Statistical Distributions of Ambient Air Pollutants in Shanghai, China[1]

HAI-DONG KAN[2] AND BING-HENG CHEN

Department of Environmental Health, School of Public Health, Fudan University, Box 249, 138 Yixueyuan Road, Shanghai 200032, China

Objective To determine the best statistical distribution of concentration data of major air pollutants in Shanghai. **Methods** Four types of theoretic distributions (lognormal, gamma, Pearson V and extreme value) were chosen to fit daily average concentration data of PM_{10}, SO_2 and NO_2 from June 1, 2000 to May 31, 2003 in Shanghai by using the maximum likelihood method. The fit results were evaluated by Chi-square test. **Results** The best-fit distributions for PM_{10}, SO_2 and NO_2 concentrations in Shanghai were lognormal, Pearson V, and extreme value distributions, respectively. **Conclusion** The results can be further applied to local air pollution prediction and control, e.g., the probabilities exceeding the air quality standard and emission source reduction of air pollutant concentration to meet the standard.

Key words: Air pollution; Statistical distribution; Shanghai; Maximum likelihood

INTRODUCTION

Ambient air pollution consists of a mix of different pollutants. In Shanghai, one of the largest cities in China, particulate matter (PM), sulfur dioxide (SO_2), and nitrogen oxides (NO_x) are the major air pollutants and routinely monitored. Since June 1, 2000, PM_{10} (those aerodynamic diameter less than 10 microns) and NO_2 have taken the place of TSP (total suspended particle) and NO_x as indicator pollutants. During the past decade, Shanghai has undergone the most rapid development and urbanization in her history, and ambient air pollution in Shanghai has gradually changed from the conventional coal combustion type to the mixed coal combustion/motor vehicle emission type. As shown in Fig. 1, the traditional coal combustion-related air pollution (e.g. PM, SO_2) in Shanghai improved substantially from 1990 to 1999, while vehicle-originated air pollution (e.g. NO_x) became a serious public health concern in the meantime[1].

Statistical distributions have been used in the analysis of air pollution data[2-5]. The concentrations of air pollutants are usually random variables, which are influenced by emission levels, meteorological conditions and geography. Although it is largely agreed that

---

[1]This work was supported by Shanghai Municipal Committee of Science and Technology through grant 03DZ05052.

[2]Correspondence should be addressed to Hai-Dong KAN, Tel: 86-21-54237908. Fax: 86-21-64046351. E-mail: hdkan@shmu.edu.cn

Biographical note of the first author: Hai-Dong KAN, male, born in 1974, Ph.D. in environmental health science, majoring in environmental epidemiology and health risk assessment.

0895-3988/2004
CN 11-2816

there is no priori reason to expect that atmospheric distribution should adhere to a specific probability distribution, the correctly chosen distribution can be used to predict the frequency that exceeds the ambient air quality standard (AQS), and emission source reduction to meet the AQS. For example, by making use of the direct association between emission level and some parameters (e.g., the location parameter) of the statistical distributions[6], Lu successfully predicted the probabilities exceeding the air quality standard and emission source reduction of PM_{10} concentration to meet the air quality standard in Taiwan[5].

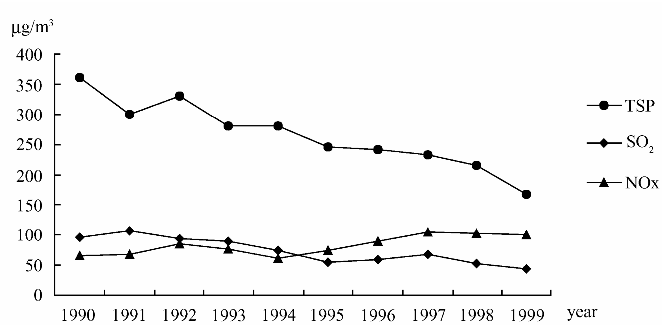

FIG. 1. Annual average concentrations of TSP, SO_2, NO_x in urban districts of Shanghai (1990-1999).

Therefore, information regarding the statistical distribution is necessary for developing air pollutant control strategies. It is also useful to examine the similarities and differences among the types of air pollution in different areas.

In this paper, the measured concentration data of PM_{10}, SO_2 and NO_2 between June 1, 2000 and May 31, 2003 in Shanghai were analyzed to simulate the frequency distribution and to estimate the distributional parameters, and the main objective was to fit the daily concentration data to determine the optimal shape of the concentration distribution.

