Malaysian Journal of Environmental Management 7 (2006): 33 - 41

Air Quality Predictions Using Log Normal Distribution Functions of Particulate Matter in Kuala Lumpur

JANNATUL NAEMAH MOHAMED SEDEK, NOR AZAM RAMLI & AHMAD SHUKRI YAHAYA

ABSTRACT

Malaysia's capital city of Kuala Lumpur has been developed from a small settlement at the confluence of the Klang and Gombak rivers into a vibrant and bustling metropolis. The associated rapid economic growths have started to impose costs in terms of industrial pollution and the degradation of urban environment. Air pollution is a major issue in this city that has been affecting human health, agricultural crops, forest species, and ecosystems. This paper reports findings of a research aimed at developing a tool needed to analyze the statistical characters of air quality data in Kuala Lumpur, Malaysia. In this study, the theoretical distribution Log Normal was used to fit the parent distribution of PM_{10} for 1998 and 2002 for Kuala Lumpur. Two estimating methods namely method of maximum likelihood and method of moments, were used to estimate the parameters of the theoretic distributions. From the statistical properties of air pollutants, the probabilities of air pollutant concentration exceeding Malaysia Ambient Quality Standards can be predicted. The results of this work provide useful information for air quality management, and can also be used to develop the corresponding air pollution control strategy.

ABSTRAK

Bandar utama Malaysia, Kuala Lumpur, telah berkembang daripada sebuah petempatan kecil di pertemuan Sungai Klang dan Sungai Gombak kepada sebuah bandar metropolis yang sibuk dan maju. Pertumbuhan ekonomi yang pesat telah mula membebankan bandar ini dengan kos dalam bentuk pencemaran industri dan kehakisan alam sekitar bandar. Pencemaran udara ialah isu penting di bandar ini yang menjejaskan kesihatan manusia, tanaman, sepsis hutan dan ekosistem. Kertas ini melaporkan dapatan satu penyelidikan bertujuan membangunkan satu kaedah untuk menganalisis ciri-ciri statistik kualiti udara Kuala Lumpur. Dalam kajian ini, taburan teoretis Log Normal telah digunakan untuk menyuaikan taburan paren PM_{10} untuk tempoh 1998 ke 2002 bagi Kuala Lumpur. Dua kaedah penganggaran iaitu kaedah kehampiran maksimum dan kaedah momen telah digunakan untuk menganggar parameter-

parameter taburan teoretis itu. Ciri-ciri statistik bahan pencemaran udara digunakan untuk meramalkan kebarangkalian kepekatan bahan pencemaran melebihi Piawaian Kualiti Ambien Malaysia. Dapatan kajian ini menyediakan maklumat yang berguna untuk pengurusan kualiti udara dan boleh digunakan untuk membangunkan strategi kawalan pencemaran udara.

INTRODUCTION

Air pollution is defined as the presence in the atmosphere of one or more pollutants in such quantities and of such duration as may tend to be injurious to human, plant, or animal life, or to property, or which may unreasonably interfere with the comfortable enjoyment of life or property, or the conduct of business (Canter 1996). Among different environmental pollution problems, air pollution is reported to cause the greatest damage to health and loss of welfare from environmental causes in Asian countries (Hughes 1997). Although data and study on air pollution in Malaysia are very limited, one review that had been done by Afroz et al. (2003) indicated that suspended particulate matter (SPM) and nitrogen dioxide (NO_2) are among the predominant pollutants.

The impact of air pollution is broad, especially for human being where it can cause several significant effects (including carcinogenic effects). State-of-the-art epidemiological research has found strong links between air pollution and morbidity and mortality rates, with air pollution can cause a wide range of human cardio respiratory health problems especially in children, old, and ill people. Such health problems include cardiac arrhythmias, reducing lung function, asthma, chronic bronchitis, and increasing respiratory symptoms, such as sinusitis, sore throat, dry and wet cough, and hay fever (World Health Organization 1998).

There are several form of distributions used to fit the parent distribution of air pollutant concentration data. In fact, several frequency distribution functions have been proposed in literature and used to fit air quality measurements. These distributions are log normal distribution (Mage & Ott 1984) Weibull distribution (Georgopoulos & Seinfeld 1982) and type V Pearson distribution.

The lognormal distribution was more widely used to represent the type of air pollutant concentration distribution. This distribution was not only used for ambient air quality data but also for indoor air quality and dissolved solids in groundwater (Ott 1990). From the study that has been done by Hadley and Toumi (2003), the two parameter log normal distribution can be a very good description of annual mean daily sulphur dioxide concentrations for a wide range of ambient levels, time periods and monitoring site types. The log normal distribution has a consistently better fit to the data than the normal distribution.

