ELSEVIER

The Science of the Total Environment 298 (2002) 119–130

the Science of the Total Environment
An International Journal for Scientific Research into the Environment and its Relationship with Man

www.elsevier.com/locate/scitotenv

Estimating the frequency distributions of PM_{10} and $PM_{2.5}$ by the statistics of wind speed at Sha-Lu, Taiwan

Hsin-Chung Lu[a,]*, Guor-Cheng Fang[a]

[a]*Department of Environmental Engineering, Hungkuang Institute of Technology, Sha-Lu, Taichung, Taiwan*

Received 15 November 2001; accepted 25 March 2002

Abstract

The frequency distribution of air pollutant concentration varies with the meteorological conditions and pollutant emission level. There exists a simple relation between the frequency distribution of wind speed and frequency distribution of air pollutant concentration. The concentration of air pollutant, C, at cumulative probability, p, is inversely proportional to the wind speeds, u, at probability of $(100-p)$ when the distributional types and shape factors of both data are the same. The relationship is shown as $K=C_p u_{(100-p)}$, where K is constant. In this study, three theoretical distributions (log-normal, Weibull and type V Pearson distributions) are selected to fit the measured data of PM_{10}, $PM_{2.5}$ and wind speed. The frequency distributions of air pollutants can be estimated from the simple relationship of air pollutant concentration and wind speed. The results show that the log-normal distribution is the best one to represent the data of PM_{10}, $PM_{2.5}$ and wind speed. The K values of PM_{10} and $PM_{2.5}$ are nearly constant from the 30–80th percentiles. It was also found that the distributions of PM_{10} and $PM_{2.5}$ can be successfully estimated from the distribution of wind speed. The Kolmogorov–Smirnov (K–S) test shows that there is no significant discrepancy between the estimated and measured distribution of PM_{10} and $PM_{2.5}$ at the 95% confidence level. Therefore, the distribution of air pollutants is easily estimated when the wind speed data are known.

Keywords: Frequency distribution; Wind speed; PM_{10}; $PM_{2.5}$; Kolmogorov–Smirnov test; Sha-Lu

1. Introduction

The frequency distribution of air pollutant concentration is useful in understanding the statistical characteristics of air quality. It can be used to estimate how frequently a critical concentration level is exceeded (Seinfeld and Pandis, 1998). Knowledge of the frequency distribution is necessary for developing air pollutants control strategies at different areas. However, the concentrations of air pollutants usually vary randomly and are correlated with the emission levels and meteorological conditions.

Many types of distributions have been used to fit the air pollutant concentration data. These distributions are log-normal (Mage and Ott, 1984; Kao and Friedlander, 1995), Weibull (Georgopoulos and Seinfeld, 1982), gamma (Berger et al., 1982), and type V Pearson distributions (Morel et al., 1999). It was found that the frequency distri-

*Corresponding author. Fax: +886-4-26525245.
E-mail address: hclu@sunrise.hkc.edu.tw (H.-C. Lu).

PII: S0048-9697(02)00164-X

bution of PM_{10} and the gas-phase pollutants is approximately a log-normal distribution (Kao and Friedlander, 1995). The results of Burkhardt et al. (Burkhardt et al., 1998) showed that the NH_3 concentration frequency distribution could be fitted well with a log-normal distribution. Other research pointed out that the Weibull distribution has a good fitting result to hourly averaged oxidant data (Seinfeld and Pandis, 1998). Moreover, the type V Pearson distribution was successfully used to fit the measured data of PM_{10}, $PM_{2.5}$, SO_2, NO_2 in Santiago (Morel et al., 1999). Therefore, there is no one universal distribution that can represent the frequency distribution of air pollutants.

The distributions of air pollutant concentrations are influenced by meteorological conditions, such as wind speed. A simple relationship exists between the statistics of the wind speed and air pollution data. Simpson et al. (1984) linked air quality data and wind speed to the Atmospheric Turbulence and Diffusion Laboratory (ATDL) model. A modified model that relates air pollutant concentration, C, and wind speed, u, is obtained. The concentration of air pollutant data at cumulative probability, p, is inversely proportional to the wind speeds at probability $(100-p)$ when the frequency distributions of both data are log-normally distributed and the geometric standard deviation (shape factor) is the same (Simpson et al., 1984). This modified model was demonstrated and used to predict the maximum level of air pollutant (Simpson, 1990).

In this work, 42 random sampling data of PM_{10}, $PM_{2.5}$ and wind speed were measured from July to December, 2000, at Sha-Lu, Taiwan. Three theoretical frequency distributions, namely log-normal, Weibull and type V Pearson, were used to fit the measured data. The distributional parameters were estimated by the method of maximum-likelihood. The K–S test was used to determine which type of distribution is appropriate to represent the frequency distribution of air pollutants and wind speed. The purposes of this paper were to demonstrate the relationship between the frequency distributions of wind speed and air pollutants, and to determine the K value. Finally, the frequency distribution of pollutants was estimated from the frequency distribution of wind speed.

2. Methods

2.1. Probability density functions of distributions

Three theoretical distributions, log-normal, Weibull, and type V Pearson, are selected to fit the measured data of the PM_{10}, $PM_{2.5}$ and wind speed. The probability density functions (p.d.f.) of these distributions are shown as follows.

2.1.1. Log-normal distribution

$$p_l(x_i)=\frac{1}{\sqrt{2\pi}x_i\ln(\sigma_g)}\exp\left[-\frac{(\ln x_i-\ln\mu_g)^2}{2(\ln\sigma_g)^2}\right] \quad (1)$$

where x_i is the pollutant concentration of species i, μ_g and σ_g are the scale and shape parameters of the log-normal distribution and represent the geometric mean and the standard geometric deviation, respectively.