METHODS

Data

Daily average PM_{10}, SO_2 and NO_2 concentrations in urban area of Shanghai were collected from the database of Shanghai Environment Online (http://www.envir.online.sh.cn/index.asp), from June 1, 2000 to May 31, 2003. According to Shanghai Environmental Monitoring Center, air pollutant concentrations were averaged from the results monitored by six fix-site stations in the urban areas of Shanghai. The six sites scatter in different functional areas of Shanghai and their monitoring results could represent the average air pollution level in Shanghai. No missing data were found in the variables described above.

Statistical Distributions Used in Representing the Distributions of Air Pollutant Concentrations

The distributions chosen to fit the data, including lognormal, gamma, Pearson V and extreme value distributions, are described in Table 1[6]. Extreme value distributions are the limiting distributions for the minimum or the maximum of a very large collection of random observations from the same arbitrary distribution, sometimes also known as the Fisher-Tippett distribution or log-Weibull distribution.

TABLE 1

Probability Distribution Functions Used to Fit the Air Pollutant Data

Distributions	Formula	Description
Lognormal	$f(x)=\frac{e^{-(\ln((x-\theta)/m)^2/(2\sigma^2))}}{(x-\theta)\sigma\sqrt{2\pi}}$	σ is the shape parameter, θ is the location parameter and m is the scale parameter.
Pearson V	$f(x)=\frac{x^{-(\alpha+1)}\exp(-\frac{\beta}{x})}{\beta^{-\alpha}\Gamma(\alpha)}$	Γ is the gamma function, and α, β are the parameters of Pearson V distribution
Gamma	$f(x)=\frac{(\frac{x-\mu}{\beta})^{\gamma-1}\exp(-\frac{x-\mu}{\beta})}{\beta\Gamma(\gamma)}$	γ is the shape parameter, μ is the location parameter, β is the scale parameter, and Γ is the gamma function
Extreme Value	$f(x)=\frac{1}{\beta}e^{\frac{x-\mu}{\beta}}e^{-e^{\frac{x-\mu}{\beta}}}$	μ is the location parameter and β is the scale parameter

Note. *x* Denotes the Actual Distribution Data.

Distribution Estimation and Goodness-of-fit Test

Distribution parameters were estimated using the method of maximum likelihood. This method always gives a minimum variance estimate of parameters. Chi-square test was used to test the goodness-of-fit.

RESULTS

Variability of Measured Data With Time

Table 2 summarizes the basic statistics of PM_{10}, SO_2 and NO_2 concentration data in Shanghai from June 1, 2000 to May 31, 2003. The daily PM_{10}, SO_2 and NO_2 concentration variability with time is shown in Fig. 2. It can be seen that the concentrations of the three pollutants had seasonal variability. Compared with the second level of China National Ambient Air Quality Standard, which was 150, 150 and 80 μg/m^3 for daily average concentration of PM_{10}, SO_2 and NO_2 respectively, the frequency of exceedance with the standard was higher for PM_{10} than those for SO_2 and NO_2, suggesting that particulate air pollution has become the major environmental problem in Shanghai.

TABLE 2

Basic Statistics of Daily Average PM_{10}, SO_2 and NO_2 Concentrations in Shanghai

	Number	Mean	SD	Minimum	P (25)	Median	P (75)	Maximum
PM_{10}	1090	100.5	1.9	17.0	29.0	81.0	216.0	534.0
SO_2	1090	39.8	0.6	8.0	12.0	36.0	72.0	146.0
NO_2	1090	59.9	0.6	14.4	24.0	56.0	99.2	211.2

Note. June 1, 2000 to May 31, 2003 (μg/m^3).

