Jurnal
Teknologi

Full Paper

SPATIAL ANALYSIS OF THE CERTAIN AIR POLLUTANTS USING ENVIRONMETRIC TECHNIQUES

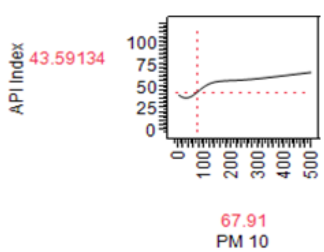
Mohammad Azizi Amran[a], Azman Azid[a], Hafizan Juahir[a*], Mohd Ekhwan Toriman[a,b], Ahmad Dasuki Mustafa[a], Che Noraini Che Hasnam[a], Fazureen Azaman[a], Mohd Khairul Amri Kamarudin[a], Ahmad Shakir Mohd Saudi[a], Kamaruzzaman Yunus[c]

[a]East Coast Environmental Research Institute (ESERI), Universiti Sultan Zainal Abidin (UniSZA), Gong Badak Campus, 21300 Kuala Terengganu, Terengganu Darul Iman, Malaysia
[b]School of Social, Development and Environmental Studies, Faculty of Sciences and Huminities, National University of Malaysia, 43600 Bangi, Selangor Darul Ehsan, Malaysia
[c]Kulliyyah of Science, International Islamic University Malaysia, 25200 Kuantan, Pahang Darul Makmur, Malaysia

Article history
Received
15 January 2015
Received in revised form
10 February 2015
Accepted
15 June 2015

*Corresponding author
hafizanjuahir@unisza.edu.my

Graphical abstract



Abstract

This study aims to identify the spatial variation of air pollutant and its pattern in the northern part of Peninsular Malaysia for four years monitoring observation (2008-2011) based on the seven air monitoring stations. Air pollutant variables that used in this study were Nitrogen Dioxide (NO_2), Ozone (O_3), Carbon Monoxide (CO), and Particulate Matter (PM_{10}) data and had been supplied by Department Of Environment Malaysia (DOE). ANOVA, environmetric techniques (HACA and Descriptive Analysis) and Artificial Neural Network (ANN) approach were used in data analysed. According to ANOVA single test, significance p-value of PM_{10} ($p= 2.5E^{-268}$) is smaller than significance alpha level ($p=0.05$) and it suitable parameter for further analysis in construct the prevention actions compared to O_3, NO_2 and CO. HACA categorized seven air monitoring station into three cluster group of station such as High Concentrated Site (HCS), Moderate Concentrated Site (MCS), and Low Concentrated Site (LCS). Descriptive statistics show the 25th percentile, median, and 75th percentile boxplot and identified the greater (>500 µg/m^3) and smaller (<0.05ppm) outliers, and comparing distributions between each air pollutant. The findings from ANN have verified that the R^2 and RMSE value (0.7981 and 5.734, respectively) were categorized as a significant value for the future prediction. In contrast, PM_{10} levels in Air Pollutant Index equal to 43.59 were 67.91 ug/m^3, O_3 (0.038 ppm), NO_2 (0.019 ppm), and then CO (1.27 ppm) concentration values. This proved that the PM_{10} concentration was categorized as a main contributor to the air pollutant measurement of statistical method compared with other pollutants.

Keywords: ANOVA, environmetric techniques, descriptive analysis, artificial neural network, air pollutant index

Abstrak

Kajian ini bertujuan untuk mengenal pasti variasi spatial bahan pencemar udara dan corak di bahagian utara Semenanjung Malaysia selama empat tahun pemantauan pemerhatian (2008-2011) berdasarkan tujuh stesen pemantauan udara. Pembolehubah pencemar udara yang digunakan dalam kajian ini ialah Nitrogen Dioksida (NO_2), Ozon (O_3), Karbon Monoksida (CO), dan Zarah Habuk (PM_{10}) serta data telah dibekalkan oleh Jabatan Alam Sekitar Malaysia (JAS). ANOVA, teknik environmetric (HACA dan Analisis deskriptif) dan pendekatan Rangkaian Neural Buatan (ANN) telah digunakan dalam data analisis. Menurut ujian tunggal ANOVA, kepentingan p-nilai PM_{10} (p = 2.5E-268) adalah lebih kecil daripada tahap alfa signifikan (p = 0.05) dan ia parameter yang sesuai untuk analisis selanjutnya dalam membina tindakan pencegahan berbanding O_3, NO_2 dan CO. HACA dikategorikan tujuh stesen pemantauan udara ke dalam tiga kumpulan kelompok stesen seperti Lokasi Kepekatan Tinggi (HCS), Lokasi Kepekatan Sederhana (MCS) dan Lokasi Kepekatan Rendah (LCS). Statistik deskriptif menunjukkan persentil ke-25, median, dan 75 persentil plot kotak dan dikenal pasti (> 500 μg/m^3) yang tertinggi dan terendah (<0.05 ppm) titik terpencil, dan pengagihan membandingkan antara setiap pencemar udara. Penemuan daripada ANN telah mengesahkan bahawa R^2 dan nilai RMSE (0,7981 dan 5,734 masing-masing) dikategorikan sebagai nilai penting bagi ramalan masa depan. Sebaliknya, tahap PM_{10} dalam Indeks Pencemaran Udara sama dengan 43,59 ialah 67,91 ug / m^3, O_3 (0,038 ppm), NO_2 (0,019 ppm), dan kemudian CO (1.27 ppm) Nilai kepekatan ini membuktikan bahawa kepekatan PM_{10} yang dikategorikan sebagai penyumbang utama kepada pencemaran udara melalui pengukuran kaedah statistik berbanding dengan pencemaran lain.