2.1.2. Weibull distribution

$$p_w(x_i)=\frac{\lambda}{\sigma_w}\left(\frac{x_i}{\sigma_w}\right)^{\lambda-1}\exp\left[-\left(\frac{x_i}{\sigma_w}\right)^{\lambda}\right],$$
$$x_i\geq 0;\ \sigma_w,\ \lambda>0 \quad (2)$$

where σ_w and λ are the scale and shape parameters of the Weibull distribution, respectively.

2.1.3. Type V Pearson distribution (*Morel et al., 1999*)

$$p_t(x_i)=\frac{(\rho)^{\rho+1}}{x^{eq}\Gamma(\rho+1)}\left(\frac{x_i}{x^{eq}}\right)^{-(\rho+2)}\exp\left(-\frac{\rho x^{eq}}{x_i}\right) \quad (3)$$

where x^{eq} and ρ are the scale and shape parameters of the type V Pearson distribution, respectively. Γ is the gamma function.

2.2. Estimation of distributional parameters

The method of maximum-likelihood is used to estimate the distributional parameters. This method always gives a minimum variance for estimating the parameters. The p.d.f. of a theoretical distribution is $p(x_i,\ \theta_1,\ \theta_2)$, where θ_1, θ_2 are the

Table 1
The equations used for estimating parameters of three theoretical distributions

Distribution	Equations for estimating distributional parameters
Log-normal	$\ln\mu_g = \frac{1}{n}\sum_{i=1}^{n}\ln x_i,\ (\ln\sigma_g)^2 = \frac{1}{n}\sum_{i=1}^{n}(\ln x_i - \ln\mu_g)^{2*}$
Weibull	$\lambda = \left[\left(\sum_{i=1}^{n} x_i^{\lambda}\ln x_i\right)\times\left(\sum_{i=1}^{n} x_i^{\lambda}\right)^{-1} - \frac{1}{n}\sum_{i=1}^{n}\ln x_i\right]^{-1},\ \sigma_w = \left(\frac{1}{n}\sum_{i=1}^{n} x_i^{\lambda}\right)^{1/\lambda *}$
Type V Pearson	$\frac{d\ln(\Gamma(\rho+1))}{d\rho} = \ln(\rho x^{eq}) - \frac{1}{n}\sum_{i=1}^{n}\ln(x_i),\ x^{eq} = \frac{\rho+1}{\rho}\frac{1}{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i}}$

* See reference (Jakeman et al., 1986).

parameters of the distributions. Then the likelihood function L is:

$$L = \prod_{i=1}^{n} p(x_i, \theta_1, \theta_2) \tag{4}$$

It is more convenient to work with the logarithm of L. The parameters, θ_1 and θ_2, are obtained by differentiating $\ln(L)$ with respect to θ_1 and θ_2 and setting the result equal to zero (Georgopoulos and Seinfeld, 1982). These equations for calculating the distributional parameters by the method of maximum-likelihood are shown in Table 1.

2.3. Kolmogorov–Smirnov test

The K–S statistic is defined as the maximum difference between the sample cumulative probability and the expected cumulative probability.

$$D_{max} = \max|F(x_i) - S(x_i)| \tag{5}$$

where $F(x_i)$ and $S(x_i)$ are the expected and observed cumulative frequency functions.

The D_{max} will be compared to the largest theoretical difference, $D_{\alpha/2}$, acceptable for the K–S test at a certain significance level, α. When D_{max} is less than $D_{\alpha/2}$, then we say that the hypothesis that there is no difference between observed and expected distributions has been accepted at the α level of significance.

2.4. Relation between wind speed and air pollutant concentration

The relationship between air pollutant concentration, C, and wind speed, u, is derived from the ATDL model (Hanna, 1971) as:

$$C = \frac{AQ}{u} \tag{6}$$

where Q is the average area source strength of pollutant and A is an atmospheric stability factor and varies with atmospheric conditions.

A modified formula for the ATDL model is obtained by Daly and Steele (1976) as:

$$C = \frac{K}{u} \tag{7}$$

where $K = AQ$.

If the relationship of Eq. (7) holds and the frequency distribution of wind speed and air pollutant concentration are both log-normally distributed, then the relations between distributional parameters are (Bencala and Seinfeld, 1976):

$$\mu_{g,c} = K/\mu_{g,u} \tag{8}$$

$$\sigma_{g,c} = \sigma_{g,u} \tag{9}$$

where $\mu_{g,c}$ and $\sigma_{g,c}$ are the geometric mean and geometric standard deviation of the air pollutant concentration, respectively. $\mu_{g,u}$ and $\sigma_{g,u}$ are the geometric mean and geometric standard deviation of wind speed, respectively. If wind speed and pollutant data have the same type of distribution, such as the Weibull, log-normal or gamma distribution, then Eqs. (8) and (9) also hold.

When the frequency distributions of wind speed and air pollutant concentration are both log-normally distributed, the general form of Eq. (7) is described as (Simpson et al., 1984):

$$K = C_p u_{100-p} \tag{10}$$

where C_p is the air pollutant concentration corresponding to the p-percentile, and u_{100-p} is the value of wind speed corresponding to the $(100-p)$ percentile. This equation reveals that the concentration of air pollutants at cumulative probability p is inversely proportional to the wind speeds at probability $(100-p)$ when the distributional types and shape factors of both data are the same. The constancy of the product in the higher and lower percentiles is not as good as in the middle ranges, in general, since both u and C values are becoming small and experimental error is to be expected (Simpson et al., 1984). In addition, the log-normal distribution is inappropriate for the extreme percentiles (Berger et al., 1982). Therefore, only the data at the 30–80th percentile has been used in this work.

