Bulletin of Environment, Pharmacology and Life Sciences
Bull. Env. Pharmacol. Life Sci., Vol 4 [1] December 2014: 139-145

Online ISSN 2277-1808
Journal's URL:http://www.bepls.com
CODEN: BEPLAD
Global Impact Factor 0.533
Universal Impact Factor 0.9804

ORIGINAL ARTICLE

Study of the Effects of Climatic Parameters on the Pollutants CO, SO_2, NOx, and O_3 through Regression Models

Forouzan Farrokhian , Maryam Bahri*
Department of Environmental Management, Ahvaz Branch, Islamic Azad University, Ahvaz, Iran
***Email:** bahrimaryam@gmail.com

ABSTRACT
Thousands of people lost their lives due to pollution or suffer from fatal diseases in different parts of the world every year. According to the world health statistics, the risk of air pollution has been more than what has been assumed. The aim of this study was to evaluate the correlation between climatic parameters (temperature, pressure, humidity, wind direction and speed) and indices of air pollutants (SO2, NOX, CO, and O3) and present the related regression models. The study was conducted at the industrial area in the southwestern of Iran, which has a five-year interval with an average of 24 hours from 06.21.2009 to 21.06.2014 and by statistical analysis methods. The findings indicated that a negative correlation existed between NOx with atmospheric parameters such as temperature, humidity, pressure, and wind direction. All the atmospheric parameters were important factors affecting O3 pollution and there was a positive correlation between pressure, temperature, humidity, and wind speed with ozone wind but a negative correlation existed between wind direction and ozone. Carbon monoxide was positively correlated with atmospheric parameters such as temperature, humidity, pressure, and wind speed but Sulfur dioxide was negatively correlated with atmospheric parameters such as temperature, humidity, pressure, and wind speed and direction. In this regard, some studies are suggested on modeling the point spread of pollutants in order to develop policies and programs of the air pollution control and prevention.
Keywords: *Air pollution, Climate parameters, Regression equations, Petrochemical special economic zone of Mahshahr, Iran*

Received 23.08.2014 **Revised 01.11.2014** **Accepted 03.12.2014**

INTRODUCTION

The March 2014 report of the World Health Organization (WHO) shows that about 7 million have lost their lives due to air pollution in 2012. This is accounted for one-eighth of total global mortality. That is two times the previous estimates and introduces air pollution as the largest risk to environmental health in the world. Moreover, the March 2009 report has regarded policymakers' awareness of the impact of climate changes on health as an important issue who could take action to reduce the effects of global climate changes through priorities and investments.

In a climate, there is a wide range of different factors or variables, including climate conditions and geographical factors, man-made environmental conditions, and so forth that can contribute to the air pollution and dispersion of air pollutants in any area. Distribution and concentration of contaminants can be predicted by statistical models in a variety of places and situations through the study on the relationship between climate variables and various factors in each region. This is important because it identifies the various pollutants and their accumulation centers and consequently sensitive and vulnerable sites and also predicts the necessary actions to prevent the occurrence of adverse consequences.

Given the importance of estimation in fluctuations in the concentration of pollutants NOx and NO2 in cities of 36 regions in Europe, Beelen *et al*. [1] presented the regression models that could explain these small changes. However, the local predictions were different for each study area. Fujii *et al*. [3] analyzed the management of air pollution in China's industrial sectors from 1998 to 2009 and conducted their studies using three pollutants of sulfur dioxide (SO2), the dust and smoke.

Chen *et al*. [3] presented a regression model to investigate the aspects of air pollution for the pollutants SO2, NO2, and PM10 in Tianjin, China. This study aimed to improve understanding of the structure of dimensions of air pollution in urban areas, and assess and support of policies. In another study by Lu et al. [4], the researchers have examined pollution index system (API) consisting of five major air pollutants, including sulfur dioxide (SO2), respirable suspended particulates (RSP), carbon monoxide (CO), and ozone (O3) in order to address the air pollution.

