Hindawi Publishing Corporation
Journal of Probability and Statistics
Volume 2014, Article ID 864965, 6 pages
http://dx.doi.org/10.1155/2014/864965

Research Article

A Study of Probability Models in Monitoring Environmental Pollution in Nigeria

P. E. Oguntunde,[1] O. A. Odetunmibi,[1] and A. O. Adejumo[2]

[1] *Department of Mathematics, Covenant University, Ota, Ogun State, Nigeria*
[2] *Department of Statistics, University of Ilorin, Ilorin, Nigeria*

Correspondence should be addressed to A. O. Adejumo; aodejumo@gmail.com

Received 29 January 2014; Revised 17 April 2014; Accepted 18 April 2014; Published 5 May 2014

Academic Editor: Zhidong Bai

Copyright © 2014 P. E. Oguntunde et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In Lagos State, Nigeria, pollutant emissions were monitored across the state to detect any significant change which may cause harm to human health and the environment at large. In this research, three theoretical distributions, Weibull, lognormal, and gamma distributions, were examined on the carbon monoxide observations to determine the best fit. The characteristics of the pollutant observation were established and the probabilities of exceeding the Lagos State Environmental Protection Agency (LASEPA) and the Federal Environmental Protection Agency (FEPA) acceptable limits have been successfully predicted. Increase in the use of vehicles and increase in the establishment of industries have been found not to contribute significantly to the high level of carbon monoxide concentration in Lagos State for the period studied.

1. Introduction

It is common knowledge that population growth and globalization have become the major drivers of pollution. Out of the various forms of pollution, a large number of studies that investigated the relationship between air quality and health effects cited air pollution as the major environmental issue of concern to the community. Increase in hospitalization, emergency room attendance, and decreased lung function have been associated with the following common air pollutants: carbon monoxide (CO), nitrogen oxides (NO_x), inhalable particles (measured as PM_{10}), photochemical oxidants (measured as ozone), and sulphur dioxide SO_2.

Air pollution is defined as the presence in the outdoor atmosphere of one or more pollutants in such quantities and of such duration that may tend to be injurious to human, plant, or animal life or property or which may unreasonably interfere with the comfortable enjoyment of life or property or the conduct of business [1, 2].

In this research work, emphasis will be on one of these criteria pollutants which is carbon monoxide because of the major threats it poses to human health.

Carbon monoxide is a colourless, odourless, and highly poisonous gas produced in large quantities as a result of incomplete combustion of fossil fuels. It is known that the main source of carbon monoxide is from motor vehicle exhaust (vehicular emission); about two-thirds of the pollutant emissions come from transportation sources, while other sources include industrial processes and open burning activities [3, 4].

The natural concentration of carbon monoxide in air is around 0.2 ppm, and that amount is not harmful to humans, while exposure to the pollutant emission at 100 ppm or greater can be dangerous to human health. Carbon monoxide endangers humans specifically by its tendency to combine with haemoglobin in the blood. Their combination produces carboxyl haemoglobin (COHB), thus reducing the capacity of the blood to carry oxygen [5]. The acute effects produced by exposure to carbon monoxide (in parts per million) are given in Table 1.

Probability models have been applied successfully in many physical phenomena such as wind speed, rainfall, river discharges, and air quality. It has been applied to fit the data of vehicular emission in Chennai, India, for predicting the concentration of carbon monoxide in the ambient atmosphere [6, 7]. In their research, ten standard probability models were fitted to the data and goodness of fit was assessed using Kolmogorov-Smirnov test and Anderson-Darling test.

Table 1: Carbon monoxide poisoning: signs and symptoms [15].

