A Study of Probability Models in Monitoring Environmental Pollution in Nigeria.

Oguntunde P.E[1], Odetunmibi O.A[2] and Adejumo, A. O[3]

[1,2]Department of Mathematics, Covenant University, Ota, Ogun State, Nigeria.

[3]Department of Statistics, University of Ilorin, Ilorin, Nigeria.

Abstract

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, Log-normal, and Gamma distributions were examined on the carbon monoxide observations to determine the best. 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.

Keywords: Carbon Monoxide, Gamma, Log-normal, Pollution, Probability, Weibull.

1.0 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 as 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 shall 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-third 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 0.2ppm, and that amount is not harmful to humans, while exposure to the pollutant emission at 100ppm 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 exposures to carbon monoxide (in parts per million) are given in the table below;

Table1: Carbon monoxide poisoning; Signs and Symptoms [6]

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

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 [7],[8]. In their research, ten standard probability models were fitted to the data and goodness of fit assessed using Kolmogorov-Smirnov test and Anderson-Darling test.

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],[9].

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 were also collected on the number of newly registered vehicles and number of newly registered industries in Lagos state between the years 2004 and 2010.

2.0 Methodology

Weibull Distribution

Let X denotes a random variable, the two-parameter Weibull density function [10] is given by;

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

Where α is the shape parameter,

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 [11] is given by;

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

where μ is the location parameter and as well the mean of the distribution,

σ is the scale parameter and as well the standard deviation of the distribution.

Gamma Distribution

Let X denotes a random variable, the two parameter Gamma density function [12] with parameters α and β is given by;

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

Where α is the shape parameter,

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), Method of Moments (MOM) among others. In this paper, the Method of Likelihood Estimation shall 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 $x_1, \dots \dots \dots, x_n$ 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 [13]. Thus,

$L = \prod_{i=1}^{n} f(x_i;\theta)$ is the likelihood function. The maximum likelihood estimator (MLE) of θ, say $\hat{\theta}$, is the value of θ that maximizes L or equivalently, the logarithm of L. The MLE of θ is a solution of

$$\frac{d \log L}{d\theta} = 0$$

According to [13], the maximum likelihood estimators $\hat{\alpha}$ and $\hat{\beta}$, of the shape and scale parameters of Weibull distribution are the solution of the simultaneous equations:

$$\hat{\alpha} = \frac{n}{\left(1/\hat{\beta}\right)\sum_{i=1}^{n} x_i^{\hat{\alpha}} \log x_i - \sum_{i=1}^{n} \log x_i}$$

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

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

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

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

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

$$\log \hat{\alpha} - \psi(\hat{\alpha}) = \log\left[\bar{x} / \left(\prod_{i=1}^{n} x_i\right)^{1/n}\right]$$

$$\hat{\beta} = \bar{x}/\hat{\alpha}$$

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

$$\psi(\alpha) = \frac{d}{d\alpha}[\log \Gamma(\alpha)] = \frac{d\Gamma(\alpha)/d\alpha}{\Gamma(\alpha)}$$

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.

$$w_i \; \alpha \; \frac{1}{s_i^2}$$

Since the sample sizes n_i of the data also varies, the weight used in this research work is $w_i = \frac{n_i}{s_i^2}$.

The WLS estimate of β is given by;

$$\beta = (X'WX)^{-1}X'WY$$

The matrix of W is given by;

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

Fitting this model is equivalent to minimizing

$$SSE = \sum_{i=1}^{n} w_i((Y_i - \beta_0 - \beta_1 x_1 - \beta_2 x_2)^2)$$

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, the closer the fitted distribution appears to match the data. The hypothesis for the tests is given as;

H_0 : The data follow a specified distribution.

Versus

H_1 : The data do not follow the specified distribution.

Given 'N' ordered data points $X_1, X_2, \dots, X_N$, the test statistic for the Kolmogorov – Smirnov test is given as;

$$D = \max_{1 \le i \le N}\left(F(X_i) - \frac{i-1}{N}, \frac{i}{N} - F(X_i)\right)$$

The test statistic for Anderson-Darling is given by;

$$A^2 = -N - S$$

where

$$S = \sum_{i=1}^{N} \frac{(2i-1)}{N} [\ln F(X_i) + \ln(1 - F(X_{N+1-i}))]$$

F(.) is the cdf of the continuous distribution being tested

X_i are the ordered data.

