TEST Engineering & Management

November-December 2019
ISSN: 0193-4120 Page No. 826 - 831

Investigating the Air Pollutants Contaminants using a Statistical Technique

Haslina Zakaria[1], Shamshuritawati Sharif[2], Basharoh Abdul Karim[3], Tun Mohd Firdaus Azis[4]

[1]Institute of Engineering Mathematics, Universiti Malaysia Perlis, Arau, Perlis, Malaysia.
[2]School Of Quantitaive Sciences, Universiti Utara Malaysia, Kedah, Sintok, Kedah, Malaysia.
[3]Kolej Poly-Tech MARA Alor Setar, Alor Setar, Kedah, Malaysia.
[4]Faculty of Applied Sciences, Universiti Teknologi Mara Cawangan Perlis, Arau, Perlis, Malaysia.

Article Info
Volume 81
Page Number: 826- 831
Publication Issue:
November-December 2019

Article History
Article Received: *3 January 2019*
Revised: *25 March 2019*
Accepted: *28 July 2019*
Publication: *25 November 2019*

Abstract
Air is contaminated due to the occurrence of many gas and dust particles. Air pollution is caused by the existance of the contaminants in the air. It is very serious issue that required awareness from all significant authorities because it can alarms health of public, agricultural activities and industries, forest and non-forest species and our environments. Air pollutants are included solid particles, gasses, and liquid droplets in the air. Thus, it is significant to analyze the current situation of air quality. The station in Klang Valley which are Klang Station is our subject of interest because it is one of unhealthy area. Therefore, three years contaminants data containing five contaminants are used. In performing the analysis, two multivariate statistical analysis which are correlation analysis and principal component analysis (PCA) are applied. As a results, nitrogen oxides (52%) is the most influential contaminant of air pollutants. Furthermore, particulate matter (43.9%) and carbon monoxide (43.3%) are the second and the third most important, respectively.

Keywords: *Air quality, Multivariate statistics, Pollution, Principal component analysis.*

I. INTRODUCTION

Air is contaminated due to the occurrence of many gas and dust particles. The existence of the contaminants in the air is defined as air pollution. It is very serious issue that required awareness from all significant authorities because it can alarms health of public, agricultural activities and industries, forest and non-forest species and our environments. When there are a solid particles, gasses, and or liquid droplets occur in the air [1], it brings the pollution to our environment. In 2015, from the whole world's population, there is about 92% of us lived in unhealthy air condition [2]. One of the main reason to the air pollution issues in Malaysia is the fast transformation of this country into a wide urban country [3]. Generally, air pollutions are affected by the industrial and activities in the factories, using dry cleaners, motor vehicles dust, wind debris, and wildfires. In addition, the removal of dust from construction site and quarry mines, open burning, power plants, industrial waste incinerators are another factor that dominant to the air pollution in Malaysia [4]. Air pollutants such as solid particles, gasses, and liquid droplets can affect the ecosystems and humans health. In details, gases consist of five chemical component which are ozone (O_3), nitrogen dioxides (NO_2), volatile organic compounds (VOCs), carbon monoxide (CO), and the particulate matter (PM_{10}). The pollutants in the air may spread over the area and distillate during varied time period because the distributions of the pollutant are influenced by wind and geographical factors [5]. In Malaysia, a measurement of pollution is refer to the air pollutant index. Most of the time, it is known as API where monitored by the Act of Clean Air. This is a Malaysian Air Quality Guidelines that issued and controlled by the Department of Environment (DOE) [6]. What is the benefit of API? The API is employed to produce an air quality index where the public can used it to recognized our current air status [7], In specific, the API is daily produced by DOE and it categorized into six status indicators which are good (between 0 and 50), moderate (between 51 and 100), unhealthy (between 101 and 200), very unhealthy (between 201 and 300), hazardous (between 300 to 500), and emergency (more than 500) as shown in Table 1. Essentially, the API is computed on the basis of five pollutants consist of particulate matter (PM_{10}), ozone (O_3), carbon monoxide (CO), sulphur dioxide (SO_2), and nitrogen dioxide (NO_2). The lower scale of API present good quality of air whereas the larger scale

November-December 2019
ISSN: 0193-4120 Page No. 826 - 831

present the bad quality [8]. To be in good condition or under controlled, the API should be below than 100 scale. Otherwise, public will faced with the risk of unhealthy air.