Concentration	Symptoms
35 ppm (0.0035%)	Headache and dizziness within six to eight hours of constant exposure.
100 ppm (0.01%)	Slight headache in two to three hours.
200 ppm (0.02%)	Slight headache within two to three hours; loss of judgment.
400 ppm (0.04%)	Frontal headache within one to two hours.
800 ppm (0.08%)	Dizziness, nausea, and convulsions within 45 min; insensible within 2 hours.
1,600 ppm (0.16%)	Headache, tachycardia, dizziness, and nausea within 20 min; death in less than 2 hours.
3,200 ppm (0.32%)	Headache, dizziness, and nausea in 5 to 10 min. Death within 30 minutes.
6,400 ppm (0.64%)	Headache and dizziness in one to two minutes. Convulsions, respiratory arrest, and death in less than 20 minutes.
12,800 ppm (1.28%)	Unconsciousness after 2-3 breaths. Death in less than three minutes.

When the parent probability distribution of air pollutants is correctly chosen, the specific distribution can be used to predict the mean concentration and probability of exceeding a critical concentration [5, 8].

The objectives of this paper are to fit the three probability distributions afore-mentioned to the concentration of carbon monoxide in Lagos State, Nigeria, to determine the "best" distribution to describe the data, and to establish the distribution of carbon monoxide concentration with a view of predicting the probability that the concentration would exceed a critical or an acceptable concentration.

To this effect, observations on the pollutant concentration were collected (as available) between the years 2004 and 2010. As vehicular exhaust (emission) is the major source of carbon monoxide, information was also collected on the number of newly registered vehicles and the number of newly registered industries in Lagos State between the years 2004 and 2010.

2. Methodology

Weibull Distribution. Let X denote a random variable; the two-parameter Weibull density function [9] is given by

$$f(x;\alpha,\beta) = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1} e^{-(x/\beta)^{\alpha}}; \quad \alpha > 0,\ \beta > 0, \tag{1}$$

where α is the shape parameter and β is the scale parameter.

Lognormal Distribution. A random variable X is log-normally distributed if $\ln(X)$ is normally distributed. Its probability density function [10] is given by

$$f\left(x;\mu,\sigma^2\right) = \frac{1}{x\sqrt{2\pi\sigma^2}} e^{-(\ln x-\mu)^2/2\sigma^2}; \quad x > 0, \quad -\infty < \mu < \infty, \quad \sigma^2 > 0, \tag{2}$$

where μ is the location parameter and as well the mean of the distribution and σ is the scale parameter and as well the standard deviation of the distribution.

Gamma Distribution. Let X denote a random variable, the two parameter gamma density function [11] with parameters α and β is given by

$$f(x;\alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta}; \quad x > 0,\ \alpha > 0,\ \beta > 0, \tag{3}$$

where α is the shape parameter and β is the scale parameter.

2.1. Methods of Parameter Estimation. The parameters of the distributions can be estimated using various methods like the method of maximum likelihood estimation (MLE) and method of moments (MOM) among others. In this paper, the method of likelihood estimation will be used because it is commonly used and it always gives a minimum variance estimate of parameters.

The MLE is widely and commonly used because it has many desirable properties; the maximum likelihood estimator is consistent, asymptotically normal, and asymptotically efficient. Let $x_1, x_2, \ldots, x_n$ be a random sample of size "n" drawn from a p.d.f, $f(x,\theta)$, where θ is an unknown parameter. The Likelihood function of this random sample is the joint density function of the "n" random variables and it is a function of the unknown parameter [12]. Thus, $L = \prod_{i=1}^{n} f(x_i;\theta)$ is the likelihood function. The maximum likelihood estimator (MLE) of θ, say $\widehat{\theta}$, is the value of θ that maximizes L or, equivalently, the logarithm of L. The MLE of θ is a solution of

$$\frac{d\log L}{d\theta} = 0. \tag{4}$$

According to [12], the maximum likelihood estimators $\widehat{\alpha}$ and $\widehat{\beta}$ of the shape and scale parameters of Weibull distribution are the solution of the simultaneous equations

$$\widehat{\alpha} = \frac{n}{\left(1/\widehat{\beta}\right)\sum_{i=1}^{n} x_i^{\widehat{\alpha}} \log x_i - \sum_{i=1}^{n} \log x}, \tag{5}$$

$$\widehat{\beta} = \left[\left(\frac{1}{n}\sum_{i=1}^{n} x_i^{\widehat{\alpha}}\right)\right]^{1/\widehat{\alpha}}. \tag{6}$$

