Mathematical Modelling in Civil Engineering
Vol. 9-No. 2-2013
Doi: 10.2478/mmce-2013-0007

STATISTICAL ANALYSIS FOR PROGNOSIS OF A PHOTOCHEMICAL SMOG EPISODE

CAMELIA GAVRILĂ– Assoc. Prof., PhD, Department of Thermal-Hydraulic Systems and Atmosphere Protection, Technical University of Civil Engineering Bucharest, e-mail: cgavrila2003@yahoo.com
FLORINELA ARDELEAN – Assist. Prof., PhD, Department of Thermal-Hydraulic Systems and Atmosphere Protection, Technical University of Civil Engineering, e-mail: florinela.ardelean@gmail.com
ADRIANA COMAN – Researcher, PhD, Laboratoire Inter-Universitaires des Systèmes Atmosphériques, UMR7583, CNRS, UPEC, UPD, Paris, France, e-mail: adicom_2000@yahoo.com
ELENA BURCHIU – Assist. Prof., PhD, Department of Thermal-Hydraulic Systems and Atmosphere Protection, Technical University of Civil Engineering, e-mail: eburchiu@gmail.com

Abstract: In this paper we describe the evaluation of various climatic parameters in establishing their prognostic value in a photochemical smog episode. Our application was validated using real data from the "Cercul Militar National" and "Sos. Mihai Bravu nr. 47-49", from April 2008 to May 2008. The study was performed on hour averages of pollutant concentrations and meteorological parameters and the statistical analysis was based on multiple regressions. We concluded by using mathematical and statistical methods, [1], that an accurate Global Solar Radiation is one of the most important and essential information in the pollution report.

Keywords: multiple regressions, photochemical smog, global solar radiation

1. Introduction

The intensification of traffic based upon an unfit territorial assembly as well as elevated emissions due to pollutants from different sources has lead to the deterioration of exterior air quality in urban centers and implicitly to the growth of pollution levels within buildings.

Bucharest city is one of the great urban congestions, and the state of air quality is in permanent degradation. In previous years other studies [2-3] were performed for Bucharest. Given the great diversity and variation of pollution sources in the city, like the special dispersion conditions of pollutants, an experimental study was established that would emphasize the pollutant levels of this metropolis. The next pollutant species were analyzed through increased toxicity: nitrous oxides (NO/NO2), carbon monoxide (CO), and ozone (O3) [4].

A statistical analysis has been performed in this article, using SPSS software, to verify if and to what extent a series of pollutants and weather factors like TEMP (temperature), SR (global solar radiation), RH (relative humidity), CO, NO2, NO, influence O3 variations. To achieve this, the hierarchic method of multiple regressions has been used.

For controlling the ever increasing complexity of these fields, a proper understanding of their processes is important as is the ability to reason about them, [5].

The characteristics of the processes vary widely; however, typically only part of all the factors by which they are governed can be observed in practice, [6-8].

Moreover the processes include the effects of individual as well as random variation.

Essentially they are uncertain; the uncertainties involved render an overall understanding hard to achieve and reasoning a daunting task.

Models capturing these processes and methods for using these models are thus called for to support decision making in real-life practice.

We need statistics because we want to draw more valid conclusions from limited amounts of data and significant differences are often masked by climatological variability or experimental imprecision. It is a set of styles and page layout settings that determine the appearance of a document. On the other hand, the human mind excels at finding patterns and relationships and tends to generalize too much.

2. Statistical Analysis

The purpose of many research projects is to assess relationships among a set of variables and regression techniques often used as statistical analysis tools in the study of such relationships. Research designs may be classified as experimental or observational. Regression analyses are applicable to both types; yet the confidence one has in the results of a study can vary with the research type. In most cases, one variable is usually taken to be the response or dependent variable, that is, a variable to be predicted from or explained by other variables.

The other variables are called predictors, explanatory variables, or independent variables. Choosing an appropriate model and analytical technique depends on the type of variable under investigation, [9]. In this paper we considered the case with several independent variables or covariates (multiple regressions).

