Development & Application of Computer Techniques to Environmental Studies VII,
C.A. Brebbia, P. Zannetti & G.Ibarra-Berastegi (Editors) © 2000 WIT Press, www.witpress.com, ISBN 1-85312-819-8

Multiple linear regression modelling for short-term real-time prediction of hourly ozone, NO_2 and NO levels in the area of Bilbao

G. Ibarra[1], A. Elías[2], M.V. Albizu[3] & E. Agirre[4]
[1]Fluid Mechanics Department. University of the Basque Country. Bilbao. Spain. [2]Environmental Engineering Department. University of the Basque Country. Bilbao. Spain. [3]Environmental Department. Basque Government. Bilbao. Spain. [4]Applied Mathematics Department. University of the Basque Country. Bilbao. Spain.

Abstract

The objective of this paper is to show that by means of simple statistical Multiple Linear Regression (MLR) models, an air pollution and meteorological network, initially designed for diagnosis purposes, can be used to forecast up to eight hours ahead hourly levels of O_3, NO_2, NO. The network can be used as a prognostic tool at a given location if O_3, NO_2, NO and meteorological parameters are measured jointly. The performance of the models was determined by comparison of hourly predictions and real observations for a one year period (1994) and was compared with the simplest prediction possible: persistance of levels. The performance of the models has been evaluated using the set of statistical standard parameters included in the so called Model Validation Kit. The results show that models yield a significantly better prediction up to 8 hours ahead for O_3, NO_2, and NO than persistance of levels.

1 Introduction

Bilbao is an industrial city located in a complex terrain area of North Central Spain, with a population of about 1 million.

Development & Application of Computer Techniques to Environmental Studies VII,
C.A. Brebbia, P. Zannetti & G.Ibarra-Berastegi (Editors) © 2000 WIT Press, www.witpress.com, ISBN 1-85312-819-8

For monitoring and immission control purposes, an air pollution and meteorological network is operated in the whole area by the Basque Government since 1977. In its 1993-94 configuration, the network measured several meteorological parameters at 13 locations and air pollution variables at 26. It was originally designed as a diagnostic network to describe the evolution of the different pollutants.
It is well known that ozone is originated mainly due to the interaction of NO_X and VOC`s on the one hand and meteorological effects on the other hand. Several works (Bloomfield, P. *et al.,* [1]; Flaum, J. *et al.* [2]; Spitchinger *et al.*, [3]; Abdulwahab, S. *et al.* [4]) have shown the importance in ozone formation of the following meteorological variables:temperature, radiation, relative humidity and wind speed.

In the frame of a research project, it was decided to explore the prognostic capabilities of Bilbao´s network to use it as a forecasting tool for photochemical smog. The idea was to find an inexpensive and easy-to-use tool, that could be easily be applied along with the usual network management activities. For that reason, it was decided to build statistical prognostic models adapted to the area of study, based on Multiple Linear Regression. These models can be fitted with historical data and can be used to yield predictions of O_3, NO_2 and NO. The objective is to obtain mathematical expressions which will relate immisions (output) with a set of input variables. A strategy also known as the "*black box*" approach.

A deep analysis of data from the network, showed that Deusto was the only location reliable for the elaboration of a statistical prognostic model, mainly because many relevant variables known to be involved in the formation of ozone (O_3, NO_2, NO, temperature, radiation, wind speed and thermal contrast) were being measured jointly at this or a nearby location during the 1993-1994 period.
Any prognostic model should perform better than the most simple prediction: persistance of levels. Several statistical tools can be used to build statistical prognostic models to forecast photochemical smog. Some of the models are elaborated for diagnosis purposes (Abdulwahab, S., [4]) but most models predict the daily O_3 maximum and use either time series analysis (Kuang-Jung Hsu, [5]); Simpson R.W. and Layton A.P., [6]) or multiple linear regression (MLR). For daily maximum ozone prediction, models based on MLR yield better results than models based on time series analysis (Robeson S.M. and Steyn D.G., [7]) which seems not to be able to improve persistance. Models based on Multiple Linear Regression have also been used since late 1970`s to forecast ozone (Cassmassi, J.C., [8]; Cassmassi, J.C., [9]) at the South Coast Air Quality Management District.
However, a clearly establishment of which model performs better for a given situation is not a simple task, mainly because authors tend to give the results of their models using parameters or graphics which are not comparable. After previous works, with the aim to overcome this problem (Hanna, S.R. *et al.,* [10]) several workshops have been conducted since 1991 with the support of COST 710 and COST 615 actions, in the frame of the European Union. As a result of this work, a Model Validation Kit (European Commission, 1996 [11]) has been

