BORNEO SCIENCE 6: 57-62, DEC 1999

# STATISTICAL ANALYSIS OF AIR POLLUTION IN THE KLANG VALLEY

**Ahmad Shukri Yahaya[1] and Noormala Mohd Yusoff[2]**

[1]Unit Perkhidmatan Akademik, Universiti Sains Malaysia, Kampus Cawangan Perak, 31750 Tronoh, Perak.
[2]Jabatan Komputeran Industri, Fakulti Teknologi & Sains Maklumat, Universiti Kebangsaan Malaysia, 43600 UKM, Selangor.

**ABSTRACT.** *The first part of this paper looks at the level of air pollution in the Klang Valley by considering two major components of air pollutants that is suspended particulate matter and lead particles. The level of air pollution in traffic areas, commercial areas, industrial areas and residential areas were compared. In the second part of this study , the analysis was concentrated on air pollution by heavy metals like plumbum (Pb), ferrum (Fe), cuprum (Cu), sodium (Na), calcium (Ca), aluminium (Al) and nickel (Ni). Factor analysis by principal component method was used to identify the source of pollution by these heavy metals. The result shows that there exist three factors which contribute to air pollution that is composite, traffic and industrial sources.*

**KEYWORDS.** Air pollution, factor analysis, suspended particulate matter, varimax rotation.

## INTRODUCTION

Air pollution is a global problem due to the process of urbanisation and rapid population growth. In Malaysia, air pollution has been a major problem over the past few years due to the rapid growth of the industrial sector, the increase in traffic volume and the expansion in petroleum and the timber sectors. In particular, air pollution in Kuala Lumpur has reached a critical level. This is found true in a few stretches of roads like Setapak Road and Sungei Besi Road (Juhaidi Yean Abdullah 1993), where the air pollution level has reached the danger zone.

On 11th August 1993, haze has appeared around the Klang Valley, in Southern Perak and the inner areas of Pahang. In Subang, visibility on that day was only two kilometers (Foo Yee Ping 1993). The occurrence of this haze was due to smoke emmitted from automobiles and industrial chimneys.

Painter (1974) has defined air pollution as the presence of dirty substances in the air that has reached a certain concentration level and has reached a duration which can harm living things, plants and animals in the environment.

Air pollutants can be classified into two categories: (I) primary pollutants, that is, pollutants which are emitted directly from certain sources and (II) secondary pollutants, that is, pollutants which are formed from a chemical process with other particulates from the air. As an example sulphur dioxide is a primary pollutant which is emitted by charcoal whereas ozone is a secondary pollutant which is the result of the reaction between nitrogen dioxide ($NO_2$) and oxygen ($O_2$) under the influence of the sun.

Air pollutants can also be classified into two categories according to their physical size as particulates and gas. A particulate is a small particle which can take the form of a solid or liquid. Plumbum (Pb), ferrum (Fe), cuprum (Cu), chromium (Cr), aluminium (Al) are examples of particulates. The size of a particulate is between $0.1\mu$ (micron) and $500\mu$. ($1\mu = 1/1000$ mm). Particulates of sizes less than $500\mu$ can be suspended in the air for a few seconds up to a few months. Particulates of between $0.1\mu$ and $1\mu$ can be suspended in the atmosphere for a longer time, and is called suspended particulate matter (SPM). An example of a suspended particulate matter is smoke which is the result of burning or photochemical aerosol process. Examples of particulates of sizes between $1\mu$ and $10\mu$ are dust and smog which are emitted by industrial areas. Particulates of sizes greater than $10\mu$ are the results of mechanical processes such as roads and soil erosions by wind.

The study of air pollution is concentrated mainly in the Klang Valley because of its rapid growth in population and industry. The aim of this study is to compare the density level of SPM and Pb in the Klang Valley from the year 1986 to 1992. Four areas which are classified as industrial, commercial, traffic, and residential areas were studied and a comparison of the level of air pollution due to SPM and Pb was carried out. Factor analysis by principal component method was also used to identify the source of air pollution due to heavy metals like Pb, Fe, Cu, Na, Ca, Al and Ni.

## DATA

The data were obtained from the air quality unit, the Department of Environment from the year 1986 to 1992. The area involved is the Klang Valley and the data were collected from four areas namely industrial, traffic, commercial and residential areas.