Kata kunci: ANOVA, teknik environmetric, analisis deskriptif, rangkaian neural buatan, indeks pencemaran udara

1.0 INTRODUCTION

Malaysia is located in the centre of South East Asia, which getting experience and influence with the rapid growth development. Development is a major contributor in rises of local or regional air pollution cases. Air pollution is made up of a mixture of particles and gases in destructive amounts that is released into the atmosphere due to natural or human activities [1]. Particulate matter (PM_{10} and $PM_{2.5}$) was considered as the main pollutant of air pollution compared to other pollutants [2]. Instead of PM_{10} value, there are some air pollutant sources such as Ozone (O_3), Carbon Monoxide (CO), and Nitrogen Dioxide (NO_2), which important in measuring the air pollutant index (API). Air pollution is a fundamental problem in worldwide and it conveys two significant concerns; the serious damage in health problems, and towards the loss of welfare from environmental occurred in Asian countries [3]. Urban and industrial area is the most targeted areas where the concentrated pollution has been detected. The air pollutant level will be concentrated and dispersed during certain time such peak hour caused by high concentrated of population density and manufacturing industries [4]. From this targeted area, the movement and dispersed of polluted air quality affected by seasonal wind pattern and local condition factors include land and sea breeze and transference of wind to rural or surrounding area [5].

A long distance transport of microbial life also can be affected by chemical component's content of dust [6]. Besides, highly concerned about air pollution will decrease the mortality, acute injury, pose or potential hazard to human respiratory health [1]. This environmental problem has attracted worldwide researchers in studying and getting evidence as decisions making tools to overcome respiratory health problems and loss of welfare. The data from the air monitoring program and studies on ambient air quality found that the amount of air pollutant level in many large cities was arise over time exceeding the limit of national ambient air quality standard guideline [7]. All of the analysis process was evaluated and processed using Environmentric techniques. This technique was applied for solving the environmental problems for better results [7,8]. Environmentric technique is a valuable tool in developing appropriate strategies for the effective management of natural resources and air controlling monitoring (Environmental Modelling) [9]. Otherwise, this technique also helps to lessen the complexity of large datasets of air pollutant, so that a better understanding and interpretation of data can be achieved [7]. Besides, the analysis of variance (ANOVA) was applied due to the important of this analysis in creating the hypothesis testing problems [10, 11, 12, 13, 14]. It includes both parametric and nonparametric approaches to identify the significant value of data that can be used for future analysis [10]. For more detail of air pollution data analysis, the involvement of ANN modelling was applied. This is because, the ANN has a great flexibility, efficiency, and accuracy for identifying non-linear patterns between input and output data and faster in solving the complex air quality problem for better prediction of air quality [4]. A previous study was conducted by some researchers, which the concentration of PM_{10} is always the highest in API value compared to other source of air pollutant [8].

The objective of this study is to identify the spatial variation of a major group of air pollutant such as PM_{10}, O_3, CO and NO_2 in the northern region of Peninsular Malaysia for four years monitoring period (2008-2011). Next is to evaluate the relationships between each variable of air monitoring data and identification of the major variables in giving the huge impact in an API value. This study focused more on selected pollutant concentration due to determine its spatial variation and relationship among air pollutant.

Seven air monitoring station that located in the northern region of Peninsular Malaysia (Table 1 and Figure 1) were selected in this study. The selection of monitoring station location is based on variety of land use changes and development: urban, suburban, academic institution, workplace, sport centre, and schools. This variety of station locations will produce the dissimilar reading of air pollutant data in northern region. In additional, this selection included with the probability of industrial effluent and transportation changes on the road.

2.0 EXPERIMENTAL

2.1 Description of Study Area

Seven air monitoring station that located in the northern region of Peninsular Malaysia (Table 1 and Figure 1) were selected in this study. The selection of monitoring station location is based on variety of land use changes and development: urban, suburban, academic institution, workplace, sport centre, and schools. This variety of station locations will produce the dissimilar reading of air pollutant data in northern region. In additional, this selection included with the probability of industrial effluent and transportation changes on the road.

2.2 Data Collection

The data used in this study was daily concentration of Carbon Monoxide (ppm), Nitrogen Dioxide (ppm), Ozone (ppm), and Fine Particulate Matter (µg/m^3) from seven air monitoring stations based on four years observation (2008-2011). All the data were secondary data that provided by the Department of Environment (DOE) completed with API value. The secondary data from DOE was supplied by Alam Sekitar Malaysia Agency (ASMA) that working on collection of primary monitoring data. The equipment used by ASMA to monitor air quality were Teledyne Technologies Inc. USA, and Met One Instrument Inc. USA for air quality monitoring, BAM-1020 Beta Attenuation Mas Monitor in Met One Instrument, Inc. USA for PM_{10} and monitoring. Then, Teledyne API Model 100A/100E, Teledyne API Model 200A/200E, Teledyne API Model 300A/300E and Teledyne API Model 400A/400E were used in monitoring NO_2, CO, and O_3. For statistical analysis, the hourly data were used to form a daily average, which compromises 51,135 datasets (7,305 observations per stations x 7 stations).

