PERGAMON

Chemosphere 52 (2003) 1079–1088

CHEMOSPHERE

www.elsevier.com/locate/chemosphere

Atmospheric lead concentration distribution in Northern Taiwan

Hsin-Chung Lu [a,*], Chuen-Jinn Tsai [b], I-Fu Hung [c]

[a] *Department of Environmental Engineering, Hungkuang University, 34 Chung-Chi Road, Sha-Lu 433, Taichung, Taiwan*
[b] *Institute of Environmental Engineering, National Chiao Tung University, 75 Poai Street, Hsin-Chu 300, Taiwan*
[c] *Department of Atomic Science, National TsingHua University, 101 Section 2 Kuang Fu Road, Hsin-Chu 300, Taiwan*

Received 13 February 2002; received in revised form 26 February 2003; accepted 26 February 2003

Abstract

Atmospheric lead concentrations were measured randomly, approximately once per week, at five traffic sites in northern Taiwan from September 1994 to May 1995. Three types of theoretical distributions, lognormal, Weibull and gamma were selected to fit the frequency distribution of the measured lead concentration. Four goodness-of-fit criteria were used to judge which theoretical distribution is the most appropriate to represent the frequency distributions of atmospheric lead.

The results show that atmospheric lead concentrations in total suspended particulates fit the lognormal distribution reasonably well in northern Taiwan. The intervals of fitted theoretical cumulative frequency distributions (CFDs) can successfully contain the measured data when the population mean is estimated with a 95% confidence interval. In addition, atmospheric lead concentration exceeding a critical concentration is also predicted from the fitted theoretical CFDs.

Keywords: Goodness-of-fit; Lognormal; Weibull; Gamma; Cumulative frequency distributions

1. Introduction

Lead pollution is a problem in urban cities because of its detrimental effects on human health. About 80% of atmospheric lead results from automobile exhaust (Simpson and Xu, 1994). Knowledge of the frequency distribution of lead concentration is important to assist in identifying the potential human exposure to lead and developing corresponding control strategies.

The concentrations of atmospheric lead are usually random variables, and are correlated with the emission levels, meteorological conditions and geography. Many types of distributions have been used to fit the air pollutant concentration data. These distributions are lognormal distribution (Georgopoulos and Seinfeld, 1982; Mage and Ott, 1984; Jakeman et al., 1986; Kao and Friedlander, 1995), Weibull distribution (Georgopoulos and Seinfeld, 1982; Jakeman et al., 1986), type V Pearson distribution (Morel et al., 1999) and gamma distribution (Berger et al., 1982; Holland and Fitz-Simons, 1982).

The results of previous works showed the lognormal distribution could represent the frequency distribution of atmospheric lead (Simpson and Xu, 1994; Kao and Friedlander, 1995). One research study pointed out that the Weibull distribution is a good fit result to 1971 hourly averaged oxidant data from Pasadena, California (Georgopoulos and Seinfeld, 1982). The type V Pearson distribution was used to fit the measured data

* Corresponding author. Tel.: +886-4-26318652x4106; fax: +886-4-26525245.
E-mail address: hclu@sunrise.hk.edu.tw (H.-C. Lu).

0045-6535/03/$ - see front matter © 2003 Elsevier Science Ltd. All rights reserved.
doi:10.1016/S0045-6535(03)00264-9

of PM_{10}, $PM_{2.5}$, SO_2, NO_2 in Santiago (Morel et al., 1999). Therefore, there is no universal distribution that can represent the air pollutant frequency distribution.

The frequency distribution of air pollutants obtained from continuous monitoring can be used to estimate the mean and maximum concentrations of pollutants. However, continuous monitoring is expensive. Random sampling is an alternative to estimate the arithmetic mean and maximum concentration of pollutants with an accepted level of confidence. The random sampling results of CO, total suspended particulates (TSP) and SO_2 have been found to agree well with the results of continuous monitoring for estimating the mean and maximum concentration of pollutants (Ott and Mage, 1981; Simpson, 1984). However, it is noted that the random sampling must be taken over a sufficiently long time to include the cycle of diurnal and seasonal fluctuations of the sample period to obtain representative data.

Therefore, the purposes of this work are (1) To fit the random sampling data (Pb in TSP) to a theoretical frequency distribution and judge which theoretical distribution is the most appropriate to represent the data. In this study, three types of distributions, lognormal, Weibull and gamma were selected to fit the frequency distribution of the measured lead concentration in TSP. The RMSE, index of agreement (d), log likelihood and K-S test were used as the goodness-of-fit criteria. (2) To predict the days when atmospheric lead concentration exceeds a critical concentration from the most appropriate theoretical distribution.