For lognormal distribution, the maximum likelihood estimates for μ and σ^2 are given by

$$\hat{\mu} = \left(\frac{1}{n}\right)\sum_{i=1}^{n}\log x_i, \tag{7}$$

$$\hat{\sigma}^2 = \left(\frac{1}{n-1}\right)\sum_{i=1}^{n}\left[\log(x_i - \hat{\mu})\right]^2. \tag{8}$$

Lastly, the maximum likelihood estimators $\hat{\alpha}$ and $\hat{\beta}$ for gamma distribution are solutions of the simultaneous equations

$$\log\hat{\alpha} - \psi(\hat{\alpha}) = \log\left[\frac{\overline{x}}{\left(\prod_{i=1}^{n}x_i\right)^{1/n}}\right], \qquad \hat{\beta} = \frac{\overline{x}}{\hat{\alpha}}, \tag{9}$$

where $\psi(\hat{\alpha})$ is a digamma function with an argument $\hat{\alpha}$ defined as

$$\psi(\hat{\alpha}) = \frac{d}{d\alpha}\left[\log\Gamma(\alpha)\right] = \frac{d\Gamma(\alpha)/d\alpha}{\Gamma(\alpha)}. \tag{10}$$

2.2. Weighted Least Squares. Weighted least squares is an efficient method that makes good use of small data sets. The main advantage that WLS enjoys over other methods is the ability to handle regression situations in which the data points are of varying quality. If the standard deviation of the random errors in the data is not constant across all levels of the explanatory variables, using WLS with *weights that are inversely proportional to the variance* at each level of the explanatory variables yields the most precise parameter estimates possible. Consider

$$w_i \alpha \frac{1}{s_i^2}. \tag{11}$$

Since the sample sizes n_i of the data also varies, the weight used in this research work is

$$w_i = \frac{n_i}{s_i^2}. \tag{12}$$

The WLS estimate of β is given by

$$\beta = \left(X^1 W X\right)^{-1} X^1 W Y. \tag{13}$$

The matrix of W is given by

$$W = \begin{pmatrix} w_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & w_n \end{pmatrix}. \tag{14}$$

Fitting this model is equivalent to minimizing

$$SSE = \sum_{i=1}^{n} w_i\left(y_i - \beta_o - \beta_1 x_1 - \beta_2 x_2\right)^2. \tag{15}$$

2.3. Test of Goodness of Fit. In order to verify the goodness of fit of the models to the carbon monoxide data observations, the Kolmogorov-Smirnov (K-S) and Anderson-Darling (A-D) tests are used. The lower the value of these statistics is, the closer the fitted distribution appears to match the data. The hypothesis for the tests is given as follows:

H_o: the data follow a specified distribution

versus

H_1: the data do not follow the specified distribution.

Given "N" ordered data points $X_1, X_2, \ldots, X_N$, the test statistics for the Kolmogorov–Smirnov test are given as

$$D \equiv \max_{1\le i\le n}\left(F(X_i) - \frac{i-1}{N}, \frac{i}{N} - F(X_i)\right). \tag{16}$$

The test statistics for Anderson-Darling are given by

$$A^2 = -N - S, \tag{17}$$

where

$$S = \sum_{i=1}^{n}\frac{(2i-1)}{N}\left[\ln F(X_i) + \ln\left(1 - F\left(X_{N+1-i}\right)\right)\right]. \tag{18}$$

$F(\cdot)$ is the CDF of the continuous distribution being tested and X_i are the ordered data.

