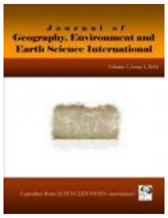

Journal of Geography, Environment and Earth Science International
4(2): 1-6, 2016; Article no.JGEESI.21931
ISSN: 2454-7352

SCIENCEDOMAIN *international*
www.sciencedomain.org

Estimation of Air Quality in Aba Urban, Nigeria Using the Multiple Linear Regression Technique

N. Agbanyim Akuagwu[1], E. N. Ejike[2] and A. U. Kalu[3*]

[1]Department of Chemistry, Abia State Polytechnic, Aba, Abia State, Nigeria.
[2]Department of Chemistry, Federal University of Technology, Owerri, Imo State, Nigeria.
[3]Department of Mathematics, Abia State Polytechnic, Aba, Abia State, Nigeria.

Authors' contributions

This work was carried out in collaboration between all authors. Author ENE designed the study and supervised data collection. Author NAA collected the data and managed the literature searches. Author AUK carried out the statistical analysis of the data and presented the results. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/JGEESI/2016/21931
Editor(s):
(1) Fombe Lawrence Fon, Regional Planning and Development, Faculty of Social and Management Sciences, University of Buea, Cameroon.
Reviewers:
(1) Mieczys Baw Szyszkowicz, Population Studies Division, Health Canada, Canada.
(2) Hideharu Shintani, Chuo University, Japan.
(3) João Gomes, Lisbon Polytechnic, Portugal.
Complete Peer review History: http://sciencedomain.org/review-history/11982

Short Research Article

Received 9th September 2015
Accepted 9th October 2015
Published 26th October 2015

ABSTRACT

The need to continuously monitor the air quality in our environment has continued to grow due to the ever increasing level of atmospheric pollutants. Indoor air pollution and urban air pollution are listed as two of the world's worst toxic pollution problems in the 2008 Blacksmith institute world's worst polluted places report. According to the 2014 WHO report, air pollution in 2012 caused the deaths of around 7 million people worldwide. Air pollution can be natural or man-made. The major primary man-made pollutants are Sulphur dioxide (SO_2), Nitrogen dioxide, (NO_2), Carbon monoxide (CO) and particulate matter (PM). The Air quality Index (AQI) is an index or a number used to characterize the quality of air at a given location. It is used for local and regional air quality reporting and management in many metropolitan or urban cities of the world. Aba is a commercial and industrializing urban city, South East of Nigeria. The multiple linear regression technique is used to estimate the air quality of the city at any given time. The model for the estimation

**Corresponding author: Email: ugwason@yahoo.com;*

was established as:

$$AQI = -24.87 + 0.17SO_2 + 0.042NO_2 + 1.68CO + 3.60PM$$

The Analysis of Variance (ANOVA) conducted on the model shows that the model is statistically significant and so, good as an estimating technique in that area. The major assumptions of the model were checked and none was found to have been violated.

Keywords: Air pollution; air quality index; multiple linear regression technique; analysis of variance; correlation coefficient; coefficient of determination.

1. INTRODUCTION

Air pollution is the introduction of particulates, biological molecules, or other harmful materials into the Earth's atmosphere, causing disease, death to humans, and damage to other living organisms such as food crops, or the natural or built environment. Air pollution may come from anthropogenic (man-made) or natural sources. The atmosphere is a complex natural gaseous system that is essential to support life on planet Earth [1].

The major primary pollutants produced by human activities include: Sulphur Oxides, Nitrogen Oxides, Carbon Monoxide, and Particulate Matter [2].

Sulphur oxides (SO_X) - particularly Sulphur dioxide, a chemical compound with the formula SO_2 is produced naturally by volcanoes and in various industrial process. Coal and petroleum often contain sulphur compounds, and their combustion generates sulphur dioxide. Usually in the presence of a catalyst such as Nitrogen dioxide ($N0_2$), further oxidation of $S0_2$ forms Sulphuric acid (H_2S0_4), and thus acid rain [3]. This is one of the causes of concern over the environmental impact of the use of these fuels as power sources [4]. Nitrogen Oxides ($N0_X$) – particularly Nitrogen dioxide, are expelled from high temperature combustion, and are also produced during thunderstorms by electric discharge. They can be seen as brown haze dome above or a plume downwind of cities.

