Malaysian Journal of Computing, 4 (2): 325-334, 2019

eISSN: 2600-8238 online

CLASSIFICATION OF HIGH AND LOW LEVEL OF PM_{10} CONCENTRATIONS IN KLANG AND SHAH ALAM, MALAYSIA

Hasfazilah Ahmat, Nor Syahida Musa, Nurhanina Nazamid and Nursyahirah Amirah Zaharin

Centre of Statistical and Decision Science Studies,
Faculty of Computer and Mathematical Sciences,
Universiti Teknologi MARA,
40450 Shah Alam, Selangor, Malaysia
hasfazilah@fskm.uitm.edu.my, norsyahida_96@yahoo.com,
ninaaiens250996@gmail.com and [4]ezemirazaharin@gmail.com

ABSTRACT

Particulate matter (PM) comprises of a complex mixture of small solid or liquid particles of organic and inorganic elements that floats freely in air. PM_{10} is defined as a particulate matter with an aerodynamic diameter of 10 μm or less. The main objective of this paper is to classify the level of PM_{10} in selected locations in Peninsular Malaysia using discriminant analysis. Two important components considered in this study, namely; the meteorological factors and pollutant factors. The meteorological factors comprise of wind speed, wind direction, humidity and temperature while pollutant factors consist of Carbon Monoxide (CO), Sulphur Dioxide (SO_2), Nitrogen Dioxide (NO_2) and Ozone (O_3). The classification of high or low level of PM_{10} concentrations was based on the Malaysia Ambient Air Quality Guideline (MAAQG). The findings indicated that the classification equation differs from location to location due to different levels of PM_{10} concentrations, location of monitoring stations and factors affecting air pollution in that location. The simulation data also verified that the classification of PM_{10} concentration was almost similar to the real condition that occurred in Klang in October 2015.

Keywords*: Air Pollution, Discriminant Analysis, PM_{10}.*

Received for review: 21-08-2019; Published: 19-12-2019

1. Introduction

Air pollution can be defined as the presence of unwanted chemical or other elements in air that affects the quality of air and human health (World Health Organization, 2018). In 2015, over 90% of the world's population lived in air-polluted areas (HEI International Scientific Oversight Committee, 2017). One of the most vital causes of the deterioration in air quality is particulate matter (PM) and it instigates some adverse health effects (Capasso *et al.*, 2015). Exposure to air pollutants for both short and long-term period has been associated with health effects (World Health Organization, 2018).

Five major risk factors for total deaths in the world are high blood pressure, smoking, high fasting plasma glucose, high total cholesterol and ambient particulate matter (HEI

International Scientific Oversight Committee, 2017). The particulate matter consists of tiny solid or liquid particles that float freely in the air. PM_{10} refers to the particles which have sized up to 10 microns (μm). The smaller the particles' size such as PM_1, the more severe it will affect human health if the particles are inhaled excessively into lungs (Beh *et al.*, 2013). The dominant pollutant in Malaysia is PM_{10} (Department of Environment Malaysia, 2018). A study by Elhadi *et al.* (2018) stated that vehicles' exhaust and non-exhaust, industrial emission, resuspension dust and oil combustion were the most dominant sources of PM_{10}. PM_{10} may cause adverse effect on the environment, increase the risk of health problems to individuals with asthma or cardiopulmonary diseases, the elderly and children as well as reduced in visibility (Abd Rahman, 2013 and Weinmayr *et al.*, 2010).

There are quite a number of statistical analyses which involve PM_{10} in Malaysia. Some of the statistical analyses that were of interest of past researchers are the regression, used in the studies by Abdullah *et al.* (2017), Juneng *et al.* (2011), Mert Cubukcu & Sinem Ozcan (2015) and Ul-saufie *et al.* (2012), correlation analysis in Biancofiore *et al.* (2017), How & Ling (2016) and Wie & Moon (2017), path analysis in Sahanavin *et al.* (2018) and Markov Chain Model in a study by Mohamad *et al.* (2017).