2.1 Multiple Linear Regression Analysis

Regression analysis serves two major purposes:

1. Control or intervention,
2. Prediction.

In many studies one important objective is measuring the strength of a statistical relationship between the binary dependent variable and each independent variable or covariate measured; findings may lead to important decisions in public health interventions.

The effect of some factors on a dependent or response variable may be influenced by the presence of other factors through effect modifications (i.e., interactions). Therefore, to provide a more comprehensive analysis, it is very desirable to consider a large number of factors and sort out which ones are most closely related to the dependent variable.

In multiple linear regressions, the prediction of a dependent variable, dependent on multiple independent variables, is achieved with the help of a regression equation. Y is the dependent variable, $X_1, X_2, ..., X_n$ independent variables, $b_1, b_2, ..., b_n$ regression coefficients, corresponding to the independent variables and a constant. The regression equation will be:

$$Y = a + b_1 \cdot X_1 + b_2 \cdot X_2 + ..., + b_n \cdot X_n$$

This method, which is a multiple linear regression analysis, involves a linear combination of the explanatory or independent variables; the variables must be quantitative with particular numerical values. A covariate or independent variable may be dichotomous or continuous (categorical factors will be represented by dummy variables). Quantity variables and errors (residue) must be normally distributed.

In this article, the hierarchal method was chosen as an analysis method for multiple linear regressions because of the fact that the order in which the independent variables are going to be introduced is known and we want to know if introducing a new independent variable in the model brings extra clarity or prediction concerning the dependent variable.

3. Results and Discussions

Our application uses real data from "Cercul Militar National", [10], see table 1

Table 1

Real data from "Cercul Militar National"							
Time	**NO**	**NO2**	**CO**	**RH**	**SR**	**TEMP**	**O3**
00:00	0,3096	0,6266	0,3456	0,6102	0	0,4174	0,2362
01:00	0,334	0,5236	0,2834	0,673	0,0001	0,3414	0,3141
02:00	0,1894	0,3058	0,1038	0,7563	0,0004	0,2615	0,4
03:00	0,0892	0,1933	0,0535	0,8164	0,0004	0,192	0,462
04:00	0,0175	0,0622	0,0171	0,8676	0,0007	0,129	0,4702
05:00	0	0	0	0,9263	0,0004	0,0702	0,4724
06:00	0,108	0,1418	0,0524	0,9816	0,0013	0,0238	0,2581
07:00	0,2688	0,2855	0,1743	1	0,0185	0	0,0844
08:00	0,6747	0,3656	0,5754	0,9297	0,0518	0,0589	0
09:00	1	0,4141	1	0,7918	0,0898	0,1821	0,0394
10:00	0,8964	0,5224	0,9166	0,6696	0,1279	0,3173	0,2454
11:00	0,785	0,4853	0,8428	0,4983	0,1711	0,471	0,4525
12:00	0,7244	0,4776	0,8546	0,3707	0,2168	0,6116	0,6299
13:00	0,6768	0,423	0,8611	0,2006	0,8471	0,7802	0,7583
14:00	0,6836	0,4201	0,8225	0,1003	1	0,8837	0,8686
15:00	0,6538	0,453	0,7979	0,0382	0,9337	0,9491	0,9245
16:00	0,6007	0,4482	0,8075	0	0,8398	0,9934	1
17:00	0,6102	0,547	0,8834	0,0191	0,6616	1	0,9181
18:00	0,6392	0,5512	0,8802	0,0184	0,3613	0,9906	0,9083
19:00	0,6244	0,693	0,8556	0,086	0,2324	0,9222	0,7307
20:00	0,6715	0,842	0,8845	0,2232	0,0873	0,8043	0,5257
21:00	0,643	1	0,7476	0,3249	0,0821	0,6951	0,2783
22:00	0,6715	0,9503	0,6342	0,4345	0,0002	0,6001	0,1345
23:00	0,5034	0,7313	0,5369	0,5529	0,0003	0,5063	0,2169
00:00	0,3096	0,6266	0,3456	0,6102	0	0,4174	0,2362

The data was done using the SPSS software (Statistical Package for the Social Sciences), [1]. This application was initially developed in 1965, at Stanford University in California. Since then, it has become the most used data analysis application in the field, getting to a point in which it dominates the field. Given its popularity and universal character, using the SPSS application is the easiest transferable aptitude of all necessary skills in research. SPSS application is constantly updated, from a pallet of presented statistical techniques point of view, as well as output detail.

