Environ Monit Assess (2010) 163:655–667
DOI 10.1007/s10661-009-0866-0

Monsoonal differences and probability distribution of PM_{10} concentration

Noor Faizah Fitri Md Yusof · Nor Azam Ramli · Ahmad Shukri Yahaya · Nurulilyana Sansuddin · Nurul Adyani Ghazali · Wesam al Madhoun

Received: 17 October 2008 / Accepted: 10 March 2009 / Published online: 14 April 2009

Abstract There are many factors that influence PM_{10} concentration in the atmosphere. This paper will look at the PM_{10} concentration in relation with the wet season (north east monsoon) and dry season (south west monsoon) in Seberang Perai, Malaysia from the year 2000 to 2004. It is expected that PM_{10} will reach the peak during south west monsoon as the weather during this season becomes dry and this study has proved that the highest PM_{10} concentrations in 2000 to 2004 were recorded in this monsoon. Two probability distributions using Weibull and lognormal were used to model the PM_{10} concentration. The best model used for prediction was selected based on performance indicators. Lognormal distribution represents the data better than Weibull distribution model for 2000, 2001, and 2002. However, for 2003 and 2004, Weibull distribution represents better than the lognormal distribution. The proposed distributions were successfully used for estimation of exceedences and predicting the return periods of the sequence year.

N. F. F. Md Yusof (✉) · N. A. Ramli ·
A. S. Yahaya · N. Sansuddin ·
N. A. Ghazali · W. al Madhoun
Clean Air Research Group, Environmental and Sustainable Development Section, School of Civil Engineering, Universiti Sains Malaysia, Engineering Campus, 14300 Nibong Tebal, Pulau Pinang, Malaysia
e-mail: myfaizah@hotmail.com

Keywords Exceedences · Lognormal distribution · Return period · Weibull distribution

Introduction

In countries with temperate climate, high pollution episode occurs mainly during wintertime (Marcazzan et al. 2001), up to twice the concentration in summer (Corani 2005). This is due to higher entropic emissions such as heating emissions from building and the unfavorable dispersion conditions. A country with tropical climate (Malaysia) experiences uniformed temperature and continuous high humidity. Seasons in this country are distinguished according to the changes in wind flow patterns and rainfall. As it is located near the equator, the wind over the country is generally light and variable. However, there are some uniform periodic changes in the wind flow patterns that explain the four seasons experienced by the country namely, north east monsoon (November to March), transitional period (April to May), south west monsoon (June to September), and another transitional period from October to November (Meteorological Department 2008).

The first aim of this study is to show the monsoonal differences influencing PM_{10} concentration in Seberang Perai, Penang (Fig. 1) located in the northern part of Malaysia during the north east

Fig. 1 The location of monitoring station

and south west monsoons. Malaysia receives substantial amount of rainfall all year round. However, there is a distinct peak wet season during the north east monsoon. Otherwise, Malaysia receives less rainfall during the south west monsoon in the west coast of Peninsular Malaysia and the western part of Sarawak. Generally known as the northern hemisphere summer monsoon, the PM_{10} concentration during south west monsoon is expected to be high in the area of study as a result of dry weather condition. The time series plot of the average daily and monthly of PM_{10} concentration for each year was used to determine the season with maximum value of PM_{10}.

The second aim of this study is to estimate the exceedences and return period of PM_{10} concentration in Seberang Perai using probability distribution. The air pollutants data are usually random variables (Lu 2002; Ott 1995) and statistical model such as probability distribution is a useful tool to model the air pollutants data. There are many types of probability distribution that have been used to fit the air pollutants data such as Weibull distribution (Wang and Mauzerall 2004; Lu 2002; Fitri et al. 2005), Gamma distribution (Lu 2003; Karaca et al. 2005; Rumburg et al. 2001), and lognormal distribution (Lu and Fang 2003; Sedek et al. 2006; Hitzenberger and Tohno 2000). However, only lognormal and Weibull distribution will be used in this study. The best distribution will represent the PM_{10} concentration data and will be utilized to predict the exceedences and return period.

Area of study

Seberang Perai, Penang is geographically located at latitude 5°22′58″ north of the equator and longitude 100°22′58″ east of the prime meridian. Seberang Perai covers an area of 738 km^2 from 1,031 km^2 of the state of Penang. It is located in the northern part of Malaysia and is a heavily industrialized area with several petrochemical complexes. Climatically, Penang represents a typical tropical monsoon characterized by uniformly high temperatures between 27–30°C in the daytime and 22–24°C at night. Besides experiencing heavy annual precipitation throughout the year with the mean annual rainfall of 267 cm, the relative humidity is also high, between 70% and 90%.

There were 647 industrial projects (14% of the national total) which have been approved by the Penang state government between 2000 and 2004 (Penang State Government 2008). These projects include electronic and electrical industry, machinery manufacturing, fabricated metal products, petrochemical industries, and food industries. The growth in industrial sectors had caused the increase in the problem of air pollution in Seberang Perai and subsequently affecting the people exposed to it. The main pollutant of concern was SO_2 (DoE 2004). However, in 2004 the pollutant of concern was PM_{10} emitted by industrial activities, motor vehicle emissions and transboundary pollutants which normally occurred during the south west monsoon.

The growth in industrial activities is normally accompanied by the rapid increase in population and traffic density. The population distributions in Seberang Perai also grow every year due to the increase in vacancies offered by the industries. The population distribution increased by 2.22% and 2.06% in 2002 and 2003. While in 2004, it went up by 1.96%. The total number of population estimated in Seberang Perai for 2004 is about 775,200. In addition to that, the number of mobile sources including passenger cars, taxis, buses, motorcycles, vans, and lorries has also increased every year. In 2001, the number of mobile sources increased 7.11% compared to the previous year. In 2002, it increased by 6.53% and in 2003 and 2004, it increased by 6.46% and 7.21%, respectively. The total number of mobile sources in 2004 is 1 461,274.

