



Techno Science Africana Journal

Journal homepage: www.technoscienceafricana.com



APPLICATION OF ARTIFICIAL NEURAL NETWORK (ANN) AND PRINCIPAL COMPONENT ANALYSIS (PCA) TO MODEL THE COMPLEX CHARACTERISTICS AND SOURCES APPORTIONMENT OF ATMOSPHERIC AIR POLLUTION IN MALAYSIA PENINSULAR

Hamza Ahmad Isiyaka¹, Adamu Mustapha¹ Hafizan Juhair²

¹Department of Geography, Faculty of Earth and Environmental Sciences, Kano University of Science and Technology Wudil, Nigeria

²East Coast Environmental Research Institute (ESERI), Universiti Sultan Zainal Abidin, Gong Badak Campus, 21300 Kuala Terengganu, Terengganu, Malaysia.

ARTICLE INFO

Keywords:

Artificial neural network;
Input combination model;
Source apportionment;
Principal Component analysis,
air pollution.

ABSTRACT

This study aims to assess the dynamic characteristics of atmospheric air pollutants, identify the major sources of pollution as well as develop the best input combination model for predicting air pollution index (API). Data from five (5) monitoring stations from 2007-2011 were collected from Department of Environment (DOE) Malaysia. The result for Principal Component Analysis (PCA) accounted for more than 39% and 15% of the total variance of the datasets attributed to anthropogenic induced emission from point source and non-point source. Further, Artificial Neural Network (ANN) was used to model the percentage contribution of each pollutant and apportion the mass source category and to develop three input combination model (A, B and C) in order to test their predictive capabilities. The result indicates that ozone is the highest contributor to API by 63% followed by particulate matter 15%. Based on source apportionment, gas and secondary pollutants contributes 92% and 8% by gas and meteorological condition. Moreover, ANN statistically proven that model A has the capability to predict API with a high level of significance at $R^2 = 0.934$ and $RMSE = 4.871$ compared with model B and C. This finding justify that ANN is a robust and important tool for air pollution modeling, decision making and environmental management.

INTRODUCTION

Air is the most important natural resources which sustain life of plants, animals and general ecosystem regulation (Botkin et. al 2012). Air quality measures the composition and concentration level of gaseous and particulate toxic elements (WHO 2002; Bernstein et. al. 2004). Urban air quality is polluted when atmospheric pollutants build up in concentration sufficiently high enough to cause direct or indirect damage to plants, animals and the ecosystem structure

(Jacobson 2002). Atmospheric air pollutants are made up of primary particles emitted from various sources and secondary particles from aerosols produced by chemical reactions (Salma et. al. 2004). The struggle for industrial development especially in developing countries to compete economically with developed nations do not give priority to pollution control due to the adoption of obsolete technologies and incompetent regulatory systems to enforce laws (Raven et. al. 2013).

This poses a serious threat to the quality of life worldwide making it an area that demands serious and urgent attention (Genc et al 2010). Malaysia is a fast developing nation within the continent of Asia in industrial, commercial agricultural and tourist attraction (DOE, 2011). Razak et. al (2013) and Ismail et al (2010) revealed that Malaysia produces 5.4 metric tons of CO₂ which exceeds the bench mark of 3.9 metric ton per capital as provided by the world resource institute and most urban cities within the Peninsular faces serious air pollution (Azmi et al. 2010). Although, the overall air quality status in Malaysia fall within good and moderate base on air pollution index (API), but still show a slight fluctuation in trend from 2008 to 2011 (DOE, 2012). In 2008, the status of good air quality fall around 59%, 55.6% in 2009, 63% in 2010 and 55% in 2011 (DOE, 2012). In Malaysia, air is polluted from both point source and non-point sources. The mobile sources originate mostly from motor vehicle (82% of pollution). Stationary pollutants are emitted from a static position (manufacturing activities, industrial fuel burning, power stations and construction sites). Tran boundary pollutants are transported by wind from neighboring location (Mutalib et. al. 2013; Dominick et al. 2012) and the worst haze episode in Malaysia was recorded in 1997 (Jamal et al. 2004). According to Ministry of Transport (2010), the number of newly registered motor vehicles in Malaysia has increased by 4.42% from 934,367 vehicles in 2004 to 1,160,082 vehicles in the year 2010. A report by Department of Statistics Malaysia (2011) shows that Malaysia population increased from 23.3 million in 2000 to 28.3 million in 2010.

However, a report by Department of Statistics Malaysia (2010) on Compendium of Environment Statistics, Malaysia experiences tropical climate with a uniform temperature, high humidity as well as copious rainfall throughout the year. The pattern of rainfall is influenced by northeast and southwest monsoon wind. Places located in the east coast of Peninsular Malaysia, northeast coast of Sabah and western Sarawak are characterized with heavy rainfall during the northeast monsoon season. Contrary to this notion, other parts of the country experiences

less rainfall except for Sabah during southwest monsoon (Department of Statistics 2009 and Department of Statistics 2010).

An efficient prediction and understanding of the percentage contribution of pollutants to the API value will provide a reliable countermeasure to prevent air pollution catastrophe in public health and well-being) (Kolehmainen et al. 2001). Therefore, arriving at the best prediction model requires the adoption of more sophisticated modeling techniques due to the multiplicity of source of pollution, complexity and dynamic nature of the physical and chemical characteristics of their formation and transportation within the atmosphere (Arhami et al 2013). However, the development of a prediction performance model, best selection of the input variables and the type of model structure pose a major challenge in modeling air pollution (Arhami et al 2013). Recent advancement in the use of artificial neural network (ANN) serves as a more accurate, reliable and modeling technique due to its ability to generalize nonlinear pattern in a large data sets when trained, it can be used for forecasting and pattern recognition which are too complex for conventional statistical methods (Chelani et al. 2002). Furthermore, the accuracy of the prediction level largely depends on the quality of input variables used. The leave one-out cross-validation method for percentage contribution and source apportionment have been applied in air pollution modeling by (Domenick et. al. 2013; Osman et al. 2012; Wahid et. al. 2013; Razak et. al. 2014; ,Mustaffa et. al. 2014). ANN is a distributed information processing system and a powerful general purpose technique composed of many simple computational elements integrating across weighted connections (Hakimpoor et al. 2011). Prediction of air quality using ANN will enable the implementation of proper planning and actions as well as strategies to control the level of pollution (Hakimpoor et al. 2011). The objectives of this study are to model the individual percentage contribution and source apportionment of air pollution, identify the major possible sources of pollutants and identify the best input combination model in predicting API.