The concentration of pollutant at any percentile can be obtained as long as the statistics of wind speed data and the K value are known. Therefore, the distribution of pollutant can be estimated.

2.5. Measured data

In this work, 42 sampling data of daily average PM_{10} and $PM_{2.5}$ were measured simultaneously by a Universal air sampler (Model 310, MSP Corp., Minneapolis, USA, 1996) from July to December, 2000 at Sha-Lu city, Taiwan. Sha-Lu city is located in central Taiwan. The sampling site is located near a road with heavy traffic volume. There is a coal-fired electric plant approximately 20 km to the east from the sample site. Da Du mountain is located 4 km west of the sampling site.

The Universal sampler [Model 310 Universal Air Sampler™ Instruction Manual (USA™), 1996] is a dichotomous sampler for size fraction of airborne particles in the 0–2.5 μm and 2.5–10 μm aerodynamic size ranges. The sampler has a design inlet sampling flowrate of 300 l/min. It includes two virtual impactors for size fractionation of airborne particles. The sampler is provided with an omni-directional inlet, a PM_{10} virtual classifier and a $PM_{2.5}$ virtual classifier. The particles in the 2.5–10 μm range (coarse particle) are collected on a 62×165 mm glass microfiber filter, and those smaller than 2.5 μm are collected on a 200×250 mm glass microfiber filter. These samples are collected for a 24-h period. After the sample is collected on the filter, the samples are conditioned in the moisture-proof box at 25 °C, 40% R.H. (relative humidity) for 24 h. Then the filters are weighed three times. For the universal, there is only one sample collected in a sampling day. The PM_{10} concentration is the sum of fine and coarse particles. Therefore, the 24-h average PM_{10}, coarse particle and $PM_{2.5}$ can be obtained from this sampler. The error of the measured particle concentrations from the weighing procedure is approximately 0.38–1.62 $\mu g/m^3$. However, the experimental error of mass concentrations will be higher, especially for the small concentrations with their smaller fraction of particle weight on the filter. If there is experimental error in the sampling and weighing procedure, it will cause a large error for small mass concentrations. Therefore, only the particle concentrations at the 30–80th percentile of the cumulative probability are taken to calculate the K value in this study.

The meteorological data, consisting of 1-h average wind speeds were recorded simultaneously by a local meteorological measurement system (model 26700, Young Inc., MI, USA). The accuracy of the system can measure the wind speed to the first digit number after decimal point (0.1 m/s). The recorded data have been used to compute the daily average wind speed during the 24-h sampling period. The meteorological measurement system was approximately 10 m away from the Universal sampler, and approximately 6 m high. The sampling dates were chosen randomly during the sampling period.

After the measured data were fitted to the three theoretical distributions, the goodness of fit was judged by the K–S test. A larger D_{max} (maximum values of difference between actual and theoretic accumulated probability) value means the fitting is better. In addition, the relationship between the distributions of wind speed and air pollutant concentration was also demonstrated and the K value was calculated. Then the cumulative frequency distributions of PM_{10} and $PM_{2.5}$ were estimated from the wind speed data.

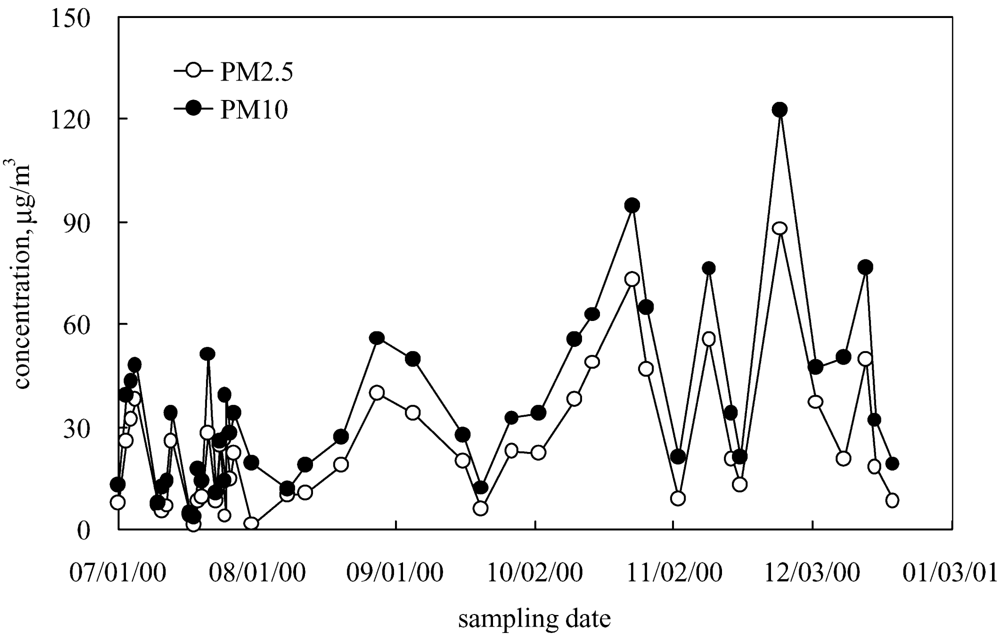


Fig. 1. The seasonal variation of PM_{10} and $PM_{2.5}$.