In a study conducted by Beelen et al. [5], the performance of regression modeling and dispersion modeling were compared in estimating the small changes in the long-term concentrations of air pollution in a Dutch urban area. Mölter et al. [6] investigated the air pollution modeling for infectious diseases and presented new ways to combine regression and air dispersion. Pey et al. [7] examined monitoring the sources and processes of controlling air quality of Mediterranean urban air. They studied the control parameters of air quality, including concentrations of mass PM, distribution and size of extremely fine particles, levels of gaseous pollutants, complete chemical properties of particles PM10 and PM2 0.5 in urban air and their reactions in various sources of emissions or atmospheric processes affecting the Mediterranean urban area.

Different pollutants in different seasons or different times have different values according to their correlation with environmental factors. However, according to what we are looking for in this study, based on the study of Zeldin and Meisel [8] in urban environments, changes in pollutant concentrations are summarized in Table 1.

Table (1): The concentration of gaseous pollutants in different seasons according to environmental conditions [8]

Pollutant	Spring	Summer	Fall	Winter
SO_X	Medium	Low	Medium	High
NO_X	Medium	Low	Medium	High
CO	Medium	Low	Medium	High
O^3	Medium	High	Medium	Low

This study sought to examine the effect of climatic parameters such as temperature, humidity, pressure, wind speed and direction on the air pollutants indices, including nitrogen oxides (NOx), carbon monoxide (CO), sulfur dioxide (SO2), and ozone (O3). It was also aimed to understand whether a logistic regression model could examine the presence or absence of a correlation between climatic parameters and the air pollutants indices. Studying and understanding these factors by experts and scholars in any area could aid legislators in setting local legislation to control and reduce pollution.

The research data was collected in the special petrochemical economic zone of Mahshahr. Special Petrochemical economic zone is located in an area of approximately 2,600 hectares in the southwest of Iran and on the coast of the Persian Gulf in Mahshahr, Imam Khomeini (RA) sector at 5 sites from north to south. The data was extracted from the bases of special zone of Mahshahr in a five years interval with an average of 24 hours. This period began from 21/06/2009 to 21/06/2014.

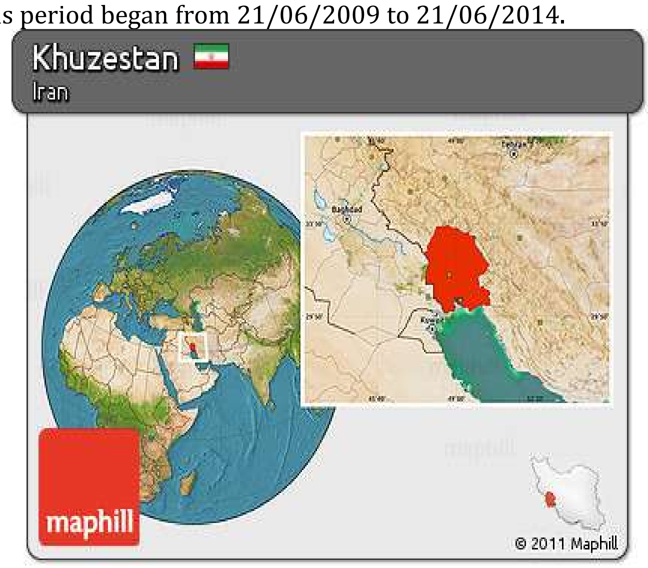

Figure 1: Position of Mahshahr Special Economic Zone in Iran and in the world

Hypotheses and modeling

The assumptions used for this analysis are:

- A lack of correlation between the independent variables (lack of collinearity): Eigenvalue and Condition Index was used in the study. Eigenvalues close to zero indicated that internal correlation of predictions is large, and small changes in the amounts of data led to large changes in the estimated coefficients of the regression equation. Condition Index value of greater than 15 indicated the probability of co-linearity between the independent variables and of greater than 30 was indicative of a serious problem in regression. If proven correlation between the independent variables, different methods could be used, one of such methods was differential of data. Data differential means to reduce the independent variable data from the mean value of the same data and use the new data instead of the independent variable.