Other studies that applied multivariate analysis were Hama *et al.* (2018) and Dominick *et al.* (2012) which utilized principal component analysis (PCA); and Isiyaka & Azid (2015) and Shah Ismail *et al.* (2017) which used discriminant analysis but focusing only on meteorological factors. Meanwhile, some researchers applied time series analysis as in Latif *et al*. (2014), Wan Mahiyuddin *et al.* (2013), Sharma *et al.* (2018) and Gupta *et al.* (2018). Some other researchers used the classical probability distribution (Md Yusof, 2009; Md Yusof *et al.*, 2011; Mohamed Noor *et al.*, 2011) and extreme value distributions (Ahmat *et al.*, 2014; Ahmat *et al.*, 2015, 2016).

Though discriminant analysis has been applied in some air pollution studies, the focus of the studies was only on gaseous pollutants (Isiyaka & Azid, 2015; Shah Ismail *et al.*, 2017). The study of PM_{10} which incorporates both gaseous pollutants and meteorological factors, however, is still lacking. In addition, none of the studies classifies the PM_{10} concentrations into high and low category based on the national guideline. The majority of the studies conducted focused only on the prediction or the forecast of the PM_{10} concentrations but not on the classification of the PM_{10} concentrations. In view of this situation, this research was carried out to classify low or high level of PM_{10} concentrations based on an interim guideline by the Department of Environment, Malaysia (DOE) which incorporated both gaseous pollutants and meteorological factors.

2. Materials and Methods

2.1 Scope of study

This study utilized the hourly data of meteorological parameters and pollutants in urban areas (Klang and Shah Alam) for a period of 17 years i.e. from 2000 to 2016. The data was furnished by the Department of Environment (DOE), Malaysia. The selection of these two locations i.e. Klang and Shah Alam were made due to the factor that these two locations constantly experienced high level of PM_{10} concentrations.

This research examined the effects of meteorological parameters (temperature, humidity, wind speed and wind direction) and gaseous pollutants (SO_2, NO_2, O_3 and CO) on PM_{10} concentrations. The level of PM_{10} was classified as high or low based on the Malaysia new interim guideline by the Department of Environment, Malaysia (DOE) of $150\mu g/m^3$. For the purpose of discriminant analysis, these data were divided into two parts with 80% of the data were used for training (to find the discriminant functions) and another 20% were used for validation.

2.2 Statistical Analysis

2.2.1 Missing Value Treatment

Missing data is not a rare problem in air quality datasets as it is usually due to some unavoidable problems such as failures of machines, changes on the setting of air station monitors or human error in handling the datasets. There are three types of missing data which are missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR) (Gelman & Hill, 2006). The multiple imputation technique was used in this study to overcome the problem of missing data. Multiple imputation can lead to consistent, efficient and normal estimates when the data is MAR (Soley-Bori *et al.,* 2013).

2.2.2 Discriminant Analysis

Discriminant analysis is a statistical technique that can be used to classify or separate individuals into different groups (dependent variable) based on a set of quantitative independent random variables. The main objective of discriminant analysis is to predict group membership based on a set of quantitative variables. Discriminant function analysis is used to determine which continuous variables discriminate between two or more naturally occurring groups and could be used to determine which variables are the best predictors. The two-step processes involved were (Poulsen & French, 2008) ;

i. testing significance of a set of discriminant functions, and
ii. the classification.

In this study, the data was carefully checked and cleaned so that it did not violate all the assumptions needed for the discriminant analysis to be carried out. The statistical method used for the selection of the significant factors to be included in the discriminant equation was the stepwise model. The statistical method used for the selection of the significant factors to be included in the discriminant equation was the stepwise model. The equation for cases with an equal sample size for each group the classification function coefficient $\left(D_j\right)$ is equal to the sum as shown in Eq. (1):

$$D_j = c_{j0} + c_{j1}x_1 + c_{j2}x_2 + ... + c_{jp}x_p \tag{1}$$

For the j^{th} group, j is 1...k, x is a raw score of each predictor and c_{j0} is a constant. If M is a column matrix of means for group j, then the constant $c_{j0} = \left(-\frac{1}{2}\right)D_j M_j$.