prepared which includes a set of FORTRAN routines to calculate a model's performance according to objective statistical parameters.
The objectives of the present work were:

1) To build statistical prognostic models for O_3, NO_2 and NO based on Multiple Linear Regression, adapted to Bilbao. These models would yield predictions of photochemical smog from 1h to 8h ahead.
2) To define according to the objective criteria of the Model Development Kit how well they perform and under which circunstamces
3) To find out if joint measurements at one given location (like Deusto) of O_3, NO_2, NO and meteorological parameters was a reliable strategy to provide the network with prognostic capabilities for O_3, NO_2 and NO. If so, it could be applied to more locations of the network and these equations could be incorporated to the daily management activities of the network.

2 Database

The idea was to use data from a diagnosis network in order to explore its prognostic capabilities by means of a stochastic approach. The database used for this study were hourly data measured in the Bilbao air pollution and meteorological network during 1993 and 1994. O_3, NO_2 and NO were measured at Deusto. Wind speed was measured at Feria (700 m away), global radiation (direct+diffuse radiation : units cal/cm^2-h) at Sondika (6 km away) and temperature at Feria and Banderas (200 m. above sea level). It was considered that radiation 6 km away and temperature and wind speed 700 m away from Deusto was reliable enough to be used in the prognostic models. Because no real data of the true vertical profile of temperatures in the atmosphere were available, the thermal contrast between two stations located one at sea level (Feria) and other one 200 m.a.s.l. (Banderas), was used as a descriptive tool of the presence or not of a thermal inversion (Ibarra, [12]) . However, the thermal contrast between these two locations is not the real difference of temperatures between Feria and the atmospheric layer located 200 m above it on its vertical and it cannot be used as a "true" value of the thermal gradient, as accurate as that obtained from a rawinsonde. In the case of Bilbao when compared with the Dry Adiabatic Gradient, the thermal contrast between Banderas and Feria always follows its main trends and can give an idea of the presence or not of a thermal inversion. The Dry Adiabatic Gradient is –1°C/100m, i.e. under dry adiabatic conditions temperature decreases by 1° C every 100 m increment in height.

3 Methodology

Many statistical air quality forecast techniques have been developed in the past years (Wilks, [13]) some of them based on MLR. There are two ways of building MLR statistical prognostic models (Zannetti, [14])

1. **Deterministic models.** The MLR coefficients are calculated by the least-squares method from a development sample and applied to the test and all future samples.

Development & Application of Computer Techniques to Environmental Studies VII,
C.A. Brebbia, P. Zannetti & G.Ibarra-Berastegi (Editors) © 2000 WIT Press, www.witpress.com, ISBN 1-85312-819-8

2. **Mixed deterministic-statistical models.** The MLR coefficients are also calculated by least squares but they are continuosly recalculated using most recent observations and the equations are used to forecast next cases.