TABLE 1: THE SCALE OF AIR POLLUTION INDEX (API)

Scale of API	Quality of Air
Between 0 and 50	Good
Between 51 and100	Moderate
Between 101 and 200	Unhealthy
Between 201 and 300	Very Unhealthy
More than 300, Less than 500	Hazardous
More than 500	Emergency

Remarkably, the air pollution be able to interrupt public health in both circumstances either short range or long range. In the short range, the bad quality of air pollution can influence everyone in severe condition, in example, the sunlight blockage may growth the spreading of anxious bacteria and several viruses [9]. For the time being, the long range health effects can impact us poorly, in example, chronic respiratory infection, lung cancer, heart infection, and failures of nerves, liver and kidney, or possible to damage our brain as well. Besides, the air contaminants may damage the respiratory passageway, and also a delicate tissues in the eyes [5]. There are so many bad effect to the health. One study in 1997 showed that acute respiratory infection or disease, asthma problems, and conjunctivitis incident were increase during the topmost smoke haze period [4]. Besides that, air pollution give a very significant impact to the health of our children, elderly group, youngsters group, and people with preexistent problem of respiratory and lung where their level of immune cannot protect the disease and symptom very well [10]. As a result, the total health injury cost was increased during the smoke haze time period. Additionally, the haze problem also distressing the country persistent productivity [4]. Therefore, identification of the most influential contaminants become important subject since there are few number of contaminants been measured by DOE. In this paper, the air-pollutant data is investigated. Multivariate statistical method is used to express the exact situation of air pollution in Klang Station. Identifying the significant contaminants is a must when involving more than one pollutants. In the Section II, we deliver the details of Klang Station. In what follows, the research methodology including the correlation analysis, and principal component analysis are presented in Section III. Lastly, the statistical results and summarization of the study are presented in Section IV, and in the last section we delivered concluding remarks.

II. KLANG STATION

In June 2013, Malaysia experiencing worst levels of haze which enforced government to declare emergency status and to closed schools in Kuala Lumpur, Selangor and Pahang as smoke from neighboring Indonesia pollutes the Malaysia atmosphere. Approximately, there is about 60% of residents suffering acute respiratory infection during the period, which are cough, difficulty of breathing, sore throats and the rest symptoms of acute respiratory infections. Unfortunately, the highest number of residents with acute respiratory infection during haze was located at Klang, due to the highest API reading at Klang comparing to other area. As reported, Port Klang was listed as the highest reading of 214 [5]. The statistics report that published by Pusat Perubatan Universiti Kebangsaan Malaysia shown that respiratory diseases has listed among the 10 highest reasons of ward hospital admission and reasons of death.

In this study, Klang station was selected based on the previous history as one of the unhealthy area that highly affected by the air pollution The station is placed in the area of Klang Valley. The Klang Valley is recognized as the center of Malaysia that contains many factories, industries and commercial activities. Geographically, the spot can be found at latitude of N03° 00.620, and longitude of E101° 24.484. The Klang Valley is somewhere nearby to Titiwangsa Mountains and Strait of Malacca. The Klang station is located in greatly industrial and commercial areas that have very congested main road links and highways. This station also very nearby to the habour that located in Port Klang and having busy power plant activities. The station is close to the main roads to Kuala Lumpur which are suffering with congested traffic flow, especially in the morning time period [11]. Thus, the atmosphere in Klang is contaminated with toxic and non-toxic air pollutant especially CO, PM_{10}, and SO_2 which are very harmful to human health and environment

November-December 2019
ISSN: 0193-4120 Page No. 826 - 831

[12]. Besides, the level of air pollutant is also influenced by the environment of Klang Valley where nearby to Klang Station. The location of Klang Station contributes a motionless condition or known as stagnant, where all the contaminants of air pollutants are static to the valley area [10]. Moreover, it is crucial because Klang lies in rapid residential and school area. Therefore, this study is carried out to investigate the atmosphere in the Klang.