2.4. Probability of Exceedance. The probability that carbon monoxide observations would exceed a specified standard or limit is based on the distribution that has been chosen as the best distribution for Carbon monoxide concentration in Lagos State for the period studied.

Mathematically, the probability of exceeding a critical concentration [13, 14] is given by

$$\Pr(X > x) = 1 - \Pr(X \le x) = 1 - \int_{-\infty}^{x} f(x)\,dx. \tag{19}$$

3. Summary of the Data Collected

In this section, we provide and describe the information gathered on carbon monoxide concentration, number of newly registered vehicles, and industries.

3.1. Data on Carbon Monoxide Concentration. This section provides information on the secondary data collected on the concentration of carbon monoxide measured in parts per million (ppm) in Lagos State (as available) from August 2004 to August 2010. The data was collected as daily data but we could only gather 412 data points for the years considered (e.g., there was no record at all for the year 2007, as shown in Table 7). The data has been summarized in Table 2 giving the minimum and maximum values of the measurement recorded, the standard deviation, mean, and the mode of the observations.

The diagrammatic representation of the data on carbon monoxide concentration (ppm) is given in Figure 1.

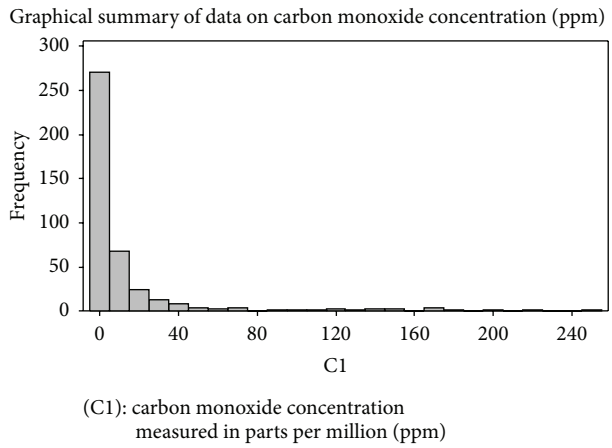


Figure 1: Histogram of data on carbon monoxide concentration.

It can be deduced from Figure 1 that the information on the carbon monoxide concentration (as collected) is positively skewed and mode occurs at 0 ppm. This justifies our reason for using positively skewed theoretical distributions to model the data set in this paper.

3.2. Data on Registered Vehicles. In this section, we provide information on the number of vehicles (trucks, buses, and cars) that were registered in Lagos State each year between the years 2004 and 2010. The data is provided in Table 3.

The graphical representation of the number of newly registered vehicles is given in Figure 2.

It can be observed from Figure 2 that there was a little decline in the number of registered vehicles in 2006 and a sharp increase in year 2007 and the highest registration was recorded in year 2008.

3.3. Data on Registered Industries. Table 4 shows the summary of the information collected on the number of newly registered industries (manufacturing industries) in Lagos State between 2004 and August, 2010. It should be noted that there are more industries in Lagos State apart from the ones captured in this paper but we only consider manufacturing industries that are registered.

The graphical representation of the number of newly registered industries is given in Figure 3.

It can be noticed in Figure 3 that only few manufacturing industries are registered. Besides, the records keep increasing from year 2004 to year 2010 except in 2009 where there was a little decline.

4. Analysis and Results

The parameters of the distributions under study (Weibull, lognormal, and gamma) were estimated by fitting the distributions to the data of carbon monoxide concentration collected using Easy-Fit statistical package.

4.1. Test of Goodness of Fit. In an attempt to choose the "best" probability model to describe the concentration of carbon monoxide in Lagos State for the period studied, Kolmogorov-Smirnov goodness of fit test was conducted. The summary of the analysis is given in Table 6.

The graph for the Cumulative Density Function (CDF) of the three distributions is shown in Figure 4.

This graph shows how well Weibull, lognormal, and gamma distributions fit the data. It can be seen that the CDF of the gamma distribution is closer to the true CDF of the carbon monoxide concentration.