2. Methods

2.1. Experimental method

The 24 h averaged lead concentration in TSP was measured randomly using a TSP high volume sampler (Model: GS2310, Graseby Inc., USA), approximately once per week, from September 1994 to May 1995 at five sites in northern Taiwan. All of the five sites are located at traffic monitoring stations. The number of samples obtained by the TSP sampler are 33, 39, 37, 32 and 35 at the Ji-Long, Da-Tong, Shan-Chung, Yong-Ho and Chung-Li sites, respectively. In order to determine the relationship between Pb in TSP and Pb in PM_{10}, the high volume PM_{10} samplers (Model: ASI/GMW 1200, Andersen Instruments, Inc., USA) were operated simultaneously at the Da-Tong site during the sampling period, and the number of PM_{10} samples was 25. The Ji-Long site is located about 12 m high and about 30 m away from a heavy-traffic main road. Ji-Long City is at the most northern part of Taiwan. The Da-Tong and Shan-Chung sites are located in Taipei City, which is a typical urban area that has high population density and heavy traffic. The Da-Tong site is located at the intersection of Min-Chen Road and Chung-Chim Road at the city center. The site is about 4.6 m high and 2 m away from the Min-Chen Road. The Shan-Chung site is located at the intersection of the highway exit and Shan-Ho Road. There is a park in the northeastern direction. The site is 4.4 m high and 3 m away from the highway exit. The Yong-Ho site is located in Taipei County, 4.4 m high and 5.6 m away from a main traffic road. The Chung-Li site is located in Tau-Yuan County, 4.4 m high and 4.5 m away from a main road. All the sampling sites are surrounded by buildings except the Shan-Chun site which has a park in the northeastern direction.

Quartz filter papers (Model 2500 QAT-UT, Pall Inc., USA) were used to sample Pb in TSP and in PM_{10}. These samples were pre-treated by microwave digestion (Model MDS-2000, CEM Inc., USA) before being analyzed by a flame AA (atomic absorption spectroscopy, Model 171-8001, HITACHI Inc., Japan). QA/QC results showed that the detection limit of flame AA is 0.024 ppm. The recovery of lead was between 89 and 98% and the replicate AA analysis precision was determined to be 3.6%.

2.2. Probability density functions of distributions

The air pollutant concentration data are often autocorrelated and may randomly fluctuate due to the changes of meteorological conditions and pollutant emissions. Nevertheless, the assumption of independent and identically distributed (i.i.d) pollutant data (stationary process) is useful in simplified statistical analysis. Although the i.i.d assumption is not a strictly valid one, it can be applied to real situations and often leads to satisfactory agreement with observations (Georgopoulos and Seinfeld, 1982).

Three theoretical distributions, namely lognormal, Weibull and gamma are considered in this study. The distribution properties are as follows.

2.2.1. Lognormal distribution

The probability density function (p.d.f.) of the lognormal distribution, $p_l(x_i)$, is (Devore, 2000)

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

where x_i is the pollutant concentration of species i, and μ_g and σ_g are the parameters of the lognormal distribution and represent the geometric mean and the standard geometric deviation, respectively.

2.2.2. Weibull distribution

The p.d.f. of the Weibull distribution, $p_w(x_i)$, is (Devore, 2000)

$$p_w(x_i) = \frac{\lambda_w}{\sigma_w}\left(\frac{x_i}{\sigma_w}\right)^{\lambda_w-1}\exp\left[-\left(\frac{x_i}{\sigma_w}\right)^{\lambda_w}\right] \tag{2}$$

where λ_w and σ_w are the parameters of the Weibull distribution.

2.2.3. Gamma distribution

The p.d.f. of the gamma distribution, $p_g(x_i)$, is (Devore, 2000)

$$p_g(x_i) = \frac{1}{\sigma_{ga}\Gamma(\lambda_{ga})}\left(\frac{x_i}{\sigma_{ga}}\right)^{\lambda_{ga}-1}\exp\left(-\frac{x_i}{\sigma_{ga}}\right) \tag{3}$$

where Γ is the gamma function, and σ_{ga} and λ_{ga} are the parameters of the gamma distribution.

By finding the most appropriate distribution to fit the measured data, then the mean of the pollutant concentration and percentiles exceeding a critical concentration can be directly estimated from the theoretical distribution. The main disadvantage of these statistical models is that theoretical distributions do not fit the extreme concentrations well (Berger et al., 1982). Therefore, they must be used carefully in predicting the probability at higher concentrations.

2.3. Estimation of distribution parameters

There are many techniques that can be used to estimate the distribution parameters. The maximum likelihood estimate is nearly unbiased and has the minimal variances possible for an unbiased estimate for large samples (Holland and Fitz-Simons, 1982; Mage and Ott, 1984). Therefore, the maximum likelihood estimate was chosen to estimate the distribution parameters of lead. If the p.d.f. of a theoretical distribution is $p(x_i, \theta_1, \theta_2)$, where θ_1, θ_2 are the parameters of distribution, 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 logarithms 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; Mage and Ott, 1984). These equations for estimating distribution parameters by the method of maximum likelihood are shown in Table 1. The equations for the Weibull and gamma distributions can be solved by numerical methods.

2.4. Goodness-of-fit criteria

Statistical indicators were used to evaluate the goodness-of-fit for the theoretical distributions. These statistical indicators are index of agreement (d), root mean square error (RMSE), log likelihood ($\ln(L)$) and K-S test.

A measure of error called the index of agreement, d, was calculated according to Eq. (5) (Willmott, 1982; Kolehmainen et al., 2001). The index is a relative and bounded measure which can make cross-comparisons between models and is limited to the range of 0–1. The index of agreement is

$$d = 1 - \frac{\sum_{i=1}^{N}(P_i - O_i)^2}{\sum_{i=1}^{N}(|P_i'| + |O_i'|)^2} \tag{5}$$

where $P_i' = P_i - \overline{O}$ and $O_i' = O_i - \overline{O}$, N is the number of data, O_i the observed pollutant concentration, P_i the predicted pollutant concentration, and $\overline{O}$ is the arithmetic mean of the observed concentration. When the index of agreement is closer to 1, then the model is more appropriate to simulate the experimental data.