2.2.3 Performance Indicators

The performance of the classification function is assessed via its error rates (probabilities of the misclassification). The error rate and the percentage of the observations misclassified by the discriminant functions are used to measure the performance of any discriminant function (Helwig, 2017).

The Apparent Error Rate (APER) was used to identify the goodness of fit of the function in this study and calculated using the fraction of observations in the training sample that are misclassified by the sample classification functions as shown in Table 1.

Table 1. Confusion matrix

		π_1	π_2	
Actual membership	π_1	n_{1c}	$n_{1M} = n_1 - n_{1c}$	n_1
	π_2	$n_{2M} = n_2 - n_{2c}$	n_{2c}	n_2

n_{1c}	Number of π_1 items classified as π_1 items
n_{1M}	Number of π_1 items misclassified as π_2 items
n_{2c}	Number of π_2 items correctly classified
n_{2M}	Number of π_2 items correctly misclassified.

The Apparent Error Rate (APER) is calculated as shown in Eq. (2),

$$APER = \frac{n_{1M} + n_{2M}}{n_1 + n_2} \tag{2}$$

which is recognized as the proportion of items in the training set that are misclassified (Johnson & Wichern, 2014). Table 2 provides the sample calculation of the *APER*.

Table 2. Sample calculation of the *APER*

		π_1	π_2	
Actual membership	π_1	$n_{1c} = 10$	$n_{1M} = 2$	$n_1 = 12$
	π_2	$n_{2M} = 2$	$n_{2c} = 10$	$n_2 = 12$

The Apparent Error Rate (APER) as expressed as a percentage, is

$$APER = \left(\frac{2+2}{12+12}\right)100\% = 16.7\% .$$

In order to verify the goodness of fit of the discriminant function, Cross Validation Rate was used (Bian, 2012) :

i. Sample is split into training and validation.
ii. Training sample is used to build the discriminant function.
iii. Validation sample is used to evaluate the performance of the discriminant functions.
iv. Cross validation error rate is the percentage of observations in the validation data, which are misclassified by the classification functions.
v. Cross validation rate can overcome bias problem, but it requires large sample.

2.2.4 Software

IBM SPSS statistics version 25.0 was used in this research for the discriminant analysis. SPSS was used to understand and interpret the results of research.

3. Results

3.1 Pollutants and Meteorological Factors that affect PM_{10} Concentrations

Table 3 and Table 4 provide the significance test result for pollutant and meteorological parameters in Klang and Shah Alam respectively. The null hypothesis would be the parameters are not significant vs the alternative hypothesis that the parameters are significant. The significant *p-value* = 0.000 less than 0.05, hence, the parameters are deemed significant.

Table 3. The Significance Test Result for Parameters in Klang

Variables Entered/Removed[a,b,c,d]

		Wilk's Lambda							
						Exact F			
Step	**Entered**	**Statistic**	**df1**	**df2**	**df3**	**Statistic**	**df1**	**df2**	**Sig.**
1	CO	.944	1	1	4905.000	288.379	1	4905.00	.000
2	Humidity	.938	2	1	4905.000	163.312	2	4904.00	.000
3	SO2	.936	3	1	4905.000	111.710	3	4903.00	.000
4	Temperature	.935	4	1	4905.000	85.053	4	4902.00	.000
5	WD	.934	5	1	4905.000	69.081	5	4901.00	.000

At each step, the variables that minimizes the overall Wilk's Lambda is entered
a. Maximum number of steps is 16
b. Minimum partial F to enter is 3.84
c. Maximum partial F to remove is 2.71
d. F level, tolerance, or VIN insufficient for further computation.

