Journal of Applied Sciences

ISSN 1812-5654

Journal of Applied Sciences 13 (2): 294-300, 2013
ISSN 1812-5654 / DOI: 10.3923/jas.2013.294.300

Finding the Best Statistical Distribution Model in PM_{10} Concentration Modeling by using Lognormal Distribution

[1]Hazrul Abdul Hamid, [2]Ahmad Shukri Yahaya, [2]Nor Azam Ramli and [2,3]Ahmad Zia Ul-Saufie
[1]Department of Mathematics, Penang Matriculation College, 13200, Kepala Batas, Penang, Malaysia
[2]Clean Air Research Group, School of Civil Engineering, Universiti Sains Malaysia, (Engineering Campus), 14300, Nibong Tebal, Penang, Malaysia
[3]Faculty of Computer and Mathematical Sciences, Universiti Teknologi Mara, 13500, Permatang Pauh, Pulau Pinang, Malaysia

Abstract: Air pollution is one of the most important issues that are often discussed, it is important to carry out the study on air pollution modeling. Air pollution models play an important role and very useful because it can help local authorities to carry out suitable action to reduce the impact of air pollution. Finding the best model would allow prediction to be made accurately. Statistical distribution modeling plays an important role in predicting air pollutant concentration. Lognormal distribution is one of the distributions that widely used in environmental engineering. One of the important steps in statistical distribution modeling is parameter estimation. There are several methods can be used to estimate the parameter in fitting distribution for air pollutant concentration data. This research compared the performance of parameter estimator for two-parameter and three-parameter lognormal distribution by using PM_{10} concentration in Nilai, Negeri Sembilan, Malaysia. Two methods were used to estimate the parameters in this study which is method of moments and method of probability weighted moments. Five performance indicators are used to determine the best estimator and the best distribution to represent the PM_{10} concentration in Nilai, Negeri Sembilan from 2003 to 2009. Results show that three-parameter lognormal distribution performs better compared to two-parameter lognormal distribution.

Key words: Lognormal distribution, PM_{10}, performance indicator, prediction model, exceedences

INTRODUCTION

The impact of air pollution is noticeable, especially for human health. There are several health effects that correlate to air pollution such as asthma, chronic bronchitis and increasing respiratory symptoms, such as sinusitis, sore throat, dry and wet cough and hay fever (WHO, 1998). An issue of great concern has been the detrimental effect of low air quality onto human health, chronically or acutely. Since air pollution can be a major problem especially to human health, air quality monitoring should be done continuously.

There are many sources of air pollution such as mobile sources, stationary sources and open burning sources (Afroz *et al*., 2003). Mobile sources include personal vehicles, commercial vehicles and motorcycles. Stationary sources refer to factory and industry, power stations, industrial fuel burning processes and domestic fuel burning while open burning sources refer to burning of solid wastes and forest fires. In Malaysia, there are 52 monitoring locations throughout the country that belong to the Department of Environment (Department of Environment Malaysia, 2010). The parameters monitored include Particulate Matter (PM_{10}), Sulphur Dioxide (SO_2) and several airborne heavy metals. Three major sources of air pollution in Malaysia are mobile sources, stationary sources and open burning sources. However, for the past few years, emissions from mobile sources have been the major sources of air pollution which contribute around 70 to 75% of total air pollution in Malaysia (Afroz *et al*., 2003). Malaysia Ambient Air Quality Guidelines state that the 24 h mean for PM_{10} concentration should not exceed 150 $\mu g\ m^{-3}$.

The probability density function of concentration in an atmospheric plume is an important quantity used to describe and discuss environmental diffusion (Yee and Chan, 1997). The concentrations of air pollutants are usually correlated with the emission levels and meteorological conditions. When the parent probability distribution of air pollutants is correctly chosen, the specific distribution can be used to predict the mean concentration and probability of exceeding a critical

Corresponding Author: Hazrul Abdul Hamid, Department of Mathematics, Penang Matriculation College, 13200, Kepala Batas, Penang, Malaysia Tel: + (60)4-5996227 Fax: + (60)4-5941009

concentration (Lu and Fang, 2003). Selecting appropriate probability models for the data is an important step in environmental data analysis. These probability models may become the basis for estimating the parameters to meet the evolving information needs of environmental quality management. The developed models also can be easily implemented for public health protection by providing early warning to the respective population (Ul-Saufie *et al*., 2012).