There are many techniques to estimate the distribution parameters, namely the method of moments, percentiles and maximum likelihood estimation (MLE) (Georgopoulos & Seinfeld 1982). The method of moments was more widely used whereas the method of maximum likelihood provides the best estimate of the parameters (Mage & Ott 1984).

DATA AND METHODS

In this paper, the PM_{10} hourly concentration in Kuala Lumpur, were taken to estimate the parameters of the log normal distributions. These sets of data were taken over a one year period for year 1998. For PM_{10} data set in 1998, there were 80 missing values. Missing values were replaced as they can affect the result in statistical analysis. The methods used to replace the missing values in the data set were: all missing values were filled with average of all data available above the missing values; all missing values were filled with average of all data available above and below missing values.

The Log Normal Distribution

The probability density function (pdf) for the two parameter log normal distribution is expressed by;

$$f(x;\alpha,\beta)=\frac{1}{x\beta\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{\ln(x)-\alpha}{\beta}\right)^2\right],\alpha>0 \text{ and } \beta>0,\ x>0 \tag{1}$$

The cumulative distribution function is given by the following equation;

$$F(x;\alpha,\beta)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\frac{\ln x-\alpha}{\beta}} e^{-\frac{x^2}{2}}dx,\alpha>0 \text{ and } \beta>0,\ \ x>0 \tag{2}$$

where α is the scale parameter and β is the shape parameter.

Parameter Estimation

In this study, two methods were used to estimates α and β for the log normal distribution.

The method of maximum likelihood

For log-normal distribution, the likelihood function is defined as;

$$L = \prod_{i=1}^{n} f(x;\alpha,\beta) = \prod_{i=1}^{n} \frac{1}{x\beta\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\ln(x)-\alpha}{\beta}\right)^2\right]$$

$$= \beta^{-n}(2\pi)^{-\frac{n}{2}} \prod_{i=1}^{n} x_i^{-1} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{\ln(x_i)-\alpha}{\beta}\right)^2\right] \qquad (3)$$

and the log-likelihood function is;

$$\ln L = -n\ln\beta - \frac{n}{2}\ln(2\pi) - \ln\left(\prod_{i=1}^{n} x_i\right) - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{\ln x_i - \alpha}{\beta}\right)^2 \qquad (4)$$

Therefore, α and β are solutions of the derivatives of $\ln L$ with respect to α and β

$$\frac{\partial \ln L}{\partial \alpha} = 0 \text{ and } \frac{\partial \ln L}{\partial \beta} = 0$$

So, α is obtained as the solution of;

$$\alpha = \frac{1}{n}\sum_{i=1}^{n} \ln(x_i) \qquad (5)$$

and,

$$\beta = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[\ln(x_i) - \hat{\alpha}\right]^2} \qquad (6)$$

The method of moments

The moment estimates of the parameter of the log normal distribution are given by;

$$\mu = 2\ln\mu'_1 - \frac{1}{2}\ln\mu'_2 \qquad (7)$$

and;

$$\sigma^2 = \ln \mu'_2 - 2\ln \mu'_1 \quad (8)$$

where; $\mu'_1 = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $\mu'_2 = \frac{1}{n}\sum_{i=1}^{n} x_i^{\,2}$

Then, by substituting sample estimates, $\bar{x}$ and s^2 for the mean and variance, respectively, the parameter of the log-normal distribution can be obtained;

$$\alpha = \sqrt{\ln\left[\mu_2 + (\mu'_1)^2\right] - 2\ln(\mu'_1)} \quad (9)$$

with $\beta = \ln(\mu'_1) - \frac{\alpha^2}{2}$ (10)

where μ'_1 and μ_2 are the mean and variance of data;

$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ and $s^2 = \sum_{i=1}^{n} \frac{\left(x_i - \bar{x}\right)^2}{\left(n-1\right)}$

RESULTS AND DISCUSSION

Estimation of α and β Values

The α and β values derived from data for PM_{10} in 1998 are given in Table 1. From the Table, it can be seen that even though there were two different methods used for replacing missing values, the values obtained for α and β, are quite similar.

