

REGRESSION ANALYSIS AND URBAN AIR QUALITY FORECASTING: AN APPLICATION FOR THE CITY OF ATHENS

TH. SLINI,
K. KARATZAS*
A. PAPAPOPOULOS

*Aristotle University of Thessaloniki,
Dept. of Mechanical Engineering
Box 483, 54124 Thessaloniki, Greece*

* to whom all correspondence should be addressed
tel: + 30 2310 996011
fax: +30 2310 996012
e-mail: kostas@aix.meng.auth.gr

Received: 13/06/02
Accepted: 09/06/03

ABSTRACT

Air pollution forecasts in major urban areas are becoming a problem concerning the day to day environmental management for city authorities. This paper describes the development of an application to forecast the peak ozone levels with the aid of meteorological and air quality variables, in the Greater Athens Area. For this purpose, a number of regression models were considered, while the selection of the final model was based on extensive analysis and on literature. The model adapted includes variables that are available on a daily basis, so as daily operational maximum ozone concentration level forecast can be achieved.

KEYWORDS: Ecological statistics, air quality forecasting, environmental modelling.

INTRODUCTION

In an effort to forecast air quality (AQ) daily, many different kinds of models have been developed (Robeson and Steyn, 1990; Yi and Prybutok, 1996; Millionis and Davies, 1994a, b; Rao and Zurbenko, 1994; Zannetti, 1990). The worldwide concern about this continuing environmental problem reinforces the need for more accurate and, at the same time, more easily applied models. Methods for AQ forecasting are required in particular for predicting the worst air quality situation. This paper focuses on forecasting of peak photochemical pollution episodes associated to ozone. High ozone concentrations may cause adverse health effects particularly for sensitive population groups, affect ecosystems harmfully

and corrode materials. As contemporary urban air pollution is strongly related to high ozone levels, it is essential to develop forecasting models that will allow for a prediction of air quality levels related to maximum ozone concentration, being useful in everyday environmental management for city authorities. These model applications also support air quality assessment for which monitoring of meteorological and AQ variables is a necessity. The complexity in AQ assessment and management in cities lies in a series of difficulties, extending well beyond the boundaries of the atmospheric physics, that include the role of turbulence and atmospheric stability, air flow and pollution dispersion in the surrounding of buildings and street canyons, inherent uncertainties in the estimation

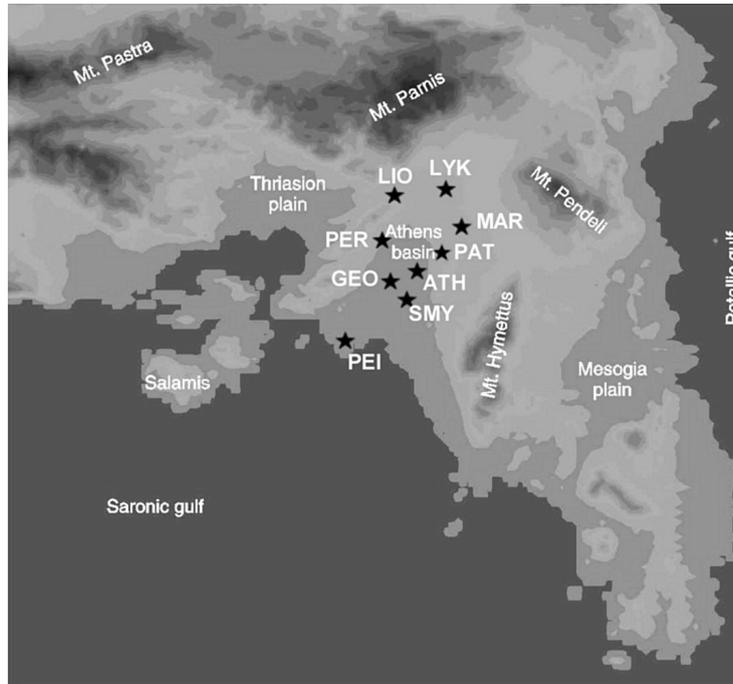


Figure 1. The Greater Athens Area and the location of the AQ monitoring network stations (asterisks). The meteorological station location, NOA, is identified with a cross (+). Contours are every 200 m.

of basic atmospheric physics parameters etc. Air pollution is a problem that can not be treated independently of the urban web. One has to consider the urban environment as a multi-dimensional, multivariable system, part of which is the AQ aspect, that also includes the layout of the city, the existence of green and un-built areas, the geometry, the architectural morphology and the thermal properties of the buildings, the vehicular traffic, the stationary thermal systems and of course the local microclimatic conditions.