Penang is one of the states affected by haze in 2004. In 2004 the haze events range between slight to moderate in June, August, and September (DoE 2004). Haze is formed by tiny particulates suspended in the atmosphere. Muraleedharan et al. (2000) conducted a chemical characterization of haze in Brunei had found that particulate matter was the dominant pollutant during haze episodes. Back in 1997, haze event in Malaysia was caused by massive biomass burning for agricultural purposes in Indonesia. The wild fires significantly increased the input of organic aerosol components to the atmosphere (Abas et al. 2004). Therefore, it is expected that PM_{10} concentration in 2004 will be high.

Method

The data

There were 51 continuous air quality monitoring stations (CAQMS) in Malaysia. The stations are strategically located in residential, urban, and industrial areas to detect any significant change in the air quality which may be harmful to human health and the environment. Seberang Perai station is located in a heavily industrialized area of Seberang Perai. The samples of PM_{10} were collected by using continuous particulate monitor BAM 1020 (Met One Instruments, Inc.). This instrument automatically measures and records hourly PM_{10} concentration levels (in milligrams or micrograms per cubic meter) using the industry-proven principle of beta-ray attenuation. The data used in this study is hourly PM_{10} data in Seberang Perai taken from the year 2000 to 2004. The data are regularly subjected to standard quality control processes and quality assurance procedures by the DoE.

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). Assume that $y_1, y_2, \ldots, y_n$ be a time series data with n number of observations and there are k missing values denoted by $y_1^*, y_2^*, \ldots, y_k$ where $k<n$. Thus, the observed data with missing values can be represented as follows:

$$y_1,\ y_2,\ \ldots,\ y_{n1},\ y_1^*,\ y_{n1+1},\ y_{n1+2},\ \cdots,\\ y_{n2},\ y_2^*,\ y_{n2+1},\ y_{n2+2},\ \cdots,\ y_k^*,\ y_n \quad (1)$$

Therefore, the first missing values in Eq. (1) occur after $n1$ observations and the second missing values occur after $n2$ observations and so on. Therefore, for y_1^* will be replaced by

$$\bar{y}_1 = \frac{y_{n1} + y_{n1+1}}{2} \quad (2)$$

and y_2^* will be replaced by

$$\bar{y}_2 = \frac{y_{n2} + y_{n2+1}}{2} \quad (3)$$

If there are large missing values in the series of data denoted $y_{k1}^*, y_{k2}^*, y_{k3}^*, \ldots, y_k^*$ that occurs after $n1$ observations, the series of observed data with missing values can be represented as follows:

$$y_1,\ y_2,\ \ldots,\ y_{n1},\ y_{k1}^*,\ y_{k2}^*,\ y_{k3}^*,\ \ldots,\\ y_k^*,\ y_{n+1},\ y_{n1+2},\ \ldots,\ y_n \quad (4)$$

Thus, y_{k1}^* will be replaced by

$$\bar{y}_{k1} = \frac{y_{n1} + y_{n1+1}}{2} \quad (5)$$

For the next missing value, y_{k2}^* will be replaced by an average of new top value ($\bar{y}_{k1}$) and bottom value (y_{n+1}). The same method goes to the next missing value y_{k3}^* that will be replaced by the average of new top value ($\bar{y}_{k2}$) and bottom value (y_{n+1}). Equation (6) depicts the formula to replace the next missing values after y^*_{k1}.

$$\bar{y}_{k,i+1} = \frac{\bar{y}_{k,i} + y_{n1+1}}{2} \tag{6}$$

Where $i = 1, 2, 3, \ldots, k$

After replacing missing values, the descriptive statistics of the data such as the mean, median, standard deviation, variance, skewness, and kurtosis were determined to describe the statistical characteristics of the data sets.

Moreover, in order to show the monsoonal differences of PM_{10} concentrations in Seberang Perai, a time series plot were generated for every year (2000 to 2004). The daily average of PM_{10} concentrations were calculated by taking the average of 24 h data from the hourly PM_{10} data. The daily data in μg/m^3 vs. days were first plotted to see the average daily concentration in 1 year. The x axis that represents the days is further divided into four seasons. The data was also presented as a monthly average to show the seasonal variation of PM_{10} concentration each year. Therefore, a monthly average was determined by calculating the average of 1 month from the hourly PM_{10} concentration. A monthly average PM_{10} concentration in 2000 to 2004 was plotted. The temporal PM_{10} concentration was divided into four seasons to determine the maximum PM_{10} concentration.

The Weibull distribution

The Weibull distribution was used to fit the hourly PM_{10} concentration in this research. The two parameter Weibull probability density function with parameters α and β is given by (Lu 2003)

$$f(x) = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1} \exp\left[-\left(\frac{x}{\beta}\right)^{\alpha}\right] \tag{7}$$

The cumulative distribution function takes the form

$$F(x) = 1 - \exp\left[-\left(\frac{x}{\beta}\right)^{\alpha}\right] \tag{8}$$

where $x \geq 0, \alpha =$ shape parameter, $\beta =$ scale parameter

The α and β are the shape and scale parameters which need to be estimated. Two methods were used to estimate the parameters of α and β for the Weibull distribution, that is the maximum likelihood estimators (MLE) and method of moments (MoM).

The maximum likelihood estimator

For the MLE, a likelihood function is defined as (Mendenhall and Sincich 1994)

$$L = \prod_{i=1}^{n} f(x, \alpha, \beta) = \prod_{i=1}^{n} \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1} \exp\left[-\left(\frac{x}{\beta}\right)^{\alpha}\right]$$
$$= \frac{\alpha}{\beta^{n\beta}} \prod_{i=1}^{n} x_i^{\alpha-1} \exp\left(-\sum_{i=1}^{n} x_i^{\alpha}\right) \tag{9}$$

and log-likelihood function is

$$\ln L = n \ln \alpha - n\alpha \ln \beta + (\alpha - 1)\sum \ln x_i - \sum\left(\frac{x_i}{\beta}\right)^{\alpha} \tag{10}$$

Therefore, α and β are the solutions of the following partial derivatives

$$\frac{\partial \ln L}{\partial \alpha} = 0 \text{ and } \frac{\partial 1nL}{\partial \beta} = 0$$