4.2. Probability of Exceeding Critical Concentrations. Since gamma distribution fits the data better than the remaining fitted distributions, the probability that the carbon monoxide concentration would exceed both the Lagos State Environmental Protection Agency (LASEPA) standard (5 ppm) and the Federal Environmental Protection Agency (FEPA) standard (10 ppm) will be calculated based on the cumulative density function (CDF) of gamma distribution.

The probability density function of a gamma distribution with parameters α and β is given by

$$f(x;\alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta}; \quad x > 0,\ \alpha > 0,\ \beta > 0. \tag{20}$$

And the cumulative density function (CDF) is

$$F(x) = \Pr(X \le x) = \int_{-\infty}^{x} f(x)\,dx. \tag{21}$$

From Table 5, the shape parameter $\alpha = 0.1463$, the scale parameter $\beta = 87.378$, and

$$\therefore P(X \le 5) = 0.699181. \tag{22}$$

Hence, the probability that the carbon monoxide concentration would exceed LASEPA standard is

$$P(X > 5) = 1 - P(X \le 5) = 1 - 0.699181 = 0.300819. \tag{23}$$

Also,

$$P(X \le 10) = 0.768379. \tag{24}$$

Then, the probability that the carbon monoxide concentration would exceed FEPA standard is

$$P(X > 10) = 1 - P(X \le 10) = 1 - 0.768379 = 0.231621. \tag{25}$$

4.3. Linear Regression Modelling. The mean yearly carbon monoxide concentration Y (in ppm) will be regressed on the number of newly registered vehicles (X_1) and the number of newly registered industries (X_2). There was no data available for carbon monoxide concentration in the year 2007; therefore, the year 2007 is automatically ignored in the regression analysis. Table 7 shows the summary of the data used for the regression analysis.

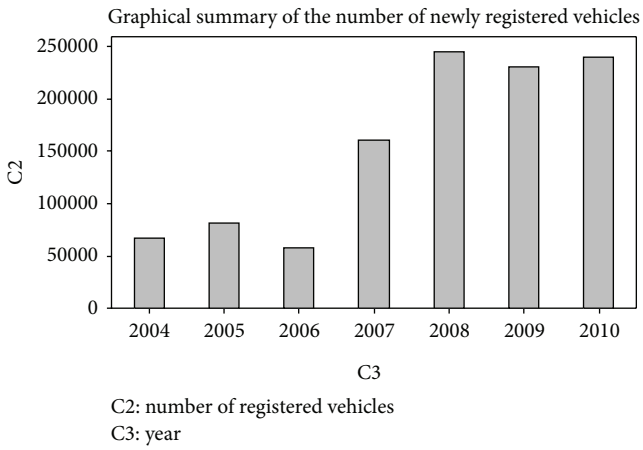


Figure 2: Bar chart of the number of newly registered vehicles.

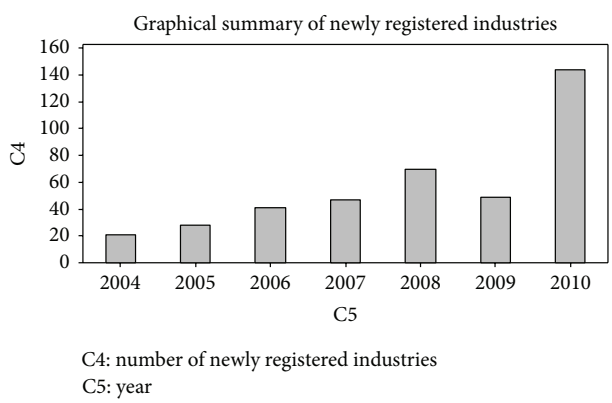


Figure 3: Bar chart of the number of newly registered industries.

Table 2: Carbon monoxide concentration (ppm). Source of data: Lagos State Environmental Protection Agency (LASEPA).