3. Results and discussion

3.1. Relationship between PM_{10} and $PM_{2.5}$

The sampling results for PM_{10} and $PM_{2.5}$ from July to December, 2000, at Sha-Lu are shown in Fig. 1. The higher particulate concentrations occurred in the wintertime and the lower concentration occurred in the summertime. It is seen that the variation of PM_{10} is closely related to the $PM_{2.5}$ data. The geometric standard deviation of the PM_{10}, coarse particulate and $PM_{2.5}$ samples are 2.14, 1.79 and 2.20, respectively. This shows that the geometric standard deviation of PM_{10} and $PM_{2.5}$ are very close. The difference of geometric standard deviation between fine ($PM_{2.5}$) and coarse particulate are larger than the value between PM_{10} and $PM_{2.5}$. The relation of linear regression for PM_{10} and $PM_{2.5}$ is $PM_{2.5}=0.74\times PM_{10}-2.64$ (unit: $\mu g/m^3$); the correlation coefficient, R, is 0.975. The $PM_{2.5}$ value appears to be a constant fraction of PM_{10} during the sampling period. The fine particulate ($PM_{2.5}$) is the dominant component of PM_{10} at Sha-Lu. This similar geometric standard deviation and highly correlated condition indicate that there may be common emission sources and similar meteorological influences for PM_{10} and $PM_{2.5}$. The discrepancy between the geometric standard deviations of $PM_{2.5}$ and coarse particulate suggests that they come from different emission sources.

The average and standard error of the samples are 35.3, 25.2 for PM_{10} and 24.0, 19.2 $\mu g/m^3$ for $PM_{2.5}$, respectively. The central limit theorem can be used to estimate the mean, μ, and variance, σ_p^2, of the population from the average of n samples, $\bar{x}_i$, within a confidence level $1-\alpha$, regardless of the type of distribution of population (Ott and Mage, 1981). The t and χ^2 distributions are introduced to calculate the confidence interval.

$$\mu=\bar{x}_i\pm t_{(1-\frac{\alpha}{2},\ n-1)}\frac{s_i}{\sqrt{n}} \quad (11)$$

$$\frac{(n-1)s_i^2}{\chi^2_{(1-\frac{\alpha}{2},\ n-1)}}\leq\sigma_p^2\leq\frac{(n-1)s_i^2}{\chi^2_{(\frac{\alpha}{2},\ n-1)}} \quad (12)$$

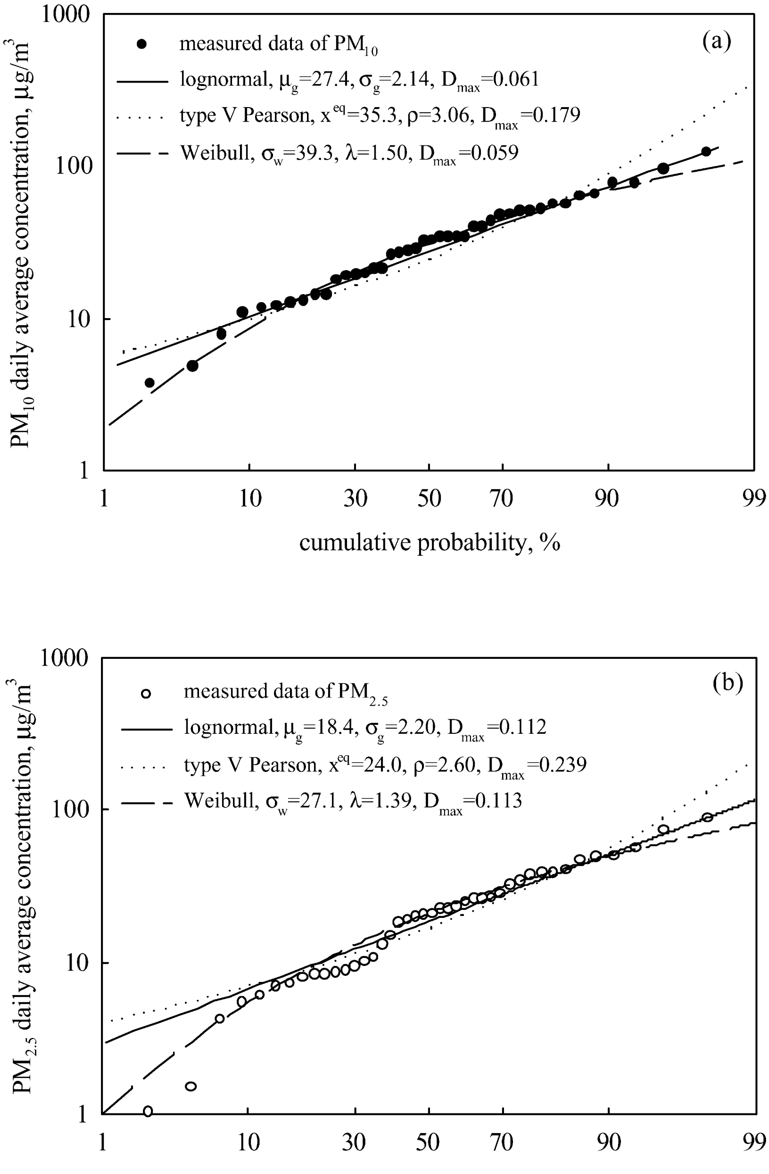


Fig. 2. The fitted results of PM_{10} and $PM_{2.5}$ for three theoretical distributions.

where s_i is the standard deviation (S.D.) of sample.

If a 95% confidence level is taken, the mean and confidence interval for PM_{10} and $PM_{2.5}$ are $\mu_{PM10}=35.3\pm 2.02\frac{25.2}{\sqrt{42}}=35.3\pm 7.8$ and $\mu_{PM2.5}=24.0\pm 6.0\ \mu g/m^3$. The ranges of S.D. of PM_{10} and $PM_{2.5}$ are $\sigma_{PM10}=20.9$–30.4 and $\sigma_{PM2.5}=16.0$–$24.8\ \mu g/m^3$, respectively. It was found that the estimated error of the mean is below 25%. If we want to increase the estimated accuracy of the confidence interval, then the sample size must increase.