Therefore, β was obtained by the solution as shown below

$$\frac{\sum_{i=1}^{n}\left(x_i^{\beta} \ln x_i\right)}{\sum_{i=1}^{n} x_i^{\beta}} - \frac{1}{\beta} - \frac{1}{n}\sum_{i=1}^{n} \ln x_i = 0 \tag{11}$$

α was then calculated using the following equation

$$\alpha = \left(\frac{1}{n}\sum_{i=1}^{n} x_i^{\hat{x}}\right)^{\frac{1}{\beta}} \tag{12}$$

The method of moments

For MoM, estimation of the parameters can be obtained using the sample mean and sample variance (Kottegoda and Rosso 1998). Using the expression for the mean and variance, β was obtained by the solution of

$$\frac{s^2}{\bar{x}} = \frac{\Gamma\left(1 + \frac{2}{\beta}\right)}{\Gamma\left(1 + \frac{1}{\beta}\right)^2} - 1 \tag{13}$$

α is calculated by the following equation

$$\alpha = \frac{\bar{x}}{\Gamma\left(1 + \frac{1}{\beta}\right)} \tag{14}$$

The lognormal distribution

The lognormal distribution was used to fit the hourly PM_{10} concentration in Seberang Perai. The two parameter lognormal probability density function with parameters α and β is given by (Lu 2003)

$$f(x) = \frac{1}{x\alpha\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\ln(x) - \beta}{\alpha}\right)^2\right] \tag{15}$$

The cdf form for the lognormal distribution is

$$F(x) = \frac{1}{2\pi} \int_{-\infty}^{\frac{\ln x - \beta}{\alpha}} e^{\frac{-x^2}{2}}\, dt \tag{16}$$

where $x \geq 0, \alpha$ = shape parameter, β = scale parameter.

Two methods were also used to estimate the parameters of α and β for the lognormal distribution that is the maximum likelihood estimators and method of moments.

The maximum likelihood estimator

For the MLE of the lognormal distribution, a likelihood function is defined as (Evans et al. 2000)

$$L = \prod_{i=1}^{n} f(x, \alpha, \beta) = \prod_{i=1}^{n} \frac{1}{x\alpha\sqrt{2\pi}} \times \exp\left[-\frac{1}{2}\left(\frac{\ln(x) - \beta}{\alpha}\right)^2\right], x > 0,\ \alpha, \beta > 0$$
$$= \alpha^{-n}(2\pi)^{-\frac{n}{2}} \prod_{i=1}^{n} x_i^{-1} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{\ln(x_i - \beta}{\alpha}\right)^2\right] \tag{17}$$

and the log-likelihood function is

$$\ln L = -n \ln \alpha - \frac{n}{2}\ln(2\pi) - \ln\left(\prod_{i=1}^{n} x_i\right) - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{\ln x_i - \beta}{\alpha}\right)^2 \tag{18}$$

Therefore, α and β are the solutions of the following partial derivatives

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

Thus, β was obtained by the solution as shown below

$$\beta = \frac{1}{n}\sum_{i=1}^{n} \ln(x_i) \tag{19}$$

and

$$\alpha = \sqrt{\frac{1}{n}\sum_{i=1}^{n}[\ln(x_i) - \beta]} \tag{20}$$

The method of moments

In this method, α and β were obtained directly from the following equations (Kottegoda and Rosso 1998)

$$\alpha = \sqrt{\ln\left(s^2 + (\bar{x})^2\right) - 2\ln(\bar{x})} \tag{21}$$

$$\beta = \ln(\bar{x}) - \frac{\alpha^2}{2} \tag{22}$$

where;

$$s^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 \text{ and } \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{23}$$

Performance indicators

In this paper, three performance indicators (PI) have been used to determine the agreement between predicted and observed PM_{10} hourly concentration data. The tests are root mean square error (RMSE), prediction accuracy (PA), and coefficient of determination (R^2) that are shown in Eqs. (24)–(26), respectively. For PA and R^2, the value would be from 0 to 1 and the PI with value that is closes to 1 gives the best fit. For RMSE, the value that is closest to zero gives the best fit.

$$\mathrm{RMSE} = \frac{1}{n-1}\sum_{i=1}^{n}(P_i - O_i)^2 \qquad (24)$$

$$\mathrm{PA} = \frac{\sum_{i=1}^{n}(P_i - \bar{P})(O_i - \bar{O})}{(n-1)\,\sigma_P\sigma_O} \qquad (25)$$

$$R^2 = \left(\frac{1}{n}\frac{\sum_{i=1}^{n}(P_i - \bar{P})(O_i - \bar{O})}{\sigma_P\sigma_O}\right)^2 \qquad (26)$$

Where n is the number of observations, P_i and O_i are the predicted and observed concentrations, respectively, $\bar{P}$ and $\bar{O}$ are the means of predicted and observed concentration, respectively. σ_p and σ_0 refer to standard deviation of predictions and observations, respectively.

Result and discussion

PM_{10} concentration level in Seberang Perai

Table 1 shows the descriptive statistics for hourly PM_{10} concentration in Seberang Perai from 2000 to 2004. The frequency distribution of hourly PM_{10} data is depicted in Fig. 2. Table 1 shows that each year the mean was higher than the median, indicating that the data was skewed to the right and there were extreme event that occurred. The skewness in Table 1 shows positive numbers that also signify the occurrences of extreme events. Moreover, the trend of positive skewness shown in Fig. 2, where the frequency distributions for all years were skewed to the right indicated high particulate emissions. The mean and median were also increasing, suggesting increasing air pollution problem. The Department of Environment reported haze event in 2004 and Seberang Perai was moderately affected (DoE 2004). However, the skewness showed high value in 2000 and 2002. After 2002, the decreasing skewness showed that the tendency of recording high PM_{10} event was reducing. The value of kurtosis in Table 1 shows that the peak of the PM_{10} distribution in Seberang Perai was higher than the normal distribution. The highest kurtosis was in 2002, followed by the years 2000 and 2001.