The RMSE is the most common indicator, which is

Table 1
Equations for estimating parameters of three theoretical distributions

Distribution type	Equations for estimating distribution parameters
Lognormal[a]	$\ln\mu_g = \frac{1}{n}\sum_{i=1}^{n}\ln x_i,\ (\ln\sigma_g)^2 = \frac{1}{n}\sum_{i=1}^{n}(\ln x_i - \ln\mu_g)^2$
Weibull[a]	$\lambda_w = \left[\left(\sum_{i=1}^{n}x_i^{\lambda_w}\ln x_i\right)\left(\sum_{i=1}^{n}x_i^{\lambda_w}\right)^{-1} - \frac{1}{n}\sum_{i=1}^{n}\ln x_i\right]^{-1},\ \sigma_w = \left(\frac{1}{n}\sum_{i=1}^{n}x_i^{\lambda_w}\right)^{1/\lambda_w}$
Gamma[b]	$\ln\lambda_{ga} = \frac{1}{\Gamma(\lambda_{ga})}\frac{d\Gamma(\lambda_{ga})}{d\lambda_{ga}} + \frac{1}{n}\sum_{i=1}^{n}\ln x_i - \ln\left(\frac{1}{n}\sum_{i=1}^{n}x_i\right),\ \sigma_{ga} = \frac{1}{\lambda_{ga}}\frac{1}{n}\sum_{i=1}^{n}x_i$

μ_g and σ_g are the parameters of the lognormal distribution, σ_w and λ_w are the parameters of the Weibull distribution, σ_{ga} and λ_{ga} are the parameters of the gamma distribution, Γ is the gamma function.

[a] See reference Georgopoulos and Seinfeld (1982).

[b] See reference Jakeman et al. (1986).

$$\mathrm{RMSE} = \left[\frac{1}{N}\sum_{i=1}^{N}(P_i - O_i)^2\right]^{1/2} \tag{6}$$

For a good model, the RMSE should approach zero. Therefore, a smaller RMSE means that the model is more appropriate.

The maximum likelihood method can also be used as a criterion for evaluating the goodness-of-fit of different distributions. The one with maximum L is considered optimal among different distributions (Holland and Fitz-Simons, 1982). The log likelihood is defined as

$$\ln L = \sum_{i=1}^{n} \ln p(x_i, \theta_1, \theta_2) \tag{7}$$

The K-S statistic ($D_{\max}$) is defined as the maximum difference between the sample cumulative probability and the expected cumulative probability, and can be written as

$$D_{\max} = \max |F(x_i) - S(x_i)| \tag{8}$$

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

$D_{\max}$ is 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 $D_{\alpha/2}$, we say that the hypothesis that there is no difference between the observed and expected distributions is accepted at an α level of significance. A smaller $D_{\max}$ indicates the fitted theoretical distribution is more adequate.

2.5. Method to estimate the number of days exceeding a critical concentration

2.5.1. Confidence interval of population mean

If the observations of random sampling are independent, the distribution of the average of the samples is normal regardless of the distribution type of population. Then the central limit theorem can be applied to estimate the confidence interval of the arithmetic mean of the population (Ott and Mage, 1981). The confidence interval of the population mean can be expressed as (Hansen et al., 1964)

$$\overline{x_i} - Z_{(1-\alpha/2)}\sqrt{\left(1-\frac{n}{N}\right)\frac{\hat{s}^2}{n}} \leqslant \mu \leqslant \overline{x_i} + Z_{(1-\alpha/2)}\sqrt{\left(1-\frac{n}{N}\right)\frac{\hat{s}^2}{n}} \tag{9}$$

where N and n are the population and sample size, μ is the arithmetic mean of the population, $\bar{x}_i$ and $\hat{s}$ are the average and standard error of the samples, and $1-\alpha$ is the confidence level of the Z-distribution.

2.5.2. Relationships between the theoretical distribution parameters and the mean and standard deviation of the population

There exist simple relationships between the theoretical distribution parameters (lognormal, gamma and Weibull) and the mean, μ, and standard deviation, σ_p, of the population. The relationships are shown in Table 2 (Devore, 2000).

If σ_p is replaced by the sample standard error, $\hat{s}$, and μ is estimated from the random sampling data within the $1-\alpha$ confidence level by Eq. (9), then the values of the distribution parameters of the population can be calculated from the equations shown in Table 2. From the estimated distributional parameters, the intervals of theoretical cumulative frequency distributions (CFDs) of the population for different distributions where the population mean is estimated within a $1-\alpha$ confidence level can be computed by integrating Eqs. (1)–(3). Finally, the intervals of probabilities and number of days, when lead concentration exceeds a critical concentration are predicted from the estimated CFDs of the population.