Minimum value (ppm)	Maximum value (ppm)	Mean value (ppm)	Mode (ppm)	Standard deviation	Total observation
0.00	249.00	12.79	0.00	33.43	412

Table 3: Data on registered vehicles. Source: Motor Vehicle Administrative Agency (M.V.A.A.), Ikeja Chapter, Lagos State.

Year	2004	2005	2006	2007	2008	2009	2010
Number of registered vehicles	67,376	81,078	57,379	160,134	244,810	230,822	239,954

Using MINITAB statistical package, regressing Y on both variable X_1 and X_2 gives the results shown in Table 8.

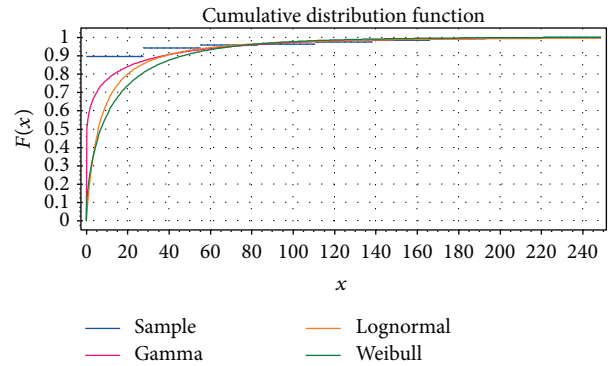


Figure 4: Graph showing the cumulative distribution function of the fitted distributions.

Table 4: Estimated number of newly registered industries. Source: Manufacturer Association of Nigeria (M.A.N), Ikeja Chapter, Lagos State.

Year	2004	2005	2006	2007	2008	2009	2010
Estimated number of newly registered industries	21	28	41	47	70	49	144

Table 5: Parameter estimates of the fitted probability models.

Distributions	Parameter estimates
Weibull	$\alpha = 0.6386$, $\beta = 12.766$
Lognormal	$\sigma = 1.4615$, $\mu = 1.777$
Gamma	$\alpha = 0.1463$, $\beta = 87.378$

Table 6: Fitted distribution type and goodness of fit statistics.

Distributions	Kolmogorov-Smirnov test	Anderson-Darling test
Weibull	0.3473	0.3692
Lognormal	0.3711	0.3904
Gamma	0.3471*	0.3466*

Note: *denotes the best fit.

Table 7: Mean yearly carbon monoxide concentration (Y), newly registered vehicles (X_1), and registered industries (X_2) for six years.

Year	2004	2005	2006	2008	2009	2010
Y (ppm)	33.90	38.96	84.00	10.99	14.13	4.82
X_1	67,376	81,078	57,379	244,810	230,822	239,954
X_2	21	28	41	70	49	144

The regression equation is

$$Y = 70.58 - 0.000276X_1 + 0.049X_2. \quad (26)$$

Equation (26) is interpreted as follows.

Table 8: Table of results from regression analysis.

Predictor	Coeff.	St. Dev.	T	P value
Constant	70.58	17.66	4.00	0.028
X_1	−0.0002757	0.0001405	−1.96	0.145
X_2	0.0491	0.2912	0.17	0.877

Table 9: Analysis of variance table.

Source of variation	DF	Sum of squares	Mean square	F	P value
Regression	2	2943.0	1471.5	3.37	0.171
Residual	3	1308.1	436.0		
Total	**5**	**4251.1**			

There will be a decrease of 0.000276 in Y for a unit change in X_1 when variable X_2 is held fixed and there will be an increase of 0.049 in Y for a unit change in X_2 when variable X_1 is held fixed.

Analysis of Variance (ANOVA)

Hypothesis

$$H_o : \beta_1 = \beta_2 = 0$$

versus

$$H_o : \beta_i \neq 0;\ i = 1, 2 \text{ for at least one } i.$$

Decision Rule. Reject H_o if P value is less than the level of significance ($\alpha = 0.05$).

Decision. We do not reject H_o since 0.171 is not less than 0.05.