Photochemical processes leading to ozone formation in the atmosphere are complex and non-linear (de Leeuw, 2000; MacDonald *et al.*, 2001; Blanchard, 2000). Still, the dispersion conditions are sensitive to meteorological changes in conditions (NRC, 1991; EPA, 1999). Indeed, it is this dependence on meteorological and AQ variables that will be exploited in the present paper.

The majority of approaches to meteorological adjustment of ozone are in some sense regression-based, with widely varying degrees of complexity. These approaches model the average behaviour of selected parameters and their structure can be divided in three broad categories: linear regression, regression trees and non-linear regression (Wilks, 1995). It should be noted that regression models are

generally site dependent. Chock *et al.* (1975) highlighted that there is no guarantee as to the reliability of the model once it is extrapolated beyond the range of the input data used to obtain it.

METHODOLOGY

In the present paper a 10-year long data set of maximum daily O_3 observations is being used (1990-99). The maximum daily values were obtained by extracting the maximum ozone levels of a dataset of hourly ozone concentrations through a 24 four hour period as it resulted from the operation of the air quality monitoring network in the city of Athens, Greece. This network consists of ten air quality monitoring stations, which have been operated by the Hellenic Ministry of Environment, Planning and Public Works, Directorate of Air and Noise Pollution from the late eighties, providing hourly values for a number of AQ parameters. All stations are carefully validated, and their observations should be considered as valid and accurate (Figure 1). More technical details on the methods and instruments implemented for the present data collection can be found on the report "Atmospheric Pollution in Athens-1996" (Directorate of Air and Noise Pollution, 1997).

The city of Athens is situated on the west coast of the Attica peninsula. Mountains acting as physical barriers with few gaps between them surround the whole basin, which is open towards the sea to the south. Sea breeze circulation influences AQ in the area, while the etesian wind, a synoptic scale seasonal north-north-easterly wind, is the main system that could reduce the possibility of prolonged photochemical pollution episodes during summer periods (Lalas *et al.*, 1987).

In the present paper, the data set collection consists of 10-year long daily observations of

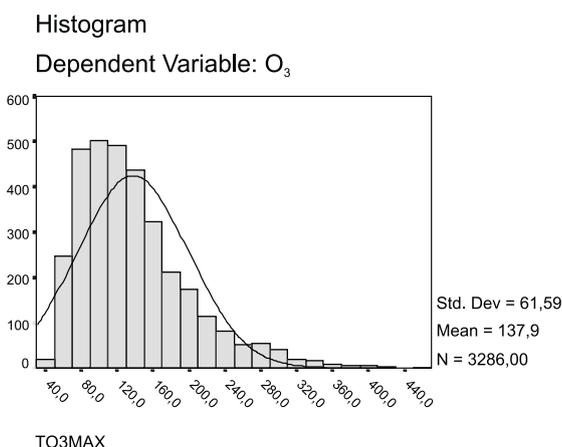
- maximum ozone concentration levels (consisting of pairs of values related to "present" and "previous" day conditions in $\mu\text{g m}^{-3}$),
- maximum ozone concentration levels until 13:00 p.m. ($\mu\text{g m}^{-3}$) coming from the AQ monitoring network,
- maximum and mean daily temperature ($^{\circ}\text{C}$),
- total solar radiation (W m^{-2}), and
- mean wind speed (m s^{-1}), coming from the meteorological observation station of the National Observatory of Athens for the same time period.

The choice of meteorological variables was based upon selecting those that were generally available and widely accepted as being associated to ozone formation (Gardner and Dorling, 2000; van Aalst and Leeuw, 1997; Thompson *et al.*, 2001), since this would facilitate the results of this work being applied in a forecasting mode: Temperature plays a crucial role, as with increasing temperature, the ozone concentration is expected to rise. Another favourable condition is the wind speed, as mixing