Suspended particulate matter (SPM) in the air was detected using an equipment known as high volume sampler. Using a tube, air from the atmosphere was sucked in and passed through papers. The weight of the particle was obtained by taking the difference between the weight of the paper before and after sampling. The unit used for measuring the density of SPM is $\mu g/m^3$.

## METHODOLOGY

The first part of the paper looks at the mean density level of SPM and Pb and comparisons were made between the four areas under study using the data from 1986 to 1992.

The second part of the paper uses the data from 1990 to 1992 to identify the source of air pollution due to heavy metals like Pb, Fe, Cu, Na, Ca, Al and Ni. Factor analysis was used for this study because there exists strong correlations between the heavy metals. For further details on factor analysis please refer to Johnson and Wichern (1992) or Chatfield and Collins (1994).

If there are $p$ variables $X_1, X_2, ..., X_p$ with mean vector and covariance matrix $\Sigma$, the factor analysis model postulates that $X$ has a linear relationship with the non-observed variables, $F_1, F_2, ..., F_m$ which is called the general factors and $\varepsilon_1, \varepsilon_2, ..., \varepsilon_p$ are the error terms.

The factor analysis model is :

$$\underset{(p \times 1)}{X-\mu} = \underset{(p \times m)(m \times 1)}{LF} + \underset{(p \times 1)}{\varepsilon} \qquad m<p$$

with $L$ being the loading matrix and $\varepsilon$ being the observed variance.

The specific variance can be written as

$$\mathrm{Var}(X_i) = \sigma^2 ii = l^2_{i1} + l^2_{i2} + ... \, l^2_{im}$$

with

$$h^2_i = l^2_{i1} + l^2_{i2} + ... \, l^2_{im}$$

The $i$th communality, $h^2_i$ is the sum of squares of loadings for variable i due to the m factors. To estimate the factor loadings, the principal component analysis was used. Varimax rotation was then used to obtain a clearer understanding of the parameters which form the factor loadings.

## RESULTS

Table 1 gives the mean density level for SPM and Pb from 1986-1992. From table 1, the highest level of SPM was recorded in 1991. This can be explained by the presence of haze during that year which resulted in more suspended particulates in the air. As for Pb, all the observed values are less than $1.2 \mu g/m^3$.

Table 1: Mean SPM and Pb density ($\mu g/m^3$)

| Year | SPM | Pb | % Pb in SPM |
|---|---|---|---|
| 1986 | 121.024 | 1.092 | 0.9 |
| 1987 | 105.764 | 0.867 | 0.8 |
| 1988 | 97.876 | 0.794 | 0.8 |
| 1989 | 113.191 | 1.135 | 1.0 |
| 1990 | 126.936 | 0.507 | 0.4 |
| 1991 | 132.368 | 0.337 | 0.25 |
| 1992 | 97.593 | 0.246 | 0.25 |

Table 2 gives the mean density level for Pb and SPM in the industrial, commercial, traffic and residential areas.

Table 2 : Mean density level for SPM and Pb by areas (1992)

| Area | SPM | Pb | % Pb in SPM |
|---|---|---|---|
| Industrial | 133.558 | 0.292 | 0.2 |
| Commercial | 118.522 | 0.359 | 0.3 |
| Traffic | 135.821 | 0.397 | 0.3 |
| Residential | 82.339 | 0.141 | 0.17 |

Three areas namely the industrial, commercial and traffic areas each has mean density level for SPM greater than the residential area. For the mean density of Pb, all four areas have small values.

An analysis of variance test was carried out to test whether there exist differences in the four main areas. The result of the test shows that there is a significant difference among the four main areas. ( $p$ value is 0.000). Therefore there exists at least two means which are different. To identify the means which are significantly different, a contrast test was carried out.

All the contrasts are significant except for the industrial and traffic areas ($p$ value is 0.5100). Thus there exist differences in mean density of SPM between commercial, residential and industrial/traffic areas. There is no difference between industrial and traffic areas.

The second part of the analysis looks at the data from 1990-1992 which have a complete data on SPM, Pb, Fe, Cu, Na, Ca, Al and Ni. The correlation matrix is given below.