Nitrogen dioxide (NO_2) is one of several Nitrogen oxides and one of the most prominent air pollutants. It is a reddish-brown toxic gas with a characteristic sharp, biting odor.

Carbon monoxide (CO) is a colourless, odorless, toxic yet non-irritating gas that is produced by incomplete combustion of fuel such as natural gas, coal or wood. Vehicle exhaust is a major source of carbon monoxide.

Particulate matter (PM) or just particulates are tiny particles of solid or liquid suspended in a gas. Some particulates occur naturally, originating from volcanoes, dust storms, forest and grassland fires, living vegetation, and sea spray. Human activities, such as the burning of fossil fuels in vehicles, power plants and various industrial processes also generate significant amount of PM. Increased level of fine particles in the air are linked to health hazards such as heart disease, altered lung function and lung cancer [5].

The Air Quality Index (AQI) is an index for reporting air quality in a location. It tells how clean or polluted the air around us is, and what associated health effects might be a concern for us. AQI as a yardstick runs from 0 to 500. The higher the value, the greater the level of air pollution and the greater the health concern [6]. The AQI is divided into six categories as shown in Table 1 below.

Table 1. Categories of AQI

AQI value	Level of health concern	Colour
0 – 50	Good	Green
51 – 100	Moderate	Yellow
101 -150	Unhealthy for sensitive groups	Orange
151 – 200	Unhealthy	Red
201 – 300	Very unhealthy	Purple
301 – 500	Hazardous	Maroon

2. MATERIALS AND METHODS

Aba is a commercial and industrializing urban city in Abia State, South East, Nigeria with a population of about 3 million people. It is located on latitude $05º18^1$N and longitude $07º35^1$E and has a tropical climate and an average annual temperature of about 28ºC. Two major seasons are experienced: A rainy season between April and October and a dry season between November and March. Eight different locations within the high activity area of the city were chosen for the study. Within each of the locations, five different points were chosen to collect data on four major air pollutants: Sulphur dioxide (SO_2); Nitrogen dioxide (NO_2), Carbon monoxide (CO), and Particulate Matter (PM).The air pollutants were measured with automatic standard samplers, in-situ, in all the locations. SO_2 and NO_2 were measured with Crowncon Gasman SO_2 and NO_2 automatic gas monitors, model CE89/336/EEC. Carbon monoxide was measured with an Extech instrument, Carbon (II) oxide meter, model C010. Particulate Matter was measured with Crowncon Gasman monitor model No. 1000.

In all a total of 132 samples were collected in March, May, August and December 2014 to reflect the two seasons in Aba.

AQI for the four pollutants were calculated by using the pollutant concentration data, Table 2 and equation1 (linear interpolation) below. The average of the AQI for each point was calculated to represent the AQI for the point.

$$I_P = \frac{I_{HI-}I_{LO}}{BP_{HI-}BP_{LO}}(C_P - BP_{LO}) + I_{LO} \quad (1)$$

Where I_P is the index for pollutant P, C_P is the rounded concentration of pollutant P, BP_{HI} is the breakpoint that is greater than or equal to C_P, BP_{LO} is the breakpoint that is less than or equal to CP, I_{HI} is the AQI value corresponding to BP_{HI} and I_{LO} is the AQI value corresponding to BP_{LO} [7].

2.1 The Model

The Multiple Linear Regression (MLR) technique, being one of the most popular and widely used techniques of multivariable analysis is used to develop a model that can be used to forecast the air quality in Aba urban. The general form of the multiple linear Regression model is:

$$Y = b_0 + b_1X_1 + b_2X_2 + \cdots + b_kX_k + e \quad (2)$$

(Where Y is the dependent variable, b_o,b_1,b_2,…, b_k are linear regression parameters or coefficients that measure the direction and magnitude of the relationship between the associating independent variable and the dependent variable, X_1, X_2,…, X_k are the independent variables, e is the error term of the linear regression).