The correlations between variables are presented in the table below. It is determined that between RH, SR, TEMP and $O_{3,}$ the correlation level is high.

Table 2

Correlations

		O3	**RH**	**SR**	**TEMP**	**CO**	**NO**	**NO2**
Pearson Correlation	O3	1,000	-,806	,810	,789	,381	,134	-,053
Sig. (1-tailed)	O3	.	,000	,000	,000	,030	,262	,401

In order to establish what mathematical instruments can apply, for the beginning, the allocation of these parameters has been studied (to see if variables are allocated normally or not). The bar charts of standardized residue presented below indicate the fact that they are distributed normally.

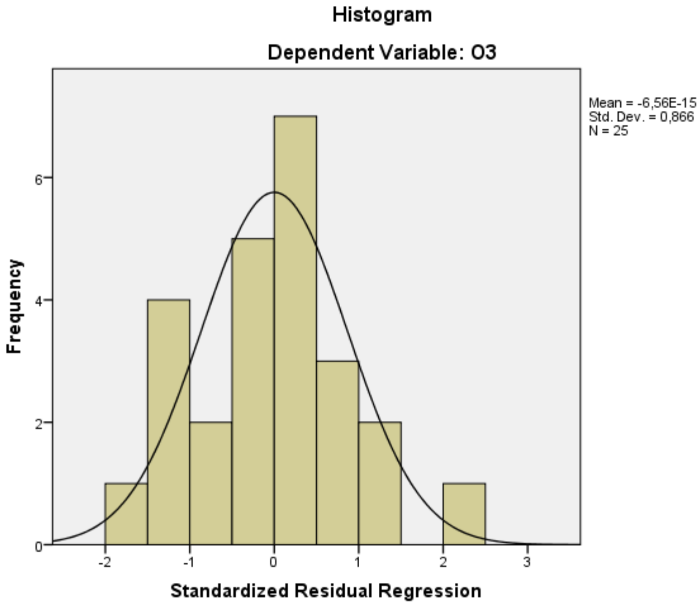

Fig. 1 - The histogram of standardized residual regression

In order to apply the hierarchal regression method, 3 models in SPSS are chosen, and from the obtained Model Summary table, we can ascertain that model 3 has the highest Adjusted R Square coefficient (0.985).

Table 3

Model Summary[d]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	,865[a]	,747	,711	,1633914	,747	20,722	3	21	,000	
2	,955[b]	,911	,894	,0991591	,164	37,018	1	20	,000	
3	,995[c]	,989	,985	,0366354	,078	64,259	2	18	,000	2,100

a. Predictors: (Constant), TEMP, SR, RH
b. Predictors: (Constant), TEMP, SR, RH, CO
c. Predictors: (Constant), TEMP, SR, RH, CO, NO2, NO
d. Dependent Variable: O3

Following the parameter variations in table 1 and the obtained results in tables 2 and 3, it can be observed that in order to realize the best ozone concentration evaluation, identifying some functions (mathematical models) is mandatory for the following time frames: [0÷7], [8÷17] and [18÷24]. In table 4 we present the coefficients for each model regression.

Table 4

Coefficients[a]

Model		[0÷7]	[8÷17]	[18÷24]
3	NO	0.2200585	- 0.2913930	0.3589665
	NO2	- 0.5152055	0.2433068	- 0.7943421
	CO	- 0.7889867	- 0.1152879	0.9454855
	RH	- 4.7260437	- 3.235572	- 1.8999056
	SR	- 2.2991443	- 0.0568894	0.3018651
	TEMP	- 3.4204421	- 2.1322516	- 1.759181
	(Constant)	5.0777646	3.3130033	2.4119156

a. Dependent Variable: O3

Our application was validated using real data from “Cercul Militar National” and “Sos. Mihai Bravu nr. 47-49”, from April 2008 to May 2008 with good results, table 5.