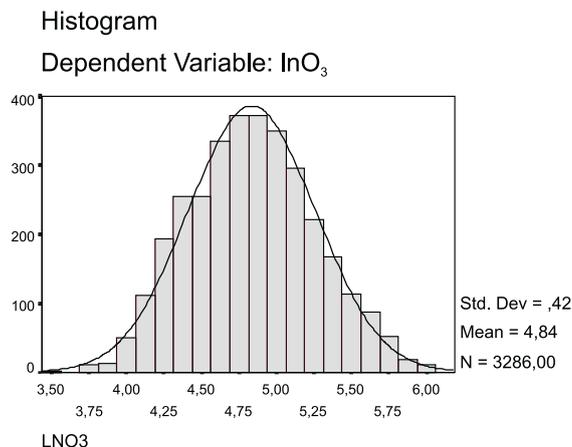
drops with lower wind speeds, while photochemical reactions, which represent one source of ozone formation, depend on the amount of solar radiation (Rappengluck *et al.*, 1993).

SPSS for Windows software (version 10.0) was used throughout this paper to fit a regression model as well as to manipulate many other "goodness of fit" tests (www.spss.com). List wise treatment was applied in cases of missing data, which means that a case would be eliminated if it had a missing value for any variables on the list. The proportion of missing data used in the development of the desirable equation is about 2%, that is about 65 cases should be excluded from the dataset.

Data analysis begins by examining each variable independent of the others. A first insight of the variables is obtained by considering the dot plots, where each dot represents a measurement of the variable plotted, and provides the location, range and distribution of each one separately. This information is reinforced by some basic summary statistics, such as mean – the arithmetic average, median, standard deviation from the mean, skewness – the departure from the mean in a continuous distribution – and kurtosis – concerning the distribution's tails in the graphical form. Examination of these results reveals that the distribution of maximum ozone concentration is not normal (Figure 2 (a)). Because of the fact that linear regression analysis is a method that requires the data to be normally distributed, the logarithmic transformation may be employed in ozone (EPA, 1999), in order to fulfil this basic assumption. In Figure 2 (b) is obvious that the transformed variable of O_3 ,



(a)



(b)

Figure 2. Normality test for the O_3 (a) and transformed $\ln\text{O}_3$ (b) data.

$\ln O_3$, fits a normal distribution. It is of course desirable to reduce the likelihood of data errors, which can lead to misleading results.

The distribution of original and transformed observations, indicated by bars, compared to a normal distribution.

Secondly, scatter plots of the response variable against each predictor variable can aid in determining the nature and strength of the relationship between the coupled variables and in identifying gaps in the data points as well as outlying data points. Scatter plots between couples of predictor variables are helpful for studying the bivariate relationships among each other, finding gaps and detecting outliers, extremely high or low cases that do not follow the usual pattern. In addition, a complement to the scatter plot is the Pearson Product Moment correlation matrix. This matrix contains the correlation statistic, which measures the strength of linear relationship between variables. The analysis of the specific data suggests that there is an almost linear relation between the response-ozone and predictor variables that becomes more evident taking the log-ozone and the predictor variables instead. Moreover, it seems that maximum and mean day temperatures are highly correlated, and so are maximum ozone concentrations of the previous day and maximum value until 13:00 of the day of interest. Maximum day temperature and previous day maximum ozone concentration levels are also correlated, therefore one of them may be included in the model. The total solar radiation does not seem to contribute efficiently to the model building, as Munn (1975) has already suggested. Moreover, because of the high correlation of temperature and solar radiation, the former is expected to act as a natural surrogate for the latter. Of all meteorological variables examined, air temperature has the strongest correlation with maximum ozone concentration, with the previous day ozone to follow. Finally, all variables except wind speed are positively correlated, meaning that increase of their level results in higher maximum ozone concentration, while in windy weather, the developing conditions increase the dilution and the dissipation of the pollutant.

However, rarely are assumptions not violated, one way or another, in regression analysis and other statistical procedures. Carrying out regression without considering possible violations of the

assumptions can lead to results that are difficult to interpret and apply (SPSS, 1993). Besides, for some problems the relationship between the response and the variables is too complicated to be explained by a simple linear model of this type. The attempt to find a satisfactory forecasting model included the development of a wide variety of models and the detailed examination of potential non-linear data transformations and interactions between predictor variables, in order to better capture ozone behaviour. When considering such kind of models the selection may become problematic too, due to their complex requirements. In addition, Gardner and Dorling (1998) pointed out that the relation between the meteorological variables analysed and the daily maximum ozone concentration may be represented satisfactorily by a linear model. Linear regression modelling combining ozone concentrations and meteorological values has repeatedly being suggested in McCollister and Wilson (1975), Robeson and Steyn (1990), Chaloulakou *et al.* (1999).