RESEARCH METHODOLOGY

Description of the Study Area

The study area is located in the east coast (Terengganu) and western part (Klang Valley) of Peninsular Malaysia covering five selected air quality monitoring stations. Kemaman station (ST1) is characterized with port, residential and industrial activities; Paka-Kertih (ST2) is a petrochemical industrial area linking the entire range of oil and gas value chain; Kuala Terengganu (ST3) is located in the city of Terengganu state with series of commercial

activities, primary industries, high traffic, port and residential areas. Putrajaya (ST4) is the main administrative center of Malaysia with high population density. The city center of Kuala Lumpur (ST5) is associated with series of commercial and industrial activities, residential houses, high population density and serious traffic congestion Department of Statistics, 2011). The location of monitoring stations comprising of their latitudes and longitude is shown in both Table 1 and Figure 1.

Table 1: Location and coordinates of monitoring station

Station ID	Location	coordinates	
		Latitude	Longitude
ST 1	Kemaman	N04° 16.260	E103° 25.826
ST2	Paka-Kertih	N04° 35.880	E103° 26.096
ST3	Kuala Terengganu	N05° 18.455	E103° 07.213
ST4	Putra Jaya	N02° 55.915	E101° 40.909
ST5	Kuala Lumpur	N03° 06.376	E 101°43.072

Source: Department of Environment (DOE) Malaysia

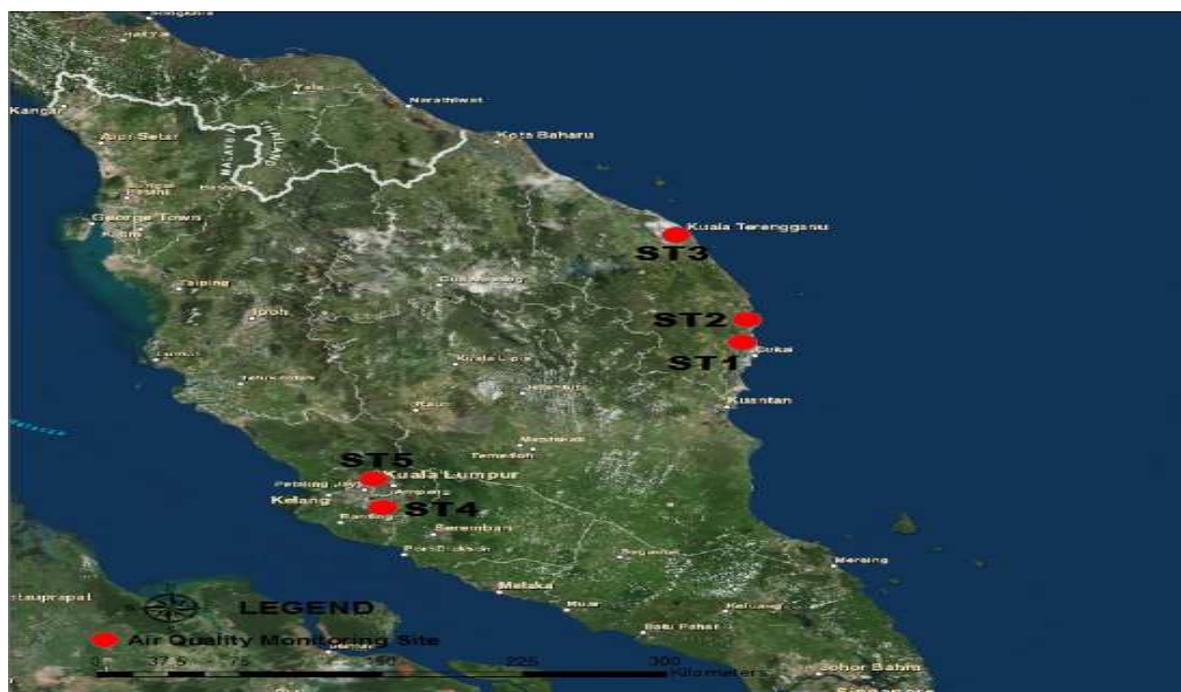


Figure 1: Map of the study area showing air quality monitoring stations

Data collection and treatment

The data use for this study comprises of five air pollutants and two meteorological parameters including carbon monoxide (CO), ozone (O₃), particulate matter with diameter less than 10 μm (PM₁₀), sulfur dioxide (SO₂), nitrogen dioxide (NO₂), ambient temperature

and wind speed. The average time varies from 1-24 hours which represents the period over which measurement are monitored and reported in order to assess the possible health impacts of specific air pollutants (DOE 1997 and DOE 2009).

The data were provided in the form of hourly readings by DOE Malaysia from 2007-2011 (5years) and were converted to daily average observations which gave a total of 64036 data sets (9148 observations × 7 parameters). Furthermore, to facilitate this analysis, the nearest neighbor method was applied using XLSTAT add-in software to estimate missing values (Dominicks et al. 2012). The missing data recorded is 4% of the original data sets. The nearest neighbor method can provide a simple scheme, where the endpoint of the gaps is used as estimates for all missing values (Junninen et al. 2004). The equation can be written as;

$$y = y_1 \text{ if } x \leq x_1 + [(x_2 - x_1) / 2] \text{ or } y = y_2 \text{ if } x \geq x_1 + [(x_2 - x_1) / 2] \tag{1}$$

Where *y* represents the interpolate, *x* is the time point of the interpolate, *y₁* and *x₁* are the coordinates of the starting point of the gap and *y₂* and *x₂* are the end points of the gaps.