The data of five air quality contaminants is obtained from the DOE. The data from year 2011 to year 2014 are collected by authorized agency known as Alam Sekitar Sdn. Bhd. (ASMA). The data is documented in hourly time period.

III. AIR QUALITY CONTAMINANTS AND ITS EFFECTS

There are five major pollutants involved in this study CO, NO_2, PM_{10}, SO_2, and O_3. CO is a colorless and adorless gas, which is produced by fossil fuel, particularly when combustion is not appropriate, as in burning coal and wood. The CO emission which released into the atmosphere is dominantly caused by transportation activities include both private cars and businesses owned vehicles [13]. The higher concentration of CO can be seen at area which has an industrial background [11]. Symptoms of CO poisoning may include headache, dizziness, weakness, nausea, vomiting, and finally loss of consciousness. The symptoms are very similar to those of other illness, such as food poisoning or viral infections.

The second pollutant is Sulphur Dioxide (SO_2). It is a colorless, highly reactive gas, which is mostly emitted from fossil fuel consumption, natural volcanic activities, and industrial processes. The main source is come from motor vehicles, diesel-engined trucks, lorries and busses [13]. SO_2 was at the high levels during the haze period. SO_2 is very harmful for plant growth, animal and human health. People with lung disease, children, older people, and those who are more exposed to SO_2 are at higher risk of the skin and lung disease [14].

Next, Nitrogen Dioxide (NO_2) levels are indicated by the industrial and urban environments. Usually, the high level is due to the burning of fossil fuels. The area with an industrial background has higher concentrations of NO_2 compared to the urban sites because CO are the primary pollutants that are mostly emitted from motor vehicles [13]. Nitrogen oxides are important ambient air pollutants which may increase the risk of respiratory infections. Coughing and wheezing are the most common complication of nitrogen oxides toxicity, but the eyes, nose or throat irritations, headache, dyspnea, chest pain, diaphoresis, fever, bronchospasm, and pulmonary edema may also occur. In another report, it is suggested that the level of nitrogen oxide between 0.2 and 0.6 ppm is harmless for the human population [15].

Moreover, O_3 is also the contaminants in the atmosphere. It is a gas that resulted from the mixture of hydrocarbon, nitrogen oxides, and sunlight. The concentration of NO which causes by motor vehicles in the urban area has the potential to interact with O_3 to form NO_2 in atmosphere [13]. It has an association with the increase in risk of respiratory disease [16].

Lastly, the air pollution contaminant is PM_{10} which is the main parameter during the haze [4]. During the haze episode, the smoke from the Sumatra will moves to the Peninsular Malaysia especially in the area of Klang Valley [11]. Wheezing, cough, dry mouth, and limitation in activities due to breathing problems are the most prevalent clinical symptoms of respiratory disease resulted from air pollution [17].

IV. MULTIVARIATE STATISTICAL TECHNIQUES

A. The Correlation Analysis

In statistical analysis, the Pearson correlation coefficient (*r*) is used to indicate the strength and direction of association among two variables [18]. In this study, it used to ensure that two contaminants of air pollutants having the relationship before conducting the next analysis which is Principal Component Analysis (PCA). If the coefficient is positive, it present both contaminants having same direction, but negative correlation present that both contaminants having inverse relationship. In terms of strength, it can be seen from the coefficient of correlation. We expect both contaminants under study must have strong relationship does not matter either positive or negative sign.

The correlation coefficient quantifies the degree of linear relationship between *i*-th and *j*-th contaminants. By definition, $\rho_{ii} = 1$ for all *i* and ρ_{ij} can vary from – 1 to 1 for all $i \neq j$ where, $\rho_{ij} = 1$

means that two variables is perfectly positive linear relationship, ρ_{ij} = -1 means that two variables is perfectly negative linear relationship, and ρ_{ij} = 0 means that two variables is perfectly no linear relationship. To identify the information contained, matrix among those contaminants is constructed. Thus, the symmetric matrix of size 5 x 5 where the element in the i-th row and j-th column is defined as,

$$\rho_{ij} = \frac{\langle R_i R_j \rangle - \langle R_i \rangle \langle R_j \rangle}{\sqrt{\left(\langle R_i^2 \rangle - \langle R_i \rangle^2\right)\left(\langle R_j^2 \rangle - \langle R_j \rangle^2\right)}} \qquad \text{(I)}$$

where representing the correlation coefficient between *i*-th and *j*-th contaminants.