Following this general form of the MLR model, the model to forecast air quality in Aba is:

$$AQI = \beta_o + \beta_1SO_2 + \beta_2NO_2 + \beta_3CO + \beta_4PM + e \quad (3)$$

(Where AQI is the Air Quality Index, β_i ($i = 1,2,3,4$) are the linear regression parameters or coefficients to be estimated.

The statistical package for social sciences (SPSS) software was used to analyse the data collected from a total of 132 points in the 8 locations of the study.

3. RESULTS AND DISCUSSION

The result of the data analysis using the SPSS software is as presented and discussed below:

A Coefficient of Correlation (R) is a coefficient that illustrates a quantitative measure of some type of correlation and dependence, meaning statistical relationships between two or more random variables or observed data values." It is a measure of the relationship or association between the observed and the predicted values. It lies between -1 and +1 (i.e $-1 \le R \le 1$). A value of -1 means negative perfect correlation, a value of 0 means no correlation and a value of +1 means positive perfect correlation. [2]. The value of R for our model is found in the output Table 3, labeled Model Summary. The value of 0.975 indicates a high positive correlation between the observed and the predicted values.

The Coefficient of Determination (R^2) is the square of the Coefficient of Correlation and determines the proportion or percentage of the variance in the dependent variable that is explained by the model. In other words, it measures how well the model fits the data. The information for our model is also found in the output table 1, labeled Model Summary. A value of 0.951 means that 95.1% of the variance in Air Quality Index (AQI) is actually explained by the model. This means that the model is a good fit for the data.

From the column titled Unstandardized Coefficients in Table 2, labeled Coefficients, we get the values of the regression coefficients of our regression equation as: b_0 = -24.87; b_1 = 0.17; b_2 = 0.04; b_3 = 1.68; b_4 = 3.60. Therefore our regression equation is:

$$AQI = -24.87 + 0.17SO_2 + 0.042NO_2 + 1.68CO + 3.60PM$$

With standard error of estimation of 19.71.

The Collinearity Statistics also shown in Table 4 shows that none of the independent variables has Tolerance value close to zero and none of their Variance Inflation Factor (VIF) values were greater than 5. These shows that our assumption of no perfect multicollinearity amongst the independent variables has not been violated.

From this table we find out the relative contribution of each of the pollutants to the variation in Air Quality Index. The contributions from SO_2, CO and PM were significant (sig. value of each = 0.000 < 0.05). SO_2 contributed the most with a Beta value of 0.800 followed by CO with a Beta value of 0.342. Particulate matter has a Beta value of 0.184. The contribution from NO_2 was insignificant with a sig. value of 0.376 > 0.05 and a Beta value of 0.022.

The result of the analysis of variance (ANOVA) conducted on the model is found in the output Table 5, labeled ANOVA. This tests the null hypothesis that multiple R in the population equals 0 (i.e P = 0). The significant value of 0.000 indicates that our model reaches statistical significance. This really means P < 0.0005.

From the Normal P-P Plot Of The Standardized Residual and the Scatter Plot, we find out that our assumptions of Normality, Linearity, Homoscedasticity and Independence of the residuals have not been violated as the Normal P-P plot, Fig. 1, shows that all the points lie in a reasonably straight diagonal line from bottom left to top right. In the scatter plot of the standardized residuals, Fig. 2, the residuals are roughly rectangularly distributed with most of the points concentrated at the centre. There is no clear or systematic pattern to the residuals. This also shows that the assumptions of Normality, Homoscedasticity, and Independence have not been violated.