Table 1 Location of air monitoring station in the northern region of Peninsular Malaysia

Station No.	Station Name	Location
S1	Kangar (CA0033)	N06° 25.296 E100° 11.021
S2	Sungai Petani (CA0017)	N05° 37.430 E100° 28.191
S3	Langkawi (CA0032)	N06° 19.545 E99° 51.311
S4	Alor Setar (CA0040)	N06° 07.248 E100° 22.056
S5	Perai (CA0003)	N05° 22.251 E100° 23.277
S6	Seberang Jaya (CA0009)	N05° 23.522 E100° 24.151
S7	USM (CA0038)	N05° 21.218 E100° 18.090

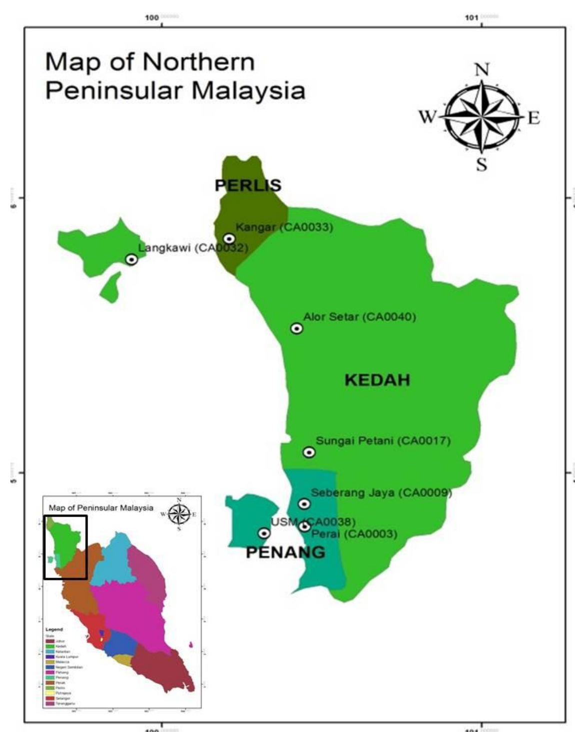


Figure 1 Location of air monitoring station in the northern region of Peninsular Malaysia

2.3 Spatial Analysis of PM_{10}, O_3, NO_2, and CO Concentration in Atmosphere

2.3.1 Application of Analysis of Variance (ANOVA)

Worldwide of the interest of practitioners has applied the functional analysis of variance methods. The importance of ANOVA in specially discussed, the functional of references and estimation of components in air monitoring quality [10]. The raw data of PM_{10}, NO_2, CO, and O_3 concentration were analysed in the ANOVA single factor formulae in order to get the p-value. Finding of p-value will determine the selection of the best parameter, whether the significance of air data to be used in further analysis, prediction, estimation and modelling. Due to PM_{10} is the major element in API calculation, then the selection of this pollutant should be considered in this study [4,5]. The single factor ANOVA can be written as:

$$SS_{total} = \sum y^2 - \frac{(\sum y)^2}{n} \tag{1}$$

$$SS_{regression} = \frac{\left(\sum xy - \frac{\sum x \sum y}{n}\right)^2}{\sum x^2 - \frac{(\sum x)^2}{n}} \tag{2}$$

where, *x* is explanatory variable of each air pollutant concentration reading, *y* is the dependent variable which is API value, *n* is the total number of air monitoring data of PM_{10}, NO_2, CO, and O_3 concentration.

2.3.2 Classified of Air Monitoring Station by Hierarchical Agglomerative Cluster Analysis (HACA)

HACA was employed in this study to identify the grouping of the air monitoring stations. The HACA was grouped according to the similarity classes based on the value of PM_{10}, O_3, CO and NO_2 since this parameter were used in the API calculation [15]. In this study, cluster analysis had been analysed by using Ward's method. The Ward's method uses Euclidean distance as a measure of similarity between each parameter. Then, the HACA results of PM_{10}, O_3, CO, and NO_2 for each station were illustrated by a dendrogram. Dendrogram functioning for separate the cluster or group of air monitoring station based on air quality data. This HACA was applied in this study to classes the seven air monitoring stations based on these air pollutant parameter values that resulting in three different statuses of classes such as Highly Concentrated Site (HCS), Moderate Concentrated Site (MCS), and Lower Concentrated Site (LCS).

2.3.3 Box-and-Whisker Plots using Descriptive Statistics

In this study, the descriptive statistics was used to identify the outliers and comparing distributions between each air pollutant. This statistical analysis was performed using XLSTAT 2014 software. The top and the bottom of the boxes represent 75th and 25th percentile (quartile) and the band near the middle boxes was the median [16] such in Figure 2. Each air pollutants (PM_{10}, O_3, CO, and NO_2) will be analysed according to their classes of HCS, MCS, and LCS. This plots is important to gives a brief picture of the other important distribution values.