The air quality status in Seberang Perai recorded by the DoE Malaysia is described in Fig. 3. The air quality status for Seberang Perai was based on Malaysia Air Pollution Index scale which closely follows the Pollutant Standard Index system of the United States (DoE 1996). Based on Fig. 3, Seberang Perai experienced moderate air quality status most of the time. There was also good air quality status recorded from 2000 to 2002. The numbers of good air quality days were 59 and 55 for 2000 and 2001, respectively, and only 2 days were recorded for 2002. No good air quality status was recorded in 2003 and 2004. However, the unhealthy air quality status was recorded every year indicates that the level of air pollution in Seberang Perai is below the acceptable level (150 μg/m^3) at certain time of the year.

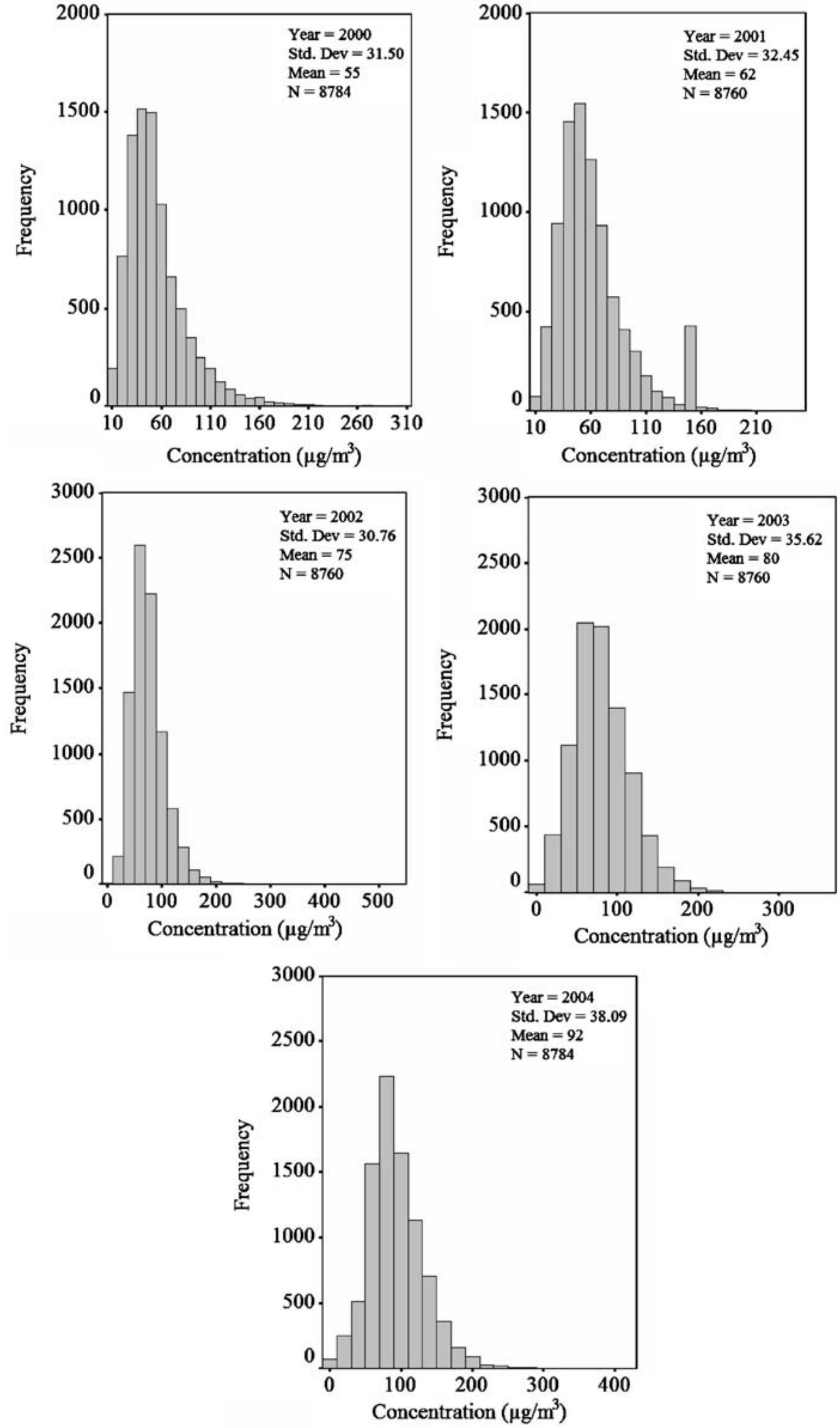
Monsoonal differences

Figure 4 shows the plot of the daily average concentration of PM_{10} in Seberang Perai. It was clear

Table 1 Descriptive statistics of hourly PM_{10} data in Seberang Perai

	2000	2001	2002	2003	2004
No.	8,784	8,760	8,760	8,760	8,784
Mean	54.82	61.73	75.03	80.13	92.31
Median	48.00	54.00	70.00	76.00	87.00
Standard deviation	31.50	32.45	30.76	35.62	38.09
Variance	992.02	1,053.06	946.39	1,268.91	1,450.51
Skewness	1.79	1.38	1.41	0.71	0.73
Kurtosis	5.40	1.90	7.86	1.31	1.72
Minimum	5	6	7	5	5
Maximum	310	248	540	362	421

Fig. 2 Hourly PM_{10} concentration data for Seberang Perai from 2000 to 2004



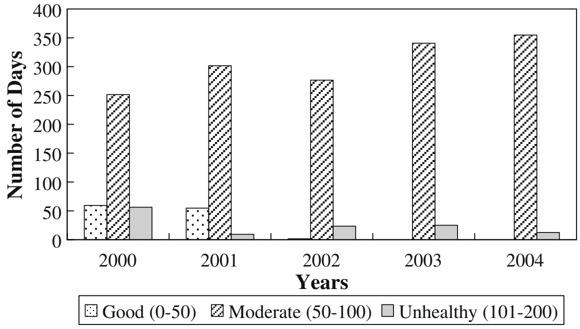


Fig. 3 Air pollution index recorded for Seberang Perai (Department of Environment 2000, 2001, 2002, 2003, 2004)