3. Results and discussion

3.1. Measured results of atmospheric Pb concentrations

The time series of samples for Pb in TSP at five sites is shown in Fig. 1, where it indicates no apparent seasonal variation for Pb concentration in TSP. These five sites are located at traffic stations, therefore the dispersion of Pb is affected by the traffic-induced turbulence. It is found that the Pb pollution at the Da-Tong and Shan-Chung sites is more severe than other sites since the

Table 2
The relationships between μ and σ_p and the parameters of the theoretical distributions

Distribution	Relationship between distributional parameters[a]	
Lognormal	$\mu = \exp\left(\ln\mu_g + \frac{(\ln\sigma_g)^2}{2}\right)$	$\sigma_p^2 = e^{2\ln\mu_g + (\ln\sigma_g)^2}\left(e^{(\ln\sigma_g)^2} - 1\right)$
Weibull	$\mu = \sigma_w \Gamma\left(1 + \frac{1}{\lambda_w}\right)$	$\sigma_p^2 = \sigma_w^2\left\{\Gamma\left(1+\frac{2}{\lambda_w}\right) - \left[\Gamma\left(1+\frac{1}{\lambda_w}\right)\right]^2\right\}$
Gamma	$\mu = \lambda_{ga}\sigma_{ga}$	$\sigma_p^2 = \lambda_{ga}\sigma_{ga}^2$

[a] See reference Devore (2000).

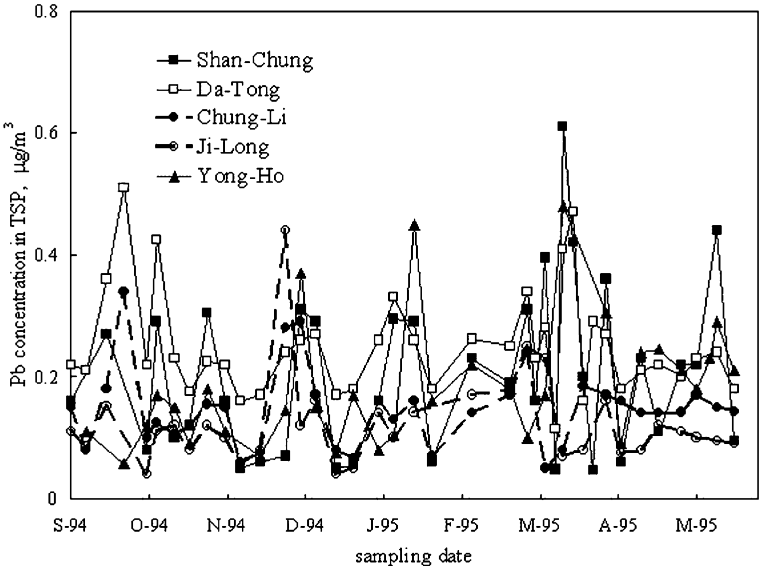

Fig. 1. The variation of Pb in TSP at five sampling sites.

traffic flowrates at these two sites are heavier. The Pb concentrations are all less than the Taiwan EPAs ambient air quality standard for lead, 1.0 μg/m^3 on a monthly average basis, indicating that the lead pollution in northern Taiwan is not severe.

The Pb concentration in PM_{10} was also measured at the Da-Tong site and was found to be closely related to the Pb in TSP. The geometric standard deviation of the Pb in PM_{10} and Pb in TSP are 1.50 and 1.37, respectively. The regression between Pb in PM_{10} ($Pb_{PM_{10}}$) and in TSP (Pb_{TSP}) gives $Pb_{PM_{10}} = 0.8 \times Pb_{TSP}$–0.02, $R^2 = 0.88$, suggesting that $Pb_{PM_{10}}$ is a constant fraction of Pb_{TSP} during the sampling period. The average ratio of $Pb_{PM_{10}}/Pb_{TSP}$ is 0.72 ± 0.10, indicating that the major part of the Pb exists in the PM_{10} (particle sizes that are more easily respirable). This result is expected since most atmospheric lead is generated from the automobile exhaust consisting mainly of fine particles.

3.2. Frequency distributions of sample Pb

In order to fit the theoretical distributions to the sample data, the assumption of i.i.d. sample data is necessary and the run test is used to evaluate the randomness of the samples. The runs (R) of samples at five sites can be obtained from Fig. 1. When a sequence of n observations is generated by a random process (stationary and independent), the sampling distribution of R has the mean and variance (Neter et al., 1988) as

$$E\{R\} = \frac{2n-1}{3}, \qquad \sigma^2\{R\} = \frac{16n-29}{90} \tag{10}$$

where n is the sample size, $E\{R\}$ the mean of R, and $\sigma^2\{R\}$ the variance of R.

When n is greater than 20, the distribution of R is approximately normal. The standardized test statistic, Z^*, is

$$Z^* = \frac{R - E\{R\}}{\sigma\{R\}} \tag{11}$$

If $Z^* \leqslant Z_\alpha$, we conclude that the observations have been rejected at an α significance level. Table 3 provides the statistics of the run test for the atmospheric Pb concentration at five sites. The run test results indicated that the null hypothesis at all sites is accepted at the 0.05 significance level except at the Chung-Li site. Therefore, the i.i.d. assumption for Pb concentrations at all sites except the Chung-Li site is valid. Although, the assumption of i.i.d. pollutant data at Chung-Li is not valid, it can also be applied to real situations and often leads to satisfactory agreement with observations (Georgopoulos and Seinfeld, 1982). The effect of not i.i.d. condition only increases the variance in distribution parameter estimates and should not introduce additional bias. Thus, the distribution parameters of samples can be estimated by the equations in Table 1.