Inference. From Table 8, considering the respective P values for the parameters β_1 and β_2, it means that the regression parameters are not significantly different from zero with an R-Sq = 69.2% and R-Sq (adjusted) = 48.7%.

Also, from Table 9, based on the P value (0.171), we conclude that the regression model is not significant at ($\alpha = 0.05$).

5. Conclusion

In this paper, we have been able to establish (based on the data collected) that the distribution of the carbon monoxide observations in Lagos State between the periods studied is positively skewed as shown in Figure 1. Gamma distribution is considered the best distribution for modeling carbon monoxide concentration in Lagos State as confirmed by the Kolmogorov-Smirnov and Anderson-Darling tests in Table 6. The carbon monoxide concentration in Lagos State exceeds the Lagos State Environmental Protection Agency (LASEPA) and the Federal Environmental Protection Agency (FEPA) standards with probabilities 0.300819 and 0.231621, respectively. Increase in the use of vehicles and increase in the establishment of industries in Lagos State do not contribute significantly to the high carbon monoxide concentration levels. Perhaps, further researches could be focused on the age of the car engines, quality of the fuel used for vehicles and machineries, then the smoking activities in the state.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] F. I. Abam and G. O. Unachukwu, "Vehicular emissions and air quality standards in Nigeria," *European Journal of Scientific Research*, vol. 34, no. 4, pp. 550–560, 2009.

[2] L. W. Canter, *Environmental Impact Assessment*, McGraw Hill, New York, NY, USA, 2nd edition, 1996.

[3] E. L. Kanyode, *Pollutants emissions measured [LUMES thesis]*, Lund University, Lund, Sweden, 2004.

[4] C. A. Ibrahim, *Overview of Air Quality Management in Malaysia*, University Science of Malaysia, 2004.

[5] G. Fellenberg, *The Chemistry of Pollution*, John Wiley & Sons, Oxford, UK, 2000.

[6] H.-D. Kan and B.-H. Chen, "Statistical distributions of ambient air pollutants in Shanghai, China," *Biomedical and Environmental Sciences*, vol. 17, no. 3, pp. 366–372, 2004.

[7] M. Harikrishna and C. Arun, "Stochastic analysis for vehicular emissions on urban roads—a case study of Chennai," in *Proceedings of the 3rd International Conference on Environmental and Health*, M. J. Bunch, V. M. Suresh, and T. V. Kumaran, Eds., Chennai, India, December 2003.

[8] H.-C. Lu and G.-C. Fang, "Predicting the exceedances of a critical PM_{10} concentration—a case study in Taiwan," *Atmospheric Environment*, vol. 37, no. 25, pp. 3491–3499, 2003.

[9] Y. Lei, "Evaluation of three methods for estimating the Weibull distribution parameters of Chinese pine (*Pinus tabulaeformis*)," *Journal of Forest Science*, vol. 54, no. 12, pp. 566–571, 2008.

[10] B. F. Ginos, *Parameter estimation for the lognormal distribution [M.S. thesis]*, Brigham Young University, Provo, Utah, USA, 2009.

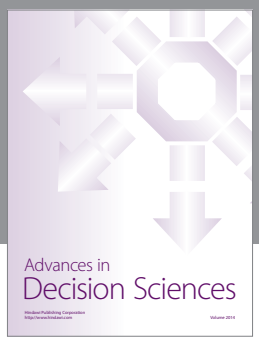
[11] C. Forbes, M. Evans, N. Hastings, and B. Peacock, *Statistical Distributions*, John Wiley & Sons, Hoboken, NJ, USA, 4th edition, 2011.

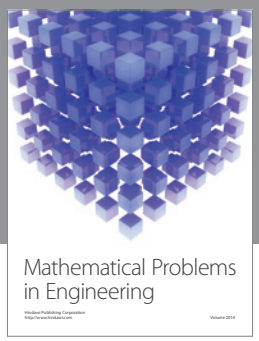
[12] M. A. Stephens, "EDF statistics for goodness of fit and some comparisons," *Journal of the American Statistical Association*, vol. 69, no. 347, pp. 730–737, 1974.