In this work, *deterministic* models were built on the basis of multiple linear regression with coefficients calculated by least squares and hourly data from the network at Deusto, Banderas, Feria and Sondika. Each equation was built using hourly data of the variables mentioned above, and in principle, appart from NODATA values, each year had 24*365=8760 cases. The group of models were *deterministic models* (from now onwards Fixed Coefficients Models, FCM models) and to calculate their coefficients, data from year 1993 were used. In the line of previous works (Takakatsu Inoue *et al.*, [15]) available data was divided into a «development sample» (year 1993) and a «test sample» (year 1994). Data from year 1993 was chosen to be the development sample and 1994 the test sample. The coefficients calculated in the development sample were used in the test sample and considered to be valid for the 8760 cases of year 1994.

3.1. Building the models.

The candidate independent variables proposed to build the models were chosen taking into account known photochemical smog production mechanisms. It is well known (Finlayson-Pitts B.J. and Pitts J.N., [16]) the mutual influence of O_3, NO_2 and NO and also the importance of meteorological variables available from the network at Deusto: temperature, radiation, wind speed and thermal inversion. With this in mind the MLR models were built as follows: If current hour is T, in both groups of models O_3, NO_2 and NO hourly average levels at time T+K (K=1, 2....8) are predicted jointly using multiple linear regression. When each of the three photochemical smog variables (O_3, NO_2 and NO) at time T+K is the dependent variable, the other two at time T+K, plus O_3, NO_2, NO, thermal inversion, radiation, wind speed and temperature at time T-Z (with Z ranging from 0 to P) are the candidate independent variables as can be seen in equations 1-2-3. The set of equations 1-2-3 is the core of each of the FCM models for prediction K hours ahead.

$$O_{3(T+K)} = AO3(T+K)_0 + AO3(T+K)_1 \times NO_{2(T+K)} + AO3(T+K)_2 \times NO_{(T+K)} +$$
$$+\sum_{Z=0}^{P} BO3(T+K)_Z \times O_{3(T-Z)} + \sum_{Z=0}^{P} CO3(T+K)_Z \times NO_{2(T-Z)} + \sum_{Z=0}^{P} DO3(T+K)_Z \times NO_{(T-Z)} +$$
$$+\sum_{Z=0}^{P} EO3(T+K)_Z \times GT_{(T-Z)} + \sum_{Z=0}^{P} FO3(T+K)_Z \times RADT_{(T-Z)} +$$
$$+\sum_{Z=0}^{P} GO3(T+K)_Z \times VX_{(T-Z)} + \sum_{Z=0}^{P} HO3(T+K)_Z \times VY_{(T-Z)} + \sum_{Z=0}^{P} IO3(T+K)_Z \times TR_{(T-Z)} \quad \textbf{(EQ. 1)}$$

$$NO_{2(T+K)} = ANO2(T+K)_0 + ANO2(T+K)_1 \times O_{3(T+K)} + ANO2(T+K)_2 \times NO_{(T+K)} +$$
$$+\sum_{Z=0}^{P} BNO2(T+K)_Z \times O_{3(T-Z)} + \sum_{Z=0}^{P} CNO2(T+K)_Z \times NO_{2(T-Z)} + \sum_{Z=0}^{P} DNO2(T+K)_Z \times NO_{(T-Z)} +$$
$$+\sum_{Z=0}^{P} ENO2(T+K)_Z \times GT_{(T-Z)} + \sum_{Z=0}^{P} FNO2(T+K)_Z \times RADT_{(T-Z)} +$$
$$+\sum_{Z=0}^{P} GNO2(T+K)_Z \times V_{X(T-Z)} + \sum_{Z=0}^{P} HNO2(T+K)_Z \times V_{Y(T-Z)} + \sum_{Z=0}^{P} INO2(T+K)_Z \times TR_{(T-Z)} \quad \textbf{(EQ. 2)}$$