The actual CFDs of the Pb samples at the five sites are shown in Fig. 2, with the estimated distribution parameters shown in Table 4. The fitted probability density functions (p.d.f.) for the measured data at the Yong-Ho and Shan-Chung sites are shown in Fig. 3. The fitted probability density functions of other sites are similar. Fig. 3 illustrates that the lognormal and gamma distributions seem to be more appropriate to represent the Pb distributions. The appropriateness must be further judged by the goodness-of-fit criteria as described in the later section.

Table 3
The run test for randomness for atmospheric Pb at five sample sites

Sampling site	Sample size	Run	$E\{R\}$	$\sigma^2\{R\}$	Z^*	
Chung-Li	35	18	23.00	5.90	−2.058	<−1.645*
Ji-Long	33	19	21.67	5.54	−1.132	>−1.645
Da-Tong	39	25	25.67	6.61	−0.259	>−1.645
Shan-Chung	37	24	24.33	6.26	−0.133	>−1.645
Yong-Ho	32	19	21.00	5.37	−0.863	>−1.645

$^*Z_{0.05} = -1.645$.

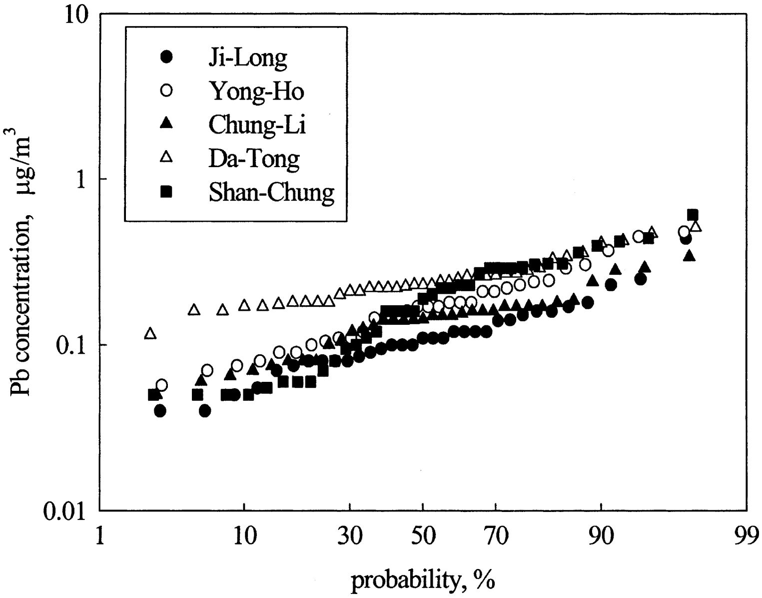

Fig. 2. The cumulative distribution function of Pb in TSP at five sampling sites.

Table 4
The estimated distribution parameters from sampling data by the method of maximum likelihood (parameters of the theoretical distributions)

Sampling site	Lognormal	Weibull	Gamma
Chung-Li	$\mu_g = 0.13$, $\sigma_g = 1.55$	$\sigma_w = 0.17$, $\lambda_w = 2.41$	$\sigma_{ga} = 0.027$, $\lambda_{ga} = 5.50$
Ji-Long	$\mu_g = 0.11$, $\sigma_g = 1.64$	$\sigma_w = 0.14$, $\lambda_w = 1.82$	$\sigma_{ga} = 0.031$, $\lambda_{ga} = 3.93$
Da-Tong	$\mu_g = 0.24$, $\sigma_g = 1.37$	$\sigma_w = 0.28$, $\lambda_w = 3.00$	$\sigma_{ga} = 0.025$, $\lambda_{ga} = 10.0$
Shan-Chung	$\mu_{g1} = 0.07$, $\sigma_{g1} = 1.44$	$\sigma_{w1} = 0.08$, $\lambda_{w1} = 2.35$	
(bimodal)	$\mu_{g2} = 0.27$, $\sigma_{g2} = 1.46$	$\sigma_{w2} = 0.31$, $\lambda_{w2} = 3.16$	
Yong-Ho	$\mu_g = 0.16$, $\sigma_g = 1.69$	$\sigma_w = 0.21$, $\lambda_w = 1.95$	$\sigma_{ga} = 0.049$, $\sigma_{ga} = 3.78$

It is also found that the frequency distributions of Pb are all unimodal except the data of the Shan-Chung site, which is bimodal. Kao and Friedlander (1995) showed that the Pb in PM_{10} at Durate area also has a unimodal lognormal distribution and its geometric standard deviation is about 2. The shape factors (σ_g) of the theoretical lognormal distributions at different sites are almost the same (1.37–1.69, see Table 4). This means that meteorological conditions and emission sources for the Pb concentration are similar at different sites except the Shan-Chung site where the traffic condition and surrounding environment are different. Unlike other sites where only main roads are surrounding the sites, in the Shan-Chung site, there is a park in the northeastern