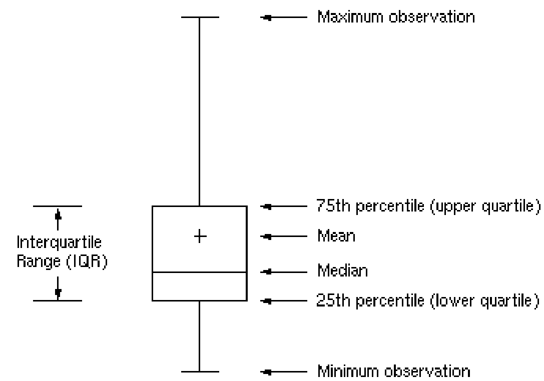


Figure 2 Boxplots characteristics

XLSTAT 2014 software approach was used to perform the descriptive analysis to gives resulting as plot location, where the plots that included in the box (within the range) and non-included in the box (out of range). The plot that out of boxes called as outliers that bring the highest or lowest values compared with other data values.

2.3.4 Artificial Neural Network (ANN) in Determine Relationship of Variables

Application of ANN in this study was used to identify the relationship between air pollutant concentrations with other pollutant concentration. It was performed using JMP10 software, which it offers flexibility and ease of application. When undergoes modelling performance of ANN, correlation of determination (R^2) and root mean square error (RMSE) value were applied [8,17]. ANN is a useful tool in prediction of the future impact and monitoring of air pollutant concentration. The equations of R^2 and RMSE can be referred as:

$$R^2 = \frac{SS_{reg}}{SS_{tot}} = \frac{SS_{reg}/n}{SS_{tot}/n} \tag{3}$$

$$RMSE = \sqrt{\frac{\sum_{j=1}^{n}(y_j - \hat{y}_j)^2}{n}} \tag{4}$$

where, y_j stand for the measured value, $\hat{y}_j$ is the estimated value of the dependent variable and, *n* is the number of observations. ANN also used to display the relationship and approving which the most parameter gives greater impact in API value. All the secondary air quality data have been undergone by using the ANOVA, environmetric techniques, and ANN analysis of spatial variation of PM_{10}, O_3, CO, and NO_2 in the northern part of Peninsular Malaysia, where it can be displayed in table, graph, and maps. There are several analyses that give their own characteristics and explanation in order to get the best results and understandings.

3.0 RESULTS AND DISCUSSION

3.1 ANOVA Single Factor

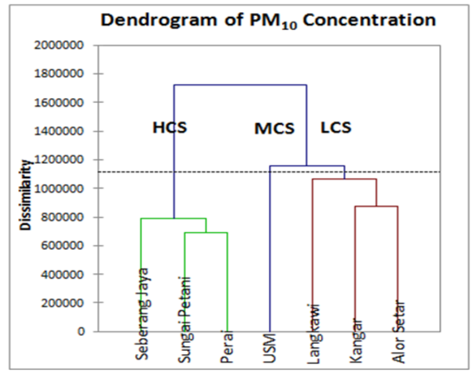
Table 2 represents each concentration of air pollutant from each air monitoring station that undergoes ANOVA single factor test. From the analysis, PM_{10} shows that the p-value is smaller than significance alpha level (p=0.05) with the *p* value is 2.458 E^{-268} while the other pollutants such as O_3, NO_2 and CO level, the *p* values give the result of zero. Therefore, the value of the variance gives the significant and suitable of PM_{10} to be used in predicting of air quality. Instead, the reason in getting the significance variation of data that indicated the functional in explaining multiple problems of environmental conditions. The effect from PM_{10} into the environment will cause climatic variations, phoneme recognition, land use estimation, ozone prediction, and spatial prediction of ozone levels [18]. Each pollutant concentration is used in categorizing the monitoring stations with using of cluster analysis. When the PM_{10} concentration keeps rising, the precaution and prevention actions need to reduce more impact to the environment. Although, to determine the significant differences between the observed concentration levels of PM_{10}, O_3, CO, and NO_2 on each air monitoring station, the environmetric techniques should be applied frequently.

Table 2 ANOVA single-test based on each air pollutant concentration

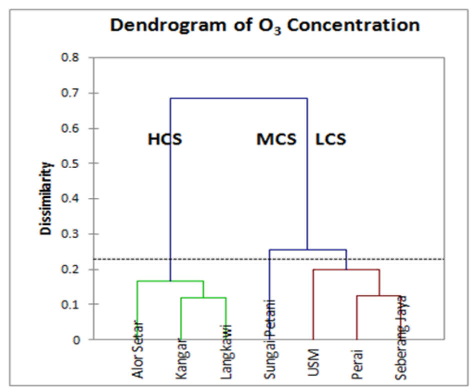
Parameter	p-value
PM_{10}	2.458×10^{-268}
O_3	0
NO_2	0
CO	0

3.2 Cluster Classification

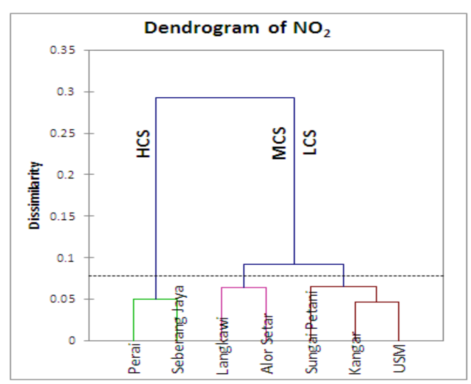
The dendrogram of PM_{10} (Figure 3) and map (Figure 4) of three locations (CA0003, CA0009, and CA0017) of air monitoring station has been classified as HCS, due to the heavy industries area, heavy transportation pathway to transport product from industries and high traffic contribute the release of particulate matter into the surrounding atmosphere and giving highly reading of air quality taken. The region of MCS was found in the station CA0038 due to located around the green campus area, and expected coming from congested urban area in Penang with the present of various activities such as transportation and industrial activities. While, for the station CA0032, CA0033, and CA0040, it were classified as a LCS due to the sustainability of urban development and some sources are coming from burning of fossil fuel by transportation activities [19].