Table 4. The Significance Test Result for Parameters in Shah Alam

Variables Entered/Removed[a,b,c,d]

		Wilk's Lambda							
						Exact F			
Step	**Entered**	**Statistic**	**df1**	**df2**	**df3**	**Statistic**	**df1**	**df2**	**Sig.**
1	CO	.983	1	1	4970.000	87.444	1	4970.000	.000
2	Humidity	.982	2	1	4970.000	46.110	2	4969.000	.000
3	SO2	.981	3	1	4970.000	32.910	3	4968.000	.000

At each step, the variables that minimizes the overall Wilk's Lambda is entered
a. Maximum number of steps is 16
b. Minimum partial F to enter is 3.84
c. Maximum partial F to remove is 2.71
d. F level, tolerance, or VIN insufficient for further computation.

As summarized in Table 5, both the pollutant factors (CO and SO_2) affected PM_{10} concentrations since both locations are located nearby and affected by similar pollutants. However, different meteorological factors affected PM_{10} concentrations in these two locations. It was found that only humidity affected PM_{10} concentrations in Shah Alam compared to three significant meteorological factors in Klang (windspeed, humidity and temperature).

Table 5. Pollutants and Meteorological Factors that affect PM_{10} Concentrations for Klang and Shah Alam

Locations	Pollutant	Meteorological
Klang	CO and SO_2	Windspeed, humidity, temperature
Shah Alam	CO and SO_2	humidity

3.2 Classification of High and Low concentrations of PM_{10}

The concentrations of PM_{10} was classified into high or low using discriminant analysis based on the Malaysia Ambient Air Quality Guideline (MAAQG). The daily maximum PM_{10} concentration with value more than 150 μg/m^3 will be classified as high while the daily maximum PM_{10} concentration with value less than 150 μg/m^3 will be initially classified as low.

Table 6 tabulates the discriminant equations for Klang and Shah Alam. The SO_2 was identified as the most significant factor affecting the level of PM_{10} concentrations in Klang and Shah Alam.

Table 6. Discriminant equations for Klang and Shah Alam

Location	Discriminant Equations
Klang	$D_{Low} = 2.824CO + 574.903SO_2 - 0.006WD + 0.377Temp + 0.052Humidity - 11.548$ $D_{High} = 5.975CO + 465.720SO_2 - 0.010WD + 0.470Temp - 0.000Humidity - 19.998$
Shah Alam	$D_{Low} = 4.079CO + 2958.180SO_2 + 8.273Humidity - 391.479$ $D_{High} = 6.930CO + 2784.723SO_2 + 8.141Humidity - 388.593$

After the discriminant equations have been identified, the classification of PM_{10} concentrations of either high or low can be done via classification scores. The concentrations will be classified into the group for which it has the highest classification score.

3.3 Performance Indicator

As shown in Table 7, the discriminant functions were considered good since all the misclassification rate were less than 5%. In general, the acceptable misclassification rate is about 30%.

Table 7. Misclassification Rate

	Klang	Shah Alam
Training data	2.40%	0.90%
Validation data	1.50%	3.00%

3.4 Simulation

For illustration, Table 8 shows the calculation of discriminant score using discriminant equations obtained in Section 3.2 and the classification using discriminant category. The illustration data used was from 1 October – 19 October 2015 for Klang. Several incidences of high PM_{10} concentrations were recorded during this period. These phenomena were due to four tropical cyclones namely "Dujuan", "Mujigae", "Koppu" and "Champi" that caused southwesterly wind and brought about substantial smoke from the burning areas in Sumatra and Kalimantan resulting in a prolonged haze in September and October 2015 (Department of Environment Malaysia, 2016). The results in Table 8 show an excellent agreement with the findings in Malaysia Environmental Quality Report 2015 (MEQR).