Table 2. Breakpoints for the AQI

PM_{10} ($\mu g/m^3$)	$PM_{2.5}$ ($\mu g/m^3$)	CO (ppm)	SO_2 (ppm)	NO_2 (ppm)	AQI	
0 – 54	0.0 - 15.4	0.0 - 4.4	0.000 - 0.034	-	0 - 50	Good
55 - 154	15.5 - 40.4	4.5 - 9.4	0.035 - 0.144	-	51 - 100	Moderate
155 - 254	40.5 - 65.4	9.5 - 12.4	0.145 - 0.224	-	101 - 150	Unhealthy for sensitive group
255 - 354	65.5 - 150.4	12.5 - 15.4	0.225 - 0.304	-	151 - 200	Unhealthy
35 5- 424	150.5 - 250.4	15.5 - 30.4	0.305 - 0.604	0.65 - 1.24	201 - 300	Very Unhealthy
425 - 504	250.5 - 350.4	30.5 - 40.4	0.605 - 0.804	1.25 - 1.64	301 - 400	Hazardous
505 - 604	350.5 - 500.4	40.5 - 50.4	0.805 - 1.004	1.65 - 2.04	401 - 500	Hazardous

NOTE: $N0_2$ has no short-term NAAQS and can generate an AQI only above a value of 200. (USEPA, 2006)

Table 3. Model summary

Model	R	R square	Adjusted R square	Std. error of the estimate
1	0.975	0.951	0.949	19.71774

Table 4. Coefficients

Model	Unstandardized coefficients (β)	Standardized coefficients (β_1)	t values	Sig. values	Tolerance values	VIF values
Constant	-24.870		-3.457	0.001		
SO_2	0.171	0.800	25.316	0.000	0.391	2.560
NO_2	0.042	0.022	0.889	0.376	0.644	1.554
CO	1.684	0.342	11.412	0.000	0.434	2.302
PM	3.597	0.184	5.981	0.000	0.413	2.424

Table 5. ANOVA

Model	Sum of squares	df	Mean square	F	Sig.
Regression	948359.848	4	237089.962	609.816	0.000
Residual	49376.236	127	388.789		
Total	997736.083	131			