Comparison of Two Parametric Estimating Methods

Figure 1 show an illustration of cdf plot for log-normal distribution using Alpha (A) and Beta (A) for PM_{10} in 1998. with different methods of parameter estimation. Two methods, maximum likelihood method and method of moments were used to estimate parameters for the log-normal distribution. The values of estimated parameter are given in Table 1. For log-normal distribution using PM_{10} data in 1998, there is not much difference between MLE and method of moments for estimating the parameter.

Table 1. Estimation of α and β values for PM_{10} in 1998

Parameter Estimation	Data			
	Alpha (A)	Alpha (B)	Beta (A)	Beta (B)
MLE	4.137	4.135	0.452	0.453
Moments	4.140	4.140	0.430	0.431

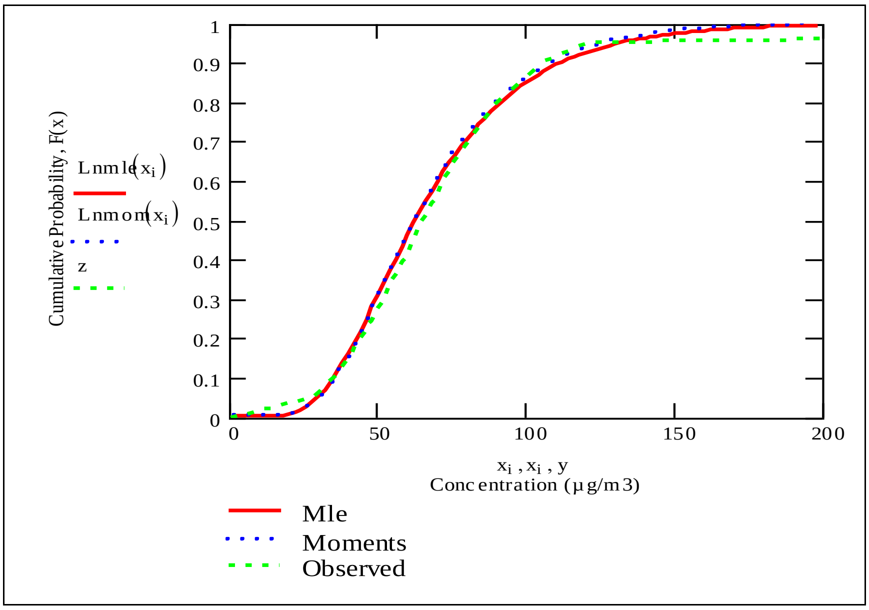

Figure 1. CDF plot for log-normal distribution with comparison between maximum likelihood method, method of moments and observed data (PM_{10} in 1998)

Prediction of the Probability Exceeding the Malaysia Ambient Quality Standards

Figures 2a and b show the pdf for PM_{10} concentration in Kuala Lumpur for the year of 1998. The values of estimated parameters are $\alpha = 4.140$, $\beta = 0.430$ for PM_{10} in 1998 using method of moments. Even though the pdf plot for PM_{10} in 1998 showed approximately similar distribution, the distribution using Method of Moment gives better fit compared to the

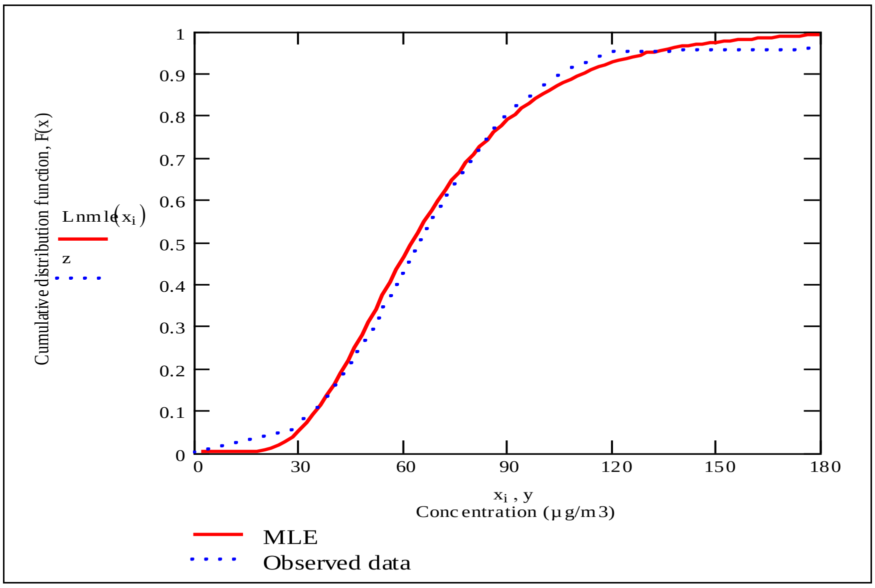

Figure 2a. CDF plot for log-normal distribution with comparison between maximum likelihood method and observed data (PM_{10} in 1998)