Given additional predictor variables, it is possible that this may no longer be the case. Specifically, for this analysis, the linear model was developed in the most simplistic manner without variable interactions, such as temperature-wind speed interaction and other combinations of meteorological variables, as the investigation showed that there was no obvious improvement concerning the particular dataset. The development set includes the first 9-year long observations (1990-1998), which correspond to 3286 values ($n=3286$), while the rest 365 (1999) were used as the validation set.

A natural measure of the effect of the meteorological variables in expressing the ozone concentration levels is the coefficient of determination, R^2 (Draper and Smith, 1981). The larger it is, the more satisfactorily the model fits the data.

In addition, through a study of residuals, many types of model inadequacies and violations of the underlying assumptions can be discovered. The model diagnostic checking can be done easily by graphical analysis of residuals, for example by plotting a normal probability plot of the residuals. If the underlying error distribution is normal, this plot will resemble a straight line. Moreover, the plot of residuals versus time or fitted values should have no structure and be randomly scattered throughout the plot.

RESULTS AND DISCUSSION

A first order regression model based on all predictors was fitted to the given dataset of the years 1990-98 to serve as a starting point. The coefficient of determination of the model equals to 0.593, indicating that all included variables accounts for the 59.30% of the original variation in the response variable, while the rest 40.70% represents the variation of the residuals. Various other first order models were fitted with different numbers of predictor variables included, and many residual plots were prepared.

The diagnostic check of many potential subsets of predictor variables suggests that maximum day temperature, maximum ozone concentrations of the previous day and the mean wind speed should be included, as they are capable to fit the data satisfactorily. The excluded variables are maximum ozone levels until 13:00 p.m., mean daily temperature and total solar radiation. According to the physical explanation, mean daily temperature is a less favourable condition to ozone formation as it is not representative of the photochemical potential of the atmosphere. Regarding the total solar radiation factor, even though it has significant physical meaning to ozone formulation, it is emitted mainly because of its large proportion of missing values in the dataset, leading to little improvement in the model. The resulting suggested model is the following:

$$[O_3] = 2.169 + 0.170 * \text{max temp} + 0.487 * \text{prev } [O_3] - 0.00618 * \text{ws} \quad (1)$$

where $[O_3]$ is the ozone concentration, max temp is the maximum daily temperature, $\text{prev}[O_3]$ is the maximum ozone concentration of the previous day, ws is the mean wind speed.

The values of the coefficients are consistent with the physical effect of the respective variables. The coefficient of the maximum ozone concentrations of the previous day is expected to have the greater positive value in accordance with the fact that it is the highest correlated with ozone episodes, followed by the maximum day daily temperature. On the contrary, the coefficient of the wind speed has a negative sign, stating its preventing role to ozone formulation.

Normality analysis shows that residuals are close

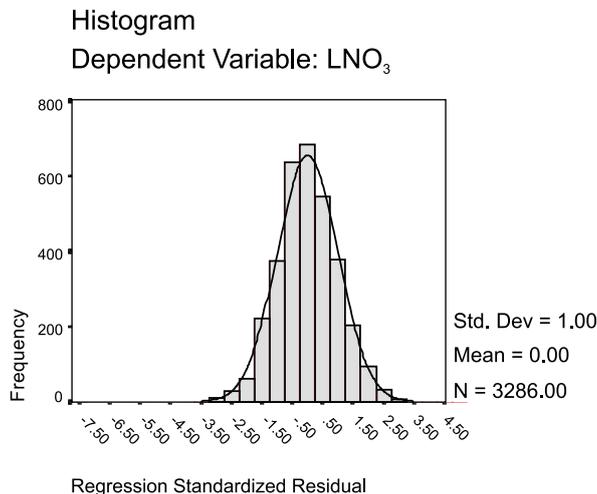


Figure 3. Normality test for the residual of O_3 regression. The distribution of Standardized Residuals, indicated by bars, compared to a normal distribution.

to a normal distribution (Figure 3), as well as the scatter plot versus time reveals no obvious pattern, meaning that the independence assumption for the errors is not violated. Moreover, the plot of the residual against the fitted values (Figure 4) does not show any unusual pattern, suggesting that the former are unrelated to any other variable including the predicted response and hence no further variance - stabilising transformation has to be applied.