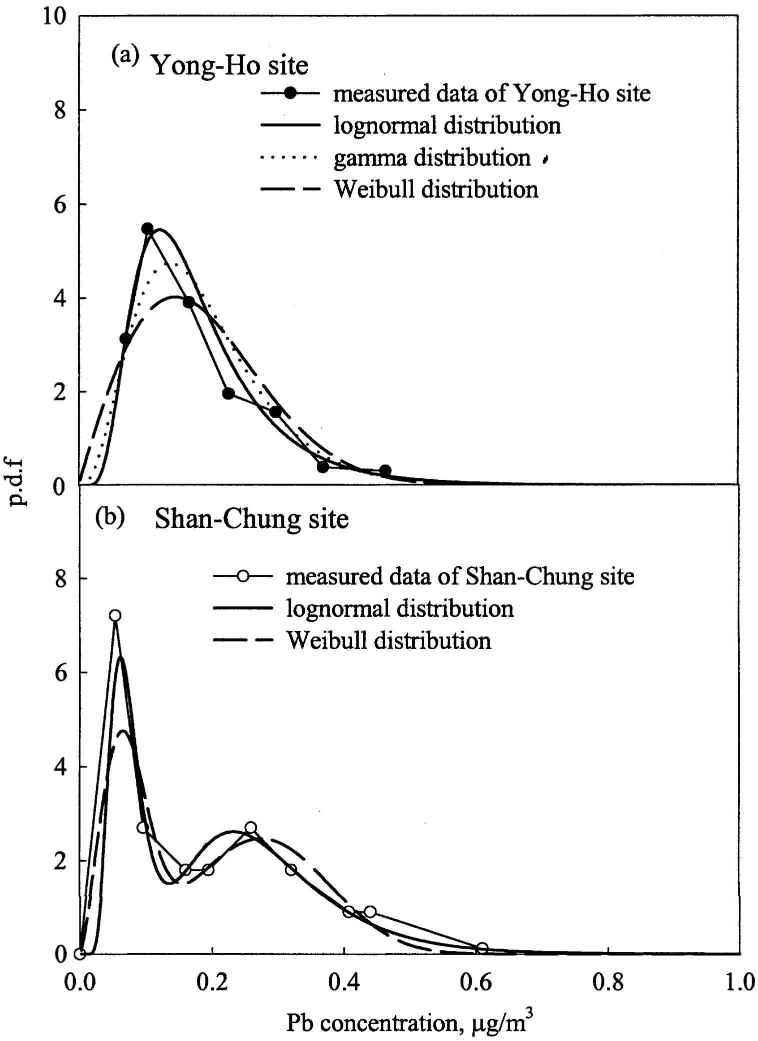

Fig. 3. Comparison of the actual and theoretical p.d.f. of Pb concentration at (a) Yong-Ho (b) Shan-Chung sites.

direction and a high way and a main road in other directions. The first mode of Pb concentration distribution of lower concentration is caused by the particles from the park, while the second mode of higher concentration is caused by the particles emitted from the vehicles.

3.3. Results of goodness-of-fit

The goodness-of-fit criteria for fitted theoretical distributions are displayed in Table 5. When the index of agreement is closer to 1, then the model is more appropriate to simulate the experimental data. A smaller RMSE means that the model is more appropriate. A smaller D_{max} indicates the fitted theoretical distribution is more adequate. For the maximum likelihood, the one with maximum L is considered optimal among different distributions. It is seen from Table 5 that all indicators show that the lognormal distribution is the best one to represent the distributions of Pb at the Ji-Long, Da-Tong and Yong-Ho sites, the gamma distribution is the next best, and the Weibull is the least appropriate. The RMSE and d values show that the lognormal distribution is adequate for Pb concentration at the Chung-Li site. However, it shows that the gamma distribution is the most adequate distribution from the D_{max} and log likelihood values. The fitted results also show that the bimodal lognormal distribution can fit the measured data successfully at the Shan-Chung site. Table 5 also reveals that the D_{max} values of the five sites are all less

Table 5
The goodness-of-fit criteria for three theoretical distributions

Site	Goodness-of-fit criteria												
	Lognormal				Weibull				Gamma				$\alpha = 0.05$
	RMSE	d	$\ln(L)$	$D_{\max}$	RMSE	d	$\ln(L)$	$D_{\max}$	RMSE	d	$\ln(L)$	$D_{\max}$	$D_{\alpha/2}$
Ji-Long	0.029 (1)	0.948 (1)	49.75 (1)	0.035 (1)	0.032 (3)	0.943 (2)	44.90 (3)	0.090 (3)	0.032 (2)	0.933 (3)	47.97 (2)	0.066 (2)	0.231
Chung-Li	0.017 (1)	0.978 (1)	49.50 (2)	0.100 (2)	0.020 (3)	0.973 (3)	48.08 (3)	0.127 (3)	0.018 (2)	0.977 (2)	49.61 (1)	0.094 (1)	0.224
Da-Tong	0.012 (1)	0.999 (1)	45.94 (1)	0.076 (1)	0.025 (3)	0.998 (3)	40.60 (3)	0.116 (3)	0.014 (2)	0.999 (2)	44.74 (2)	0.079 (2)	0.213
Shan-Chung	0.028 (1)	0.995 (1)	32.09 (1)	0.038 (1)	0.033 (2)	0.984 (2)	28.68 (2)	0.052 (2)					0.218
Yong-Ho	0.021 (1)	0.987 (1)	33.75 (1)	0.044 (1)	0.027 (3)	0.980 (3)	31.30 (3)	0.088 (3)	0.026 (2)	0.981 (2)	32.93 (2)	0.059 (2)	0.234