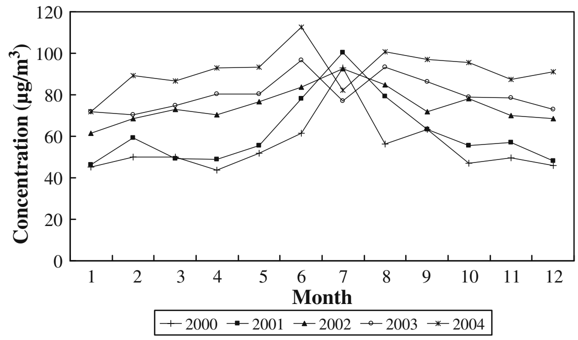
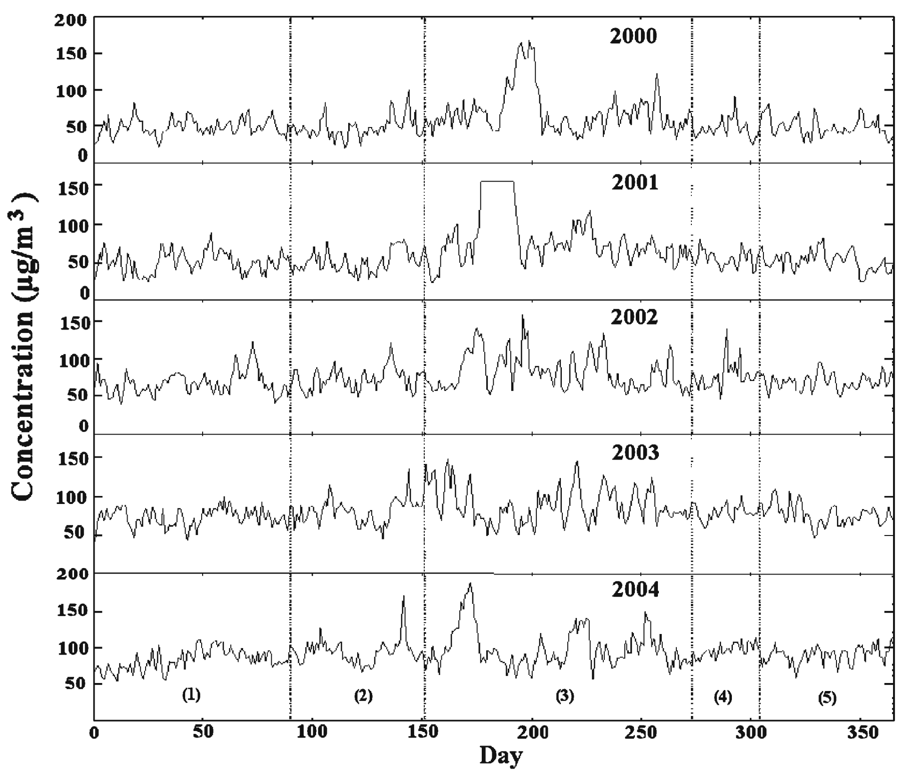


Fig. 5 Monthly average of PM_{10} concentration in Seberang Perai from 2000 to 2004

that there were periods with high PM_{10} concentration which exceed the Malaysian Ambient Air Quality Guideline (150 μg/m^3) for every year, excluding 2003. In 2000, the concentration ranged between 20 μg/m^3 to the maximum of 170 μg/m^3. The maximum concentration occurred during the south west monsoon; however, it was short-lived. The range of PM_{10} concentration in 2001 was between 20 μg/m^3 to 155 μg/m^3. The maximum concentration occurred during the south west monsoon for a longer period than the previous year. In 2003, the PM_{10} concentration began to increase during the first inter monsoon and reached its peaks during the early south west monsoon.



(1) North east monsoon
(2) First intermonsoon
(3) South west monsoon
(4) Second intermonsoon
(5) North east monsoon

Fig. 4 Daily average of PM_{10} concentration in Seberang Perai from 2000 to 2004

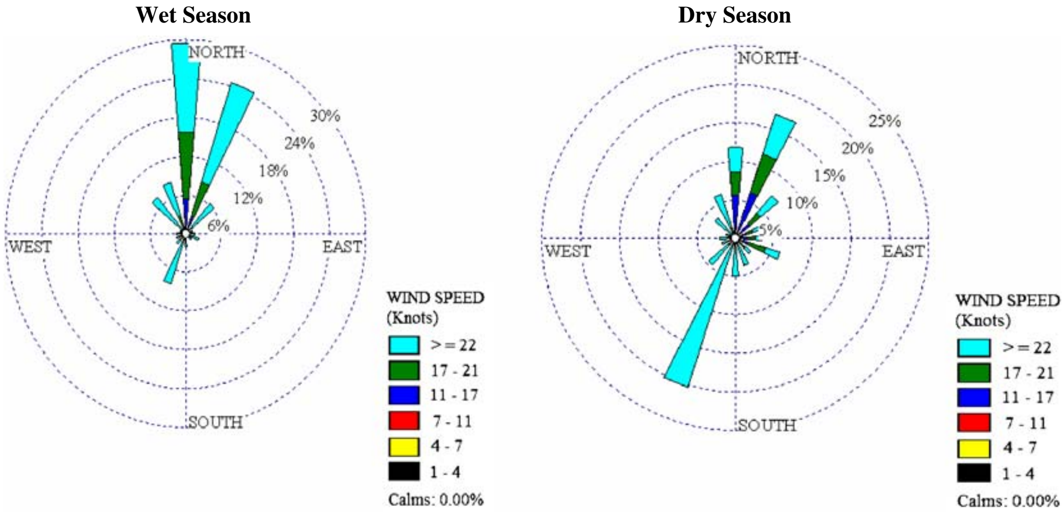


Fig. 6 Wind rose distributions in wet and dry season in Seberang Perai 2002

The year 2004 exhibited a clear peak on the first inter monsoon, but again reached its peak during the south west monsoon.