Meanwhile, the coefficient of determination is slightly reduced compared to the one of the full model, and the variation explained comes up to 58.50%. This decrease is expected when the number of variables used is reduced.

Trying to check the adequacy of the former model, the predicted ozone concentrations were calculated for the following year (1999). Figure 5 displays the sequence of the true evaluation dataset and the predicted values; although there is a slight tendency for underestimating high observed concentrations, the agreement between observed and predicted values is generally good. In addition, the simple persistence model ($[O_3] = \text{prev}[O_3]$) was developed so as to obtain a comparison parameter. Statistically, the latter model violates some basic assumptions; for example the error term is not normally and independently distributed, as some kind of pattern is present at the scatterplot against the fitted values. This should

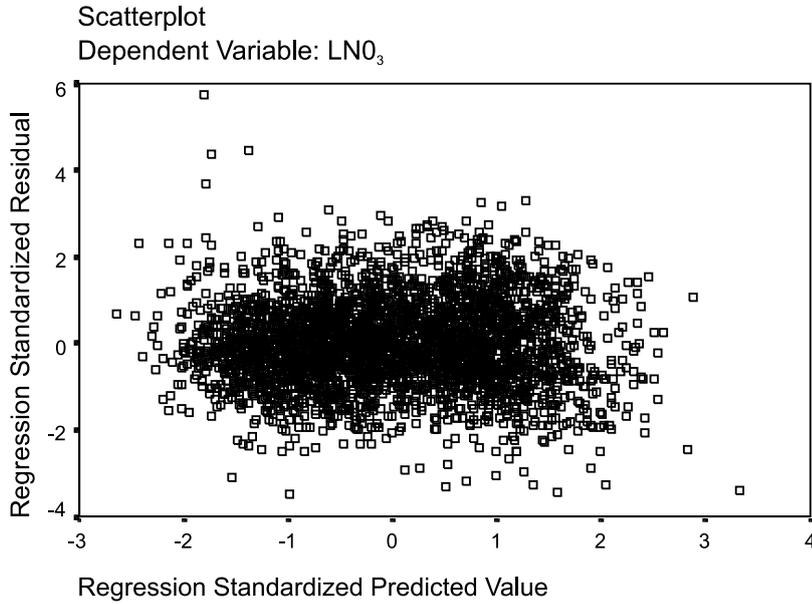


Figure 4. Scatterplot of the error term against the predicted value.

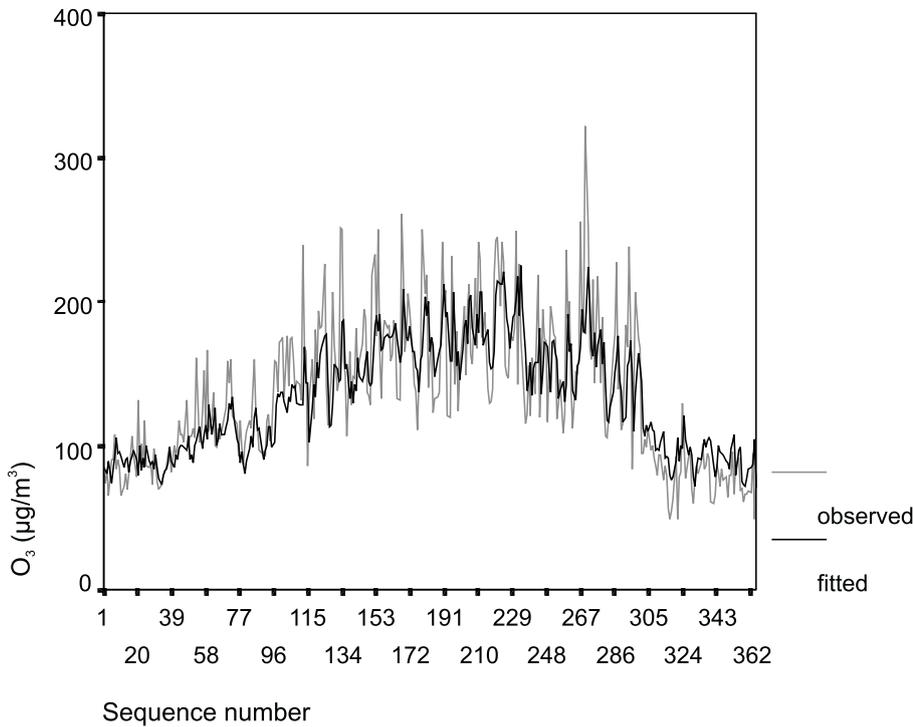


Figure 5. Time sequence of observed and fitted values using the model of eq. (1).

be expected, considering the high correlation between the predictor and the explanatory variable.