$$NO_{(T+K)} = ANO(T+K)_0 + ANO(T+K)_1 \times O_{3(T+K)} + ANO(T+K)_2 \times NO_{2(T+K)} +$$
$$+\sum_{Z=0}^{P} BNO(T+K)_Z \times O_{3(T-Z)} + \sum_{Z=0}^{P} CNO(T+K)_Z \times NO_{2(T-Z)} + \sum_{Z=0}^{P} DNO(T+K)_Z \times NO_{(T-Z)} +$$
$$+\sum_{Z=0}^{P} ENO(T+K)_Z \times GT_{(T-Z)} + \sum_{Z=0}^{P} FNO(T+K)_Z \times RADT_{(T-Z)} +$$
$$+\sum_{Z=0}^{P} GNO(T+K)_Z \times V_{X(T-Z)} + \sum_{Z=0}^{P} HNO(T+K)_Z \times V_{Y(T-Z)} + \sum_{Z=0}^{P} INO(T+K)_Z \times TR_{(T-Z)} \quad \textbf{(EQ. 3)}$$

where:

$O_{3(T+K)}$, $NO_{2(T+K)}$ and **$NO_{(T+K)}$** are the predicted values of O_3, NO_2 and NO at time T+K. When each of them is the dependent variable at time T+K the equations 1-2-3 include the other two as candidate independent variables at time T+K.

$AXX(T+K)_0$, is a constant for the prediction of the XX compound at time T+K (XX=O_3 or NO_2 or NO).

$AXX(T+K)_1$ and **$AXX(T+K)_2$** are the coefficients of the other two variables at time T+K when each of the three photochemical smog variables (O_3, NO_2 and NO) at time T+K is the dependent variable XX.

$O_{3(T-Z)}$, $NO_{2(T-Z)}$ and **$NO_{(T-Z)}$** with Z from 0 to P are the measured current (T, Z=0) and past values (T-Z, Z>0) of O_3, NO_2 and NO. **$BXX(T+K)_Z$, $CXX(T+K)_Z$, $DXX(T+K)_Z$** are, respectively, their coefficients when used to predict hourly levels of the XX compound (XX=O_3 or NO_2 or NO) at time T+K.

$GT_{(T-Z)}$, $RADT_{(T-Z)}$, $V_{x(T-Z)}$, $V_{y(T-Z)}$, $TR_{(T-Z)}$ with Z from 0 to P are the measured current (T, Z=0) and past values (T-Z, Z>0) of respectively, thermal gradient, radiation, wind speed (expressed as V_x and V_y) and temperature. **$EXX(T+K)_Z$, $FXX(T+K)_Z$, $GXX(T+K)_Z$, $HXX(T+K)_Z$,** and **$IXX(T+K)_Z$** are, respectively, their coefficients when used to predict hourly levels of the XX compound (O_3, NO_2 or NO) at time T+K.

At the coefficient calculation stage, since historical data are used, true values of $O_{3(T+K)}$, $NO_{2(T+K)}$ and $NO_{(T+K)}$ are available. However, at the prediction stage, the three equation system must be solved jointly since $O_{3(T+K)}$, $NO_{2(T+K)}$ and $NO_{(T+K)}$ are not known. To that end, initial estimations of $O_{3(T+K)}$, $NO_{2(T+K)}$ and $NO_{(T+K)}$ are computed and after an iterative process, final predictions of $O_{3(T+K)}$, $NO_{2(T+K)}$ and $NO_{(T+K)}$ are reached. K was chosen to range from 1 to 8 hours which in the case of having reliable results, can be considered to be the time ahead that short-time control strategies can be implemented by those in charge of the network. Z is the number of hours back that should be considered for the measured parameters. Each chosen variable (O_3, NO_2, NO, thermal inversion, radiation, wind speed and temperature) at any hour T-Z (0 =< Z =< P) was considered to be a variable itself, as can be seen in the above set of equations. The maximum

Development & Application of Computer Techniques to Environmental Studies VII,
C.A. Brebbia, P. Zannetti & G.Ibarra-Berastegi (Editors) © 2000 WIT Press, www.witpress.com, ISBN 1-85312-819-8

value of Z, P, was chosen to be 15 because due to the presence of NODATA values, only in 50% of the cases all candidate variables were different from NODATA value and, in principle, a prediction would be possible in year 1993. For higher values of P the percentage was even smaller. Therefore, P, was chosen to be 15 in the belief that is no use to find, maybe, a better model with Z ranging from 0 to P>15, if for practical reasons it can only be run a few times during the year. So, in principle, 130 (16*8=128; 128+2=130) candidate variables were chosen for each of the three equations.