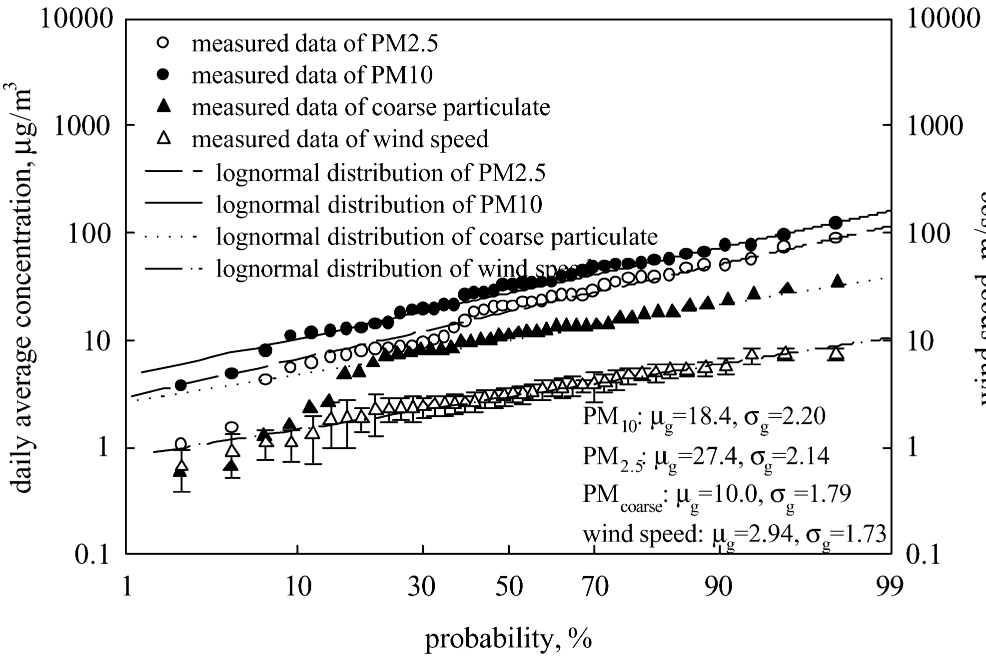


Fig. 3. The fitted results of log-normal distributions for PM_{10}, $PM_{2.5}$, coarse particulate and wind speed.

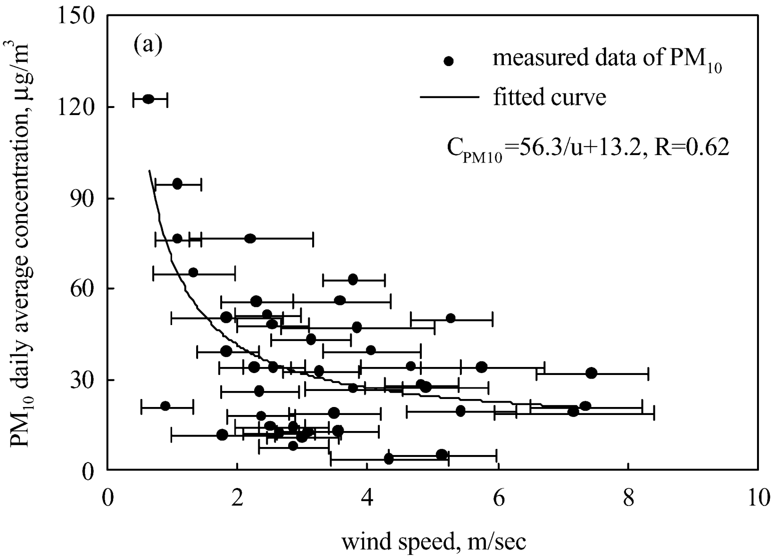
3.2. Fitted results of theoretical distribution

The fitted results of PM_{10} and $PM_{2.5}$ daily average concentrations are displayed in Fig. 2. The results of the D_{max} value for different distributions are also shown in this figure. It was found that the fitted results of the log-normal and Weibull distributions are better than the type V Pearson distribution from the D_{max} values. In the higher probability region, the type V Pearson and Weibull display overestimated and underestimated conditions, respectively. When we consider the lower probability region, the log-normal and type V Pearson show the overestimated condition. Therefore, the log-normal distribution is the most appropriate one to represent the statistical characters of PM_{10} and $PM_{2.5}$ in this work.

The D_{max} values of wind speed fitted to the log-normal, Weibull and type V Pearson distribution are 0.054, 0.068 and 0.092, respectively. It is found that the distribution of the 24-h averaged wind speed can be represented by the log-normal distribution. When the log-normal distribution is selected, the fitted results of log-normal distributions for PM_{10}, $PM_{2.5}$, coarse particulate and wind speed are displayed in Fig. 3. It is found that the shape factors of log-normal, σ_g, for air pollutant concentration and wind speed are almost the same. Therefore, the relationship between the statistical data of wind speed and air pollutant is taken to be Eq. (9). Then, this equation can be used to estimate the distributions of air pollutants from the distribution of wind speed.