Lognormal distribution have been used extensively in atmospheric sciences to describe phenomena that take on non-negative values, such as storm, daily and longer-period rain, snow and hail amounts, particle size distributions, pollutant concentrations, cloud dimensions, air velocity fluctuations, flood frequencies and radio wave amplitude fluctuations. In most cases these are empirically observed and tested fits and often other models such as the gamma distribution, are also fitted or at least cannot be excluded as being consistent with the data (Lopez, 1977).

In Malaysia, lognormal distribution is the best distribution to represent the PM_{10} concentration in the residential area (Sansuddin *et al*., 2011). Lognormal distribution was also used to fit the PM_{10} concentration in one of the industrial area in Malaysia which is Seberang Perai and result shows that the lognormal distribution agrees with the data in several years (Yusof *et al*., 2010). In environmental engineering, one of the great concerns is return period. In this study, return period is an estimate of the interval of time where the high particulate event occurs.

Particulate matter enters the body when we breathe. Large particles can be trapped in nose and throat and are removed when we cough or sneeze. In some areas, particulate matter can be very heavy because of high levels of industrial activity. This study will be concerned about coarse particles that are 10 micrometers in aerodynamic diameter (PM_{10}) or smaller because those are the particles that generally pass through the throat and nose and enter the lungs. Once inhaled, these particles can affect the heart and lungs and cause serious health effects. This research compared the performance of parameter estimator for two-parameter and three-parameter lognormal distribution by using PM_{10} concentration in Nilai, Negeri Sembilan, Malaysia and hence find the best distribution to represent the PM_{10} concentration in Nilai, Negeri Sembilan.

MATERIALS AND METHODS

Study area: Nilai is a town located in Negeri Sembilan and can be classified as an industrial area. Geographically located at latitude 2°45'N of the equator and longitude 102°15'E of the prime meridian, Nilai is a rapidly growing town due to its proximity and easy connection to Kuala Lumpur. The data used in this study is hourly PM_{10} concentration in Nilai, Negeri Sembilan taken from the year 2003 to 2009. Missing values were replaced using mean top bottom method where the data were filled with the average of data available above and below the missing values (Yahaya *et al*., 2005).

Lognormal distribution: Equation 1 shows the probability density function for the two parameter lognormal distribution (Evans *et al*., 2000):

$$f(x) = \left(\frac{1}{x\lambda\sqrt{2\pi}}\right)\exp\left\{-\frac{1}{2}\left(\frac{\ln(x)-\sigma}{\lambda}\right)^2\right\} \tag{1}$$

where, $x \geq 0$, λ represent a scale parameter and σ represent a location parameter for annual measurement of particular sites.

For three parameter lognormal distribution, probability density function given by Johnson *et al*. (1994) is as follows:

$$f(x) = \left(\frac{1}{(x-\delta)\lambda\sqrt{2\pi}}\right)\exp\left\{-\frac{1}{2}\left(\frac{\ln(x-\delta)-\sigma}{\lambda}\right)^2\right\} \tag{2}$$

where, $x>\delta$; $-\infty<\sigma<\infty$; $\lambda>0$. λ represent the scale parameter, σ represent the location parameter and δ represent the threshold parameter.

Parameter estimation: The lognormal distribution model was used to fit the hourly PM_{10} concentration observed in Nilai. There are several methods can be used to estimate the parameter of the lognormal distribution. This study only focuses on the method of moment and method of probability weighed moment to estimate the parameter. To estimate the parameters λ and σ of the lognormal distribution by using the method of moments, the formula given by Evans *et al*. (2000) is as follows:

$$\sigma = \sqrt{\ln\left[s^2 + \left(\bar{x}\right)^2\right] - 2\ln\left(\bar{x}\right)} \tag{3}$$

$$\lambda = \ln\left(\bar{x}\right) - \frac{\sigma^2}{2} \tag{4}$$

The following equations are given by Hosking (1990) where erf(x) is the error function:

$$\lambda_1 = e^{\mu+\sigma^2/2} \tag{5}$$

$$\lambda_2 = e^{\mu+\sigma^2/2}\,\mathrm{erf}(\sigma/2) \tag{6}$$

Where:

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-u^2}du = 2F\left(x\sqrt{2}\right)-1 \tag{7}$$

and F(.) is the normal distribution function which can be evaluated using the following equations:

$$F(u) = \int_{-\infty}^{u}\frac{1}{\sqrt{2\pi}}e^{-t^2/2}dt \tag{8}$$

Equation 5 and 6 can be solved for μ and σ by replacing λ_1 and λ_2 by the sample L-moments l_1 and l_2 to give the following equations:

$$\sigma = 2\mathrm{erf}^{-1}\left(\frac{l_2}{l_1}\right) \tag{9}$$

$$\lambda = \log(l_1) - \left(\frac{\sigma^2}{2}\right) \tag{10}$$

For the three-parameter lognormal distribution, the first two moments of the lognormal distribution are given by Kite (1977) as follows:

$$\mu'_1 = a + e^{\mu+\sigma^2/2} \tag{11}$$

$$\mu_2 = \left(e^{\sigma^2}-1\right)e^{2\mu+\sigma^2} \tag{12}$$

the coefficient of variation of (x-α), z_2, is also given by Kite (1977) as below:

$$z_2 = \frac{1-w^{2/3}}{w^{1/3}} \tag{13}$$

where, w is defined by:

$$w = \frac{-\gamma_1 + \left(\gamma_1^2+4\right)^{1/2}}{2} \tag{14}$$

where, γ_1 in Eq. 12 is the coefficient of skewness of the original variable x. The values of the parameters after replacing μ'_1, μ_2 and γ_1 by their sample estimates m'_1, m_2, and g_1 are given by:

$$\sigma = \left[\log\left(z_2^2+1\right)\right]^{\frac{1}{2}} \tag{15}$$

$$\lambda = \log\left(\frac{\sqrt{m_2}}{z_2}\right) - \frac{1}{2}\log\left(z_2^2+1\right) \tag{16}$$

$$\delta = m'_1 - \frac{\sqrt{m_2}}{z_2} \tag{17}$$

An approximate solution for parameter estimates by using probability weighted moment given by Hosking (1990) is as follows:

$$\sigma = 0.999281z - 0.006118z^3 + 0.000127z^5 \tag{18}$$

$$\lambda = \log\left[\frac{l_2}{\mathrm{erf}\left(\frac{\sigma}{2}\right)}\right] - \frac{\sigma^2}{2} \tag{19}$$

$$\delta = l_1 - e^{\mu+\sigma^2/2} \tag{20}$$

Where:

$$\lambda_r = r^{-1}\sum_{j=0}^{r-1}(-1)^j\binom{r-1}{j}E\left(X_{r-j:r}\right) \tag{21}$$

$$z = \sqrt{\frac{8}{3}}\,\phi^{-1}\left(\frac{1+\tau_3}{2}\right) \tag{22}$$

$$l_1 = \bar{x} \tag{23}$$

$$l_2 = 2\lambda_1 - l_1 \tag{24}$$

$$\tau_3 = \frac{l_3}{l_2} \tag{25}$$

λ_r is the r-th L-moment of the random variable X, l_1 and l_2 are the first and second sample L-moment.

Performance indicator: Five performance indicators are used to determine the best estimator. The root mean square error (RMSE) summarizes the difference between the observed and imputed concentrations and is used to provide the average error (Junninen *et al*., 2004). It is defined as:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(P_i - O_i\right)^2} \tag{26}$$

where, N is the number of imputations, O_i is the observed data point and P_i is the imputed data point.

The root mean square error method is the most common indicator. For a good estimator, the RMSE value must approach zero. Therefore, a smaller RMSE value means that the model is more appropriate (Junninen *et al*., 2004).

The Normalized Absolute Error (NAE) is a more sensitive measure of residual error than RMSE (Junninen *et al*., 2004). It is defined as:

$$\mathrm{NAE} = \frac{\sum_{i=1}^{N} \left| P_i - O_i \right|}{\sum_{i=1}^{N} O_i} \quad (27)$$

where, N is the number of imputations, O_i is the observed data point, P_i is the imputed data point. A small value for the normalized absolute error means that the model is appropriate (Junninen *et al*., 2004).