Principal component analysis for source identification (PCA)

PCA helps to eliminate the less significant parameters with minimal loss of the original variables and retain the most significant parameters within the data sets (Wang et al. 2014). This process will help to reduce the dimensionality of the data by forming new variables that are uncorrelated called principal components (PCs) (Daigle et al. 2011). The equation is expressed as:

$$Z_{ij} = a_{i1}x_{1j} + a_{i2}x_{2j} + a_{i3}x_{3j} + \dots + a_{im}x_{mj} \tag{2}$$

Where *z* is the component score, *a* is the component loading, *x* is the measured value of variables, *i* is the component number, *j* is the sample number and *m* is the total number of variables.

The PCs generated by PCA are sometimes not readily available for interpretation, therefore, it is advisable to rotate the PCs by varimax rotation with eigenvalues equal to or greater than 1 (Juahir et al.2011). The varimax rotation is considered significant in order to obtain new groups of variables called varimax factors (VFs) (Brumelis et al. 2000; Love et. al. 2004).

Furthermore, the number of varimax factors(VFs) obtained by varimax rotations is usually equal to the number of variables in accordance with the common features which can include unobservable, hypothetical and latent variables (Vega et. al. 1998; Liu et al. 2003). In addition, the VFs coefficient with a correlation greater than 0.75 are considered as strong significant factor loading, those that range from 0.50 - 0.74 have moderate, while those having correlation of 0.30 - 0.49 are classified as weak significant factor loading (Liu et al. 2003).

$$Z_{ij} = a_{f1}x_{1i} + a_{f2}x_{2i} + \dots + a_{fm}f_{mi} + e_{fi} \tag{3}$$

Where *Z* is the measured value of a variables, *a* is the factor loading, *f* is the factor score, *e* is the residual term accounting for errors or other sources variation, *i* is the sample number, *j* is the variable number and *m* is the total number of factors.

Artificial neural network for modeling air pollution

Artificial Neural Network is an information processing system that evolves from the way biological nervous system works (brain) in processing information (Moustris et al. 2010). In this research, a multilayer perceptron feed-forward artificial neural network (MLP-FF-ANN) with a non-linear activation function was used to determine the best input combination model for predicting API. Three models were developed (model A, B and C) using different input combinations. The network structure for these models consist of multiple neurons organized in layers, and information is presented to the network via an input layer (independent variable), which then send signal to the hidden layer via a system of weighted connection, where the actual processing is done and pass to the output layer (dependent variable). The network is trained severally by adjusting the weight value in order to minimize error. To achieve this, a Back propagation algorithm is introduced to the network in order to correlate the coefficient between the expected and the calculated using a supervised learning (Xie et al. 2009).

The data were divided into training (60%), testing (20%) and validation (20%) sets. The training data set is use to adjust the weight in order to estimate and learn the pattern of the parameters, testing is use to evaluate the generalization ability of the network, while validation subset is use to evaluate the performance of the trained network (Thursto et. al. 2011). Furthermore, the leave-one-out cross validation (combined PCA/ANN) method was introduced to the model in order to

apportion the mass source contribution and individual percentage contribution of pollutants (Wang et. al 2012; Sarkar et. al. 2012). Furthermore, the coefficient of the determination (R^2) and the root mean square error (RMSE) is used as a baseline to evaluate the performance of the ANN models. From the result, the model with the greatest R^2 value ($R^2=1$) and the lowest RMSE value (RMSE=0) is considered the best model for API prediction [45] as shown in equation (4) and (5)

$$R^2 = 1 - \frac{\sum(x_i - y_i)^2}{\sum y_i^2 - \frac{\sum y_i^2}{n}} \tag{4}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2} \tag{5}$$

Where x_i represent the observed data, y_i is the predicted data and n is the number of observation.

RESULTS AND DISCUSSION

Identification of major possible sources of pollution

In this study, PCA was applied on the raw data sets (PM₁₀, CO, O₃, SO₂, NO₂, Ambient temperature and Wind speed) in order to identify the source of atmospheric air pollutants within the study area. Two PCs with eigenvalues greater than 1(>1) were generated and account for more than 39% and 15% of the total variance in the data sets as shown in the PCA loading plot in Figure 2. Two varifactor (VFs) were obtained through the FA performed on the PCs. Only factor loadings greater than 0.5 were considered for discussion. The result of FA is presented in Table 3, comprising of factor loadings, eigenvalues, total variance and cumulative percentages.

Base on the result in Table 3, VF1 explains 39.9% of the total variance which indicate a

strong positive loading for CO (0.893), O₃ (0.801) and NO₂ (0.863). CO is produced by incomplete combustion of fuel in automobiles (motor vehicles, motorcycles, engine boats) especially when the engines are not turn properly (Jansen et. al. 2001; Morawska et al. 2002). Study conducted by Dominick et al (2012) reveals that CO accounts for 82% of pollution load in Malaysia. The high traffic congestion in the city center of Kuala Terengganu, Kuala Lumpur, and Putra Jaya contribute immensely to the emission of CO. Kuala Lumpur being the most populated city in Malaysia, Putra Jaya the third most populated (DOS, 2010) and the city of Kuala Terengganu encourage increase in individual ownership of automobiles which in turn create a serious traffic jam round the clock. O₃ is a secondary pollutant produced by photochemical oxidation and forms the main component of smog (Atkinson 2000).

Table 3: Factor loadings after Varimax rotation:

Variables	VF1	VF2
CO	0.893	-0.124
O ₃	0.801	0.140
PM ₁₀	0.447	-0.479
SO ₂	0.346	0.563
NO ₂	0.863	0.053
Ambient temperature	-0.039	0.699
Wind speed	-0.534	-0.152
Eigenvalue	2.794	1.092
Variability (%)	39.913	15.596
Cumulative %	39.913	55.509

A report by DOS (2010) show that Malaysia is located in the tropical region near the equator (high temperature) which provide favorable condition for nitrogen oxides (NO_x) and other VOCs to react, and CO slowly helps to oxidize nitrogen monoxide (NO) to nitrogen dioxide (NO₂) which indirectly accelerate O₃ formation (Bauer and Langmann 2002; Kovac-Andric and Gvozdic 2009; DOE, 2010). Furthermore, the

strong positive loading in the emission of NO₂ can be attributed to Industrial activities in Kemaman, Kertih and Kuala Lumpur as well as vehicular emission from the major cities (Pereira et. al. 2007). An estimate of about 69% of NO₂ emission are from power stations and industrial activities, 28% from motor vehicles and the remaining 3% makes up other sources (DOE, 2009).