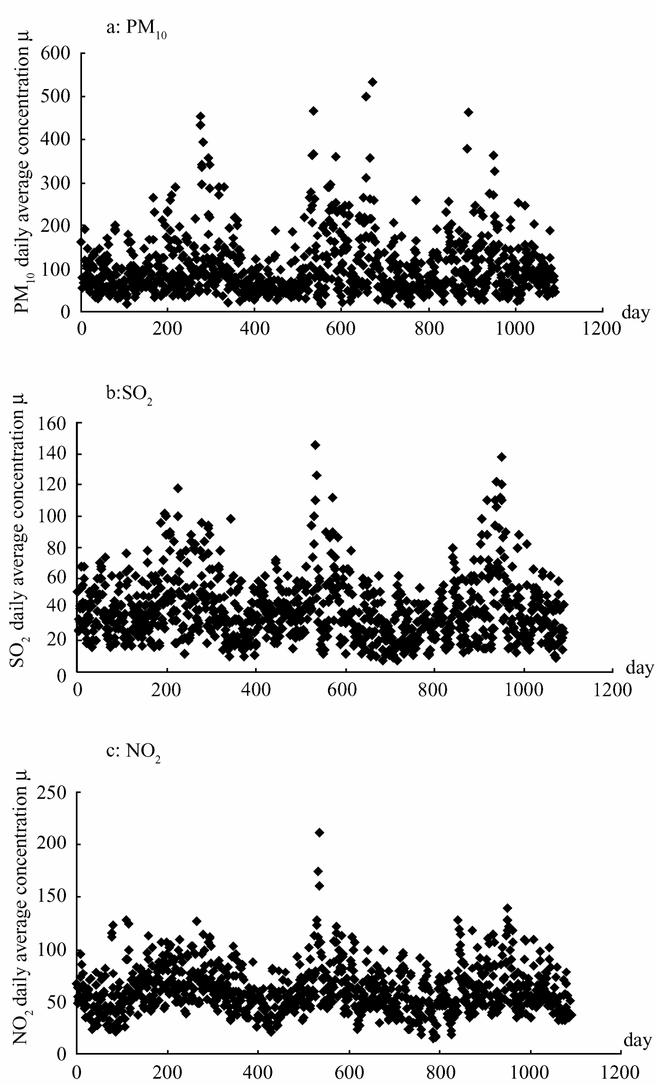

FIG. 2. Daily average concentration of PM_{10}, SO_2 and NO_2 in Shanghai from June 1, 2000 to May 31, 2003.

Fitting Results of Four Distributions and Comparisons of the Distributions

Figs. 3-5 show the compared fitting results and parameter estimations of four theoretic distributions for PM_{10}, SO_2 and NO_2, respectively.

The results of the goodness-of-fit test for the distributions are shown in Table 3. The *P* value of each distribution was all less than 0.01 for the three pollutants. Based on the Chi-square test, the best-fit distributions of PM_{10}, SO_2 and NO_2 during the period in Shanghai were lognormal, Pearson V, lognormal and extreme value distributions, respectively. Here the best distribution referred to the one with least Chi-square test value.

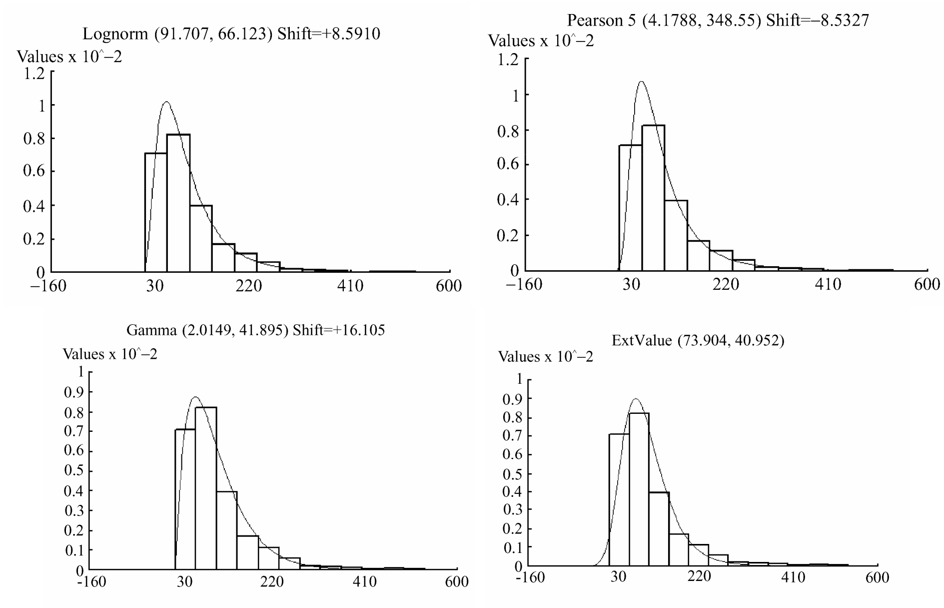

FIG. 3. Comparison of different statistical distributions of PM_{10} concentration (X-axis is the PM_{10} concentration, μg/m^3; Y-axis is the frequency value).