Index of agreement is define as follow (Junninen *et al*., 2004):

$$\mathrm{IA} = 1 - \left[\frac{\sum_{i=1}^{N} (P_i - O_i)^2}{\sum_{i=1}^{N} \left(\left| P_i - \bar{O} \right| + \left| O_i - \bar{O} \right| \right)^2} \right]^2 \quad (28)$$

where, N is the number of imputations, O_i is the observed data point, P_i is the imputed data point. For a good estimator, the RMSE value must approach the value one (Junninen *et al*., 2004).

The coefficient of determination is define as follow (Junninen *et al*., 2004):

$$\mathrm{CD} = \left[\frac{1}{N} \frac{\sum_{i=1}^{N} (P_i - \bar{P})(O_i - \bar{O})}{\sigma_P \sigma_O} \right]^2 \quad (29)$$

where, N is the number of imputations, O_i is the observed data point, P_i is the imputed data point. For a good estimator, the RMSE value must approach the value one (Junninen *et al*., 2004).

The prediction accuracy:

$$\mathrm{PA} = \frac{1}{N-1} \sum_{i=1}^{N} \frac{(P_i - \bar{P})(O_i - \bar{O})}{\sigma_P \sigma_O} \quad (30)$$

where, N is the number of imputations, O_i is the observed data point, P_i is the imputed data point. For a good estimator, the RMSE value must approach the value one (Junninen *et al*., 2004).

RESULTS AND DISCUSSION

Box plot for hourly PM_{10} concentration are shown in Fig. 1. As a simple graphical display, box plot is very ideal for comparisons. The maximum PM_{10} concentrations exceed the limit (150 μg m^{-3}) for all year with the maximum reading is 542 μg m^{-3} in 2005. Table 1 shows the box plots and also the descriptive statistics for hourly PM_{10} concentration in Nilai, Negeri Sembilan from 2003 to 2009, respectively. The mean for each year are higher than the median, which indicate the pollutants data are skewed to the right. The highest skewness is 3.59, shows that 2005 experienced high particulate events. This is likely due to the haze event that occurred that year as an effect of the transboundary movement of air pollutants emitted from forest fires and open burning activities in Indonesia (Department of Environment Malaysia, 2005). The smoke from biomass burning from regional sources also contribute to the PM_{10} concentration reading especially during dry season (Juneng *et al*., 2009). In addition, since Nilai is an industrial area, neighbouring precursory emissions which occur as a result of local societal and also industrial development also contribute to the variations of PM_{10} concentration (Juneng *et al*., 2011).

Two-parameter lognormal distribution and three-parameter lognormal distribution were used to fit the PM_{10} concentration in this study. The parameters for the distributions were estimated using method of moments and method of probability weighted moments. The best estimator and distribution was selected according to the performance of five types of goodness-of-fit criteria: Normalized Absolute Error (NAE), Prediction Accuracy (PA), coefficient of determination (R^2), Root Mean Square Error (RMSE) and index of agreement (IA). All these goodness-of-fit criteria were used to describe how well

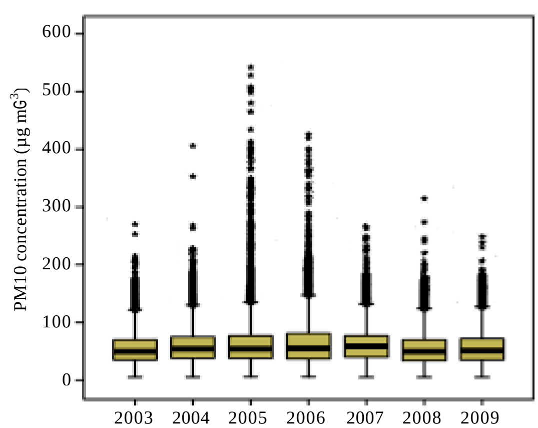

Fig. 1: Characteristic for the PM_{10} data in μg m^{-3} for the monitoring site in Nilai, Negeri Sembilan