The deterministic FCM models for each of the O_3, NO_2 and NO at T+K (K=1,....8) were fitted at the model development stage with a MLR analysis using data from year 1993 (development sample). MLR analysis was applied to each of the three equations 1-2-3 with K ranging from 1 to 8. Tolerance filtering and stepwise regression were applied to all the equations and out of 130 candidate variables, the number of chosen variables ranged from 26 for $NO_{(T+1)}$ to 42 for $NO_{2(T+4)}$. However, only a few of them accounted for the highest increases of R^2. Listwise deletion of NODATA values led to a total of valid cases to perform the calculation of coefficients ranging from 5100 to 5700 out of 8760 . This led to the calculation of the optimal number of variables, their coefficients, the determination coefficient R^2 and the standard error.

Table 2, depicts the coefficients of the 5 most important variables involved in each of the equations. Analysis of equations showed that relationships between ozone, NO_2 and NO were in agreement with known explanatory mechanisms.

Analysis of residuals showed normality, equality of variance, mean near 0 and independence of errors. At the model development stage, as a typical indicator of the goodness of fit in a classical Multiple Linear Regression (MLR) analysis, the determination coefficient R^2 was used. The determination coefficient can be understood as the proportion of the overall variability explained by the model. The values of R^2 ranged between 0.51 for NO prediction 8 hours ahead and 0.92 for ozone prediction 1 hour ahead. The coefficients obtained were used to forecast levels of $O_{3(T+K)}$, $NO_{2(T+K)}$ and $NO_{(T+K)}$ for the 8760 cases of year 1994.

4 Conclusions

The use of the *Model Evaluation Kit* allowed comparison of results between the predictions obtained with FCM and persistance. It has been employed in different model comparison exercises (Olesen, H.R., [17]) and can throw light on the performance of a model and also its field of application. Following the same nomenclature, C_p represents predicted values and C_o, observed values. A line above C_p or C_o means averages and S stands for standard deviation. The measures proposed in the kit are:

1. *Correlation coefficient R.* Customary measure of the kindness of the adjustment that quantifies the global description of the model.
2. *Normalized Mean Square Error (NMSE).* It is a version of the mean square error but normalized to the product of means of observed and predicted values.
3. *Factor of 2 (FA2).* If C_p representes predicted values and C_o, observed values, the FA2 statistic is the fraction of data which falls in the range $0.5<C_p/C_o<2$.

Development & Application of Computer Techniques to Environmental Studies VII,
C.A. Brebbia, P. Zannetti & G.Ibarra-Berastegi (Editors) © 2000 WIT Press, www.witpress.com, ISBN 1-85312-819-8

4.*Fractional Bias (FB).* It is a normalized measure used to compare between the overall mean of the model and that of observed values. A model with FB=0 is a model that represents perfectly the mean value. If absolute value of FB is 1, the error in the mean is of the same order than the real mean, that is, of the 100%. Absolute values of FB above 1 represent errors in the mean estimation over 100%. Absolute values of FB below 1 represent errors in the mean estimation below 100%. Negative values of FB represent overestimation and positive values underestimation.

5. *Fractional Variance (FS).* It is a normalized measure used to compare the overall standard deviation of the model and that of the observed values. A model with FS=0 is a model that represents perfectly the true standard deviation. If absolute value of FS is 1, the error in the standard deviation is of the same order than the real one, that is, 100%. Absolute values of FS above 1 represent errors in the estimation of the standard deviation over 100%. Absolute values of FS below 1 represent errors in the estimation below 100%. Negative values of FS represent overestimation and positive values underestimation.