Though the ability to forecast is quite satisfactory, the overall performance of the account for the

empirical pollution forecast procedure performed by PERPA (Mantis *et al.*, 1990). For comparison reasons, Table 1a and 1b present a range of statistics of the persistence model and those of the model suggested above, as recommended by

Table 1a. Statistic estimates for the comparative assessment of the model of Eq. (1) and the persistence model.

	Model of eq. (1)	Persistence model
Number of observations	365.000	365.000
Observed mean	136.126	136.126
Predicted mean	142.735	149.651
Observed standard deviation	49.781	49.781
Predicted standard deviation	42.319	55.729
Normalised mean difference (NMD)	0.048	0.099
Mean absolute error (MAE)	26.928	26.950
Mean percentage error (MPE)	2.384	3.104
Root mean squared error (RMSE)	32.594	37.756
RMSE Systematic	18.369	21.582
RMSE Unsystematic	26.925	30.979
Index of agreement (IA)	0.863	0.867
Correlation coefficient	0.769	0.779

Wilmott (1982, 1985), which show a good overall model performance for both. More specifically, the observed and predicted standard deviations quantify the amount of the variance the model is capturing compared to the observed data. Two often mentioned measures of residual error include the mean absolute error (MAE) and the root mean squared error (RMSE) which summarise the difference between the observed and modelled concentrations. The latter can be decomposed in the systematic RMSE, which is due to the model performance and the predictors included, and unsystematic RMSE that is due to residuals, factors that cannot be controlled. A good model is considered to have an unsystematic RMSE much larger than the systematic RMSE. The value of RMSE of the suggested model of equation (1), 32.594, is smaller than the value of the persistence model, 37.756, indicating a better performance of the former. A more useful measure of model performance is provided by the "index of agreement" (IA), which is defined as (Wilmott *et al.*, 1985):

$$IA = 1 - \frac{\sum_{i=1}^n |P_i - O_i|^2}{\sum_{i=1}^n (|P_i - \bar{O}| + |O_i - \bar{O}|)^2} \tag{2}$$

where P_i and O_i are the predicted and observed concentrations respectively. The index of agree-

ment ranges between 0 and 1, with 0 stating worst agreement, and best agreement given by the limiting case of 1. In this case, the index of agreement is similar for both models, expressing similar forecasting skills. Table 1b also presents statistical indices of forecasting performance related to the ozone alarm threshold.

False alarm rate, the percent of times an alert day was forecasted that did not actually occur, is reported. A reliable model requires not only having a large percentage of successful forecast but also a percentage of false positives as small as possible (Jorquera *et al.*, 1998). The critical success index (CSI or Threat score) verifies how well the high pollution events were predicted and it is unaffected by the number of correctly forecasted. This is defined as (Wilmott *et al.*, 1985):

$$CSI = \frac{A}{A + B + C} \tag{3}$$

- with A expressing the observed and forecasted alarms days
- B expressing the observed but not forecasted ones and
- C expressing the forecasted but not observed ones.

Generally, it is interesting to note that as the alarm level decreases, the forecasting performance is improved. Furthermore, model of Equation (1)

Table 1b. Forecast skill for O₃ episodes using the model of Eq. (1) and the persistence model.

	Model of eq. (1)	Persistence model
If alarm: [O ₃] > 180		
number of observed alarms	66	66
number of predicted alarms	49	107
number of days with correct alarm	25	42
percentage (%) of days with correct alarm (POD)	37.89%	63.64%
number of days with false alarm	24	65
percentage (%) of days with false alarm	48.98%	60.75%
Critical Success Index (CSI)	0.278	0.321
A	25	42
B	41	24
C	24	65
If alarm: [O ₃] > 170		
number of observed alarms	94	94
number of predicted alarms	79	128
number of days with correct alarm	51	65
percentage (%) of days with correct alarm (POD)	54.25%	69.15%
number of days with false alarm	28	63
percentage (%) of days with false alarm	35.44%	49.22%
Critical Success Index (CSI)	0.418	0.414
A	51	65
B	43	29
C	28	63
If alarm: [O ₃] > 160		
number of observed alarms	113	113
number of predicted alarms	108	146
number of days with correct alarm	80	88
percentage (%) of days with correct alarm (POD)	67.80%	74.58%
number of days with false alarm	28	58
percentage (%) of days with false alarm	25.92%	39.73%
Critical Success Index (CSI)	0.5479	0.500
A	80	88
B	38	30
C	28	58