Table 3 : Correlation matrix for the eight heavy metals

| | SPM | Pb | Fe | Cu | Na | Ca | Al | Ni |
|---|---|---|---|---|---|---|---|---|
| **SPM** | 1.0 | 0.469 | 0.514 | 0.022 | 0.029 | 0.187 | -0.051 | 0.027 |
| **Pb** | | 1.0 | 0.471 | 0.156 | 0.188 | 0.272 | 0.167 | 0.021 |
| **Fe** | | | 1.0 | 0.212 | 0.244 | 0.501 | 0.179 | 0.040 |
| **Cu** | | | | 1.0 | 0.287 | 0.220 | 0.364 | -0.024 |
| **Na** | | | | | 1.0 | 0.800 | 0.794 | 0.077 |
| **Ca** | | | | | | 1.0 | 0.499 | 0.125 |
| **Al** | | | | | | | 1.0 | 0.046 |
| **Ni** | | | | | | | | 1.0 |

From the Table 3, we can see that there is a high correlation between Na and Ca and between Na and Al. There is a fair correlation between other variables. Since there are at least two variables with high correlations, factor analysis with varimax rotation was used. Three factors with eigenvalues greater than 1 were included in the model. These three factors can explain about 71.2% of the variation in the sample. Table 4 gives the estimated values for the factor loadings, eigenvalues, communalities and the specific variance for each variable.

Table 4 : Factor analysis after varimax rotation.

| | **Estimated factor loadings** | | | | |
|---|---|---|---|---|---|
| **Variables** | 1 | 2 | 3 | Communality | Specific variance |
| Na | 0.94 | 0.08 | 0.07 | 0.891 | 0.109 |
| Al | 0.88 | -0.03 | -0.06 | 0.776 | 0.224 |
| Ca | 0.78 | 0.34 | 0.18 | 0.754 | 0.246 |
| Cu | 0.49 | 0.11 | -0.40 | 0.411 | 0.589 |
| SPM | -0.10 | 0.85 | 0.04 | 0.726 | 0.274 |
| Fe | 0.26 | 0.80 | 0.01 | 0.701 | 0.299 |
| Pb | 0.15 | 0.76 | -0.06 | 0.603 | 0.397 |
| Ni | 0.10 | 0.02 | 0.91 | 0.836 | 0.164 |
| Eigenvalues | 2.605 | 2.061 | 1.032 | | |

From Table 4, Na, Al and Ca explains the first factor, SPM, Fe and Pb forms the second factor and Ni being the third factor.

## CONCLUSION

The conclusion from the first part of the analysis is as follows. Air pollution due to SPM is quite high ( greater than $100\mu/m^3$) except for 1988 and 1992. A comparison between areas showed that the mean SPM for industrial, traffic and commercial areas are higher than the residential area. The level of Pb in the air is still low for all the four areas.

For the second part of the analysis, we can see that there are three factors which explain about 71.2% of the population variance. Na, Al and Ca can be grouped into the first factor with a variation of 32.6%; SPM, Pb and Fe can be grouped into the second factor with a variation of 25.7%; and Ni can be grouped into the third factor with a variation of 12.9%. The first factor which consists of Na, Al and Ca can be called the composite source. Pollution from Na comes from salt which is the by product of the food industry or from the sea. Al and Ca is the result of smoke from the cement industry which uses the mixture of Ca, Al and Fe. The second factor can be called the traffic factor. Fe and Pb is the result of the fuel burning process. The third factor can be classed as the pollution due to industrial sources which uses nickel. As an example, fuel burning which uses nickel as additive in engines and smelting of nickel products.

## ACKNOWLEDGMENT

The authors would like to thank the Department of Environment, Ministry of Science, Technology and the Environment Malaysia for granting permission to use the data.

## REFERENCES

Chatfield, C. and Collins, A.J. 1994. *Pengenalan Kepada Analisis Multivariat (terjemahan oleh Ahmad Shukri Yahaya dan Noormala Mohd Yusoff)*. Kuala Lumpur : Dewan Bahasa dan Pustaka.

Foo Yee Ping. 1993. Dry Weather Causes Haze In Klang Valley. In ***The Star.*** 12 August 1993.

Johnson, R.A. and Wichern, D.W. 1992. *Applied Multivariate Analysis.* New York : Prentice-Hall.

Juhaidi Yean Abdullah. 1993. Traffic Policemen In Danger Zone. *New Straits Times. 26 September 1993.*

Painter, D.E. 1974. *Air Pollution Technology.* Virginia : Reston.