RMSE: root mean square error, d: index of agreement, $\ln(L)$: log likelihood, $D_{\max}$: maximum difference between the sample cumulative probability and the expected cumulative probability, $D_{\alpha/2}$: theoretic difference acceptable for the K-S test at a certain significance level, α.

than $D_{\alpha/2}$ within a 95% confidence level. Therefore, the hypothesis that there is no difference between the observed and three theoretical distributions is accepted at a 95% confidence level.

3.4. The interval of theoretical cumulative frequency distribution of the Pb population

From the relationships between μ, σ_p and the parameters of the theoretical distributions (see Table 2), the distributional parameters of the population can be calculated when the population mean is estimated within a $1-\alpha$ confidence level. For example, the confidence interval of the population mean of Pb is $0.126 \leqslant \mu_p \leqslant 0.166$ and the $\hat{s} = 0.065$ μg/m^3 within a 95% confidence level at the Chung-Li site. If the distribution of the population is lognormal, then the distributional parameters equal $\mu_{gH} = 0.11$, $\sigma_{gH} = 1.63$ at the high bound and $\mu_{gL} = 0.15$, $\sigma_{gL} = 1.46$ at the low bound. The interval of the cumulative distribution function for the lognormal distribution can be estimated from these two sets of parameters by integrating Eq. (1) while the confidence interval of the population mean is estimated within a 95% confidence level. The predicted intervals of the cumulative distribution functions for three theoretical distributions at the Yong-Ho and Chung-Li sites are displayed in Fig. 4. The results are similar for other sites. It is found that the measured data are within the intervals of the three estimated CFDs. The intervals of the different CFDs can successfully contain the measured data.

3.5. Predicting the probability and number of days exceeding a criterion by the CFD of Pb

The probability that the Pb concentration exceeds a critical concentration can be obtained from Fig. 4. For example, the probability that the Pb concentration less than 0.4 μg/m^3 at the Yong-Ho site can be read from Fig. 4(a) and its value is between 0.938 and 0.974 for the lognormal distribution. The probability of the Pb concentration exceeding 0.4 μg/m^3 is 0.026–0.062. Therefore, the estimated exceeding day in one year is 10 to 23 days for the lognormal distribution. The estimated probabilities and number of days of the population when the Pb concentration exceeds 0.4 μg/m^3 at the five sites for different theoretical distributions are shown in Table 6. It is found that the intervals of exceeding days are large at some sampling sites (e.g. the Yong-Ho site). However, the estimated intervals of the CFD of the population are dependent on the sample size. As the sample size increases, the confidence interval of the population mean will decrease and the interval of the CFD will be smaller. Then the predicted intervals of probabilities and number of days will become narrower, and the predicted results will be more accurate.

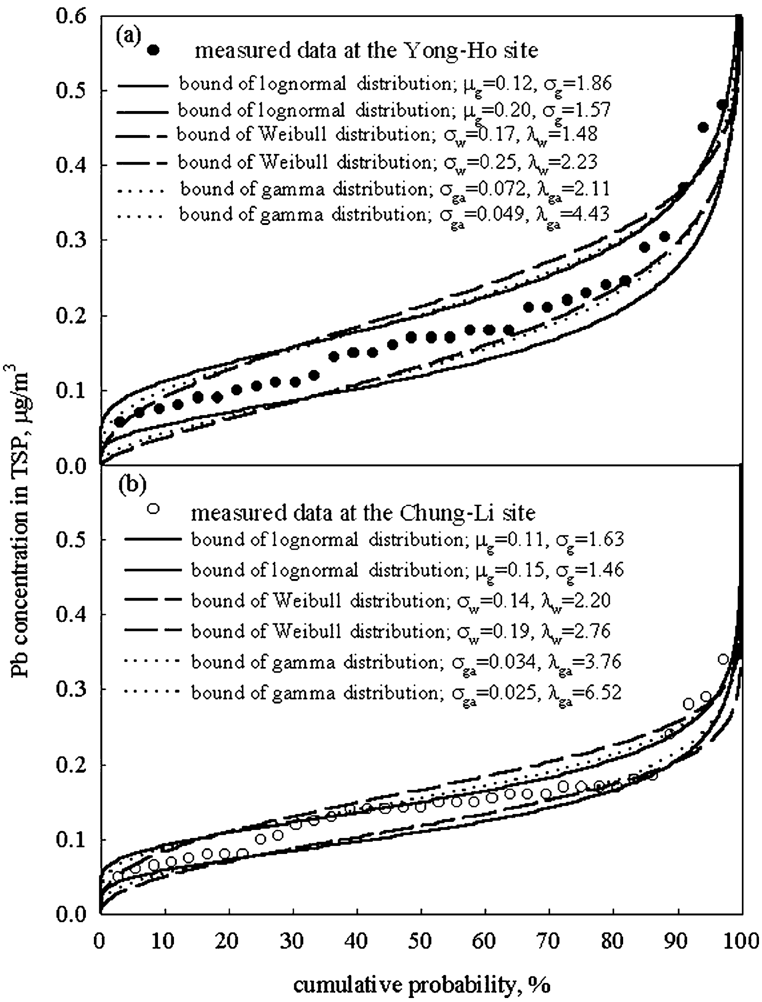

Fig. 4. The bound of the CFDs of population for lead concentration at (a) Yong-Ho (b) Chung-Li sites.