Figure 5 illustrates the plot of the monthly average of PM_{10} concentration in Seberang Perai (2000 to 2004). The plot showed clearly that the highest concentration period of PM_{10} was during the south west monsoon, between June to September. In contrast, the concentration during the north east monsoon is generally lower than the south west monsoon. This may be caused by the influence of substantial amount of rainfall during north east monsoon which had increased the removal rate of PM_{10} through precipitation. In addition, the PM_{10} concentration saw an increased every year.

During the south west monsoon, the prevailing wind was generally south westerly and light (below 15 knots) (Meteorological Department 2008). Figure 6 illustrates the wind rose for wet and dry seasons in Seberang Perai in 2002. During the dry season, more than 20% of winds were predominantly blowing from the south west with an average wind speed of 14.05 m/s. The wind from the south west carried pollutants emitted from Indonesia caused by biomass burning for agricultural purposes. Therefore, PM_{10} concentrations were high during the south west monsoon. Stable atmospheric condition and low rainfall particularly in the west coast state of Peninsular Malaysia including Penang and western part of Sarawak aggravated the high PM_{10} concentration.

The north east monsoon usually commences in early November and ends in March. During this season, steady easterly or north easterly winds of 10 to 20 knots prevail (Meteorological

Table 2 Parameters for the weibull and lognormal distribution

Year	Parameter	Weibull		Lognormal	
		MLE	MoM	MLE	MoM
2000	α	1.8690	1.8011	0.5465	0.5341
	β	62.0614	61.6465	3.8579	3.8614
2001	α	2.0306	1.9874	0.5050	0.4940
	β	69.9906	69.6438	3.9972	4.0007
2002	α	2.5123	2.6221	0.4160	0.3942
	β	84.3935	84.4503	4.2357	4.2402
2003	α	2.3667	2.3960	0.5266	0.4246
	β	90.3330	90.4050	4.2676	4.2936
2004	α	2.5305	2.6040	0.5036	0.3965
	β	103.7098	103.9237	4.4230	4.4466

Table 3 The result of performance indicators

Year	PI	Weibull		Lognormal	
		MLE	MoM	MLE	MoM
2000	RMSE	6.753746	6.429064	0.861199	0.0000
	R^2	0.952892	0.957524	0.998726	0.998901
	IA	0.987765	0.989208	0.999572	0.999778
2001	RMSE	7.622258	7.501725	6.60616	6.376834
	R^2	0.943899	0.946086	0.959305	0.960478
	IA	0.985552	0.986208	0.989536	0.989973
2002	RMSE	6.451572	6.492875	3.665362	2.870622
	R^2	0.958071	0.954669	0.989139	0.990015
	IA	0.989024	0.988461	0.996402	0.997547
2003	RMSE	3.174484	3.206679	13.70309	5.041664
	R^2	0.991236	0.990896	0.954837	0.979015
	IA	0.997833	0.99777	0.971462	0.994753
2004	RMSE	4.975957	4.985441	15.43711	4.955016
	R^2	0.983013	0.982019	0.960593	0.9822
	IA	0.995639	0.995521	0.969327	0.995565

Department 2008). From Fig. 6, it can be seen that almost 30% winds were predominantly blowing from north with an average wind speed of 14.67 m/s. During this monsoon, the primary maximum rainfall occurs between October and November while the secondary maximum generally occurs from April to May. Thus, the high rainfall during this season increases particulate washout and reduce the concentration of PM_{10}.

Parameter estimation

The Weibull and lognormal distribution models were used to fit the hourly PM_{10} concentration observed in Seberang Perai. Table 2 shows the parameters obtained for the Weibull and Lognormal distribution using the MLE and MoM methods. It shows that the results are different when different methods and probability distributions models were used. Performance indicators were used to determine the best method and distribution to represent the PM_{10} concentration. Table 3 shows the result for the performance indicators. For R^2 and IA, the closer value to 1 gives the best result that fit the data while for RMSE, the smallest value gives the best fit. Therefore, from 2000 to 2002, it shows that the lognormal distribution using MoM is the best. Lognormal distribution precludes values less than zero and the distribution has no upper bound. Therefore, the lognormal distribution's shape has considerable potential to describe environmental quality data (McBean and Rovers 1998). The Weibull distribution using the MLE represented the best PM_{10} concentration for 2003 and 2004. This might be because 2003 and 2004 show high means of 80.13 and 92.31 ug/m^3, respectively.

Pdf and cdf plots

Figure 7 shows the pdf plot of the best distributions using the final shape and scale parameters selected from the best distribution shown in

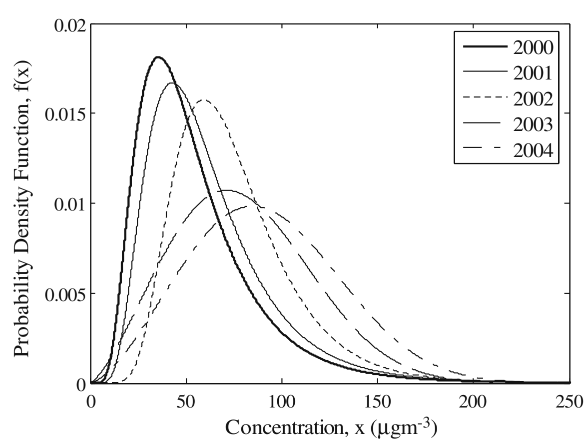


Fig. 7 The pdf plot of the selected distribution for Seberang Perai

Table 4 Final α (shape) and β (scale) parameters

Year	Selected probability distribution	α (shape)	β (scale)
2000	Lognormal (MoM)	0.5341	3.8614
2001	Lognormal (MoM)	0.4940	4.0007
2002	Lognormal (MoM)	0.3942	4.2402
2003	Weibull (MLE)	2.3667	90.3330
2004	Weibull (MLE)	2.5305	103.7098

Table 4. The pdf plot represents the observed data from 2000 to 2004. In 2000, the pdf plot shows that the PM_{10} concentrations were lower compared to other years. Conversely, extreme values had been recorded as shown by the long tail. In 2001, the density was lower and the mode of the plot skewed more to the right indicated that PM_{10} concentration had increase. In 2002, the plot showed higher density functions and the mode of the plot was more skewed than the previous year. The year 2003 and 2004 showed that the densities were becoming lower and the modes of the plots were getting more right skewed indicating that there were more pollution sources that contributed to high PM_{10} concentration over the years.