2.4 Probability of Exceedances

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 [14],[15] is given by;

$$\Pr(X > x) = 1 - \Pr(X \le x)$$

$$= 1 - \int_{-\infty}^{x} f(x)dx$$

3.0 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 a daily data but we could only gather 412 data points for the years considered (for example, there was no record at all for the year 2007; as would be discussed later in table 7). The data has been summarized in the table below; giving the minimum and maximum values of the measurement recorded, the standard deviation, mean and the mode of the observations.

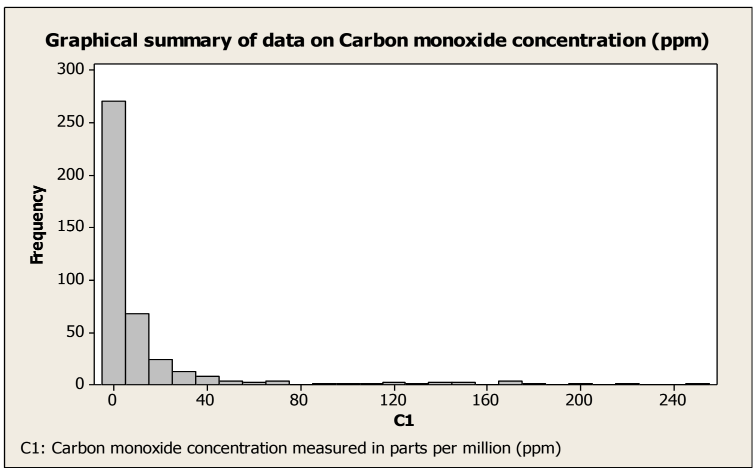
Table 2: Carbon monoxide concentration (ppm)

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

Below is the diagrammatic representation;

Figure 1: Histogram of data on Carbon Monoxide Concentration (ppm)



It can be deduced from Figure 1 that the information on the carbon monoxide concentration (as collected) is positively skewed and mode occurs at 0ppm. 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 year 2004 and 2010. The data is provided below;

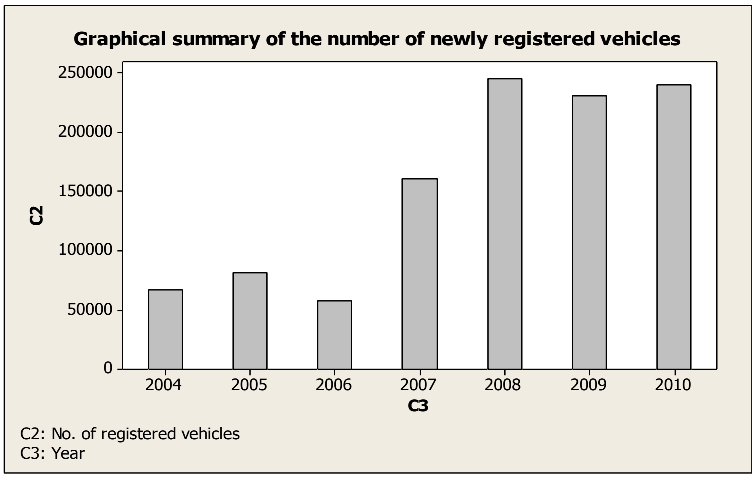
Table 3: Data on Registered Vehicles.

Source: Motor Vehicle Administrative Agency (M.V.A.A); Ikeja Chapter, Lagos State.

Year	2004	2005	2006	2007	2008	2009	2010
No. of registered vehicles	67,376	81,078	57,379	160,134	244,810	230,822	239,954

The graphical representation is given below;

Figure 2: Bar Chart of Number of Newly Registered Vehicles



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

3.3 Data on Registered Industries

Table 4 below shows the summary of the information collected on the number of newly registered industries (manufacturing industries) in Lagos State between the years 2004 till 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.