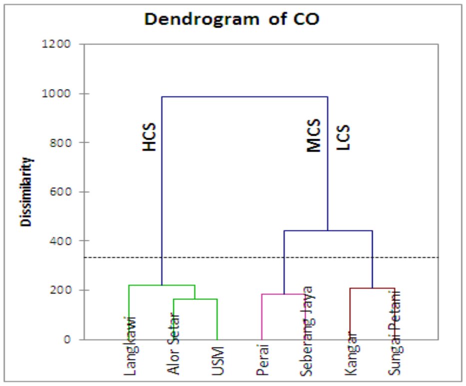
(a) Particulate Matter (PM_{10})



(b) Ground Level Ozone (O_3)



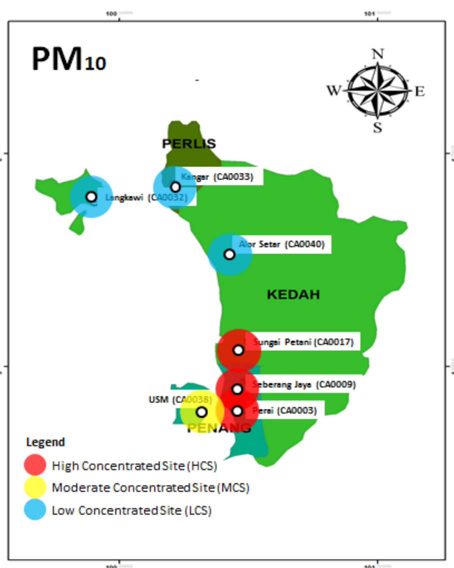
(c) Nitrogen Dioxide (NO_2)



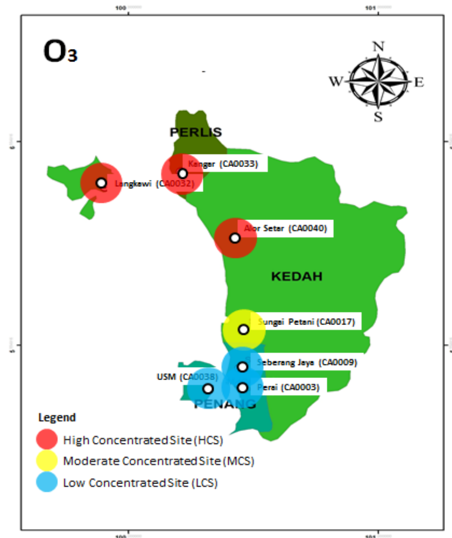
(d) Carbon Monoxide (CO)

Figure 3 Dendrogram of air monitoring station based on each air pollutant concentration

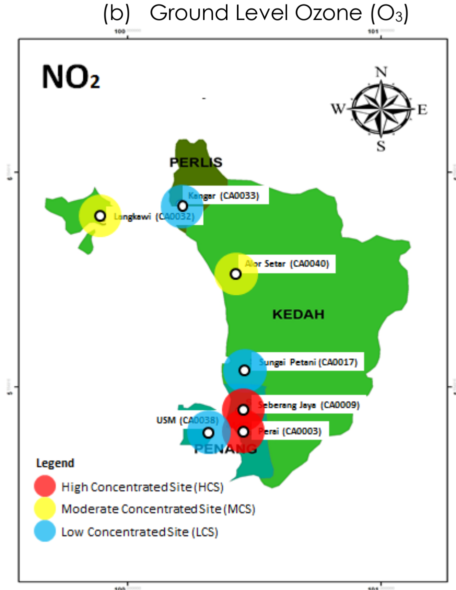
The dendrogram of O_3 concentration shows that most of the station was contrariwise proportional compared to the PM_{10}. This is because the lowest PM_{10} concentration was influenced to the highest of O_3 concentration value. The increasing of O_3 concentration influenced by the higher usage of air ventilation, freezer and air-conditioner in the office, homes and institutional buildings, then it will released the O_3 gases into the air and spread to the buffer area [20]. For NO_2 cluster group concentrations, station CA0003 and CA0009 was in HCS, station CA0032 and CA0040 in MCS, and station CA0017, CA0033 and CA0038 in LCS. While, for CO concentration, the cluster has been detected as HCS in station CA0032, CA0038 and CA0040, for MCS in station CA0003 and CA0009, and LCS in station CA0017 and CA0033.