Table 1: Descriptive statistics for PM_{10} concentration for Nilai

Parameter	2003	2004	2005	2006	2007	2008	2009
Mean	54.88	59.75	63.25	63.54	62.89	55.28	55.95
St. dev.	27.01	29.99	44.21	39.19	30.91	28.15	28.66
Median	50	54	54	55	59	50	51
Mode	33	36	40	44	73	43	42
Min.	5	5	6	6	5	5	5
Max.	270	406	542	426	266	315	249
Skewness	1.09	1.34	3.59	2.32	1.04	1.28	1.05
Kurtosis	2.25	4.95	21.72	10.93	1.98	3.22	1.63

Table 2: Performance Indicators Value for Nilai, Negeri Sembilan

Year	PI	2-parameter (MoM)	2- parameter (PWM)	3-parameter(MoM)	3-parameter(PWM)	Best estimator
2003	NAE	0.032046	0.031600	0.009883	0.009209	Probability Weighted Moments
	PA	0.996461	0.995595	0.998514	0.998585	(3-parameter)
	R^2	0.992707	0.990984	0.996803	0.996945	
	RMSE	2.337117	2.606909	1.531893	1.470031	
	IA	0.998080	0.997707	0.999181	0.999249	
2004	NAE	0.027120	0.028303	0.012727	0.008598	Probability Weighted Moments
	PA	0.994439	0.993976	0.995617	0.995408	(3-parameter)
	R^2	0.988683	0.987763	0.991028	0.990612	
	RMSE	3.209454	3.312160	2.847134	2.916620	
	IA	0.997060	0.996962	0.997697	0.997580	
2005	NAE	0.078484	0.053836	0.065684	0.053519	Probability Weighted Moments
	PA	0.973204	0.966612	0.985615	0.987333	(3-parameter)
	R^2	0.946910	0.934125	0.971214	0.979086	
	RMSE	10.181531	12.200349	12.729302	10.740664	
	IA	0.986034	0.977720	0.981836	0.983144	
2006	NAE	0.023538	0.022622	0.030302	0.017908	Probability Weighted Moments
	PA	0.992667	0.992489	0.993718	0.997014	(3-parameter)
	R^2	0.985163	0.984810	0.987249	0.989210	
	RMSE	4.834828	4.979325	6.541765	5.599359	
	IA	0.996059	0.995776	0.986509	0.984591	
2007	NAE	0.038488	0.037838	0.021642	0.022255	Probability Weighted Moments
	PA	0.995095	0.994085	0.997726	0.997570	(3-parameter)
	R^2	0.989988	0.987980	0.995229	0.994918	
	RMSE	3.114636	3.455364	2.128888	2.201656	
	IA	0.997394	0.996929	0.998793	0.998708	
2008	NAE	0.026791	0.024975	0.245043	0.012269	Probability Weighted Moments
	PA	0.996986	0.996497	0.982338	0.997860	(3-parameter)
	R^2	0.993212	0.992986	0.241133	0.995511	
	RMSE	2.269884	2.389292	19.07739	1.866611	
	IA	0.998331	0.998213	0.749677	0.998885	
2009	NAE	0.035669	0.033937	0.268758	0.010448	Probability Weighted Moments
	PA	0.991347	0.994743	0.984230	0.999240	(3-parameter)
	R^2	0.988311	0.989215	0.177483	0.998397	
	RMSE	2.611838	3.069184	20.57303	1.139683	
	IA	0.996066	0.997198	0.698608	0.999601	

the distribution fits a set of observations. Table 2 shows the value of each performance indicator for each estimator. The best distribution representing each year can be identified based on this goodness-of-fit criteria. From the goodness-of-fit criteria, three-parameter lognormal distribution fits the PM_{10} concentration better compared to two-parameter lognormal distribution for every year from 2003 to 2009. These finding supported by Taylor *et al.* (1986) where lognormal distribution is appropriate for particulate data. Norazian *et al.* (2011) also found that lognormal distribution is very suitable to represent the air pollutant data where lognormal distribution perform better for PM_{10} concentration in industrial area in Malaysia. Previous research by Sansuddin *et al.* (2011) found that gamma distribution was the best distribution to represent PM_{10} concentration in Nilai for 2002. However, previous researches only consider the two-parameter distribution. However, lognormal distribution has been widely used to model many kinds of environmental contaminant data including air quality data (Gilbert, 1987).