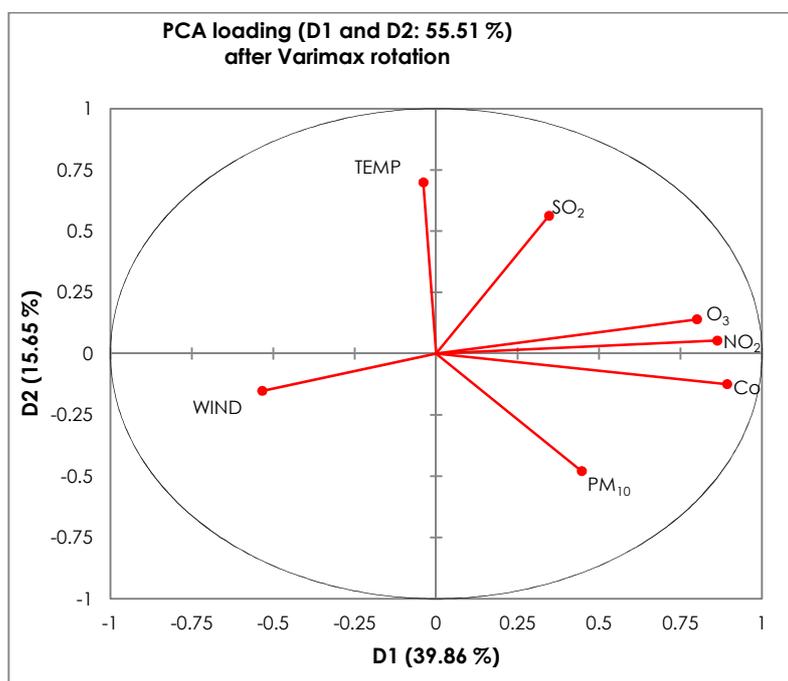


Figure 2: Plot diagram for PCA loadings after varimax rotation

The second varimax factor (VF2) accounted for 15.6% of the total variance showing a moderate loading for SO₂ and ambient temperature. The source of moderate variation in SO₂ can be attributed to combustion of fuel in automobiles, aircraft as

well as burning of fuel in factories and other industrial activities [46,47,53]. Temperature also accelerates photochemical oxidation (Bauer et. al. 2002; Kovac and Gvozdic, 2009) and provides favorable condition for nitrogen to react with oxygen (Latif et. al. 2011).

Modeling the percentage contribution of each pollutants using ANN

The leave-one-out cross-validation method was used to model the percentage contribution of individual pollutant based on the coefficient of determination developed by ANN as shown in Figure 3. It is statistically proven that O₃ is the most significant Pollutant affecting the level of air quality within the study area at R² = 63%. The second most significant pollutants contributing to the level of air pollution is PM₁₀ at R² = 15%. Other

parameters (CO, SO₂, NO₂, Ambient temperature and Wind speed) fall within the range of 3%, 5%, 4%, 5% and 5% respectively. This is because O₃ and PM₁₀ are secondary pollutants produced by photochemical oxidation of other precursors (Abdullah et al. 2012)). The heavy traffic jam, industrial activities and biomass burning in Kuala Lumpur, Putrajaya and Kuala Terengganu can be attributed to these variations (Bana et. al. 2013).

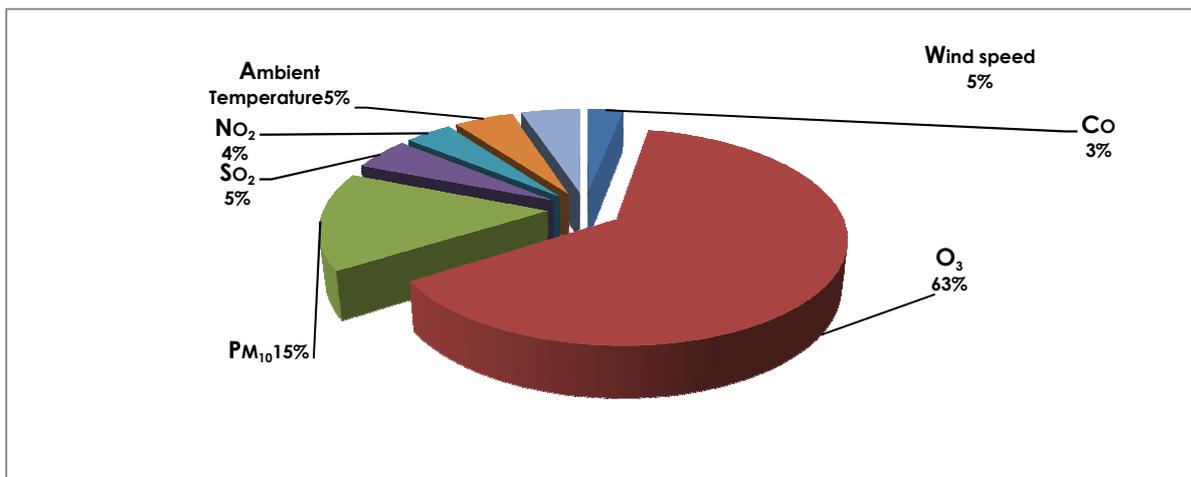


Figure 3: Percentage contribution of individual pollutants Index

Source apportionment of mass contribution using ANN

The mass source category apportionment was also modeled using a combined ANN and principal component scores after varimax rotation in Figure 4. The leave-one-out cross

validation method was also employed. The result revealed that gas and secondary pollutants affect the air pollution index by 92% followed by gas and meteorological condition 8%.