$R = \frac{\overline{(C_o - \overline{C_o}) - (C_p - \overline{C_p})}}{SC_p \times SC_o}$	$FB = \frac{\overline{C_o} - \overline{C_p}}{0.5 \times (\overline{C_p} + \overline{C_o})}$
$NMSE = \frac{\overline{(C_o - C_p)} \times \overline{(C_o - C_p)}}{\overline{C_p} \times \overline{C_o}}$	$FS = \frac{SC_O - SC_P}{0.5 \times (SC_O - SC_p)}$

Table 1. Goodness of fit indicators according to the Model Validation Kit

5 Results

Application of the models to the test sample (year 1994) led to the following results of table 2-3-4. Having joint measurements of O_3, NO_2, and NO at a given location and meteorological variables close enough, it is possible to use MLR to provide the network with forecasting capabilities of O_3, NO_2, NO at that location. MLR equations can be easily calculated and incorporated to the network management activities. Although the coefficients of the equations are likely to be influenciated by local conditions, the methodology is easily aplicable to any air pollution network and can be used as an easy-to-use and simple tool. Due to the large amount of cases used to draw these conclusions they can be considered to be robust enough. The quality of the predictions is at least, as good as that from much more sophisticated and expensive models. Computational needs and implementation costs are small since it can be run on a PC and calculation time is low. The network gains in forecasting capabilities up to 8 hours ahead only at the location where the mentioned variables are measured jointly. The model is intended to forecast photochemical smog levels 8 hours ahead. The only interest of predictions between 1 and 7 hours is, not only to give an estimation of the levels 8 hours ahead but also an hourly description of the whole episode.

Development & Application of Computer Techniques to Environmental Studies VII,
C.A. Brebbia, P. Zannetti & G.Ibarra-Berastegi (Editors) © 2000 WIT Press, www.witpress.com, ISBN 1-85312-819-8

N# HOURS AHEAD	R		NMSE		FA2	
	FCM	PERS	FCM	PERS	FCM	PERS
1	.928	.913	.11	.14	.836	.827
2	.845	.798	.24	.31	.739	.718
3	.767	.680	.36	.49	.662	.627
4	.694	.565	.47	.66	.616	.569
5	.635	.463	.55	.82	.578	.524
6	.596	.38	.6	.94	.562	.488
7	.580	.316	.61	1.03	.556	.451
8	.560	.27	.63	1.10	.544	.432

Table 2. Ozone. 1994. Model performance comparison K (1,...8) hours ahead

N# HOURS AHEAD	R		NMSE		FA2	
	FCM	PERS	FCM	PERS	FCM	PERS
1	.867	.86	.05	.06	.974	.972
2	.733	.713	.10	.12	.924	.926
3	.638	.591	.13	.18	.893	.890
4	.585	.493	.14	.22	.875	.856
5	.549	.418	.15	.25	.855	.824
6	.527	.37	.16	.27	.853	.801
7	.525	.343	.16	.29	.845	.791
8	.505	.33	.16	.29	.84	.78

Table 3. NO_2. 1994. Model performance comparison K (1,...8) hours ahead

N# HOURS AHEAD	R		NMSE		FA2	
	FCM	PERS	FCM	PERS	FCM	PERS
1	.852	.83	.48	.6	.711	.884
2	.691	.618	.87	1.36	.519	.77
3	.603	.458	1.04	1.94	.438	.698
4	.543	.332	1.17	2.4	.415	.64
5	.514	.259	1.21	2.67	.408	.614
6	.498	.231	1.25	2.79	.392	.573
7	.497	.239	1.28	2.78	.402	.56
8	.507	.274	1.26	2.67	.415	.553

Table 4. NO. 1994. Model performance comparison K (1,...8) hours ahead

Intensive application of this strategy to all the interesting locations in the network can spatially cover several areas of interest for prognostic purposes. This approach can be used as an inexpensive and useful element in the air quality management of an area where a network exists.