Table 8. Simulation using Discriminant Equations for Klang

Date	**CO**	**SO_2**	**PM_{10}**	**WD**	**Temp**	**Humid**	**PM_{10} Initial Category**	**D_{Low}**	**D_{High}**	**PM_{10} Disc. Category**
20151001	3.67	0.01	233.00	243.71	30.40	81.90	high	17.11	17.06	low
20151002	2.60	0.00	172.84	43.02	29.56	80.50	high	12.59	10.40	low
20151003	4.24	0.01	326.42	142.72	34.68	80.80	**high**	22.03	24.42	**high**
20151004	4.06	0.01	382.52	114.28	30.80	76.20	**high**	19.39	21.29	**high**
20151005	3.38	0.01	224.87	252.45	34.39	78.20	high	18.70	18.05	low
20151006	2.43	0.00	206.22	99.42	28.44	86.40	high	12.23	8.76	low
20151007	3.24	0.01	150.91	149.96	30.27	86.50	high	16.06	14.88	low
20151008	1.87	0.01	119.42	244.07	33.14	84.90	**low**	13.78	8.05	**low**
20151009	2.00	0.00	123.99	175.54	29.69	85.30	**low**	10.41	5.57	**low**
20151010	1.61	0.01	112.04	257.70	32.01	78.20	**low**	10.45	4.39	**low**
20151011	1.70	0.00	139.47	228.38	28.72	78.00	**low**	5.60	0.41	**low**
20151012	0.88	0.00	122.63	169.32	30.07	87.20	**low**	3.49	-4.17	**low**
20151013	2.53	0.01	135.07	235.83	34.31	82.20	**low**	14.27	11.23	**low**
20151014	1.76	0.01	127.34	170.22	30.96	84.40	**low**	11.33	5.69	**low**
20151015	2.61	0.01	119.01	159.70	30.17	79.90	**low**	13.27	10.51	**low**
20151018	3.35	0.00	206.29	147.80	32.25	74.40	**high**	14.79	15.11	**high**
20151019	4.17	0.01	265.62	187.14	32.38	78.60	**high**	18.85	21.06	**high**

4. Conclusion

The research had identified that the main pollutants affected the level of PM_{10} concentrations in Klang and Shah Alam were Carbon Monoxide (CO) and Sulphur Dioxide (SO_2). Klang and Shah Alam are located nearby main roads, industrial and residential areas and thus experienced high density of vehicles which contributed to high concentrations of these two pollutants (Azid *et al.*, 2015).

The misclassification rate shows that the discriminant functions obtained were good since both the misclassification rate were less than 5%. The simulation results show an excellent agreement with the real condition that occurred in Klang in October 2015 as reported in Malaysia Environmental Quality Report 2015 (MEQR). Therefore, the discriminant functions can be used to classify high and low level of PM_{10} concentrations in Klang and Shah Alam.

Acknowledgement

The authors would like to thank the Department of Environment, Malaysia for providing the air quality data in this study.

Reference

Abd Rahman, M. F. (2013). *A study on the concentration and dispersion of PM10 in UTHM by using simple modelling and meteorological factors* (Master's Thesis). Retrieved from http://eprints.uthm.edu.my/3693/

Abdullah, S., Ismail, M., & Yuen Fong, S. (2017). Multiple linear regression (MLR) models for long term PM_{10} concentration forecasting during different monsoon seasons. *Journal of Sustainability Science and Management*, *12*, 60–69.

Ahmat, H., Yahaya, A. S., & Ramli, N. A. (2014). Prediction of PM_{10} extreme concentrations using three parameters extreme value distributions (EVD). In *Sixth International Conference on Postgraduate Education (ICPE-6)*. Melaka: Universiti Teknikal Malaysia Melaka.

Ahmat, H., Yahaya, A. S., Ramli, N. A., Zia, A., Japeri, M., & Hamid, H. A. (2015). Analysis of PM_{10} using extreme value theory, *ESTEEM Academic Journal*, *11*(March), 135–143.

Ahmat, H., Yahaya, A. S., & Ramli, N. A. (2016). Prediction of PM_{10} extreme concentrations in selected industrial monitoring stations in Malaysia using extreme value distribution (EVD): classical and bayesian approaches. *ESTEEM Academic Journal*, *12*(1), 1–10.