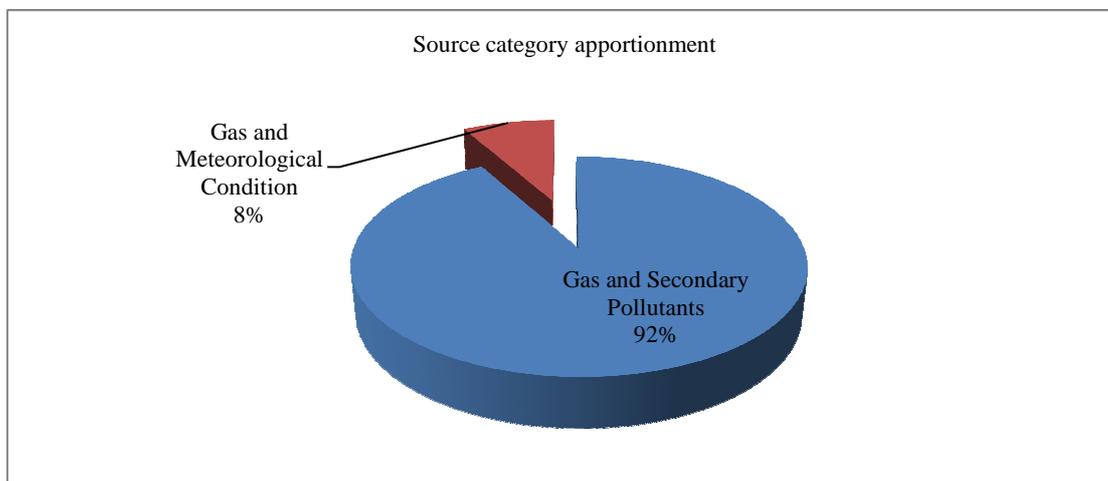


Figure 4: Source apportionment of air pollution

Determining the best input combination model for predicting API using MLP-FF-ANN

In selecting the best input combination for predicting air pollution, three models were developed (model A, model B and model C). The input combination with the highest coefficient of determination (R^2) and lowest Root mean square error (RMSE) is considered the best model for predicting API. Figure 6 below show the R^2 and RMSE performance for all the models, while Table 4 indicate the optimum gauge in identifying the prediction performance of model A, B and C.

The network structure for model A comprises of original raw data (PM₁₀, CO, O₃, SO₂, NO₂, Ambient temperature and Wind speed) as

input variables. Twenty neurons in the hidden layer with one output (API) were examined to get the optimum gauge with the highest R^2 and lowest RMSE. However, the highest $R^2 = 0.934$ with the minimum RMSE = 4.871 were gotten at node eighteen (18). On the other hand, model B gave an $R^2 = 0.921$ and RMSE = 5.050 at node eighteen (18). The input combination for model B comprises of seven principal components before varimax rotation. Model C was constructed using principal component scores after varimax rotation with engine values greater than one (> 1) base on two varifactors. The highest $R^2 = 0.551$ with the lowest RMSE = 12.410 was reached at node nineteen (19). Figure 5 shows different input combination model for MLP-FF-ANN structure.

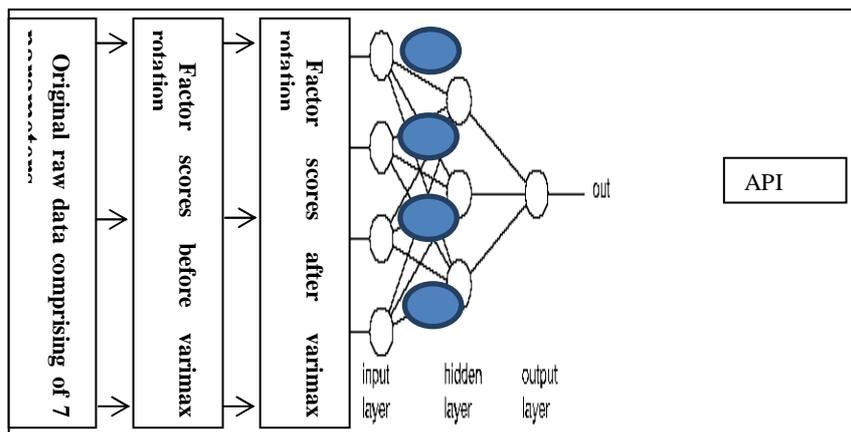
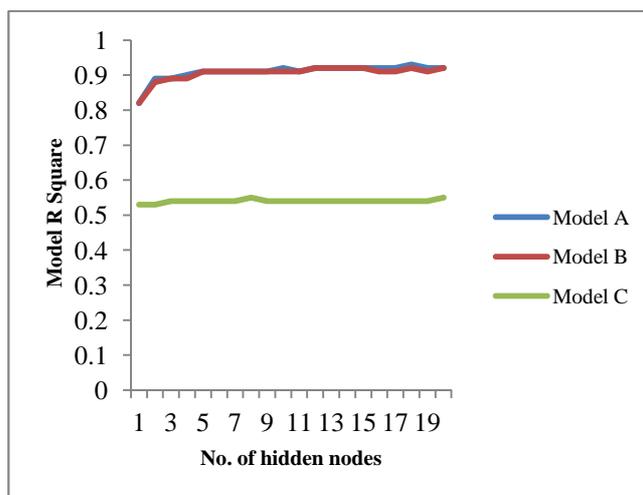