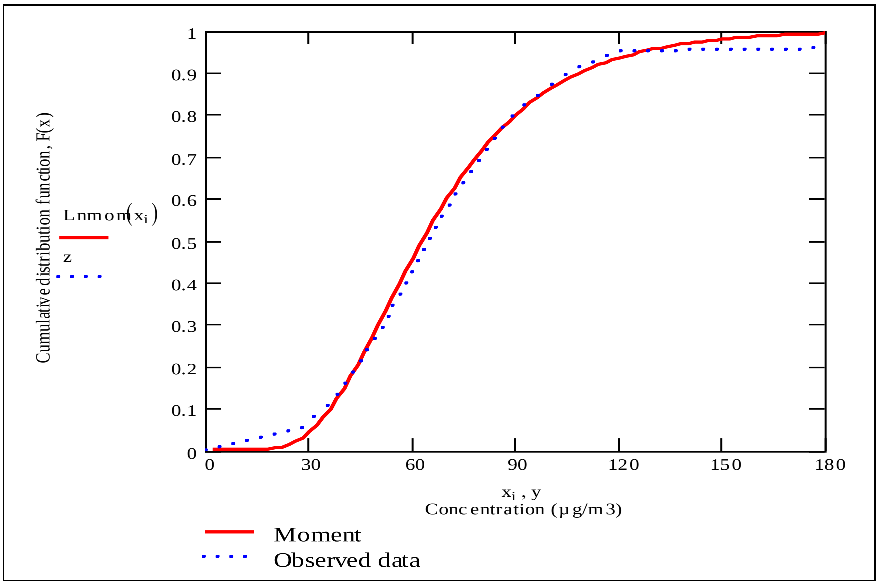

Figure 2b. CDF plot for log-normal distribution with comparison between method of moments and observed data (PM_{10} in 1998)

MLE especially at higher PM_{10} concentrations. From Figures 2a & b, it is found that the probabilities for PM_{10} concentration in 1998 exceeding Malaysia Ambient Quality Standard (>150 μg/m^3) are 0.027. It means that the number of days for PM_{10} exceeding the Malaysia Ambient Quality Standards for 1999 is 10.

CONCLUSION

The result of this work shows that the PM_{10} concentration distribution can be represented by log-normal distribution. Between the two parametric estimating methods, there is not much different. For that method of moments with $\alpha = 4.140$, $\beta = 0.430$ for PM_{10} in 1998. The probabilities exceeding Malaysia Ambient Quality Standards are predicted by the log-normal distribution for PM_{10} concentration. The probability exceeding Malaysia Ambient Quality Standard for PM_{10} concentration in 1998 is 0.027. It means that the number of days for PM_{10} exceeding the Malaysia Ambient Quality Standards for 1998 would be 10 days.

REFERENCES

Afroz, R., Hassan, M.N. & Ibrahim, N.A. 2003. Review of air pollution and health impacts in Malaysia. *Environment Research* 92(2): 71-77.

Canter, L.W. 1996. *Environmental impact assessment.* 2nd. Edition. Singapore: McGraw Hill.

Georgopoulos, P.G. & Seinfeld, J.H. 1982. Statistical distributions of air quality concentrations. *Environmental Science and Technology* 16: 401A-416A.

Hadley, A. & Toumi, R. 2003. Assessing changes to the probability distribution of sulphur dioxide in the UK using a lognormal model. *Atmospheric Environment* 37: 1461-1474.

Hughes, G. 1997. *Can the environment wait? Priority issues for East Asia*. Washington: World Bank.

Mage, D.T. & Ott, W.R. 1984. An evaluation of the method of fractals, moments and maximum likelihood for estimating parameters when sampling air quality data from a stationary lognormal distribution. *Atmospheric Environment* 18:163-171.

Ott, W.R. 1990. A physical explanation of the log-normality of pollutant concentrations. *Journal of the Air and Waste Management Association* 40: 1378-1383.

World Health Organization. 1998. Report of the Bioregional Workshop on Health Impacts of Haze Related Air Pollution. Manila: WHO.

School of Civil Engineering,
Engineering Campus,
Universiti Sains Malaysia,
14300 USM, Nibong Tebal, Penang, MALAYSIA.