has a lower proportion of correct and false alarm days, no matter of the critical maximum ozone level adapted (180 $\mu\text{g m}^{-3}$, 170 $\mu\text{g m}^{-3}$, 160 $\mu\text{g m}^{-3}$), resulting in a slightly better Critical Success Index than the persistence model only in the last two cases, where the maximum ozone threshold is 170 $\mu\text{g m}^{-3}$ or 160 $\mu\text{g m}^{-3}$. However, the proportion of false alarm days of the persistence model is much higher, indicating of the incapability of the latter to be used in a reliable operational AQ fore-

casting system. In addition, it should be mentioned that the scatter of the observations influences the degree of accuracy of the models developed (American Meteorological Society, 1978). It has also been pointed out that in situations such as the comparison of data generated by collocated air quality measurements systems, there is no real separation into dependent and independent variables, and deviations between fitted and observed data values will occur in both x and y directions

due to random measurement errors. Hence, the reduced major axis (RMA) method (Davis, 1986) recommends the linear fit to be achieved by minimising the product of the x and y deviations between the data and the fitted values. The latter was also demonstrated recently by Ayers (2001), who suggested that a reduced major axis regression, rather than standard linear regression, is likely to be the method of choice for the regression analysis of air quality data sets.

CONCLUSIONS

A linear model has been used for the regression and prediction of daily air pollution levels constructing an operational peak ozone concentration forecasting module, developed from scratch, for the Athens basin, Greece. Maximum temperature, maximum ozone level of the previous day and wind speed proved to be the most essential factors that influence ozone concentrations. Regarding the persistence model, it seems to be acceptable, but the extremely large proportion of false alarms should be considered together with the operational requirements of the system as the principal limitations.

The general inspection of this research confirms the difficulties arising in statistical atmospheric modelling. The low prediction ability of the model is not surprising, since a stochastic, random component in the regression will always exist and remain unexplained. For this reason, the RMA method is considered to be developed in the

future, as the aim of the current paper was to demonstrate the effectiveness of "standard" regression models created from scratch to support an every-day operational AQ forecasting and management application.

Considering the manifoldness of the urban landscape, the particularities of the urban microclimate and the complexity of air quality determination, the quantity of actual measurements' data needed to produce a far reaching and reliable prediction leads very fast to a problem difficult to be managed. The statistical linear regression analysis model becomes, in that sense, attractive, as it enables a fast assessment of the location-specific situation.

Even so, given the inevitable limits of regression analysis, the use of area and time covering monitoring data remains a prerequisite. It can, therefore, be only emphasised, that the statistical methodology should be guided by an understanding of the underlying physical mechanisms. Communication between statisticians and atmospheric scientists is a substantial part of the model building process.

ACKNOWLEDGEMENTS

This paper is related to the Hellenic implementation of the APNEE project (<http://www.apnee.gr>), supported by the European Commission DG B.4 under the 5th Framework Programme, IST-1999-11517, Key Action: Systems and Services for the Citizen.