- Constant variance: the plot of residuals was used to study the constant variance versus the predicted values. If the fitted model was appropriate, the graph should be symmetric to the point $e_i = 0$ and the points around this point were distributed uniformly. It showed the constant error variance. Moreover, if the points were distributed in the form of a funnel, it indicated that the error variance was not constant. In this case, the test is not possible and the problem must be resolved by changing the dependent variable data.
- Normality of residuals: Shapiro-Wilk test was used to evaluate it. The null hypothesis of this test showed that the residuals had a normal population with mean zero and one. Therefore, if the P value was greater than 0.05, we could conclude that the residuals were normal. The hypotheses are as follows.

H0: residuals are normal.
Ha: residuals are not normal.

Due to the large number of data in the study, it was very difficult to obtain a P value greater than 0.05 because this test is one of the strongest tests of normality. In this case we would say that if the value of the test statistic (w) was greater than 0.95, the null hypothesis was accepted and it could be concluded that the residuals were normally distributed.

In the next step, regression model was presented using regression techniques based on the research hypotheses, the climatic parameters, and indices of air pollutants. The analysis indicated the statistical dependencies between a dependent variable and one or more (descriptive) independent variable. Multiple regression analysis was used in this project because of some independent variables and one dependent variable and also because the aim of the study was to determine the linear relationship between variables. The significance of the relationship and correlation of climatic parameters with each pollutant were analyzed separately. The model used in this study is:

$$Y_i=\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + e_i$$

Each sign of the relation is defined as:

Y_i: the dependent variable value

β_0: intercept of the regression line origin, X_1: independent variable of wind direction, X_2: independent variable of pressure, X_3: independent variable of temperature, X_4: independent variable of humidity, X_5: independent variable of wind speed, β_i: the coefficient of i^{th} independent variable, and e_i : i^{th} error value

Data Analysis

Table (2) shows the calculation of Eigen values and the independent variables for different pollutants.

Table 2: The Eigen values and condition index of the independent variables

Model	Dimension	NOx		O3		CO		SO2	
		Eigenvalue	Condition Index	Eigen value	Condition Index	Eigen value	Condition Index	Eigen value	Condition Index
1	Wind Direction	5.645	1.000	5.630	1.000	5.642	1.000	5.633	1.000
	Baro Pressure	.244	4.808	.246	4.789	.229	4.964	.242	4.826
	Temperature	.066	9.216	.062	9.508	.066	9.276	.062	9.507
	Humidity	.036	12.493	.054	10.198	.055	10.108	.054	10.173
	Wind Speed	.008	25.926	.008	25.903	.008	25.764	.008	26.020
	e_i	5.254E-6	1036.562	5.192E-6	1041.274	5.260E-6	1035.684	5.159E-6	1044.895

As seen in Table 2, there were eigenvalues greater than 15 and close to zero. This indicated that there was a serious problem in estimating regression line. Therefore, the collinearity problem could be resolved by differential method. Table 3 shows the mean values of various contaminants to differentiate.

Table 3: Mean values for each parameter of atmospheric pollutants

Variable	Mean NOx	Mean O3	Mean CO	Mean SO2
Wind Direction	219.45	220.74	218.19	219.79
Baro Pressure	1004.5	1003.8	1004.3	1003.9
Temperature	26.018	26.913	26.391	26.897
Humidity	52.189	49.839	51.349	50.118
Wind Speed	3.3837	3.9318	3.8949	3.9572

Table 4: Eigenvalues and Condition Index for the independent variables after subtracting the mean