Table 5

Ozone concentration evaluation

Time	"Cercul Militar National"			Sos. Mihai Bravu nr. 47-49		
	O_3 measured	O_3 computed	Precision %	O_3 measured	O_3 computed	Precision %
07:00	0.0844	0.0837	99.20	0.9394	0.9621	97.64
13:00	0.7583	0.7586	99.96	0.5	0.5164	96.82
15:00	0.9245	0.9402	98.32	0.4848	0.4971	97.52
20:00	0.5257	0.5257	100	0.4394	0.5843	75.20

The trust threshold of the model (precision) has slightly dropped in another area of Bucharest metropolis (Mihai Bravu Street), according to the precision the latter is based on (National Military Circle), because orographic (topographic) differences between these two areas exist and differences regarding the pollutant sources, fact which leads to modifying physics – chemical processes that took place in the molded space, implicitly to changes in ongoing developed processes at the interface between the atmosphere and the lithosphere (ground).

5. Conclusions

In order to evaluate the ozone concentration, we have applied the hierarchal regression method, from which we can conclude that the model which contains the variables TEMP, SR, RH, CO, NO2, NO best explains its evolution. The correlations between variables are presented in table 2 and one can observe that global solar radiation has a very high importance in ozone (O3) evaluation.

The realized model is unique for the analyzed season (April-May), complies with the International Legislation regulations in the field and has been verified with good results in two air quality monitoring points, placed in Bucharest metropolis, more exactly: National Military Circle and 47-49 Mihai Bravu Street.

Using the earlier presented calculus algorithm, used for building the ozone immission prognosis model, corresponding to April-May season, other tropospheric ozone concentration evaluation models can be obtained for other time intervals in a year.

Mathematical models which have the capacity to accurately describe the correlations between weather conditions and atmospheric concentration levels of pollutant emissions are excellent instruments for analysis and interpretation of experimental data as well.

References

[1] Jaba, E., Grama A. (2004). *Analiza statistica cu SPSS sub Windows*, Bucharest, Ed. Polirom.

[2] Ardelean F. (2005). *Atmosphere pollution in the urban area – Doctoral thesis*, Technical University of Civil Engineering, Bucharest.

[3] Colda, I., Ardelean, F. (2004), *Meteorology and climatology*, Bucharest, Ed. Conspress.

[4] Angelino E., Degani R., Gnagnetti M., Mortarino S., Noekel K., Pateri R., Sucharov L.J. (2000). *Traffic induced air pollution in Milan city: a modelling study – Urban transport V: urban transport and the environment for the 21st century*, Milan.

[5] Borrego C., Tchepel O., Barros N., Miranda A.I. (2000). Impact of road traffic emissions on air quality of the Lisbon region. *Atmospheric Environment, 34*, 4683-4690.

[6] Harkonen J., Kukkonen J., Valkonen E., Karppinen A. (1998). Influence of vehicle emission characteristics and meteorological conditions on urban NO2 concentrations. *International Journal of Vehicle Design, 20*, 125-130.

[7] Ionescu A., Candau Y., Mayer E., Colda I. (2000). Analytical determination and classification of pollutant concentration fields using air pollution monitoring network data. Methodology and application in the Paris area, during episodes with peak nitrogen dioxide levels. *Environmental Modelling & Software 15*, 565-573.

[8] Chandra M. P., Praveen K. M., Jayarama R. S. (2003). Monitoring of air pollution in indian metropolitan cities: modeling and quality indexing. *International Journal of Environment and Pollution, 21*, 365-382.

[9] The Romanian Auto Registry. (2003). *Air quality monitoring – annual reports 2002 and 2003*.