096	1096	1096	1096	1096	1096	1096	1096

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).
[T] = Temperature; [SR] = Solar Radiation; [HR] = Humidity Rate; [WS] = Wind Speed

As in Table 2, although the correlation between meteorological parameter and PM_{10} concentration is found to be significant in most cases, the linear relationship between them implies that the fraction of the variance (i.e R2) rarely exceeds 25% (Table 3). The low R2 suggests that the investigation of other variables contributing to the variability of pollutants at the site should be considered simultaneously [9].

Table 3. The regression linear models of PM_{10} with each individual meteorological parameters for PasirGudang.

Pollutants	Regression Linear Model	R^2	R
PM_{10}	=66.4 – 2.1*[WS]	0.024	0.155
	=-80.8 + 4.93*[T]	0.10	0.32
	=59.8 – 0.031*[SR]	0.012	0.11
	=134 – 0.98*[HR]	0.059	0.242
	=0.64*Previous day's PM_{10} concentration +19	0.41	0.64

Thus, taken this into consideration the multi linear regression equation considering all the meteorological parameters as predictor variables affecting the PM_{10} concentration is:

$$PM_{10}=90.6 - 3.26*[WS] + 2.93*[T] - 0.048*[SR] - 1.07*[HR], \quad R2=0.18$$

The above equation showed that the relationship between the meteorological parameters and PM_{10} concentration is not adequately explained (i.e small value of R2). However (see Table 3), it is oberved that linear relationship between PM_{10} and its previous day concentration revealed a higher R2=0.41, which concurred that the concentrations of PM_{10} can also be influenced by its concentration in the previous day [8].

Thus, including the previous day's concentration of PM_{10} into the multi linear regression revealed that:

$$PM_{10}=75.1-2.39*[WS]+1.13*[T]-0.019*[SR]-0.79*[HR]+0.57*PM_{10}\text{' previous day}, \quad R2=0.47$$

The inclusion of the previous day PM concentration in the model slightly improved the R2 value of the fitting but was not fully explained the relationship. Alternatively, artificial neural networks were considered to treat the data further.

3.2. *Artificial Neural Network Model*

Two techniques of neural network were applied in two scenarios. One of them is Feed Forward Network (FNN) and the other one is Elman Network. FNN is implemented by two hidden layers and one output layer, while Elman network is implemented by one hidden layer and one output layer. The overview of the parameters and their values in two different scenarios are presented in Table 5, which showed that the two applied neural networks have rather presented small errors (MSE and MAE) and high correlation coefficient (R) in predicting PM_{10} concentration, with Elman network presented more precise results. The prediction of PM_{10} concentrations for one day ahead was performed. The model performances are presented in Figures 1a and 1b which showed a good agreement between the predicted and daily observed PM_{10} concentration with R=0.69 and R=0.70 for one day ahead for FNN and Elman networks, respectively. Hence, artificial neural networks are able to be a predicting tool in estimating PM_{10} concentration to a certain degree of accuracy in this case.

Table 4.Structure and training results for the neural network models.

Net Type	Topology	Training function	Learning	Learning rate	MSE	MAE	R	Epoch	Use
FNN	8 12 1	Tansig-Tansig-Purelin	Trainlm	0.01	0.0461	0.1532	0.69	100	daily
Elm	10 1	Tansig-purelin	Trainlm	0.01	0.0446	0.1421	0.70	16	Daily