B. Principal Component Analysis (PCA)

Next, after performing the correlation coefficient, we present the analysis of principal component. This is one of the multivariate statistical analysis that can help the researcher to convert into a new, and smaller dataset of uncorrelated [19][7]. The advantage of this technique is a large dataset of correlated variables can be explained using variance and eigenvalues. PCA also be able to display the most significant contaminants. The information may help us to identify the root causes of the air pollutants occurrences.

Using this analysis, it means that the less significant contaminants will be removed from the data set by minimizing the potential loss of original data [19]. In other words, PCA is the process of investigating the linear transformations $Y_1, Y_2, \ldots, Y_p$ from the original variables $X_1, X_2, \ldots, X_p$. Linear transformation can be obtained by conducting the following step should be followed [20].

1. Performing data standardization,
2. Compute matrix of the covariance,
3. Compute the eigenvalues and its eigenvectors from the covariance matrix.
4. Compute factor loading

Basically, PCA can be used to merge several variables together by referring to the correlation coefficient between two or more air pollutants. Futhermore, the varimax rotation approach is employed because it can simplifies the factor structure as well as the interpretation is easier and reliable.

Additionally, to find the number of PCA that should be considered, we used the approach by Kaiser Criterion [6]. Therefore, the number of varifactors is depends on eigenvalues which larger than 1. Besides that, factor loadings is also important things. It is the values that used to quantify the relationship between variables. The factor loading can be classified as Strong (more than 0.75), moderate (between 0.50 to 0.75), and weak (0.30 to 0.49) factor loading [6][21]. For achieving the objective, MATLAB software is used to help the statistical computations.

V. STATISTICAL RESULTS AND SUMMARIZATION

A. Relationship and Association Among Several Air Pollutants

Table 1 represents the correlation coefficient between the pollutants of air pollution. In this paper, we found that all pollutants have a strong and positive correlations with each pollutant excluding relations for SO_2 and NO_2 (r = -0.108). The result showed that PM_{10} had a strong and positive correlation with CO (r = 0.705) and positive correlation with NO_2, O_3 and SO_2. A positive relationship exists when one pollutant increases as another pollutant increases, or one pollutant decreases while the other pollutant decreases. These findings report a high correlation between PM_{10} and CO and reliable with a study by [18] and [22]. CO emissions are greater if the engine from the boat is not set properly and therefore the fuel is not fully burned. However, improper combustion of motor vehicle engines is a main cause of CO [7]. Meanwhile, high concentrations of PM_{10} are associated with gas and particulate matter, which is caused by vehicle smoke, open burning and soil emission around Klang area. Klang is located in a greatly manufacturing area surrounded by congested major roads and near the busy port (Port Klang) [7].

November-December 2019
ISSN: 0193-4120 Page No. 826 - 831

TABLE 1: PEARSON'S CORRELATION COEFFICIENT BETWEEN CONTAMINANTS

Parameters	O_3	CO	SO_2	PM_{10}	NO_2
O_3	**1.000**	0.190	0.020	0.208	0.281
CO	0.190	**1.000**	0.149	**0.705**	0.171
SO_2	0.020	0.149	**1.000**	0.109	-0.108
PM_{10}	0.208	**0.705**	0.109	**1.000**	0.001
NO_2	0.281	0.171	-0.108	0.001	**1.000**

B. The Most Important Contaminants

According to Table 2, there have two varifactors (VFs) were retained with eigenvalues larger than 1 (>1) and worn as the contribution pollutants of air pollutants. These varifactors correspond to 62.25% of the variance of the dataset were chosen regarding to eigenvalues larger than 1. VF1 contains on 37.63% of the variance. It shows high loadings from CO (0.643) and PM_{10} (0.619). While, VF2 accounted for 24.63% of the variation in the dataset. It shows high loads of SO_2 (0.484), NO_2 (-0.687) and O_3 (-0.466). Eigenvalues with less than 1 (<1) are ignored because they are overridden by more significant factors. In this paper, the results of varifactors with total values higher than 0.4 were chosen as the selection threshold, as the values are more consistent. According to Table 2, it represent that Nitrogen Oxide (52%) was the mainly significant air pollutant at the Klang station followed by Particulate Matter (43.9%) and Carbon Monoxide (43.3%).