In general, every year have a long tail signifying that extreme events had occurred. The pdf plot illustrates the expected trend of PM_{10} concentration based on the descriptive statistics. Every year from 2000 to 2004, the right skewed distribution indicated increasing level of PM_{10} concentration was recorded.

Figure 8 shows the cdf plot for the best distribution that represents the observed data from 2000 to 2004. The plot was utilized to determine the probability of PM_{10} concentration exceeding more than the Malaysia Ambient Air Quality Guideline (MAAQG) that is 150 μg/m^3. The MAAQG for 24-h average of PM_{10} concentration follows the National Ambient Air Quality Standard (NAAQS) by the USEPA (Table 5).

The cdf plot also shows that the concentrations were increasing every year and the probability exceeding more than 150 μg/m^3 was also increasing. The probability of exceeding more than 150 μg/m^3 and the return period obtained from the cdf plot is shown in Table 6. Table 6 shows that the predicted return periods were underestimated for all years. However, the differences between the predicted and actual days were small except for the years 2000 and 2003, which have shown 68.4% and 50% differences in predicted and actual days, respectively. This might be due to many missing values in 2001 that reflected the number of days with high PM_{10} concentration in that year. In Fig. 4, it can be seen that at highest concentration, the plot had become flat indicating that there were many missing values at that point that make the replaced PM_{10} concentration using mean top bottom technique had become constant. Additionally, haze event have been recorded in 2004 that had increased the number of days with high particulate event in that year that make a large difference in the predicted value in 2003. Even though the predicted values

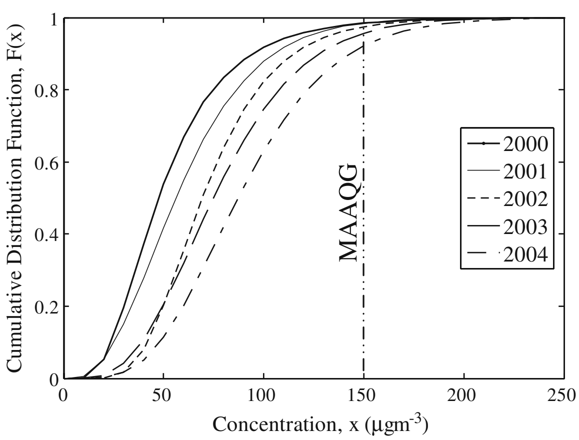


Fig. 8 The cdf plot of the selected distribution for Seberang Perai

Table 5 Ambient air quality standards—Malaysia and the United States

Air pollutant	Malaysia (μg/m^3)	USEPA (μg/m^3)
Particulate matter (PM_{10})		
Annual	50	N/S
24-h average	150	150[a]

Source DoE 2006 and USEPA 2009

N/S not specified

[a]Not to be exceeded more than once per year on average over 3 years

Table 6 Exceedences and return period of $PM_{10} > 150\ \mu g/m^3$ in Seberang Perai

Years	Probability > 150 $\mu g/m^3$	Return period (days)	
		Predicted	Actual
2000	0.015774	6	19
2001	0.020519	7	8
2002	0.025385	9	14
2003	0.036120	13	26
2004	0.078527	29	34

were influenced by the quality of the observed data and the haze event, the trends for the next years were successfully predicted. Parallel with time series plot in Fig. 5, both the predicted return period and the plot showed that the concentration of PM_{10} in Seberang Perai is increasing over the years. The years 2000 to 2001 show only slight increases in the return period that are between 1 and 2 days. On the other hand, the year 2003 had started to show a significant increase of return period, which is 13 days, 4 days more than the previous year while the next year had recorded 29 days, that is 16 days more than in 2003. The predicted days with high PM_{10} concentration could happen continuously or at different time. Based on monsoonal differences in Seberang Perai, it is predicted that the high PM_{10} concentration will occur mostly during the south west monsoon.

Conclusion

This study showed the monsoonal differences in PM_{10} concentration clearly by using the daily average and monthly average plot. From both plots, the south west monsoon that is also known as the northern hemisphere summer monsoon recorded the highest PM_{10} concentration every year. Therefore, the effects of differences in weather condition experienced in different monsoon have been shown successfully.

This study had revealed that the air quality status in Seberang Perai were not good all the time. Over the years, the air pollution index has been increasing. The PM_{10} concentrations are also increasing and trigger worries due to the bad effects on human health as well as the environment. The seasonal variation for Seberang Perai showed the highest PM_{10} concentration recorded during south west monsoon every year proving that wind speed and wind direction do affect the concentration of PM_{10}. Thus, the predicted return period might occur during this season.

From the probability distribution used to fit the observed data, the lognormal distribution agrees with the data in 2000, 2001, and 2002 better than the Weibull distribution. However, the Weibull distribution fit the observed data better in 2003 and 2004. These results also indicate the increasing trend of PM_{10} concentration as Weibull is better to fit the high concentration emission. It is further proved by the return periods that were increasing over the years. The return period for 2000 and 2001 is 6 and 7 days, respectively. 2002 recorded 9 days, 2003 13 days, and 2004 29 days. Therefore, the Weibull distribution might be better than the lognormal distribution to represent the high PM_{10} concentration in the future if no actions were taken to reduce the PM_{10} concentration.

There are many local sources in Seberang Perai that contribute to the increase in PM_{10} concentration. Even haze event that have been recorded in 2004 might influence the high PM_{10} concentration in 2004, but local sources should be given more emphasis as they are under the control of local authority and actions could be taken to penalize the responsible persons or organizations. Furthermore, the pdf plot showed an increasing trend over the years, meaning that the PM_{10} concentration in 2004 might be higher than the previous year even without haze event.