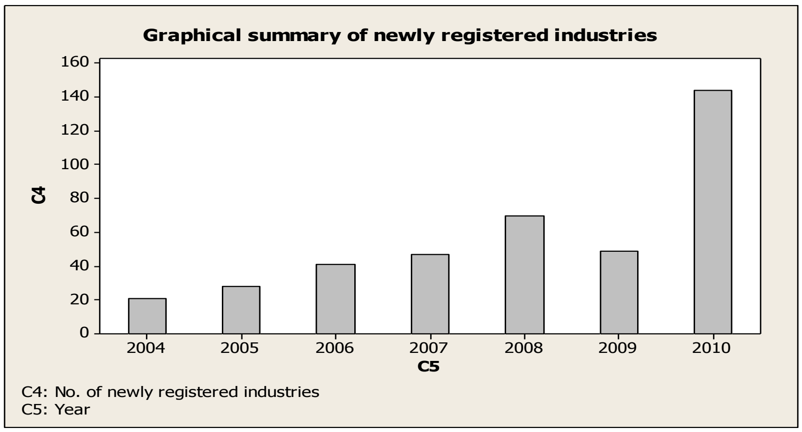
Table 4: Estimated number of newly registered industries.

Source: Manufacturer Association of Nigeria (M.A.N), Ikeja Chapter, Lagos State.

Year	2004	2005	2006	2007	2008	2009	2010
Estimated no. of newly registered industries	21	28	41	47	70	49	144

The graphical representation is given below;

Figure 3.0: Bar Chart of Number of Newly Registered Industries



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

4.0 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.

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	$\alpha = 0.1463$, $\beta = 87.378$

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. Below is the summary of the analysis;

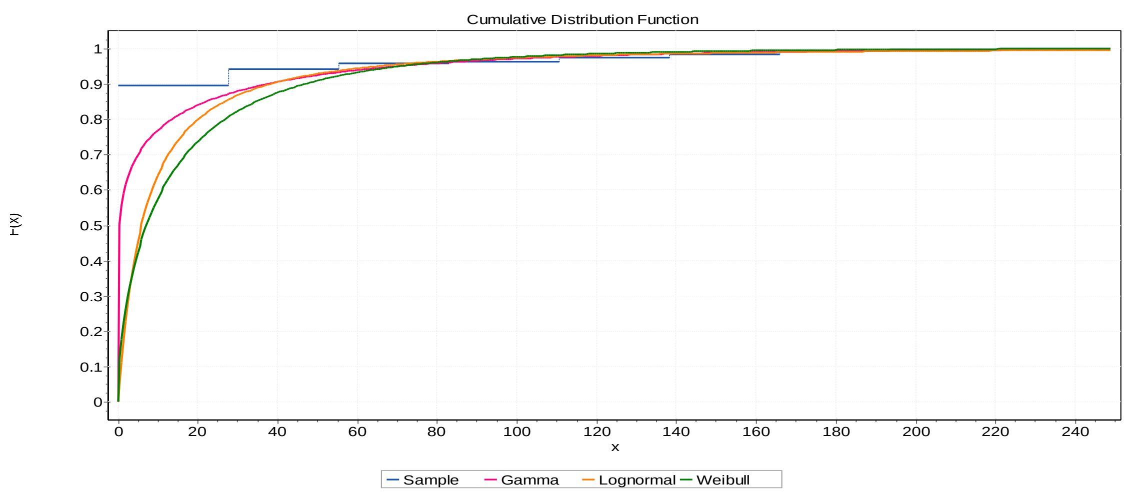
Table 6: Fitted Distribution Type and Goodness-of-fit Statistics.

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

The graph for the cumulative density functions (cdf) of the three distributions is shown in Figure 4 below;

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



This graph shows how well Weibull, Lognormal and Gamma distributions fit the data. It can be seen that the cdf of 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 (5ppm) and the Federal Environmental Protection Agency (FEPA) standard (10ppm) shall 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;

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

And the cumulative density function (cdf) is;

$$F(x)=P(X\leq x)$$
$$=\int_{-\infty}^{x}f(x)dx$$

From Table 5.0, the shape parameter $\alpha = 0.1463$ and the scale parameter $\beta = 87.378$

$$\therefore P(X\leq 5)=0.699181$$

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

$$P(X>5)=1-P(X\leq 5)$$
$$=1-0.699181$$
$$=0.300819$$

Also,

$$P(X\leq 10)=0.768379$$

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

$$P(X>10)=1-P(X\leq 10)$$
$$=1-0.768379$$
$$=0.231621$$

4.3 Linear Regression Modelling

The mean yearly carbon monoxide concentration Y (in ppm) shall 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, year 2007 is automatically ignored in the regression analysis. Table 7 shows the summary of the data used for the regression analysis;