Figure 5: ANN structure with different input combination model



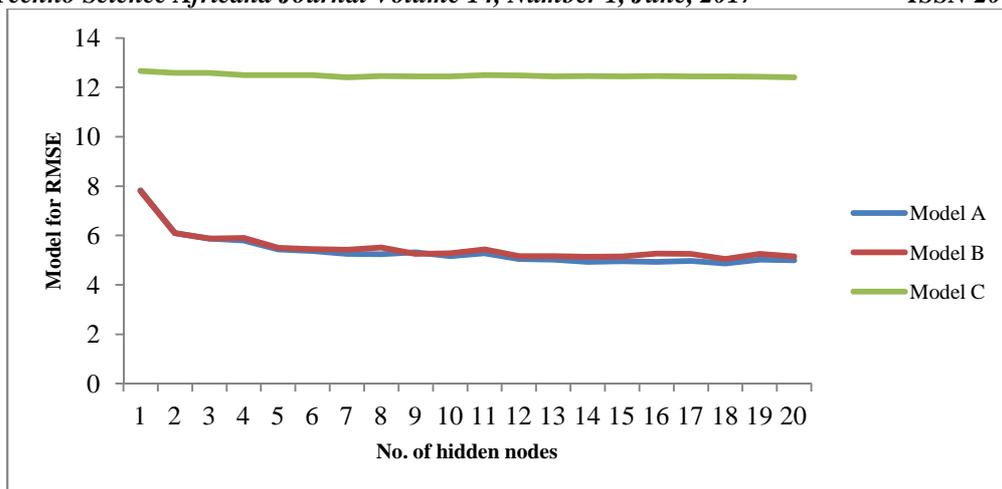


Figure 6: R² and RMSE performance for model A, B and C

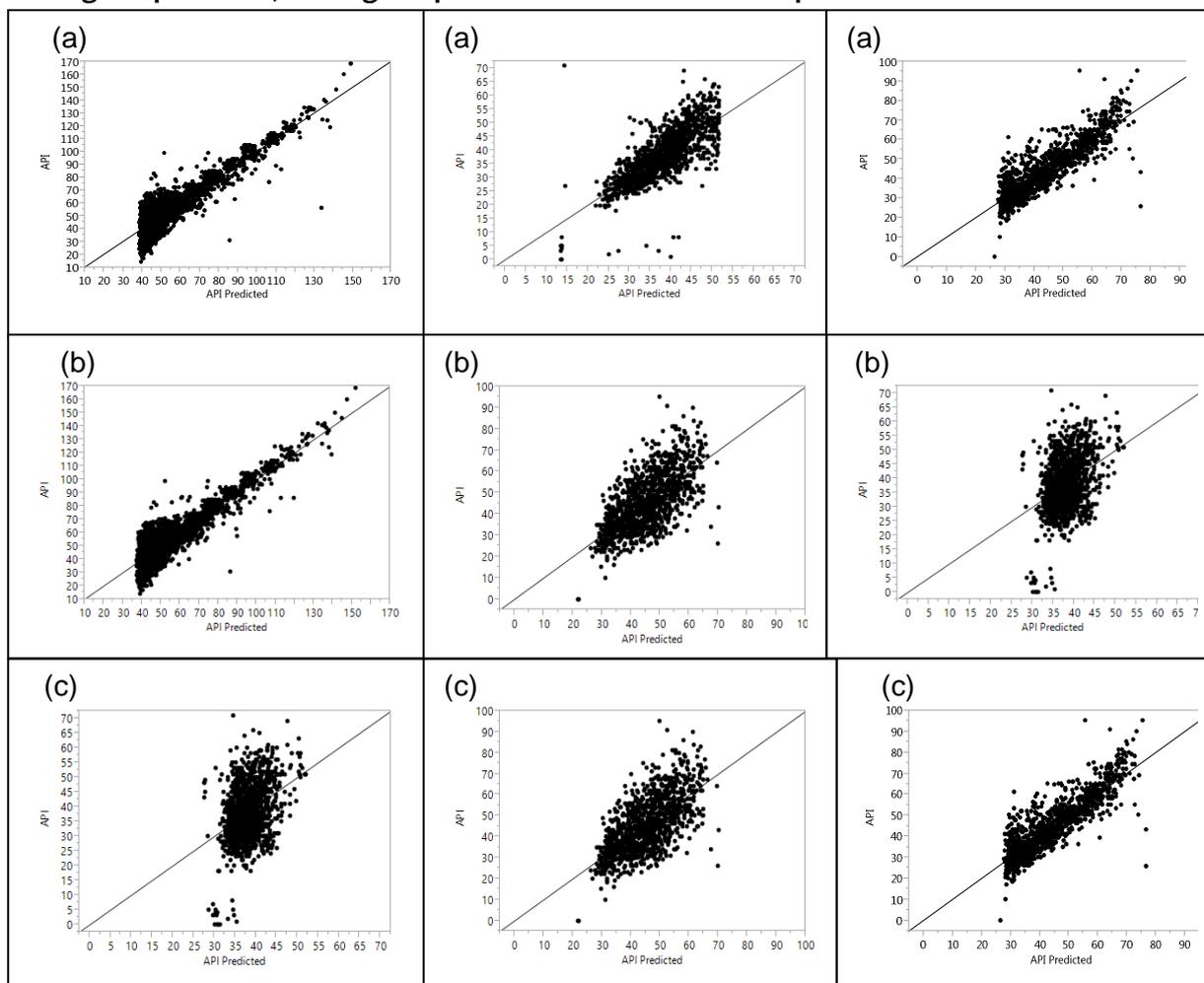
Table 4: Prediction performance in selecting the best input combination model

Models	Hidden nodes	R ²	RMSE
MLP-FF-ANN Model A	[7,18,1,1]	0.934	4.871
MLP-FF-ANN Model B	[7,18,2,1]	0.921	5.050
MLP-FF-ANN Model C	[7,19,3,1]	0.551	12.410

From the above results gotten in Table 4 and Figure 6, we can statistically ascertain that model C did not perform well in the prediction of API using factor scores after varimax rotation due to the reduction in the number of input variables compared with other models. The result for model B using seven principal components before varimax rotation as input variables is quite encouraging. The R² = 0.921 and RMSE = 5.050 is close to model A

performance and can be used as a good input combination to predict API. Furthermore, the best input combination for prediction is model A using the raw data with the highest R² = 0.934 and lowest RMSE = 4.871. This model (A) is reliable and has more capability to predict API compared with model B and C. The scatter plots for API prediction performance for training, testing and validation (model A, B and C) are shown in Figure 7 below.

Training API predicted, Testing API predicted and Validation API predicted



sources originate by northeast and southwest monsoon wind.

Figure 7: Scatter plot diagram for predicting API during training, testing and validation (a) model A (b) model B and (c) model C.