TABLE 2: RESULTS OF PCA

Parameters	VF1	VF2	VF3	VF4	VF5	Communalities
Ozone (O_3)	0.358	-0.466	0.389	0.703	0.095	0.346
Carbon Monoxide (CO)	0.643	0.142	-0.206	-0.213	0.692	0.433
Sulphur Dioxide (SO_2)	0.166	0.484	0.829	-0.212	-0.072	0.262
Particulate Matter 10 (PM_{10})	0.619	0.236	-0.296	0.097	-0.681	0.439
Nitrogen Oxides (NO_2)	0.219	-0.687	0.176	-0.637	-0.207	0.520
Eigenvalues	1.881	1.231	0.910	0.710	0.266	
Variability (%)	37.627	24.626	18.209	14.209	5.329	

VI. CONCLUSION

The correlation analysis and principal component analysis (PCA) were used in this paper. The larger the correlation value, the stronger the correlation between the two pollutants under study. It will also reflect the PCA's results. For example, the correlation value for CO and PM_{10} is large (0.705), so in PCA, both pollutants are maintained in the same component with a large factor load. Basically, PCA is a technique used to emphasize inequality and generate strong patterns in datasets. In addition, PCA also has the ability to determine the source of pollutants to indicate the most important parameters. In this study, the results show that NO_2 is the most significant air pollution followed by PM_{10} and CO at Klang station. In practice, higher NO_2 concentrations are due to the source of motor vehicle smoke and industrial emission, while high sources of PM_{10} are often contributed by gas and particles. In particular, CO is formed by the incomplete combustion of the engine fuel. Meanwhile, NO_2 is largely due to industrial activity and heavy traffic congestion and one of the main causes of PM_{10} concentration is the hot weather factor. According to data obtained from the Ministry of Transport Malaysia, one of the major contributing factors to the structure of the pollutants is the smoke of motor vehicles in Malaysia. To control and address pollution issues, the governing bodies, legislative and regulatory bodies need to take action by designing and formulating effective plans for reducing emissions from motor vehicles and other anthropogenic activities.

ACKNOWLEDGEMENT

We are very grateful to the Department of Environment Malaysia (DOE) for their use of air quality data. Thanks to Universiti Malaysia Perlis for financial support and Universiti Utara Malaysia for providing research tools and moral support. We would also like to thank the reviewers for their constructive comments and suggestions.

REFERENCES

1. B. R. Gurjar, T. M. Butler, M. G. Lawrence and J. Lelieveld, Atmospheric Environment 42(7), 1593-1606 (2008).
2. World Health Organization (WHO). Ambient air pollution: A global assessment of exposure and burden of diseases. Geneva: WHO (2016).
3. A. M. Abdullah, M.A. A. Samah and T. Y. Jun, Open Environmental Sciences 6, 13-19 (2012).
4. R. Afroz, M. N. Hassan and N. A. Ibrahim, Environmental Research 92, 71-77 (2003).
5. N.A. Mabahwi, O. L. H. Leh and D. Omar. Social and Behavioral Sciences 170, 282 – 291 (2015).
6. A. Azid, H. Juahir, M. E. Toriman, M. K. A. Kamarudin, A. S. M. Saudi, C. N. C. Hasnam and M. R. Osman, Water, Air, & Soil Pollution 225(8), 2063 (2014).
7. D. Dominick, H. Juahir, M. T. Latif, S. M. Zain and A. Z. Aris, Atmospheric Environment 60, 172-181 (2012).
8. Department of Environment Malaysia (DOE) Main sources of air pollution in Malaysia. Kuala Lumpur: Ministry of Science, Technology and Environment (2014).
9. R. Beardsly, P. A. Bromberg, D. A. Costa, R. Devlin, D. W. Dockery, M. W. Frampton, W. Lambert, J. M. Samet, F. E. Speizer and M. Utell. Sci. Am, 24-25 (1997).