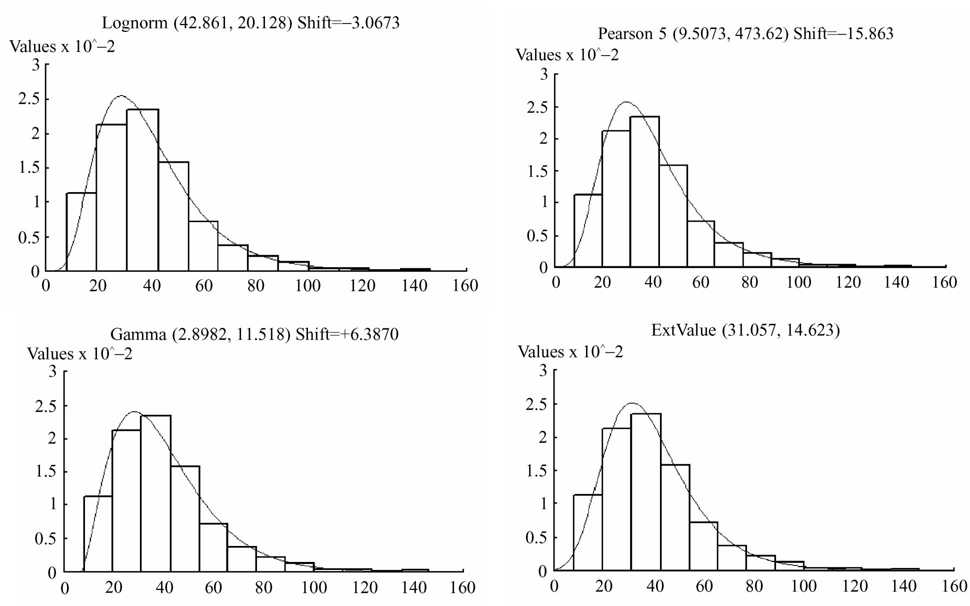

FIG. 4. Comparison of different statistical distributions of SO_2 concentration (X-axis is the SO_2 concentration, μg/m^3; Y-axis is the frequency value).

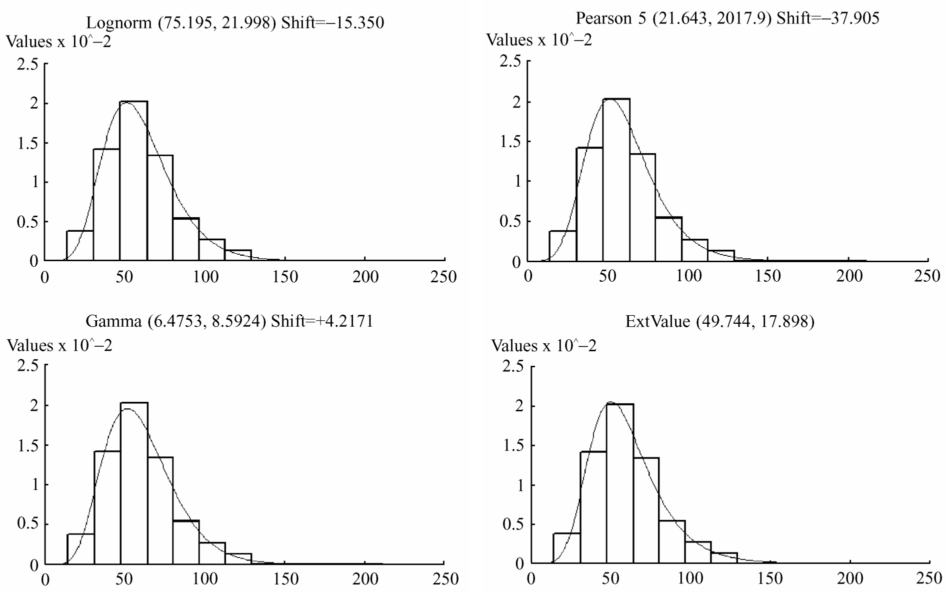

FIG. 5. Comparison of different statistical distributions of NO_2 concentration (X-axis is the NO_2 concentration, μg/m^3; Y-axis is the frequency value).

TABLE 3

Fitted Distribution Type and Goodness-of-fit Statistics Using the Chi-square Test

Distribution	PM_{10}	SO_2	NO_2
Lognormal	65.76*	129.62	138.84
Pearson V	83.94	99.72*	138.84
Gamma	98.67	111.23	215.78
Extreme Value	191.93	123.68	114.78*

Note. *The best fit.