Model	Dimension	NOx		O3		CO		SO2	
		Eigen value	Condition Index	Eigen value	Condition Index	Eigen value	Condition Index	Eigen value	Condition Index
1	Wind Direction	2.650	1.000	2.682	1.000	2.601	1.000	2.680	1.000
	Baro Pressure	1.075	1.570	1.075	1.579	1.101	1.537	1.070	1.582
	Temperature	1.000	1.628	1.000	1.638	1.000	1.613	1.001	1.637
	Humidity	.956	1.665	.917	1.710	.957	1.649	.924	1.703
	Wind Speed	.249	3.265	.253	3.258	.264	3.141	.251	3.269
	e_i	.071	6.124	.073	6.058	.076	5.838	.074	6.009

Table 4 shows the resolution of the problem of collinearity of the independent variables after differentiation of the data because none of the numerical values of the Condition index was greater than 15.

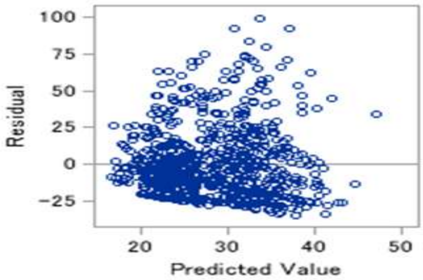

Figure a: NOx

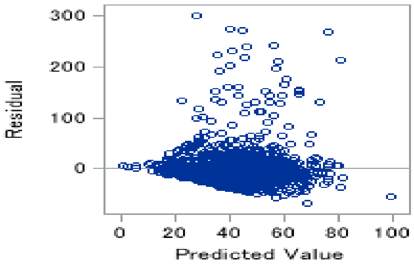

Figure B: O3 pollution

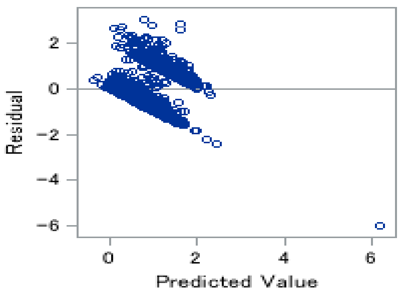

Figure c: pollutant CO

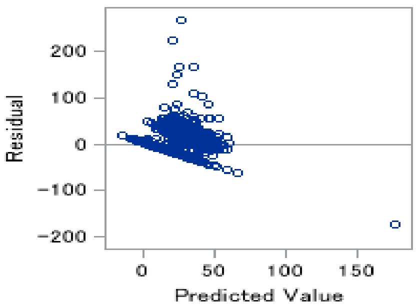

Figure D: SO2 emissions

Figure 2: Variance of the residuals of independent variables

As illustrated by examining the residuals variance, a funnel plot shows instability of variances and if the points are uniformly distributed, the constant variance could be inferred. Figure 2 shows that:

Figures (a), (b), and (d) of the variance of residuals of independent variables had a funnel-like state for the pollutants NOx, O3, and SO2. Thus, an appropriate conversion should be applied to stabilize the variance of the data.

Figure (c): variance of the residuals of independent variables for the pollutant CO; this figure was not in a funnel shape, we could conclude that the variance was constant.
Box-Cox method was used to find a suitable transformation. Such a conversion showed a numeric value so that the dependent variable data should reach to its power in order to be equal to the variance.
According to the conversion, the data of dependent variables should reach to the powers 0.5, 0.3, and 0.16 respectively for the pollutants NOx, O3, and SO2. Graph of residuals versus predicted values after conversion of the data indicates that the variance is constant.