10. M.B. Awang, A. B. Jaafar, A. M. Abdullah, M. B. Ismail, M. N. Hassan, R. Abdullah, S. Johan and H. Noor. Respirology 5, 183-196 (2000).
11. A. A. A. Mohtar, M. T. Latif, N. H. Baharudin, F. Ahamad, J. X. Chung, M. Othman and L. Juneng, Geoscience Letters 5(1), 21 (2018).
12. N. D. Mohamad, Z. H. Ash'aari and M. Othman, Procedia Environmental Sciences, 30, 121-126 (2015).
13. S. Z. Azmi, M.T. Latif, A. S. Ismail, L. Juneng and A. A. Jemain. Air Quality, Atmosphere & Health, 3(1), 53-64 (2010).
14. T. M. Chen, J. Gokhale, S. Shofer and W. G. Kuschner. Am J Med Sci., 333, 249-56 (2007).
15. T. W. Hesterberg, W. B. Bunn, R. O. McClellan, A. K. Hamade, C. M. Long, and P. A. Valberg. Crit Rev Toxicol, 39, 743-81 (2009).
16. A. K. Gorai, F. Tuluri and P. B. Tchounwou. Int. J. Environ Res Public Health. 11, 4845-69 (2014).
17. Y. Gao, E. Y. Chan, L. Li, P.W. Lau, T.W. Wong. BMC Public Health. 14:105 (2014).
18. S. R. A. Rahman, S. N. S. Ismail, M. F. Ramli, M.T. Latif, E. Z. Abidin and S. M. Praveena, World Environment 5(1), 1-11 (2015).
19. A. Azid, H. Juahir, M.A. Amran, Z. Suhaili, M. R. Osman, A. Muhamad, I. Z. Abidin, N.H Sulaiman and A.S M. Saudi. *Malaysian Journal of Analytical Sciences*, *19(*6), 1415-1430 (2015).
20. F. Hussain, Y. Z. Zubairi and A. G. Hussin. Scientific research and essays, 6(15), 3172-3181 (2011).
21. C. W. Liu, K. H. Lin and Y. M. Kuo. Science of the Total Environment, 313, 77 – 89 (2003).
22. B. Mansouri, E. Hoshyari and A. Mansouri. Iranian International Journal of Environmental Science 1, 1440-1447 (2011).

AUTHORS PROFILE

Haslina Zakaria completed her Decision Science (Hons) first degree in Universiti Utara Malaysia, 2004. She had started her career in Engineering of Mathematics in 2006, and her teaching experience had exceeded 13 years in Universiti Malaysia Perlis (UniMAP). She had obtained her Master of Science (Data Analysis) from Universiti Utara Malaysia in 2019. Her interest in this profession motivates her to work hard in the teaching, research, publications and consultancy. Her main expertise is in engineering mathematics and statistics.

Shamshuritawati Sharif currently a senior lecturer at Universiti Utara Malaysia. She obtained a PhD in Mathematics at Universiti Teknologi Malaysia. Her Master degrees in Decision Science were from Universiti Utara Malaysia in 2003. Her diploma and Bachelor of Science inStatistics were obtained from MARA Institute of Technology (ITM) in 2000. She is interested in multivariate hypothesis testing, industrial statistics, statistical quality control, network analysis and centrality measure.

Basharoh Abdul Karim completed her Bachelor of Science (Honors) in Mathematics at Universiti Putra Malaysia (UPM) on February 9, 2002. She has started her career as a lecturer in Kolej Poly-Tech MARA Alor Setar since June 2002, and her teaching experience had exceeded 17 years. She had obtained her Master of Science (Data Analysis) from Universiti Utara Malaysia in 2019. Her interest in this profession motivates her to work hard in the teaching, research and publications. Her main expertise is in statistics and business mathematics.

Tun Mohd Firdaus Bin Azis currently a lecturer at Universiti Teknologi MARA Cawangan Perlis. Her Master degrees in Entomology were from Universiti Kebangsaan Malaysia in 2013. He is interested in study of insect, and Biostatistics.