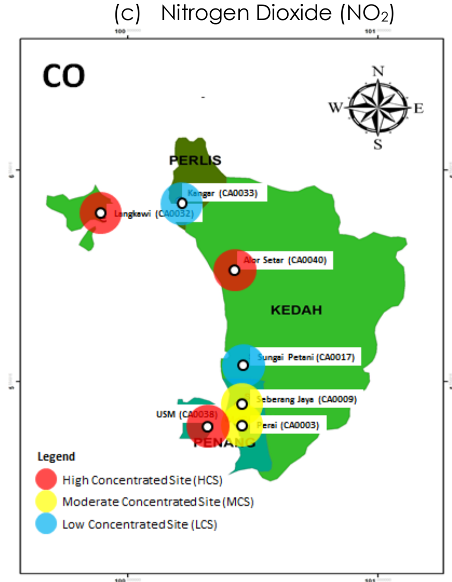
(a) Fine Particulate Matter (PM_{10})



(b) Ground Level Ozone (O_3)



(c) Nitrogen Dioxide (NO_2)



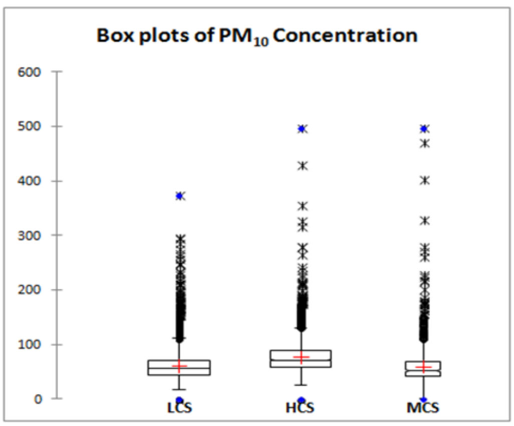
(d) Carbon Monoxide (CO)

Figure 4 Map of classification of (a) PM_{10}, (b) O_3, (c) NO_2 and (d) CO concentration

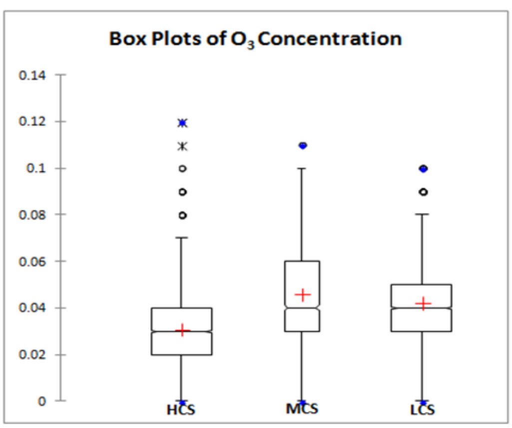
These three types of classes (HCS, MCS, and LCS) were used as a references station in grouping the similarity of variation of between seven air monitoring station can be reduced to just three air monitoring station (CA0009 represent of HCS, CA0038 represent of MCS, and CA0033 represent of LCS) for future monitoring. By simplify and reduce the monitoring station number for air quality monitoring, it will give efficient impact in terms of saving the cost and time consuming.

3.3 Descriptive Analysis

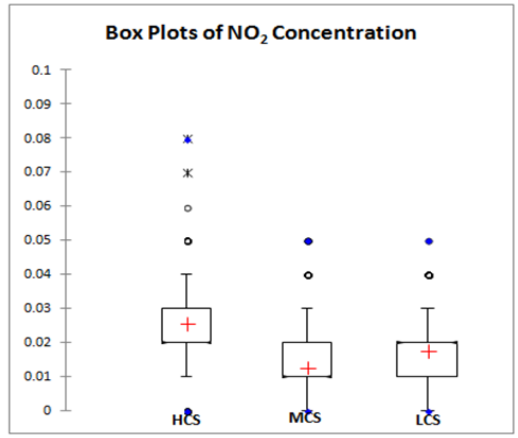
Based on the Figure 5, it shows that the boxplots in determine every reading of air pollutant concentration included in the boxes and outliers. The boxes contain of 25th percentile, median, the 75th percentile value of data and less or more than that located outside of the boxes. Each data have shown as a plot.



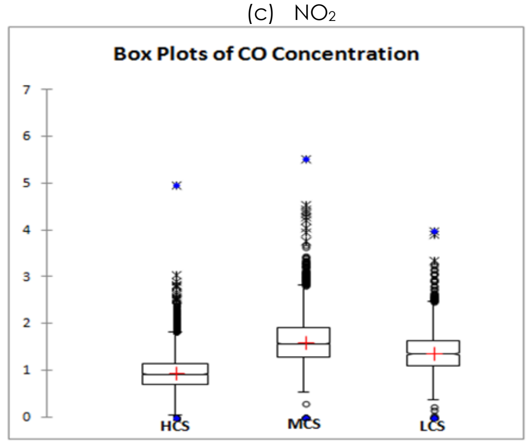
(a) PM_{10}



(b) O_3



(c) NO_2



(d) CO

Figure 5 Boxplots of each Air Pollutant (a) PM_{10}, (b) O_3, (c) NO_2 and (d) CO based on Cluster Classes for four years monitoring (2008-2011)