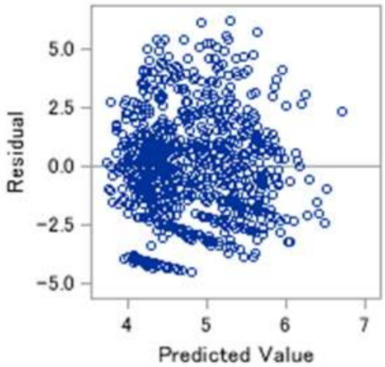

Figure a: pollutant NOx

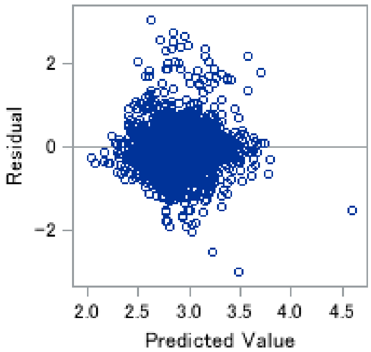

Figure b: pollutant O_3

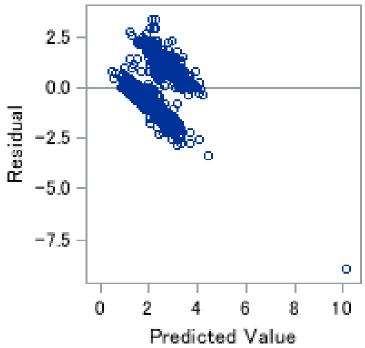

Figure c: pollutant SO2

Figure 3: Variance of the residuals of independent variables after data transformation
Table (5) shows the research data normalization through Shapiro-Wilk test. The normality test, after transforming the data (since the Shapiro Wilk test value was greater than 0.95 and the number of data is large) indicates normality of the residuals.

Table 5: Test of normality after the data transformation

Tests for Normality																
	NOx pollution				O3 pollution				CO pollution				SO2 pollution			
Test	Statistic		p Value		Statistic		p Value		Statistic		p Value		Statistic		p Value	
Shapiro-Wilk	W	0.989415	Pr < W	<0.0001	W	0.956387	Pr < W	<0.0001	W	0.9502	Pr < W	<0.0001	W	0.965089	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.040951	Pr > D	<0.0100	D	0.065276	Pr > D	<0.0100	D	0.107967	Pr > D	<0.0100	D	0.083511	Pr > D	<0.0100
Cramer-von Mises	W-Sq	0.211576	Pr > W-Sq	<0.0050	W-Sq	1.762722	Pr > W-Sq	<0.0050	W-Sq	5.047286	Pr > W-Sq	<0.0050	W-Sq	3.119224	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	1.614708	Pr > A-Sq	<0.0050	A-Sq	12.68054	Pr > A-Sq	<0.0050	A-Sq	27.34752	Pr > A-Sq	<0.0050	A-Sq	16.52948	Pr > A-Sq	<0.0050

Finally, Table (6) shows the coefficients of regression model based on all independent variables for the equation of pollutants.
For pollutants Nox, the independent variable is the square root of values and new dependent variables are the centralized values of major independent variables. Independent variables of p-values less than 0.05 were significantly associated with contaminants Nox but they did not have a significant relationship with wind speed with a p-value greater than 0.05. hence, the only variable that could be removed from the model and had no effect on the dependent variable of the second square root of Nox was the variable wind speed.
As shown in Table (6), all p-values of the pollutant O3 were less than the significance level of 05/0. Thus, it could be concluded that the pollutants of O3 had a significant relationship with the independent variables (wind direction, pressure, temperature, humidity, and wind speed), and were the factors affecting the pollutants.

Table 6: Estimates of the parameters of all independent variables of the pollutants