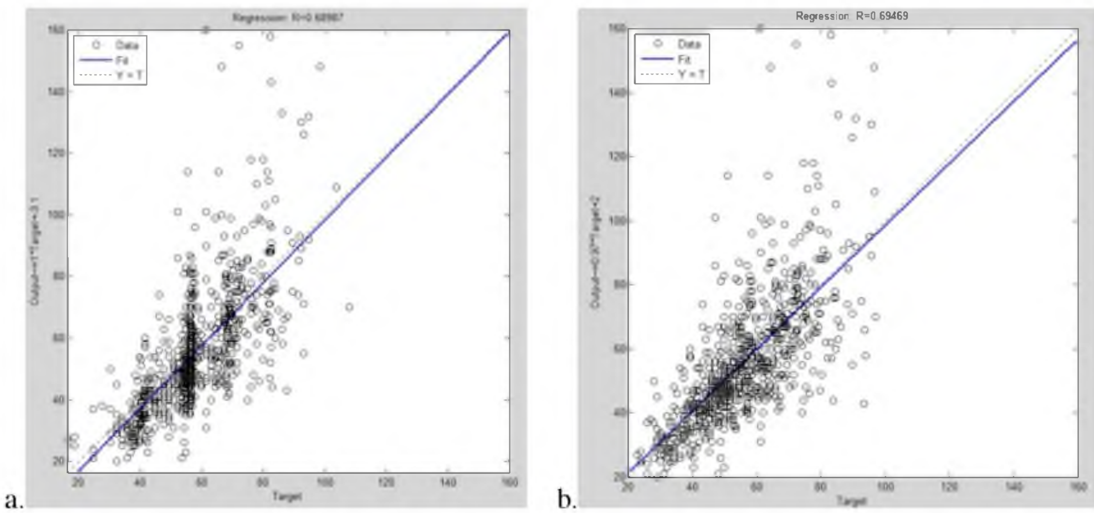

Figure 1. Comparison between the daily predicted and observed PM_{10} concentrations using a. FNN network and b. Elman network.

4. Conclusion

The concentration of PM_{10} is associated with the meteorological parameters. However the results from this study show that the relationship between ground level concentration of PM_{10} and each meteorological parameter is weak, while the statistical developed model for PM_{10} and the meteorological parameters shows that the value of R2 is 0.18. Inclusion of the previous day's PM_{10} concentration improved the model significantly but not fully explained the relationship. However, the use of artificial neural networks significantly improved the model fitting with good accuracy.

References

[1] Grivas G and Chaloulakou A 2006 Artificial neural network models for prediction of PM_{10} hourly concentrations, in the Greater Area of Athens, Greece Atmospheric Environment. 40 1216- 1229

This reference has two entries but the second one is not numbered (it uses the 'Reference (no number)' style.

[2] Carnevale C, Pisoni E and Volta M 2010 A non-linear analysis to detect the origin of PM_{10} concentrations in Northern Italy Science of the Total Environment. 409 182- 191More references

[3] Kartal S and Ozer U 1998 Determination and parameterizationof some air pollutants as a function of meteorologicalparameters in Kayseri, Turkey Air and Waste Management Association. 48 853-859

[4] Kovač-Andrić E, Brana J and Gvozdić V 2009 Impact of meteorological factors on ozone concentrations modelled by time series analysis and multivariate statistical methods Ecological Informatics 4 117–122

[5] Perez P and Reyes J 2006 An integrated neural network model for PM_{10} forecasting Atmospheric Environment 40 2845- 2851

[6] Voukantsis D, Karatzas K, Kukkonen J, Räsänen T, Karppinen A and Kolehmainen M 2011 Intercomparison of air quality data using principal component analysis, and forecasting of PM_{10} and $PM_{2.5}$ concentrations using artificial neural networks, in Thessaloniki and Helsinki Science of the Total Environment. 409 1266- 1276

[7] Widrow B and Sterns S D 1985 Adaptive signal processing. New York: Prentice-Hall

[8] Akpinar S, Oztop H F and Akpinar E K 2008 Evaluation of relationship between meteorological parameters and air pollutant concentrations during winter season in Elazığ, Turkey Environ Monit Assess. 146 211-224

[9] Zeri M J, Oliveira-Ju´nior J F and Lyra G B 2011 Spatiotemporal analysis of particulate matter, sulfur dioxide and carbon monoxide concentrations over the city of Rio de Janeiro, Brazil Meteorol Atmos Phys. 113 139-152

[10] Azmi S Z, Latif M T, Ismail A S, Juneng L and Jemain A A 2010 Trend and status of air quality at three different monitoring stations in the Klang Valley, Malaysia Air Qual Atmos Health. 3 53-64

[11] Gvozdić V, Kovač-Andrić E and Brana J 2011 Influence of Meteorological Factors NO_2, SO_2, CO and PM_{10} on the Concentration of O_3 in the Urban Atmosphere of Eastern Croatia Environ Model Assess. 16 491-501