Although only PM_{10} data in Seberang Perai were analyzed, the methods can be extended to other stations in Malaysia. The results of this study provide useful information on air quality status in Seberang Perai and can be used for air quality management as well as to develop air pollution control strategy.

Acknowledgements Thanks to Universiti Sains Malaysia for providing financial support to carry out this work under USMST grant 6035197 and Ministry of Science, Technology and Innovation e-science grant 305.PA.6013404 and the Department of Environment, Malaysia.

References

Abas, M. R. B., Rahman, N. A., Omar, N. Y. M. J., Maah, M. J., Samah, A. A., Oros, D. R., et al. (2004). Organic composition of aerosol particulate matter during a haze episode in Kuala Lumpur, Malaysia. *Atmospheric Environment, 38*, 4223–4241.

Corani, G. (2005). Air quality prediction in Milan: feed-forward neural networks, pruned neural networks and lazy learning. *Ecological Modelling, 185*, 513–529.

Department of Environment, Malaysia (1996). *Malaysia environmental quality report 1996*. Kuala Lumpur: Department of Environment, Ministry of Sciences, Technology and the Environment, Malaysia.

Department of Environment, Malaysia (2000). *Malaysia environmental quality report 2000*. Kuala Lumpur: Department of Environment, Ministry of Sciences, Technology and the Environment, Malaysia.

Department of Environment, Malaysia (2001). *Malaysia environmental quality report 2001*. Kuala Lumpur: Department of Environment, Ministry of Sciences, Technology and the Environment, Malaysia.

Department of Environment, Malaysia (2002). *Malaysia environmental quality report 2002*. Kuala Lumpur: Department of Environment, Ministry of Sciences, Technology and the Environment, Malaysia.

Department of Environment, Malaysia (2003). *Malaysia environmental quality report 2003*. Kuala Lumpur: Department of Environment, Ministry of Sciences, Technology and the Environment, Malaysia.

Department of Environment, Malaysia (2004). *Malaysia environmental quality report 2004*. Kuala Lumpur: Department of Environment, Ministry of Sciences, Technology and the Environment, Malaysia.

Department of Environment, Malaysia (2006). *Malaysia environmental quality report 2006*. Kuala Lumpur: Department of Environment, Ministry of Sciences, Technology and the Environment, Malaysia.

Evans, M., Hastings, N., & Peacock, B. (2000). *Statistical distributions* (3rd ed.). New York: Wiley.

Fitri, N. F., Ramli, N. A., & Yahaya, A. S. (2005). Fitting two parameters Weibull distribution function for particulate matter in Penang. In J. M. Jahi, K. Ariffin, A. Awwang, K. Aiyub, & M. R. Razman (Eds.), *Prosiding seminar kebangsaan pengurusan persekitaran 2005*. Bangi: UKM.

Hitzenberger, R., & Tohno, S. (2000). Black carbon (BC) concentrations and size distributions at two urban sites (Uji, Japan and Vienna, Austria). *Journal of Aerosol Science, 31*(Suppl), 112–113.

Karaca, F., Alagha, O., & Erturk, F. (2005). Statistical characterization of atmospheric PM10 and PM2.5 concentrations at a non-impacted suburban site of Istanbul, Turkey. *Chemosphere, 59*, 1183–1190.

Kottegoda, N. T., & Rosso, R. (1998). *Statistics, probability and reliability for civil and environmental engineers*. Singapore: McGraw-Hill.

Lu, H. C. (2002). The statistical characters of PM_{10} concentration in Taiwan area. *Atmospheric Environment, 36*, 491–502.

Lu, H. C. (2003). Estimating the emission source reduction of PM_{10} in Central Taiwan. *Chemosphere, 54*(7), 805–814.

Lu, H. C., & Fang, G. C. (2003). Predicting the exceedences of a critical PM_{10} concentration-a case study in Taiwan. *Atmospheric Environment, 37*, 3491–3499.

Marcazzan, G. M., Vaccaro, S., Valli, G., & Vecchi, R. (2001). Characterisation of PM10 and PM2.5 particulate matter in the ambient air of Milan (Italy). *Atmospheric Environment, 35*, 4639–4650.

McBean, E. A., & Rovers, F. A. (1998). Statistical procedures for analysis of environmental monitoring data and risk assessment (pp. 71–72). New Jersy: Prentice Hall PTR.

Mendenhall, W., & Sincich, T. (1994). *Statistics for engineering and the sciences*. New Jersey: Prentice-Hall.

Meteorological Department (2008). Official website of Malaysia Meteorological Department (MMD); General climate of Malaysia. http://www.met.gov.my/english/education/climate/climate01.html. Accessed 1st February 2008.

Muraleedharan, T. R., Radojevic, M., Waugh, A. & Caruana, A. (2000). Chemical characterization of the haze in Brunei Darussalam during the 1998 episode. *Atmospheric Environment, 34*, 2725–2731.

Ott, W. R. (1995). *Environmental statistics and data analysis*. New York: Lewis.

Penang State Government (2008). Official portal of Penang State Government. http://www.penang.gov.my. Accessed 1st February 2008.

Rumburg, B., Alldredge, R., & Claiborn, C. (2001). Statistical distributions of particulates matter and error associated with sampling frequency. *Atmospheric Environment, 35*, 2907–2920.

Sedek, J. N. M., Ramli, N. A., & Yahaya, A. S. (2006). Air quality predictions using lognormal distribution functions of particulate matter in Kuala Lumpur. *Malaysian Journal of Environmental Management, 7*, 33–41.

U.S. Environmental Protection Agency(2009). National ambient air quality standard. http://www.epa.gov/air/criteria.html. Accessed 2nd February 2009.

Wang, X., & Mauzerall, D. L. (2004). Characterizing distributions of surface ozone and its impact on grain production in China, Japan and South Korea: 1990 and 2020. *Atmospheric Environment, 38*, 4383–4402.

Yahaya, A. S., Ramli, N. A., & Yusof, N. F. F. M. (2005). Effects of estimating missing values on fitting distributions. In *Proceeding international conference on quantitative Sciences and its applications 2005 (ICOQSIA)*. Penang.