Acknowledgements. This work was performed under financial support of the University of the Basque Country, UPV *Euskal Herriko Unibertsitatea.* The authors wish to thank the Environmental Department of the Basque Government for providing with data from their air pollution network

References

[1] Bloomfield P., Royle, J. Steinberg, L., and Yang Q. (1996) Accounting for meteorological effects in measuring urban ozone levels and trends. *Atmospheric Environment* **30**, 3067-3077.

[2] Flaum J., Rao, T. And Zurbenko, I. (1996) Moderating the influence of meteorological conditions on ambient ozone concentrations. *Journal of the Air & Waste Management Association* **46**, 35-46.

[3] Spitchinger N., Winterhalter, M. and Fabian, P. (1996) Ozone and grosswetterlagen. Analysis for the Munich Metropolitan area. *Environmental Science & Pollution Research* **3**, 145-152.

[4] Abdulwahab S., Bouhmara, W. Ettouney, H., Sowerby, B and Crittenden B. (1996) A statistical model for predicting ozone levels. *Environmental Science & Pollution Research* **3**, 195-204.

[5] Kuang-Jung Hsu (1992) Time series analysis of the interdependence among air pollutant. *Atmospheric Environment* **26B**, 491-503.

[6] Simpson R.W. and Layton A.P. (1983) Forecasting peak ozone levels. *Atmospheric Environment* **17**, 1649-1654.

[7] Robeson S.M. and Steyn D.G. (1990) Evaluation and comparison of statistical forecast models for daily maximum ozone concentrations. *Atmospheric Environment* **24B**, 303-312.

[8] Cassmassi, J.C., (1998). *Objective ozone forecasting in the South Coast Air Basin: Updating the objective prediction models for the late 1990's and southern California ozone study (SCOS97-NARSTO) application.* Proceedings of the 12th Conference on Numerical Weather Prediction, Phoenix, AZ. American Meteorology Society, Boston, MA. 54-58.

[9] Cassmassi, J.C., (1987). *Development of an objective ozone forecast model for the South Coast Air Basin.* Reprints 80th Annual Meeting, Air Poll. Control Assoc., paper no, 87-71.3, 21-26 June 1987, New York, N.Y.

[10] Hanna, S.R., Strimaitis, D.G. and Chang, J.C. (1991). *User's guide for software for evaluating hazardous gas dispersion models.* American Petroleum Institute. 1220 L. Street, Northwest. Washington. D.C. 20005

[11] European Commission, 1994. *The Evaluation of Models of Heavy Gas Dispersion. Model Evaluation Group Seminar.* Office for Official Publications of the European Communities. L-2985. Luxemburg.

[12] Ibarra, G. (1994). *Application of cluster analysis in optimization processes of surface meteorological and air quality networks.* AIR POLLUTION II. VOLUME 1. Computational Mcchanics Publications. pp. 515-522.

[13] Wilks, D. S. (1995). Statistical methods in the atmospheric sciences. Academic Press. 467 pp.

[14] Zannetti, P. (1990) *Air Pollution Modelling: Theories, computational methods and available software*. Computational Mechanics Publications. Van Nostrand Reinhold, New York, NY (USA).

[15] Takakatsu Inoue, Mamoru Hoshi and Masaaki Taguri (1986) Regression analysis of nitrogen oxide concentration. *Atmospheric Environment* **20**, 71-85.

[16] Finlayson-Pitts B.J. and Pitts J.N. (1986). *Atmospheric Chemistry: Fundamentals and experimental techniques*. Wiley, New York.

[17] Olesen, H.R. (1994). *Model Validation Kit for the Workshop. Compendium of Materials. Operational Short-Range Atmospheric Dispersion Models For Environmental Impact Assessments in Europe*. National Environmental Research Institute. Denmark. 1994