DISCUSSION

Many types of probability distributions have been used to fit the air pollutant concentration data. Among them, the lognormal distribution was more widely used to represent the type of air pollutant concentration distribution. This distribution was used not only for ambient air quality data but also for indoor air quality and dissolved solids in groundwater[7]. That the concentration of air pollutants tends to be a lognormal distribution has been explained by the theory of successive random dilutions[7]. That is, after the pollutants are emitted from the source, in the transport process before they reach the receptor, they undergo successive mixing and diluting, resulting in a lognormal frequency distribution. Using the lognormal pollutant concentration distribution, Saltzman showed that changes in the standard deviation of the pollutant concentrations had a major effect on the calculated health risk[8,9]. A larger lognormal standard deviation had the effect of making the distribution flatter and would result in a larger overlap of the concentration curve and the pollutant frequency distribution. This, combined with the hypothesis that the greatest health effect is related to the low-to-mid range concentrations due to the greater frequency of occurrence of these lower levels[10], suggests that the shape of the concentration distribution and how it changes with increasing emission levels have important health implications.

Pearson V is another frequently used distribution that has been derived from the theory of stochastic process. The key advantage of this distribution is that it provides the link between the emission level and air pollutant concentration. Therefore, from this distribution, it is possible to estimate the amount of abatement needed in different areas to meet air quality standard (AQS)[3]. Lu successfully employed the Pearson V distribution to predict the probabilities exceeding the air quality standard and emission source reduction of PM_{10} concentration to meet the air quality standard in Taiwan[5].

Recently, the distributions of extreme value and gamma have also been used to fit the probability density functions of daily air pollutant concentration[11].

The current study showed that the pollutants studied (PM_{10}, SO_2 and NO_2) had different statistical distribution. The difference might be due to the different diffusion characteristics of individual pollutant in the air, and the interaction of diffusion characteristics and local geography, weather conditions in Shanghai[3]. The underlying mechanisms need to be further explored.

The statistical distribution of air pollutant concentration has an impact on human health and the setting of air pollutant regulations[10]. That is why people have interest in determining the specific distribution of major air pollutants in different places, especially particulate matter. The present analysis shows that the best performing statistical distributions of various air pollutants in Shanghai are different. For PM_{10}, the best distribution is lognormal, while for SO_2 and NO_2, the best distributions are Pearson V, and extreme value respectively. The results can be further applied to local air pollution prediction and control.

REFERENCES

1. Shanghai Municipal Environmental Protection Bureau (1999). Shanghai Environmental Quality Report.
2. Larsen, R. I. (1973). An air quality data analysis system for interrelating effects, standards, and need source reductions. *Journal of Air Pollutants and Control Assessment* **23**, 933-940.
3. Morel, B., Yeh, S., and Cifuentes, L. (1999). Statistical distributions for air pollution applied to the study of the particulate problem in Santiago. *Atmospheric Environment* **33**, 2575-2585.
4. Kao, A. S. and Friedlander, S. K. (1995). Frequency distribution of PM_{10} chemical components and their source. *Environment Sciences and Technology* **29**, 19-28.
5. Lu, H. (2002). The statistical character of PM_{10} concentration in Taiwan area. *Atmospheric Environment* **36**, 491-502.
6. Georgopoulos, P. G. and Seinfeld, J. H. (1982). Statistical distribution of air pollutant concentration. *Environmental Science and Technology* **16**, 401A-416A.
7. Ott, W. R. A. (1990). Physical explanation of the lognoramlity of pollutant concentration. *Journal of Air and Waste Management Association* **40**, 1378-1383.
8. Saltzman, B. E. (1987). Lognormal model for health risk assessment of fluctuating concentrations. *American Industrial Hygiene Association Journal* **48**, 140-149.
9. Saltzman, B. E. (1997). Health risk assessment of fluctuating concentrations using lognormal models. *Journal of the Air and Waste Management Association* **47**, 1152-1160.
10. Federal Register (1997). National Ambient Air Quality Standards for Particulate Matter: *Final Rule* **40** CFR Part 50, 62, 38651-38854.
11. Rumburg, B., Alldredge, R., and Claiborn, C. (2001). Statistical distributions of particulate matter and the error associated with sampling frequency. *Atmospheric Environment* **35**, 2907-2920.

(*Received December* 16, 2003 *Accepted April* 31, 2004)