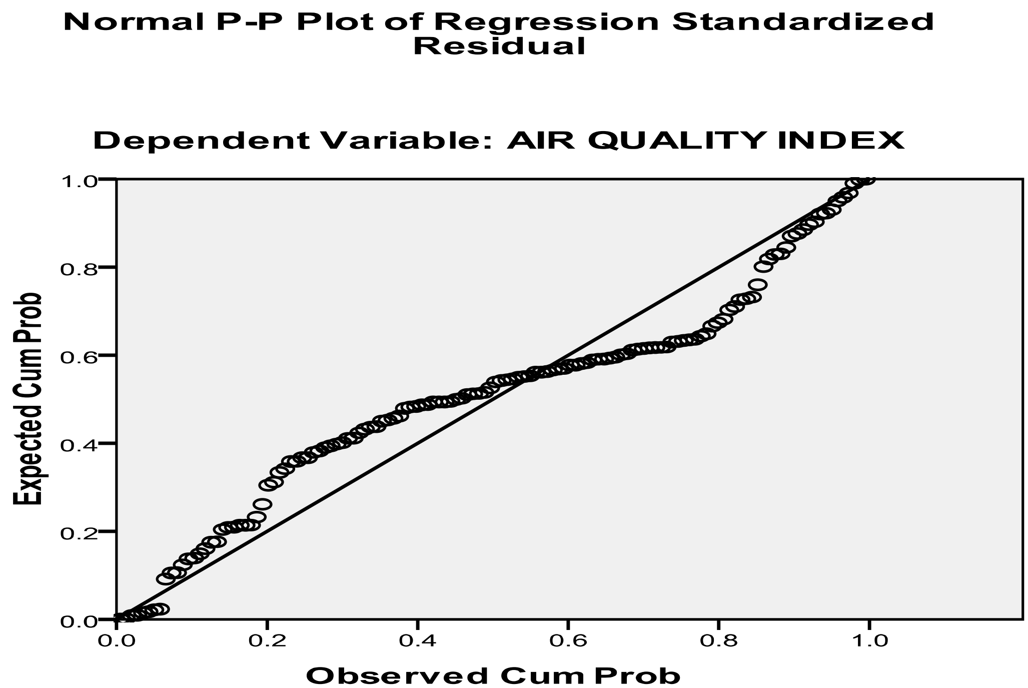

Fig. 1. The normal P- P plot of the standardized residuals

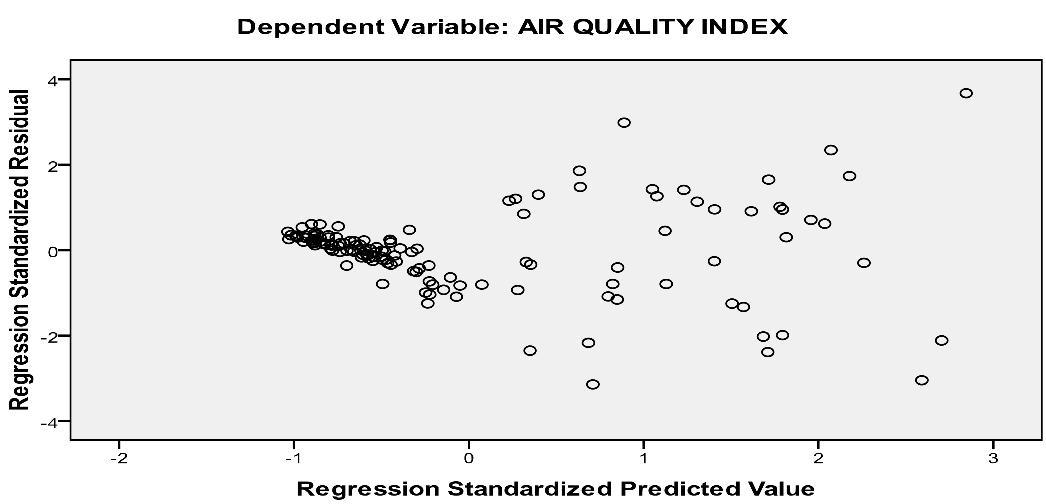

Fig. 2. The scatterplot of the standardized residuals

4. CONCLUSION

A mathematical model for the estimation of air quality in Aba urban was found to be $AQI = -24.87 + 0.17SO_2 + 0.042NO_2 + 1.68CO + 3.60PM$. The Correlation Coefficient (R) of 0.975 shows that there is a strong positive relationship between Air Quality Index (AQI) and the Pollutants- SO_2, NO_2, CO, and PM.

The Coefficient of determination (R^2) shows that 95.1% of the variance in AQI is explained by the model. In other words, the model is a very good fit for the data collected and so will be very good in forecasting the Air Quality in Aba urban at any given time.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Chang JC, Hanna SR. Air quality model performance evaluation. Meteorology and Atmospheric Physics. 2004;87:167-196.
2. Anikender K, Pramila G. Forecasting of air quality in Delhi using principal component regression technique. Atmospheric Pollution Research. 2011;2: 436-444.
3. WHO, 2014 7 Million premature deaths annually linked to air pollution. "Reports". Worst polluted. Org. Archived. From the original on 11 August 2010. Retrieved 25 March.
4. Aron R. Models for estimating current and future sulphur dioxide concentrations in Taipei, Bulletin of Geophysics. 1984;25: 47-52.
5. Chen H, Goldberg MS, Vileneuve PJ. A systematic review of the relation between long-term exposure to ambient air pollution and chronic disease. Reviews on Environmental Health. 2008;23(4):243-97.
6. Cogliani E. Air Pollution forecasting in cities by an air pollution index highly correlated with metrological variables. Atmospheric Environment. 2001;35:2871-2877.
7. United State Environmental Protection Agency (USEPA). Guideline for reporting of daily air quality – Air Quality Index (AQI). Publication No. EPA-454 IB-06-001 May, 2006.

Peer-review history:
The peer review history for this paper can be accessed here:
http://sciencedomain.org/review-history/11982