3.3. Relation between wind speed data and air pollution concentration

The concentration of air pollutants is always inversely proportional to wind speed (Duijm, 1994). Fig. 4 shows the relationship between PM_{10}, $PM_{2.5}$ and wind speed. The relationship between PM_{10}, $PM_{2.5}$ and wind speed is $C_{PM10}=56.3/u+13.2$, $R=0.62$ and $C_{PM2.5}=42.0/u+7.1$, $R=0.61$, respectively. It is found that the concentration of air pollutant is inversely proportional to wind speed. However, it is clear that the data is less correlated with the model, $C=\frac{K}{u}$. If the data



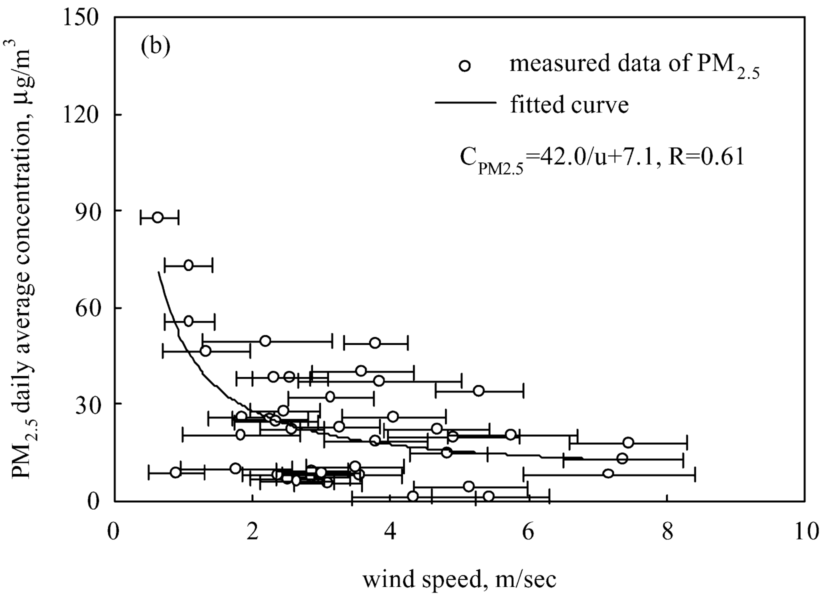


Fig. 4. The relationships between PM_{10}, $PM_{2.5}$ and wind speed data.

of wind speed are used to estimate the pollutant concentration from this relation, the estimated results will be doubtful. Therefore, the alternative method [Eq. (9)] can be taken to estimate the distributions of air pollutants from the distribution of wind speed.

Wind speed data and air pollutant concentrations are fitted to the theoretical distribution in this study. It is found that the distributions of wind speed and pollutants are all log-normal, except at the low percentile (see Fig. 3) and the geometric standard deviations are similar. Therefore, there

Table 2
The calculated K values of air pollutants at various percentiles

Percentile, P	$C_{PM2.5,P}$, ($\mu g/m^3$)	$C_{PM10,P}$, ($\mu g/m^3$)	$C_{coarse,P}$, ($\mu g/m^3$)	Wind speed, ($u_{(100-p)}$, m/s)	$K_{PM2.5}$, ($\mu g/m^2 s$)	K_{PM10}, ($\mu g/m^2 s$)	K_{coarse}, ($\mu g/m^2 s$)
5.0	1.5	4.8	0.7	7.2	10.7	34.5	4.7
10.0	5.4	10.7	1.6	5.4	29.3	58.2	8.7
15.0	6.9	12.1	2.7	5.1	35.3	62.1	13.7
20.0	7.8	12.9	5.0	4.8	37.8	62.2	24.4
25.0	8.4	17.7	7.4	4.3	36.5	76.9	32.0
30.0	9.3	19.0	8.0	3.8	35.3	71.9	30.3
35.0	10.6	20.8	8.1	3.6	38.1	74.9	29.1
40.0	14.7	25.9	9.5	3.5	51.4	90.3	33.2
45.0	18.8	27.3	9.9	3.1	58.8	85.6	31.0
50.0	20.3	31.9	10.7	3.0	61.0	95.6	32.2
55.0	22.2	33.7	11.6	2.9	63.6	96.7	33.2
60.0	22.8	33.9	11.9	2.6	58.5	87.0	30.4
65.0	25.6	39.0	13.2	2.5	64.3	97.8	33.1
70.0	26.1	43.1	13.4	2.4	62.0	102.2	31.7
75.0	34.0	49.8	14.0	2.3	78.3	114.8	32.2
80.0	38.0	51.1	15.9	1.8	70.0	94.3	29.2
85.0	39.9	55.7	17.9	1.8	70.7	98.7	31.8
90.0	48.8	64.9	20.6	1.1	53.2	70.7	22.5
95.0	55.5	76.2	23.2	0.9	51.1	70.1	21.3

In this study, only the 30–80th percentile range has been used to calculate K value since the relationship $[K = C_p u_{(100-p)}]$ holds at middle percentiles and inappropriate in extreme percentiles (Simpson et al., 1984).

must exist a constant K value for PM_{10} and $PM_{2.5}$ at Sha-Lu. The K values, calculated by Eq. (9), at various percentiles are shown in Table 2. It is noted that the relationship holds at middle percentiles and is inappropriate in extreme percentiles (Simpson et al., 1984). The K value of pollutants is almost constant from the 30th to the 80th percentiles. Therefore, only the data for the 30–80th percentile have been used.

The averaged K value can be calculated from the regression relation of C_p and $1/u_{100-p}$. The linear regression line is $y = ax$, where $y = C_p$ and $x = 1/u_{100-p}$. Therefore, the slope of the regression line, a, is equivalent to K. In this work, the data of wind speed and pollutants are taken from the 30th to 80th percentiles. The regression equations and correlation coefficients for PM_{10}, $PM_{2.5}$ and coarse particulate are: $C_{p,PM10} = 93.4/u_{(100-p)}$, $R = 0.93$; $C_{p,PM2.5} = 60.8/u_{(100-p)}$, $R = 0.89$; and $C_{p,coarse} = 31.1/u_{(100-p)}$, $R = 0.98$, respectively. The method of linear regression is least-squares. The calculated K values are 93.4, 60.8 and 31.1 $\mu g/(m^2/s)$ for PM_{10}, $PM_{2.5}$ and coarse particulate, respectively. The high correlation coefficient of the coarse particulate is due to the almost equal σ_g of wind speed and coarse particulate. The lower correlation coefficient of the PM_{10} and $PM_{2.5}$ can be caused by the small inconsistency of geometric standard deviation (σ_g) between the PM_{10}, $PM_{2.5}$ and wind speed data.