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

Using MINITAB statistical package, regressing Y on both variable X_1 and X_2 gives the following results;

Table 8: Table of results from Regression Analysis

Predictor	Coeff	StDev	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

The regression equation is;

$$Y = 70.58 - 0.000276X_1 + 0.049X_2 \quad (4)$$

Equation (4) is interpreted as follows;

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_0: $\beta_1 = \beta_2 = 0$

Versus

H_1: $\beta_i \neq 0$; $i = 1,2$ for at least one i

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			

Decision Rule: Reject H_0 if p-value is less than the level of significance $(\alpha = 0.05)$

Decision: We do not reject H_0 since 0.171 is not less than 0.05

Inference: From Table 8, considering the respective p-value 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 as 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 does 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.

References

[1] Abam F.I and Unachukwu G.O (2009): Vehicular Emission and Air Quality Standards in Nigeria. *European Journal of Scientific Research* ISSN 1450-216X Vol.34 No.4, pp.550-560.

[2] Canter, L. W (1996). ‘Environmental Impact Assessment’, Second Edition. New York: McGraw Hill.

[3] Eric Lwanga Kanyode (2004). ‘Pollutants Emissions Measured’. *LUMES Thesis, Lund University, Sweden.*

[4] Ibrahim, C. A. (2004). ‘Overview of Air Quality Management in Malaysia’, *University Science of Malaysia*, 3 September 2004.

[5] Fellenberg, G. (2000). ‘The Chemistry of Pollution’. England: John Wiley & Sons Ltd.

[6] http://en.wikipedia.org/wiki/Carbon_monoxide_poisoning “Carbon monoxide poisoning; Signs and Symptoms”

[7] Hai-Dong kan, Bing-Heng chen (2004). ‘Statistical Distributions of Ambient Air Pollutants in Shanghai, China’. *Biomedicals and Environmental Sciences*. 17(3): 366-72

[8] Harikrishna, M. and Arun, C. (2003). ‘Stochastic Analysis for Vehicular emissions on Urban Roads - A Case Study of Chennai. In: Martin J. Bunch, V. Madha Suresh and T Vasantha Kumaran (eds)’ *Proceeding of the third International Conference on Environmental and Health*, 15 - 17 December 2003, Chennai. India

[9] Lu, H. C. and Fang, G. C. (2003). ‘Predicting the Exceedences of a critical PM_{10} Concentration - A Case Study in Taiwan’. *Atmospheric Environment.* 37 (4), p. 3491 -3499.

[10] Lei. Y (2008). ‘Evaluation of three methods for estimating the Weibull distribution parameters of Chinese pine’. *Journal of Forest Science*, 54 (12): 566-571.

[11] Brenda F. Ginos (2009): ‘Parameter Estimation for the Lognormal distribution’. *Masters Thesis, Brigham Young University, Provo, United States*.

[12] Forbes, C., Evans, M., Hastings, N., and Peacock, B. (2011). ‘Statistical Distributions’ (4th Edition). *John Wiley & Sons, Inc.*, Hoboken, New Jersey.

[13] Stephens, M. A. (1974). ‘EDF Statistics for Goodness of Fit and Some Comparisons’. *Journal of the American Statistical Association*, 69, pp. 730-737.

[14] Yahaya, Ahmad Shukri and Ramli, Nor Azam (2008). ‘Modelling Of Carbon Monoxide Concentration in Major Towns In Malaysia: A Case Study In Penang, Kuching and Kuala Lumpur’. *Project Report. Universiti Sains Malaysia*.

[15] Zaharim. A, Najid. S, Razali. A and Sopian. K (2009). ‘Analyzing Malaysian wind speed data using statistical distribution’. *EE’09 Proceedings of the 4th IASME/WSEAS International Conference on Energy & Environment*. Pages 363-370.