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Parameter Estimates For NOx					
Intercept	1	4.80744	0.07469	64.37	<.0001
dWind_Direction	1	-0.00481	0.00169	-2.85	0.0045
dBaro_Pressure	1	-0.06919	0.02259	-3.06	0.0023
dTemperature	1	-0.13198	0.02328	-5.67	<.0001
dHumidity	1	-0.01240	0.00577	-2.15	0.0321
dWind_Speed	1	-0.05618	0.11142	-0.50	0.6143
Parameter Estimates For O3					
Intercept	1	2.91451	0.01503	193.93	<.0001
dWind_Direction	1	-0.00111	0.00035140	-3.17	0.0016
dBaro_Pressure	1	0.04705	0.00458	10.28	<.0001
dTemperature	1	0.06034	0.00476	12.68	<.0001
dHumidity	1	0.00781	0.00118	6.59	<.0001
dWind_Speed	1	0.15138	0.01486	10.19	<.0001
Parameter Estimates For CO					
Intercept	1	0.90401	0.02212	40.87	<.0001
dWind_Direction	1	0.00081750	0.00051073	1.60	0.1097
dBaro_Pressure	1	0.06806	0.00669	10.17	<.0001
dTemperature	1	0.06918	0.00694	9.96	<.0001
dHumidity	1	0.01027	0.00173	5.94	<.0001
dWind_Speed	1	0.41010	0.02203	18.61	<.0001
Parameter Estimates For SO2					
Intercept	1	0.71515	0.00295	242.28	<.0001
d Wind_Direction	1	-0.00016534	0.00006894	-2.40	0.0166
dBaro_Pressure	1	-0.00927	0.00090212	-10.28	<.0001
dTemperature	1	-0.00983	0.00093402	-10.52	<.0001
dHumidity	1	-0.00163	0.00023374	-6.96	<.0001
dWind_Speed	1	-0.06333	0.00290	-21.86	<.0001

In addition, the values obtained in Table (6) shows that the independent variables of pressure, temperature, humidity, and wind speed had a significant relationship with the pollutant carbon monoxide (P-values less than 0.05). However, there was no significant relationship with the wind direction (p-value greater than 0.05). Thus, the wind direction parameter was removed in the regression equation of CO. In other words, the wind direction had no effect on the rate of changes of CO in this region.

By examining the obtained values of the pollutants SO2, all independent variables (wind direction, pressure, temperature, humidity, and wind speed) had a significant with the pollutant Sulfur dioxide and were the factors affecting this pollutant (p-value less than 0.05). Thus, the final regression model for these four pollutants is as follows:

Equation 1:

$$\sqrt{Y_{NOx}} = 4.80709 - 0.00482(X_1 - \bar{X}_1) - 0.06597(X_2 - \bar{X}_2) - 0.12945(X_3 - \bar{X}_3) - 0.01221(X_4 - \bar{X}_4)$$

Equation 2:

$$(Y_{O3})^{0.3} = 2.91451 - 0.00111(X_1 - \bar{X}_1) + 0.04705(X_2 - \bar{X}_2) + 0.06034(X_3 - \bar{X}_3) + 0.00781(X_4 - \bar{X}_4) + 0.15138(X_5 - \bar{X}_5)$$

Equation 3:

$$Y_{co} = 0.90402 + 0.06699(X_2 - \bar{X}_2) + 0.06565(X_3 - \bar{X}_3) + 0.00867(X_4 - \bar{X}_4) + 0.40987(X_5 - \bar{X}_5)$$

Equation 4:

$$(Y_{SO_2})^{-16/100} = 0.71515 - 0.00016534(X_1 - \bar{X}_1) - 0.00927(X_2 - \bar{X}_2) - 0.00983(X_3 - \bar{X}_3) - 0.00163(X_4 - \bar{X}_4) - 0.06333(X_5 - \bar{X}_5)$$