CONCLUSION

In this study, different input combinations were tested in order to determine the best model for predicting API using MLP-FF-ANN. All the models were able to predict API beyond 50% but still apparent to note that any reduction or compromise in the number of input variables from the original will limit the prediction capability of MLP-FF-ANN. Both model A and B gave a very reliable prediction using seven input variables but still show that the best combination for predicting API is model A. PCA revealed that the major source of air pollution within the study area are mostly induced anthropogenic emission from automobiles, industrial activities, power plants, biomass burning, air and sea ports. The percentage contribution of individual pollutants and source apportionment was achieved using the leave-one-out cross-validation method. It shows that O₃ is the most significant parameter

affecting the API value followed by PM₁₀. Based on the source category apportionment, the API value is influenced by 92% gas and secondary pollutants and 8% gas and meteorological condition using a combined ANN/PCA scores after varimax rotation. However, the selection of the best input combination and modeling the percentage contribution of pollutants makes ANN an important tool for air pollution modeling, decision making, problem solving and environmental management.

Acknowledgement

The authors wish to express their profound gratitude to the Department of Environment (DOE) Malaysia for their kind gesture in providing the data used for this study. A special thanks to the East Coast Environmental Research Institute, Universiti Sultan Zainal Abidin, Malaysia for the constructive academic support towards putting this findings into reality.

REFERENCES

- Abdullah, A. M., Abu Samah, M. A., & Jun, T. Y. (2012). An overview of the air pollution trend in Klang Valley, Malaysia. *Open Environ Sci*, 6, 13-19.
- Arhami, M., Kamali, N., & Rajabi, M. M. (2013). Predicting hourly air pollutant levels using artificial neural networks coupled with uncertainty analysis by Monte Carlo simulations. *Environmental Science and Pollution Research*, 20(7), 4777-4789.
- Atkinson, R. (2000). Atmospheric chemistry of VOCs and NOx. *Atmospheric Environment* 34, 2063e2101
- Azid, A., Juahir, H., Toriman, M. E., Kamarudin, M. K. A., Saudi, A. S. M., Hasnam, C. N. C., ... & 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(8), 1-14
- Azmi, S.Z., Latif, M.T., Ismail, A.S., Juneng, L., Jemain, A.A., (2010). Trend and status of air quality at three different monitoring stations in the Klang Valley, Malaysia. *Air Quality Atmosphere and Health* 3, 53e64.
- Banan, N., Latif, M.T., Juneng, L., Ahmad, F. (2013). Characteristics of surface ozone concentration at stations with different background in the Malaysia Peninsula. *Aerosols and Air quality Research* 13, 1090- 1106
- Bauer, S. E., & Langmann, B. (2002). Analysis of a summer smog episode in the Berlin-Brandenburg region with a nested atmosphere-chemistry model. *Atmospheric Chemistry and Physics*, 2(4), 259-270.
- Bernstein, J. A., Alexis, N., Barnes, C., Bernstein, I. L., Nel, A., Peden, D. & Williams, P. B. (2004). Health effects of air pollution. *Journal of Allergy and Clinical Immunology*, 114(5), 1116-1123.
- Botkin, D. B., Keller, E. A., & Rosenthal, D. B. (2012). *Environmental science*. Wiley.
- Brümelis, G., Lapiņa, L., Nikodemus, O., & Tabors, G. (2000). Use of an artificial model of monitoring data to aid interpretation of principal component analysis. *Environmental Modelling & Software*, 15(8), 755-763.
- Brunekreef, B., & Holgate, S. T. (2002). Air pollution and health. *The lancet*, 360(9341), 1233-1242.
- Chelani, A. B., Gajghate, D. G., & Hasan, M. Z. (2002). Prediction of ambient PM10 and toxic metals using artificial neural networks. *Journal of the Air & Waste Management Association*, 52(7), 805-810.
- Daigle, A., Andre', S., Dan. B., Daniel, C., Loubna, B. (2011). Multivariate analysis of low-flow regimes in eastern Canadian river California to global change. *Climatic Change* 87, S273-S292.
- Department of Environment Malaysia (DOE) (1997). *A Guide to Air Pollution Index In Malaysia (API)*. Ministry of Science, Technology and Environment, Kuala Lumpur.
- Department of Environment Malaysia (DOE) (2009). *Malaysia Environmental Quality Report*, Ministry of Science, Technology and Environment, Kuala Lumpur.
- Department of Environment Malaysia (DOE) (2010). *Malaysia Environmental Quality Report*, Ministry of Science, Technology and Environment, Kuala Lumpur.
- Department of Environment Malaysia (DOE) (2011). *Malaysia Environmental Quality Report*, Ministry of Science, Technology and Environment, Kuala Lumpur.
- Department of Environment Malaysia (DOE) (2012). *Malaysia Environmental Quality Report*, Ministry of Science, Technology and Environment, Kuala Lumpur.
- Department of Statistic Malaysia (DOS) (2009). *Basic Population Characteristics by Administrative Districts 2009 Report*.
- Department of Statistic Malaysia (DOS) 2010. *Basic Population Characteristics by Administrative Districts 2009 Report*.
- Department of Statistic Malaysia (DOS) 2011. *Basic Population Characteristics by Administrative Districts 2009 Report*
- 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.
- Genc, D. D., Yesilyurt, C., & Tuncel, G. (2010). Air pollution forecasting in Ankara, Turkey using air pollution index and its relation to assimilative capacity of the atmosphere. *Environmental monitoring and assessment*, 166(1-4), 11-27.
- Hakimpoor, H., Arshad, K. A. B., Tat, H. H., Khani, N., & Rahmandoust, M. (2011). Artificial neural networks' applications in management. *World Applied Sciences Journal*
- Helena, B., Pardo, R., Vega, M., Barrado, E., Fernandez, J. M., & Fernandez, L. (2000). Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by principal component analysis. *Water research*, 34(3), 807-816.
- Ismail, A. S., Latif, M. T., Azmi, S. Z., Juneng, L., & Jemain, A. A. (2010). Variation of surface ozone recorded at the eastern coastal region of the Malaysian Peninsula. *American Journal of Environmental Sciences*, 6(6), 560.
- Jacobson, M. Z. (2002). *Atmospheric pollution: history, science, and regulation*. Cambridge University Press.
- Jamal, H.H., Pillay, M.S., Zailina, H., Shamsul, B.S., Sinha, K., Zaman Huri, Z., Khew, S.L., Mazrura, S., Ambu, S., Rahimah, A., Ruzita, M.S., (2004). *A Study of Health Impact & Risk Assessment of Urban Air Pollution in Klang Valley*. UKM Pakarunding Sdn Bhd, Malaysia, Kuala Lumpur.
- Janssen, N. A., van Vliet, P. H., Aarts, F., Harssema, H., & Brunekreef, B. (2001). Assessment of exposure to traffic related air pollution of children attending schools near motorways. *Atmospheric environment*, 35(22), 3875-3884.