The probability density function graphs were plotted using the values of parameter according to the best distribution for each year, as given in Table 3. Figure 2 shows the probability density function plot of PM_{10} concentration using the best estimator and distribution selected through the performance indicator criteria. The probability density function plot in Fig. 2 shows that the

Table 3: Parameter for the lognormal distribution using the best method

Parameter	2003	2004	2005	2006	2007	2008	2009
Shape	0.3489	0.3669	0.3891	0.3109	0.3661	0.4033	0.3657
Location	4.2621	4.2989	4.5507	4.3782	4.1726	4.1304	4.2711
Threshold	-20.5221	-18.9905	-10.900	-15.2815	-9.4599	-12.1896	-20.6006

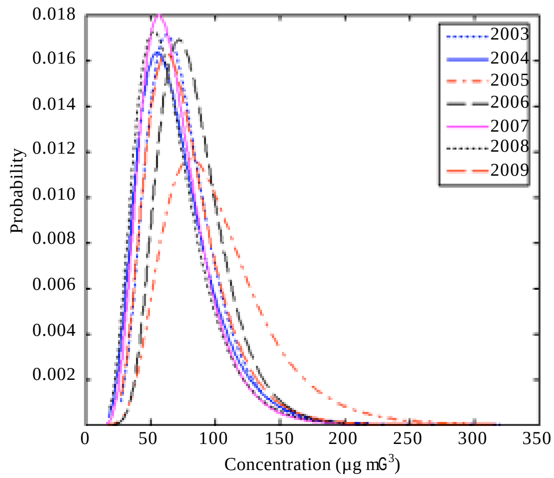

Fig. 2: The probability density function plot of PM_{10} concentration

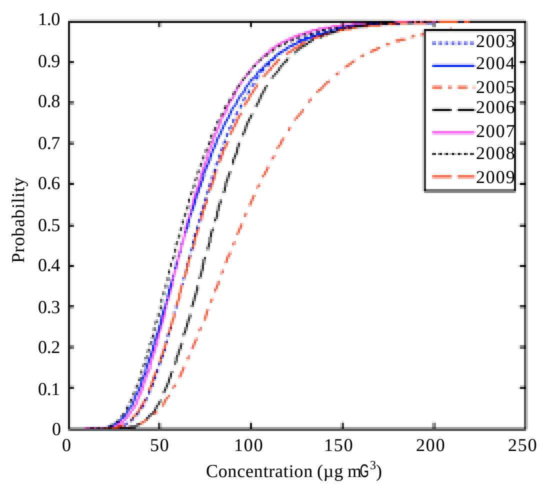

Fig. 3: The cumulative density function plot of PM_{10} concentration

distribution for each year is positively skewed with the highest skewness in 2005. This is probably caused by the high particulate event in 2005 (Department of Environment Malaysia, 2005). The probability density function plots also shown that a concentration of PM_{10} exceed the threshold limit set by Malaysia Ambient Air Quality Guidelines (150 µg m^{-3}) for each year, perhaps caused by monitoring station in Nilai are located in industrial area. In 2008, the pdf plot shows that the PM_{10} concentrations were lower compared to other years. Figure 3 shows the cumulative density function plot for the best distribution that represents the observed monitoring record from 2003 to 2009. The cumulative density function plot was used to determine the probability of PM_{10} concentration exceeding the Malaysia Ambient Air Quality Guideline (MAAQG). The probability of exceeding more than 150 µg m^{-3} were used to obtained the return period as shown in Table 4.

Table 4: The predicted and actual exceedences

Years	Predicted exceedences (days)	Actual exceedences (days)
2003	1.8	1.7
2004	3.9	3.8
2005	13.8	11.4
2006	11.9	10.9
2007	4.5	4.9
2008	2.9	2.5
2009	2.9	2.1

CONCLUSION

Based on the results from the analysis of PM_{10} concentration in Nilai, Negeri Sembilan from 2003 to 2009, every year experienced high particulate events. Two-parameter and three-parameter lognormal distribution were used to fit the PM_{10} concentration using method of moments and method of probability weighted moments to estimate the parameters. Results show that three-parameter lognormal distribution is the best distribution to represent the industrial area in Nilai, Negeri Sembilan. Predictions of the PM_{10} exceedences were estimated using the best estimator and the best distribution for each year. There are differences between predicted and actual values of exceedences but the errors are small. The results of this study provide useful information on air quality status in Nilai Negeri Sembilan and can be used for air quality management.