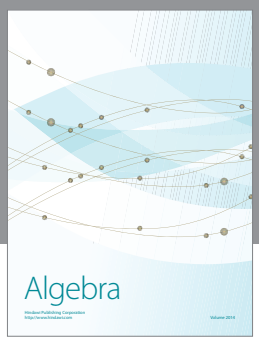
[13] A. S. Yahaya and N. A. Ramli, "Modelling of carbon monoxide concentration in major towns in Malaysia: a case study in Penang, Kuching and Kuala Lumpur," Project Report, Universiti Sains Malaysia, 2008.

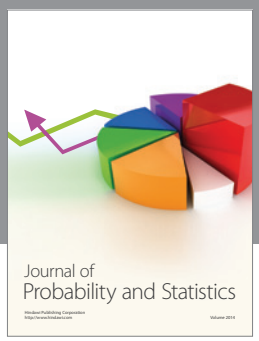
[14] A. Zaharim, S. Najid, A. Razali, and K. Sopian, "Analyzing Malaysian wind speed data using statistical distribution," in *Proceedings of the 4th IASME/WSEAS International Conference on Energy & Environment (EE '09)*, pp. 363–370, 2009.

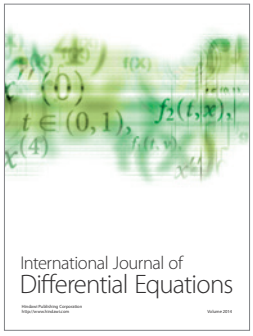
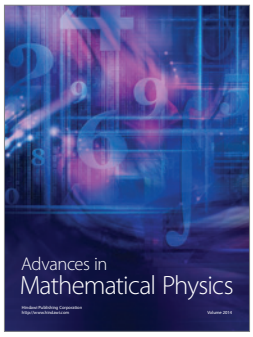
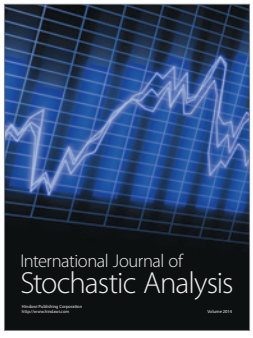
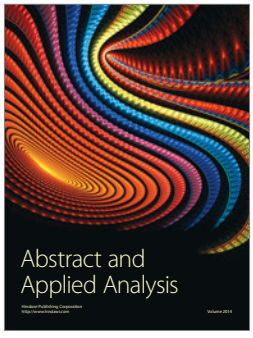
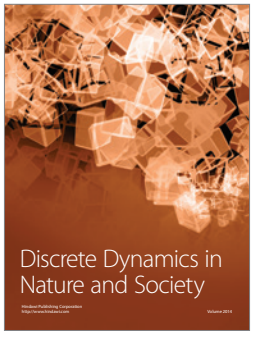
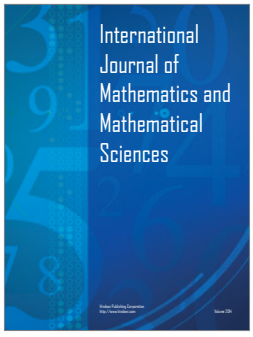
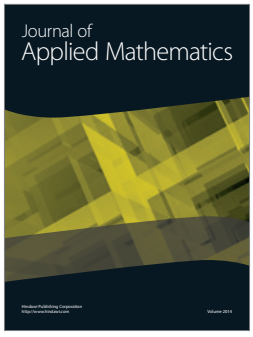
[15] "Carbon monoxide poisoning; Signs and Symptoms," http://en.wikipedia.org/wiki/Carbon_monoxide_poisoning.









Hindawi

Submit your manuscripts at
http://www.hindawi.com