CONCLUSION

Air pollution is one of the environmental problems that has challenged researchers and practitioners today. In addition to economic costs, such as costs for patients who have been affected by air pollution, postponement of workdays caused by air pollution, and so forth, this problem has severely affected the people welfare. The correlation between climatic parameters (temperature, pressure, humidity, wind direction and speed) and indices of air pollutants (SO2, NOX, CO, and O3) were investigated in this study.
As seen in the regression equation of the pollutant NO_X, the square root of NOx had a negative correlation with the independent variables (atmospheric parameters of temperature, humidity, pressure, and wind direction) meaning that amounts of NOx are reduced with the increase in each of the independent variables. This is acceptable according to the research of Zeldin and Meisel (Table 1), which shows concentration of NOx increases in the winter.
The regression equation of pollutant ozone (O3) shows that all the independent variables (wind direction, pressure, temperature, humidity, and wind speed), are factors affecting O3, and except for the wind direction, which is negatively correlated with ozone, other parameters had a positive correlation. It means that it they increase, the ozone concentration increases and according to the studies of Zeldin and Meisel and Table 1, which shows that the concentration of ozone increases in the summer, is acceptable.
According to the regression analysis for the pollutant carbon monoxide (CO), it can be said that carbon monoxide had a positive correlation with the independent variables such as temperature, humidity, pressure, and wind speed, meaning that if each of the independent variables increases, the amounts of CO increases. This is different from the research of Zeldin and Meisel (Table 1), which shows the carbon monoxide concentration is increased in the winter. Since the traffic is the main source of the production of CO in big cities, this is acceptable in the special economic zone of Mahshahr due to lack of traffic as well as the lack of carbon monoxide-producing industries such as iron and steel industries, power plants, refineries, etc.
Finally, according to the regression analysis for Sulfur dioxide (SO2), it can be said that this pollutant had a negative correlation with all independent variables (temperature, humidity, pressure, wind speed and direction), meaning that as each of the independent variables increased, the quantities of SO2 reduced. This is acceptable according to Zeldin and Meisel (Table 1), which shows that the concentration of Sulfur dioxide increases in winter.
The obtained regression equations can be used to coordinate during overhaul in order to provide an efficient and accurate management of the air pollution control and prevention in the special economic zone of Mahshahr where many of the petrochemical and lateral industries are operate in an area less than two thousand hectares.

REFERENCES

1. Beelen, R., Hoek, G., Vienneau, D., Eeftens, M., Dimakopoulou, K., Pedeli, X., et al. (2013). Development of NO2 and NOx land use regression models for estimating air pollution exposure in 36 study areas in Europe e- The ESCAPE project. *Atmospheric Environment , 72*, 10-23.
2. Fujii H., M. S. (2013). Decomposition analysis of air pollution abatement in China: empirical study for ten industrial sectors from 1998 to 2009. *Journal of Cleaner Production* , 59:22-31.
3. Chen, L. W. (2012). A land use regression model incorporating data on industrial point source pollution. *Journal of Environmental Sciences* , 24(7): 1251–1258.
4. Lu W-Z., H. H.-d. (2011). Assessing air quality in Hong Kong: A proposed, revised air pollution index (API). *Building and Environment* , 46(12): 2562–2569.
5. Beelen, R., Voogt, M., Duyzer, J., Zandveld, P., & Hoek, G. (2010). Comparison of the performances of land use regression modelling and dispersion modelling in estimating small-scale variations in long-term air pollution concentrations in a Dutch urban area., *44*, pp. 4614-4621.
6. Mölter, A., Lindley, S., Vocht, F. d., Simpson, A., & Agius, R. (2010). Modelling air pollution for epidemiologic research — Part I: A novel approach combining land use regression and air dispersion. *Science of the Total Environment , 408*, 5862–5869.
7. Pey J., A. A. (2010). Monitoring of sources and atmospheric processes controlling air quality in an urban Mediterranean environment. *Atmospheric Environment* , 44: 4879-4890.
8. Zeldin, M. &. (1978). Use of Meterological Data in Air Quality Trend Analysis. North Carolina. *US Environmental Protection Agency.*
9. *WHO.* (2009). Retrieved from World Health Organization: http://www.who.int /mediacentre/news/notes/ 2009/climate_change_20090311/en/

CITATION OF THIS ARTICLE

Forouzan F , Maryam B. Study of the Effects of Climatic Parameters on the Pollutants CO, SO_2, NOx, and O_3 through Regression Models. Bull. Env. Pharmacol. Life Sci., Vol 4[1] December 2014: 139-145