Every air monitoring data that undergo descriptive analysis will be shown as a plot for lowest and highest data reading for each air pollutant concentration. Each plot for lowest or minimum value was always having the similar value for every cluster class, such as 0 μg/m³ for PM_{10}, 0.0 ppm for O_3, 0.0 ppm for NO_2, and 0 ppm for CO concentration. The major difference of highest or maximum value of each air pollutant concentration can be detected and interpret in details. For PM_{10} concentration, the maximum plots for LCS, MCS, and HCS was <400 μg/m³, 500 μg/m³, and >500 μg/m³. While maximum plots of O_3 concentration for LCS, MCS and HCS was 0.1 ppm, 0.11 ppm, and 0.12 ppm. Meanwhile, maximum plots of NO_2 concentration for LCS, MCS and HCS was <0.05 ppm, 0.05 ppm, and >0.08 ppm. Lastly, for the CO concentration readings give that the maximum plots for LCS, MCS and HCS was <4 ppm, <6 ppm, and >5 ppm. The boxplots was contain with the air pollutant data between the recommended value that located under the controls of air quality and safely environment for respirational. Although, the extreme plots that exceeding the 75th percentile of boxplots was considered as a probability polluted, thread for health [21] and contribute to other environmental problems such as climate changes, global warming, ozone depletion, and others.

3.4 ANN Modelling

Performance indicators for ANN were used for API forecast and each parameter variation in the northern region of Peninsular Malaysia give the result of 0.7981 and 5.734 for R^2 and RMSE respectively. Figure 6 shows the linking of each parameter to the calculation for API calculation.

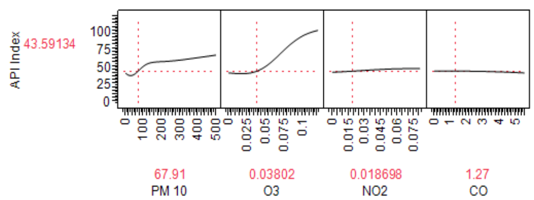
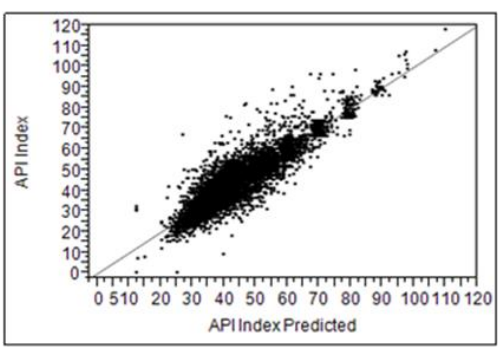
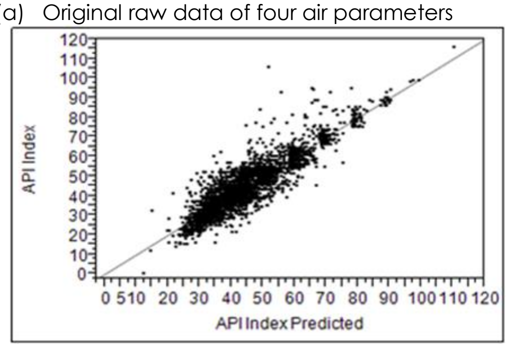


Figure 6 Graph of each air pollutant parameters at constant API value

At constant API equal to 43.59 value, different value for each parameters were analysed to identify the relationship and give significant effecting the API value. PM_{10}, O_3, NO_2, and CO levels at API value equal to 67.91, 0.038, 0.019, and 1.27 concentration value. PM_{10} show the most significance value that giving the prove that the PM_{10} concentration was the main influence to the air pollution measurement of statistical method [4,5].



(a) Original raw data of four air parameters



(b) Data of PM_{10} parameter

Figure 7 Scatter plot diagram of API prediction performance (actual by predicted plot): (a) an original raw data (4 variables) and (b) data of PM_{10} concentration in API

Figure 7(a) and (b) shows the result of scatter plot between actual and predicted API for both models using original raw data (for all 4 variables) and PM_{10} (single variable) as input variable data. The scatter plot actual API shows the efficiently directly proportional movement of data that similar compared with combined scattered predicted plot included with all variables.

4.0 CONCLUSION

The result of this study show that the function of ANOVA in determine the significance p-value= 2.5 E^{-268} is smaller than significance alpha level (p=0.05) to approve that the PM_{10} data and may others can be used for further analysis also the most important parameter in API calculation. HACA was successfully in grouped the seven air monitoring stations into three different clusters (HCS, MCS, and LCS) for each parameter based on four selected air pollutant variable recorded from the different local surroundings for four year monitoring periods and approved with the actual gas release situation or level of industrial and combustion of fossil fuels in transportation at different location of the station. By using Descriptive Analysis, the air monitoring data have been arranged in four parameter boxplots and outliers for each HCS, MCS and LCS clustered classes. Findings from ANN have proven the R^2=0.7981 and RMSE=5.734 value was the significant value for future prediction. While, the relationship between each variable (parameter) in graph of air monitoring data, the concentration of PM_{10} that give more impact on API value compare with other CO, NO_2 and O_3 concentration can be observed. Qualifying a significant air pollutant parameter of input data using statistical method, it shows a predictive ability at least as good as the one given by the standard model. It also can be defined that this technique may be seen as a good substitute for establishing ANN models in the air quality forecast compared to others. This ANN not only avoids time wasted, but it also can save the cost of monitoring purposes. Therefore, it verified that the feed-forward ANN architecture is able to predict API values from all existing input with slight precision.

Acknowledgement

The authors acknowledge the Air Quality Division of the Department of Environment (DOE) under the Ministry of Natural Resource and Environment, Malaysia, for giving us permission to utilize air quality data, and East Coast Environmental Research Institute (ESERI), UniSZA for giving advice, guidance, and support for this study.