Azid, A., Juahir, H., Ezani, E., Toriman, M. E., Endut, A., Abdul Rahman, M. N., … Umar, R. (2015). Identification source of variation on regional impact of air quality pattern using chemometric. *Aerosol and Air Quality Research*, *15*, 1545–1558.

Beh, B. C., Tan, F., Tan, C. H., Syahreza, S., Mat Jafri, M. Z., & Lim, H. S. (2013). PM_{10}, $PM_{2.5}$ and PM_1 distribution in Penang Island, Malaysia. In N. B. Ibrahim, R. Abd-Shukor, I. A. Rahman, Z. Ibarahim, & C. C. Yap (Eds.), *AIP Conference Proceedings* (Vol. 1528, pp. 146–150). Pahang, Malaysia: American Institute of Physics.

Bian, H. (2012). SPSS discriminant function analysis. Retrieved August 21, 2019, from http://core.ecu.edu/ofe/statisticsresearch/SPSS Discriminant Function Analysis.pdf

Biancofiore, F., Busilacchio, M., Verdecchia, M., Tomassetti, B., Aruffo, E., Bianco, S., … Di Carlo, P. (2017). Recursive neural network model for analysis and forecast of PM_{10} and $PM_{2.5}$. *Atmospheric Pollution Research*, *8*(4), 652–659.

Capasso, L., Longhin, E., Caloni, F., Camatini, M., & Gualtieri, M. (2015). Synergistic inflammatory effect of PM_{10} with mycotoxin deoxynivalenol on human lung epithelial cells. *Toxicon*, *104*, 65–72.

Department of Environment Malaysia. (2016). *Malaysia environmental quality report 2015*. Subang Jaya : OMR Press Sdn. Bhd.

Department of Environment Malaysia. (2018). *Environmental quality report 2017*. Shah Alam : Salz-Terachi Design Sdn Bhd.

Dominick, D., Juahir, H., Latif, M. T., Zain, S. M., & Aris, A. Z. (2012). Spatial assessment of air quality patterns in Malaysia using multivariate analysis. *Atmospheric Environment*, *60*, 172–181.

Elhadi, R. E., Abdullah, A. M., Abdullah, A. H., Ash'aari, Z. H., & Khan, F. (2018). Seasonal variations of atmospheric particulate matter and its content of heavy metals in Klang Valley, Malaysia. *Aerosol and Air Quality Research*, *18*, 1148–1161.

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models* (1st ed.). Cambridge : Cambridge University Press.

Gupta, S., Gadi, R., Sharma, S. K., & Mandal, T. K. (2018). Characterization and source apportionment of organic compounds in PM_{10} using PCA and PMF at a traffic hotspot of Delhi. *Sustainable Cities and Society*, *39*, 52–67.

Hama, S. M. L., Cordell, R. L., Staelens, J., Mooibroek, D., & Monks, P. S. (2018). Chemical composition and source identification of PM_{10} in five North Western European cities. *Atmospheric Research*, *214*, 135–149.

HEI International Scientific Oversight Committee (2017). *State of global Air 2017: A special report on global exposure to air pollution and its disease burden*. Boston : Health Effects Institute.

Helwig, N. E. (2017). Discrimination and classification. Retrieved August 21, 2019, from http://users.stat.umn.edu/~helwig/notes/discla-Notes.pdf

How, C. Y., & Ling, Y. E. (2016). The Influence of $PM_{2.5}$ and PM_{10} on air pollution index (API). In *Environmental Engineering, Hydraulics and Hydrology: Proceeding of Civil Engineering*, (pp. 132–143).

Isiyaka, H. A., & Azid, A. (2015). Air quality pattern assessment in Malaysia using multivariate techniques. *Malaysian Journal of Analytical Science*, *19*(5), 966–978.

Johnson, R. A., & Wichern, D. W. (2014). *Applied multivariate statistical analysis*. *The Mathematical Gazette* (6th ed., Vol. 72). Essex: Pearson Education Limited. http://doi.org/10.2307/3619964

Juneng, L., Latif, M. T., & Tangang, F. (2011). Factors influencing the variations of PM_{10} aerosol dust in Klang Valley, Malaysia during the summer. *Atmospheric Environment*, *45*, 4370–4378.