When the atmospheric condition is unchanged, and the K value is obtained, the concentration of air pollutant at different percentiles can be estimated from Eq. (9) ($C_p = K/u_{(100-p)}$). Therefore, the frequency distribution of air pollutants can be obtained. Fig. 5 shows the actual and estimated frequency distributions of PM_{10} and $PM_{2.5}$. The difference between the actual and estimated concentration is significant in the lower cumulative probability region. This is caused by the inconsistency of the wind speed data with the log-normality in the low probability region. The K–S test is used to judge the discrepancy between actual and estimated distributions. The results are shown in Table 3. The D_{max} of the PM_{10} and $PM_{2.5}$ distributions are smaller than the $D_{\alpha/2}$, $\alpha = 0.05$. Therefore, there are no significant differences at the 5% significance level. It is seen that the discrepancies between actual and estimated distributions are not

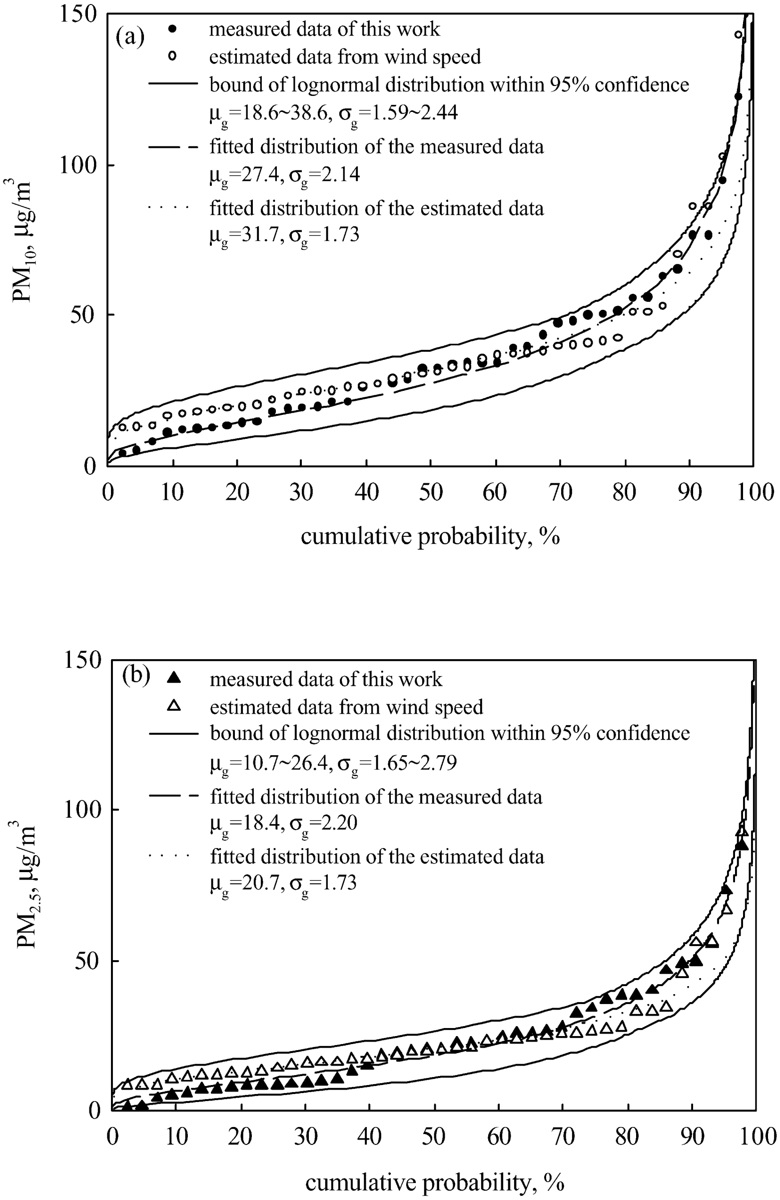


Fig. 5. The actual and estimated frequency distributions of PM_{10} and $PM_{2.5}$.

significant for PM_{10} and $PM_{2.5}$. These estimated results are acceptable.

The solid line in Fig. 5 is the bound of the lognormal distributions of population, which are estimated within a 95% confidence level. The bound line can be obtained from the following equations (Georgopoulos and Seinfeld, 1982).

$$\mu_g = \mu\left[\left(\frac{\sigma_p}{\mu}\right)^2 + 1\right]^{-1/2} \tag{13}$$

$$\sigma_g = \exp\left(\left(2\ln\left(\frac{\mu}{\mu_g}\right)\right)^{1/2}\right) \tag{14}$$

The estimated μ and σ of PM_{10} and $PM_{2.5}$ within

Table 3
Results of the K–S test of PM_{10} and $PM_{2.5}$ for the measured and estimated data by wind velocity data