References

[1] Kampa, M. and Castanas, E. 2008. Human Health Effects of Air Pollution. *Environmental Pollution*. 151: 362-367.

[2] Elbayoumi, M., Ramli, N. A., Yusof, N. F. F. M., Yahaya, A. S. B., Madhoun, W. A., Ul-Saufie, A. Z. 2014. Multivariate Methods for Indoor PM_{10} and $PM_{2.5}$ Modelling in Naturally Ventilated Schools Buildings. *Atmospheric Environment*. 94: 11-21.

[3] 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.

[4] Azid, A., Juahir, H., Aris, A. Z., Toriman, M. E., Latif, M. T., Zain, S. M., Yusof, K. M. K. K., Saudi, A. S. M. 2014. Spatial Analysis of the Air Pollutant Index in the Southern Region of Peninsular Malaysia Using Environmetric Techniques. Springer Science. 10.1007/978-981-4560-70-2_56.

[5] 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.

[6] Aenab, A. M., Singh, S. K., Lafta, A. J. 2013. Critical Assessment of Air Pollution by ANOVA Test and Human Health Effects. *Atmospheric Environment.* 71: 84-91.

[7] Mutalib, S. N. S. A., Juahir, H., Azid, A., Sharif, S. M., Latif, M. T., Aris, A. Z., Zain, S. M., Dominick, D. 2013. Spatial and Temporal Air Quality Pattern Recognition Using Environmetric Techniques: A Case Study in Malaysia. *Environmental Science Processes & Impacts*. 15: 1717.

[8] Azid, A., Juahir, H., Toriman, M. E., Kamarudin, M. K. A., Saudi, A. S. M., Hasnam, C. N. C, Aziz, N. A. A., Azaman, F., Latif, M. T., Zainudin, S. F. M., Osman, M. R., Yamin, M. 2014. Prediction of the Level of Air Pollution Using Principal Component Analysis and Artificial Neural Network Techniques: A Case Study in Malaysia. *Water Air Soil Pollution*. 225: 2063.

[9] Juahir, H., Pauzi, M., Osman, R., Saim, N. 2012. Chemometric Application in Identifying Sources of Organic Contaminants in Langat River Basin. *Environmental Monitoring & Assessment Journal*. 184(2): 1001-1014.

[10] Estévez-Pérez, G. and Vilar, J.S. 2013. Functional ANOVA Starting from Discrete Data: An Application to Air Quality data. Springer Sciences. *Environment Ecology Stat*. 20: 495-517.

[11] Huang, J. 1998. Projection Estimation in Multiple Regression with Application to Functional ANOVA Models. *Ann Stat.* 26: 242-272.

[12] Lin, Y. 2000. Tensor Product Space ANOVA Models. *Ann Stat*. 28: 734-755.

[13] Gu, C. 2002. *Smoothing Spline ANOVA Models. Springer series in statistics*. New York. Springer. Vol. XIII.

[14] Angelini, C., De Canditiis, D. and Leblanc, F. 2003. Wavelet Regression Estimation in Nonparametric Mixed Effect Models. *Journal Multivariate Analysis*. 85: 267-291.

[15] Azid, A., Juahir, H., Latif, M. T., Zain, S. M., Osman, M. R. 2013. Feed-Forward Artificial Neural Network Model for Air Pollutant Index Prediction in the Southern Region of Peninsular Malaysia. *Journal of Environmental Protection.* 4: 1-10.

[16] Langer, S., Beko, G. 2013. Indoor Air Quality In The Swedish Housing Stock and Its Dependence on Building Characteristics. *Building and Environment*. 69: 44-54.

[17] Xie, H., Ma, F. and Bai, Q. 2009. Prediction of Indoor Air Quality Using Artificial Neural Networks. *2009 Fifth International Conference on Natural Computation*. 412-418.

[18] Bernard, S. M., Samet, J. M., Grambsch, A., Ebi, K. L., Romieu, I. 2001. The Potential Impacts of Climate Variability and Change on Air Pollution-Related Health Effects in the United States. *Environmental Health Perspectives*. 109(Suppl 2): 199-209.

[19] Jamhari, A. A., Sahani, M., Latif, M. T., Chan, K. M., Tan, H. S., Khan, M. F., Tahir, N.T. 2014. Concentration and Source Identification of Polycyclic Aromatic Hydrocarbons (Pahs) In PM10 of Urban, Industrial and Semi-Urban Areas in Malaysia. *Atmospheric Environment*. 86: 16-27.

[20] Yau-Hwang Kuo and Jang-Pong Hsu. 1996. *MCFC-R: A Fuzzy Connectionist Model for Color-blindness Plate Recognition*. 2: 718-723.

[21] Rahman, S., Sidik, N. M., Hassan, M. H. J., Rom, M. T., Jauhari, I. 2005. Temperature Performance and Usage Conditions of Domestic Refrigerator-Freezers in Malaysia. *HKIE*. 12(2). DOI:10.1080/1023697X.2005.10668000.

[22] Basu, R. 2009. High Ambient Temperature and Mortality: A Review of Epidemiologic Studies from 2001 to 2008. *Environmental Health*. doi:10.1186/1476-069X-8-40.