REFERENCES

Afroz, R., M.N. Hassan and N.A. Ibrahim, 2003. Review air pollution and health impact in Malaysia. Environ. Res., 92: 71-77.

Department of Environment Malaysia, 2005. Malaysia environment quality report 2005. Department of Environment Ministry of Natural Resources and Environment, Malaysia.

Department of Environment Malaysia, 2010. Malaysia environment quality report 2010. Department of Environment Ministry of Natural Resources and Environment, Malaysia.

Evans, M., N. Hasting and B. Peacock, 2000. Statistical Distributions. 3rd Edn., Wiley, New York, ISBN-13: 978-0471371243.

Gilbert, R.O., 1987. Statistical methods for environmental pollution monitoring. Van Nostrand Reinhold, New York, ISBN: 978-0-471-28878-7, pp: 336.

Hosking, J.R.M., 1990. L-moment analysis and estimation of distributions using linear combinations of order statistics. J. Royal Stat. Soc., Ser. B, 52: 105-124.

Johnson, N., S. Kotz and N. Balakrishnan, 1994. Continuous Univariate Distributions. John Wiley and Sons, New York.

Juneng, L., M.T. Latif, F.T. Tanggang and H. Mansor, 2009. Spatio-temporal characteristics of PM10 concentration across Malaysia. Atmos. Environ., 43: 4584-4594.

Juneng, L., M.T. Latif and F. Tanggang, 2011. Factors influencing the variations of PM10 aerosol dust in klang valley, malaysia during the summer. Atmos. Environ., 45: 4370-4378.

Junninen, H., H. Niska, K. Tuppurainen, J. Ruuskanen and M. Kolehmainen, 2004. Methods for imputation of missing values in air quality data sets. Atmos. Environ., 38: 2895-2907.

Kite, G.W., 1997. Frequency and Risk Analysis in Hydrology. Water Reseach Publications, Fort Collins.

Lopez, R.E., 1977. The lognormal distribution and cumulus cloud populations. Mon. Weather Rev., 105: 865-872.

Lu, H.C. and G.C. Fang, 2003. Predicting the exceedences of a critical PM10 concentration: A case study in Taiwan. Atmos. Environ., 37: 3491-3499.

Norazian, M.N., M.M.A. Abdullah, C.Y. Tan, N.A. Ramli, A.S. Yahaya and N.F.M.Y. Fitri, 2011. Modelling of PM10 concentration for industrial area in malaysia: A case study in shah alam. Phys. Procedia, 22: 318-324.

Sansuddin, N., N.A. Ramli, A.S. Yahaya, N.F.F.M. Yusof, N.A. Ghazali and W.A.A. Madhoun, 2011. Statistical analysis of PM_{10} concentrations at different locations in Malaysia. Environ. Monit. Assess., 180: 573-588.

Taylor, J.A., A.J. Jakemen and R.W. Simpson, 1986. Modelling distributions of air pollutant concentrations: Identification of statistical models. Atmos. Environ., 20: 1781-1789.

Ul-Saufie, A.Z., A.S. Yahaya, N.A. Ramli and H.A. Hamid, 2012. Performance of multiple linear regression model for long term PM_{10} concentration prediction based on gaseous and meteorological parameters. J. Applied Sci., 12: 1488-1494.

WHO, 1998. Report of the bioregional workshop on health impacts of haze related air pollution. World Health Organization, Manila.

Yahaya, A.S., N.A. Ramli and N.F. Yusof, 2005. Effects of estimating missing values on fitting distribution. Proceeding International Conference on Quantitative Sciences and its Applications, December 6-8, 2005, Universiti Utara Malaysia, Malaysia.

Yee, E. and R. Chan, 1997. A simple model for the probability density functions of concentration fluctuations in atmospheric plumes. Atmos. Environ., 31: 991-1002.

Yusof, N.F.F.M., N.A. Ramli, A.S. Yahaya, N. Sansuddin, N.A. Ghazali and W. Madhoun, 2010. Monsoonal differences and probability distribution of PM_{10} concentration. Environ. Monitor. Assess., 163: 655-667.