Latif, M. T., Dominick, D., Ahamad, F., Khan, M. F., Juneng, L., Hamzah, F. M., & Nadzir, M. S. M. (2014). Long term assessment of air quality from a background station on the Malaysian Peninsula. *Science of The Total Environment*, *482–483*, 336–348.

Md Yusof, N. F. F. (2009). *Modeling and prediction of PM_{10} concentration during high particulate events in Malaysia* (Doctoral Thesis). Retrieved from http://eprints.usm.my/41297/1/Noor_Faizah_Fitri_Bt_Md_Yusof24.pdf

Md Yusof, N. F. F., Ramli, N. A., & Yahaya, A. S. (2011). Extreme value distribution for prediction of future PM_{10} exceedences. *International Journal of Environmental Protection*, *1*(4), 28–36.

Mert Cubukcu, K., & Sinem Ozcan, N. (2015). Evaluation of air pollution effects on asthma disease: the case of Izmir. *Procedia-Social and Behavioral Sciences*, *202*, 448–455.

Mohamad, N., Deni, S. M., & Japeri, A. Z. U.-S. (2017). Modeling of daily PM_{10} concentration occurrence using Markov Chain model in Shah Alam, Malaysia. *Journal of Environmental Science and Technology*, *10*(2), 96–106.

Mohamed Noor, N., Tan, C. Y., Abdullah, M. M. A.-B., Ramli, N. A., & Yahaya, A. S. (2011). Modelling of PM_{10} concentration in industrialized area in Malaysia : A Case Study in Nilai. In *2011 International Conference on Environment and Industrial Innovation* (Vol. 13, pp. 18–22). Singapore: IACSIT Press.

Poulsen, J., & French, A. (2008). *Discriminant function analysis*. San Francisco, CA: San Francisco State University.

Sahanavin, N., Prueksasit, T., & Tantrakarnapa, K. (2018). Relationship between PM_{10} and $PM_{2.5}$ levels in high-traffic area determined using path analysis and linear regression. *Journal of Environmental Sciences*, *69*, 105–114.

Shah Ismail, A., Makmom Abdullah, A., & Armi Abu Samah, M. (2017). Environmetric study on air quality pattern for assessment in Northern Region of Peninsular Malaysia. *Journal of Environmental Science and Technology*, *10*, 186–196.

Sharma, N., Taneja, S., Sagar, V., & Bhatt, A. (2018). Forecasting air pollution load in Delhi using data analysis tools. *Procedia Computer Science*, *132*, 1077–1085.

Soley-Bori, M., Horn, M., Morgan, J., & Min Lee, K. (2013). *Dealing with missing data: Key assumptions and methods for applied analysis*. Boston : Boston University School of Public Health.

Ul-saufie, A. Z., Yahaya, A. S., Ramli, N. A., & Hamid, H. A. (2012). Robust regression models for predicting PM_{10} concentration in an industrial area, *International Journal of Engineering and Technology*, *2*(3), 364–370.

Wan Mahiyuddin, W. R., Sahani, M., Aripin, R., Latif, M. T., Thach, T.-Q., & Wong, C.-M. (2013). Short-term effects of daily air pollution on mortality. *Atmospheric Environment*, *65*, 69–79.

Weinmayr, G., Romeo, E., de Sario, M., Weiland, S. K., & Forastiere, F. (2010). Short-Term effects of PM_{10} and NO_2 on respiratory health among children with asthma or asthma-like symptoms: A systematic review and Meta-Analysis. *Environmental Health Perspectives*, *118*(4), 449–457.

Wie, J., & Moon, B.-K. (2017). ENSO-related PM_{10} variability on the Korean Peninsula. *Atmospheric Environment*, *167*, 426–433.

World Health Organization. (2018). What is air pollution? Retrieved August 20, 2019, from http://maps.who.int/airpollution/