Table 6
The predicted exceeding probabilities and number of days with Pb > 0.4 μg/m^3 for different theoretical distributions

Sample site	The probability that Pb > 0.4 μg/m^3 for different theoretical distributions (exceeding number of days in one year)					
	Lognormal		Gamma		Weibull	
	Low bound	High bound	Low bound	High bound	Low bound	High bound
Chung-Li	0.004 [2]	0.005 [2]	<0.001 [0]	<0.001 [0]	0.002 [1]	0.002 [1]
Ji-Long	0.008 [3]	0.010 [4]	0.004 [1]	0.003 [1]	0.005 [2]	0.006 [2]
Da-Tong	0.041 [15]	0.081 [29]	0.023 [8]	0.086 [31]	0.037 [14]	0.082 [30]
Shan-Chung	0.060 [22]	0.107 [39]	0.060 [22]	0.124 [45]	0.062 [23]	0.124 [45]
Yong-Ho	0.026 [10]	0.062 [23]	0.029 [11]	0.058 [21]	0.030 [11]	0.057 [21]

4. Conclusions

The lead concentrations in TSP were measured randomly at five traffic stations in the northern Taiwan. The lognormal, Weibull and gamma distributions were chosen to fit the measured data at these five sites. Then the RMSE, degree of agreement, log likelihood and K-S test were used as the goodness-of-fit criteria to judge which distribution is adequate to represent the frequency distribution of lead. In addition, the intervals of probabilities and number of days of the population on which lead concentration exceeds a critical concentration were predicted from the estimated theoretical CFDs of the population.

It is found that atmospheric lead concentrations in TSP fit the lognormal distribution reasonably well in northern Taiwan. From the random sampling data, the intervals of different CFDs of the population can be predicted when the confidence interval of the population mean is estimated within a 95% confidence level. The estimated intervals of CFDs can successfully contain the measured data. Knowledge of the population distribution of air pollutants obtained in this study is useful in identifying the potential human exposure to air pollutants and developing corresponding control strategies.

Acknowledgement

Financial support provided by the Taiwan EPA-84-E3L1-09-09 is gratefully appreciated.

References

Berger, A., Melice, J.L., Demuth, Cl., 1982. Statistical distributions of daily and high atmospheric SO_2 concentrations. Atmospheric Environment 16 (12), 2863–2877.

Devore, J.L., 2000. Probability and Statistics for Engineering and the Science, fifth ed. Duxbury, USA.

Georgopoulos, P.G., Seinfeld, J.H., 1982. Statistical distribution of air pollutant concentration. Environmental Science and Technology 16 (7), 401A–416A.

Hansen, M.H., Harwitz, W.N., Madow, W.G., 1964. Sample Survey Methods and Theory, vol. II. Wiley, New York.

Holland, D.M., Fitz-Simons, T., 1982. Fitting statistical distributions to air quality data by the maximum likelihood method. Atmospheric Environment 16 (5), 1071–1076.

Jakeman, A.J., Taylor, J.A., Simpson, R.W., 1986. Modeling Distributions of air pollutant concentrations-II. Estimation of one and two parameters statistical distributions. Atmospheric Environment 20 (12), 2435–2447.

Kao, A.S., Friedlander, S.K., 1995. Frequency distributions of PM10 chemical components and their sources. Environmental Science and Technology 29 (1), 19–28.

Kolehmainen, M., Martikainen, H., Ruuskanen, J., 2001. Neural networks and periodic components used in air quality forecasting. Atmospheric Environment 35 (5), 815–825.

Mage, D.T., Ott, W.R., 1984. An evaluation of the method of fractiles, moments and maximum likelihood for estimating parameters when sampling air quality data from a stationary lognormal distribution. Atmospheric Environment 18 (1), 163–171.

Morel, B., Yen, S., Cifuentes, L., 1999. Statistical distribution for air pollutant applied to the study of the particulate problem in Santiago. Atmospheric Environment 33 (16), 2575–2585.

Neter, J., Wassennan, W., Whitemore, G.A., 1988. Applied Statistics. Allyn and Bacon, Inc., Boston.

Ott, W.R., Mage, D.T., 1981. Measuring air quality levels inexpensively at multiple locations by random sampling. Journal of Air Pollutants and Control Assessments 31 (4), 365–369.

Simpson, R.W., 1984. Predicting frequency distributions for ozone, NO_2 and TSP from restricted data sets. Atmospheric Environment 18 (2), 353–360.

Simpson, R.W., Xu, H., 1994. Atmospheric lead pollution in an urban area- Brisbane, Australia. Atmospheric Environment 28 (19), 3073–3082.

Willmott, C.J., 1982. Some comments on the evaluation of model performance. Bulletin American Meteorological Society 63 (11), 1309–1313.