- Juahir, H., Zain, S. M., Yusoff, M. K., Hanidza, T. T., Armi, A. M., Toriman, M. E., & Mokhtar, M. (2011). Spatial water quality assessment of Langat River Basin (Malaysia) using environmetric techniques. *Environmental Monitoring and Assessment*, 173(1-4), 625-641.
- Junninen, H., Niska, H., Tuppurainen, K., Ruuskanen, J., & Kolehmainen, M. (2004). Methods for imputation of missing values in air quality data sets. *Atmospheric Environment*, 38(18), 2895-2907.
- Kolehmainen M, Martikainen H, Ruuskanen J. (2001) Neural networks and periodic components used in air quality forecasting. *Atmos. Environ* 35:815–825
- Kovač-Andrić, E., Brana, J., & Gvozdić, V. (2009). Impact of meteorological factors on ozone concentrations modelled by time series analysis and multivariate statistical methods. *Ecological Informatics*, 4(2), 117-122.
- Latif, M. T., Azmi, S. Z., Noor, A. D. M., Ismail, A. S., Johnny, Z., Idrus, S., ... & Mokhtar, M. B. (2011). The impact of urban growth on regional air quality surrounding the Langat River Basin, Malaysia. *The Environmentalist*, 31(3), 315-324.
- Liu, C. W., Lin, K. H., & Kuo, Y. M. (2003). Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan. *Science of the Total Environment*, 313(1), 77-89.
- Love, D., Hallbauer, D., Amos, A., & Hranova, R. (2004). Factor analysis as a tool in groundwater quality management: two southern African case studies. *Physics and Chemistry of the Earth, Parts A/B/C*, 29(15), 1135-1143.
- Ministry of Transport Malaysia (MOT) (2010). Malaysia Transport Statistics Report 2010, Putrajaya, Malaysia
- Morawska, L., Jayaratne, E. R., Mengersen, K., Jamriska, M., & Thomas, S. (2002). Differences in airborne particle and gaseous concentrations in urban air between weekdays and weekends. *Atmospheric Environment*, 36(27), 4375-4383.
- Moustris, K. P., Ziomas, I. C., & Paliatsos, A. G. (2010). 3-Day-ahead forecasting of regional pollution index for the pollutants NO₂, CO, SO₂, and O₃ using artificial neural networks in Athens, Greece. *Water, Air, & Soil Pollution*, 209(1-4), 29-43.
- Mustaffa, N. I. H., Latif, M. T., Ali, M. M., & Khan, M. F. (2014). Source apportionment of surfactants in marine aerosols at different locations along the Malacca Straits. *Environmental Science and Pollution Research*, 21(10), 6590-6602.
- 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* 1-2
- Osman, R., Saim, N., Juahir, H., & Abdullah, M. P. (2012). Chemometric application in identifying sources of organic contaminants in Langat river basin. *Environmental monitoring and assessment*, 184(2), 1001-1014.
- Pereira, M.C., Santos, R.C., Alvim-Ferraz, M.C.M. (2007). Air quality improvement using European Environmental Policies: a case study of SO₂ in a coastal region in Portugal. *Journal of Toxicology & Environment Health- part A: Current issues* 70,347-351.
- Raven P.H., David M., Hassan Z., Linda R. B. (2013). *Environment, international student version*. 8 edition, John Wiley & Sons, ISBN 978-1-118-09237-8
- Razak, I. S., Latif, M. T., Jaafar, S. A., Khan, M. F., & Mushrifah, I. (2014). Surfactants in atmospheric aerosols and rainwater around lake ecosystem. *Environmental Science and Pollution Research*, 1-10.
- Razak, M. I. M., Ahmad, I., Bujang, I., Talib, A. H., & Ibrahim, Z. (2013). Economics of air pollution in Malaysia. *Int J Humanit Soc Sci*, 3(13), 173-177.
- Salma, I., Chi, X., & Maenhaut, W. (2004). Elemental and organic carbon in urban canyon and background environments in Budapest, Hungary. *Atmospheric Environment*, 38(1), 27-36
- Sarkar, A., & Kumar, R. (2012). Artificial Neural Networks for Event Based Rainfall-Runoff Modeling. *Journal of Water Resource and Protection*, 4(10), 891.
- Thurston, G. D., Ito, K., & Lall, R. (2011). A source apportionment of US fine particulate matter air pollution. *Atmospheric Environment*, 45(24), 3924-3936.
- Vega, M., Pardo, R., Barrado, E., & Debán, L. (1998). Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. *Water research*, 32(12), 3581-3592.
- Wahid, N. B. A., Latif, M. T., & Suratman, S. (2013). Composition and source apportionment of surfactants in atmospheric aerosols of urban and semi-urban areas in Malaysia. *Chemosphere*, 91(11), 1508-1516.
- Wang, L., Li, J., Lu, H., Gu, Z., Rioual, P., Hao, Q., . & Chu, G. (2012). The East Asian winter monsoon over the last 15,000 years: its links to high-latitudes and tropical climate systems and complex correlation to the summer monsoon. *Quaternary Science Reviews*, 32, 131-142.
- Wang, Y. B., Liu, C. W., Liao, P. Y., & Lee, J. J. (2014). Spatial pattern assessment of river water quality: implications of reducing the number of monitoring stations and chemical parameters. *Environmental monitoring and assessment*, 186(3), 1781-1792.
- World Health Organization (2002). *World Health Report: Reducing Risk, Promoting Healthy Life*. Geneva, Switzerland.
- Xie, H., Ma, F., & Bai, Q. (2009, August). Prediction of indoor air quality using artificial neural networks. In *Natural Computation, 2009. ICNC'09. Fifth International Conference on* (Vol. 2, pp. 414-418). IEEE.