Concentration (x, μg/m^3)	PM_{10}*			$PM_{2.5}$*		
	$S_1(x)$	$S_2(x)$	$D=\|S_1(x)-S_2(x)\|$	$S_1(x)$	$S_2(x)$	$D=\|S_1(x)-S_2(x)\|$
$0\leq x<15$	0.233	0.070	0.163	0.395	0.279	0.116
$0\leq x<20$	0.333	0.214	0.119	0.476	0.500	0.024
$0\leq x<30$	0.476	0.476	0.000	0.714	0.810	0.095
$0\leq x<40$	0.667	0.738	0.071	0.857	0.881	0.024
$0\leq x<50$	0.762	0.810	0.048	0.929	0.905	0.024
$0\leq x<60$	0.857	0.881	0.024	0.952	0.952	0.000
$0\leq x<70$	0.905	0.881	0.024	0.952	0.976	0.024
$0\leq x<80$	0.952	0.905	0.048	0.976	0.976	0.000
$0\leq x<90$	0.952	0.952	0.000	1.000	0.976	0.024
$0\leq x\leq\infty$	1.000	1.000	0.000	1.000	1.000	0.000
			$D_{max}=0.163$ $D_{\alpha/2}=0.205$, $\alpha=0.05$			$D_{max}=0.116$ $D_{\alpha/2}=0.205$, $\alpha=0.05$

* Sampling size = 42.

95% confidence are $\mu_{PM10}=35.3\pm7.8$, $\mu_{PM2.5}=24.0\pm6.0$ μg/m^3; $\sigma_{PM10}=20.9$–30.4, $\sigma_{PM2.5}=16.0$–24.8 μg/m^3 in this work, respectively. Therefore, the estimated μ_g and σ_g of PM_{10} and $PM_{2.5}$ within 95% confidence are $\mu_{g,PM10}=18.4$–38.8, $\mu_{g,PM2.5}=10.6$–26.5 μg/m^3; $\sigma_{g,PM10}=1.58$–2.44, $\sigma_{g,PM2.5}=1.65$–2.80 μg/m^3, respectively. Then the bound can be obtained from the estimated distributional parameters of the log-normal distribution. It is found that the estimated and actual data are all within the bound of the log-normal frequency distribution of population. It is noted that the estimated results in the low probability region are suspect because of the inconsistency of log-normality of wind speed.

4. Conclusions

In this work, three theoretical distributions (log-normal, Weibull and type V Pearson distributions) are selected to fit the PM_{10}, $PM_{2.5}$ and wind speed distributions. It is found that the log-normal distribution is the most appropriate one to represent the actual air pollutant and wind speed distribution. It is also found that PM_{10} and $PM_{2.5}$ have common emission source and the meteorological influence on the distributions of PM_{10} and $PM_{2.5}$ is similar. The relationship of $K=C_p u_{100-p}$ is also calculated and demonstrated. It is also found that the distributions of PM_{10} and $PM_{2.5}$ are successfully estimated from the wind speed data and there is no significant discrepancy at the 5% significance level between the estimated and measured distributions of pollutants. Moreover, the estimated and actual distributions of pollutant are contained within the bound of the log-normal frequency distribution of population within 95% confidence.

Acknowledgments

The financial support provided by the Hung-kuang Institute of Technology Research Council, through research contract (HKC-90-B-001) is gratefully appreciated.

References

Bencala KE, Seinfeld JH. On frequency distribution of air pollutant concentration. Atmos Environ 1976;10:941–950.
Berger A, Melice JL, Demuth Cl. Statistical distributions of daily and high atmospheric SO_2 concentrations. Atmos Environ 1982;16:2863–2877.
Burkhardt J, Sutton MA, Milford C, Storeton-West RL, Fowler D. Ammonia concentrations at a site in southern Scotland from 2 yr. of continuous measurements. Atmos Environ 1998;32:325–331.
Daly NJ, Steele LP. A predictive model for CO in Canberra. Symposium on air pollution diffusion modeling, Canberra, Australia 1976. p. 18–20.
Duijm NJ. Long-term air quality statistics derived from wind-tunnel investigations. Atmos Environ 1994;28:1915–1923.

Georgopoulos PG, Seinfeld JH. Statistical distribution of air pollutant concentration. Environ Sci Technol 1982;16:401A–416A.

Hanna SR. A simple method of calculating dispersion from urban area sources. J Air Pollut Control Ass 1971;21:774–777.

Jakeman AJ, Taylor JA, Simpson RW. Modeling distributions of air pollutant concentrations—II. Estimation of one and two parameters statistical distributions. Atmos Environ 1986;20:2435–2447.

Kao AS, Friedlander SK. Frequency distributions of PM_{10} chemical components and their sources. Environ Sci Technol 1995;29:19–28.

Mage DT, Ott WR. An evaluation of the method of fractiles, moments and maximum likelihood for estimating parameters when sampling air quality data from a stationary lognormal distribution. Atmos Environ 1984;18:163–171.

Model 310 Universal Air Sampler™ Instruction Manual (USA™) 1996. MSP corporation, 1313 Fifth Street, SE Suite 206, Minneapolis, MN, USA, 55415.

Morel B, Yen S, Cifuentes L. Statistical distribution for air pollutant applied to the study of the particulate problem in Santiago. Atmos Environ 1999;33:2575–2585.

Ott WR, Mage DT. Measuring air quality levels inexpensively at multiple locations by random sampling. J Air Pollut Control Ass 1981;31:365–369.

Seinfeld JH, Pandis SN. Atmospheric chemistry and physics: from air pollution to climate change. New York: John Wiley & Sons, 1998. p. 1271.

Simpson RW, Jakeman AJ, Daly NJ. The relationship between the ATDL and the statistical distributions of wind speed and pollution data. Atmos Environ 1984;19:75–82.

Simpson RW. A model to control emissions which avoid violations of PM10 health standards for both short and long term exposures. Atmos Environ 1990;24A:917–924.