REFERENCES

- American Meteorological Society (1978), Accuracy of dispersion models - a position paper, *Bulletin of American Meteorological Society*, **59**, 1025-1026.
- Ayers GP. (2001), Comment on regression analysis of air quality data, *Atmospheric Environment*, **35**, 2423-2425.
- Blanchard C. L. (2000), Ozone process insights from field experiments, Part III: extent of reaction and ozone formation, *Atmospheric Environment*, **34**, 2035-2043.
- Chaloulakou A., Assimacopoulos D. and Lekkas T. (1999), Forecasting daily maximum ozone concentrations in the Athens basin, *Environmental Monitoring and Assessment*, **56**, 97-112.
- Chock P.D., Terrell T.R. and Levitt S.B. (1975), Time series analysis of riverside, California, air quality data, *Atmospheric Environment*, **9**, 978-989.
- Davis J.C. (1986), *Statistical and Data Analysis in Geology*, 2nd Edition, Wiley, New York.
- de Leeuw A.A.M. Frank (2000), Trends in ground level ozone concentrations in the European Union, *Environmental Science & Policy*, **3**, 189-199.
- Directorate of Air and Noise Pollution (1997), *Air pollution in Athens-1996*, Ministry of Environment, Planning and Public Works.
- Draper N.R. and Smith H. (1981), *Applied Regression Analysis*, 2nd Edition, Wiley, New York.
- EPA, Guidelines for developing an ozone forecasting program, EPA-454/R-99-099, July 1999, U.S. Environmental Protection Agency, Office of Air quality Planning and Standards, Research Triangle Park, North Carolina.

- Gardner M.W. and Dorling S.R. (1998), Artificial neural networks (the multilayer perceptron) - a review of applications in the atmospheric sciences, *Atmospheric Environment*, **32**, 2627-2636.
- Gardner M.W. and Dorling S.R. (2000), Statistical surface ozone models: an improved methodology to account for non-linear behaviour, *Atmospheric Environment*, **34**, 21-34.
- Jorquera H., Pérez R., Cipriano A., Espejo A., Letelier M. V. and Acuña G. (1998), Forecasting ozone daily maximum levels at Santiago, Chile, *Atmospheric Environment*, **32**, 3415-3424.
- Lalas D. P., Tombrou-Tsella M., Petrakis M., Asimakopoulos D.N. and Helmis C. (1987), An experimental study of the horizontal and vertical distribution of ozone over Athens, *Atmospheric Environment*, **21**, 2681-2693.
- Mantis H. T., Reparis C. C., Zerefos C. S., Ziomas J.C. and Bais A. (1990), Research centre for Atmospheric Research and Climatology, Academy of Athens, Scientific papers, 77-78.
- McCollister G.M. and Wilson K.R. (1975), Linear stochastic models for forecasting daily maxima and hourly concentrations of air pollutants, *Atmospheric Environment*, **9**, 417-423.
- MacDonald C., Roberts P., Main H., Dye T., Coe D. and Yarbrough J. (2001), The 1996 Paso del Norte Ozone Study: analysis of meteorological and air quality data that influence local ozone concentrations, *The Science of The Total Environment*, **276**, 93-109.
- Millionis A.E. and Davies T.D. (1994a), Regression and stochastic models for air pollution - I. Review, comments and suggestions, *Atmospheric Environment*, **28**, 2801-2810.
- Millionis A.E. and Davies T.D. (1994b), Regression and stochastic models for air pollution - II. Application of stochastic models to examine the links between ground-level smoke concentrations and temperature inversions, *Atmospheric Environment*, **28**, 2811-2822.
- Munn R.E. (1975), The oxidant climatology of Canada, In: *Photochemical Air Pollution: Formation, Transport, and Effects*, Publ. No. NRCC 14096 of the Environment Secretariat, National Research Council of Canada.
- NRC (National Research Council) (1991), Rethinking the Ozone Problem in Urban and Urban and Regional Air Pollution, National Academy Press, Washington DC.
- PERPA (1989), Air pollution in Athens area, Vols I-IV, Pollution sources, Ministry of Environment, Physical Planning and Public Works, Athens (in Greek).
- Rao S. T. and Zurbenko I. G. (1994), Detecting and tracking changes in ozone air quality, *Journal of Air and Waste Management Association*, **44**, 1089-1092.
- Rappengluck B., Kourtidis K. and Fabian P. (1993), Measurements of ozone and peroxyacetyl nitrate (PAN) in Munich, *Atmospheric Environment*, **27B**, 293-305.
- 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.
- SPSS (1993), SPSS for Windows, Base System User's Guide Release 6.0. Marija J. Norusis, SPSS Inc.
- Thompson M. L., Reynolds J., Cox L. H., Guttorp P. and Sampson P. D. (2001), A review of statistical methods for the meteorological adjustment of tropospheric ozone, *Atmospheric Environment*, **35**, 617-630.
- Van Aalst R.M. and de Leeuw F.A.A.M. (1997), National Ozone Forecasting Systems and International Exchange in Northwest Europe, Report of the Technical Working Group on Data Exchange and Forecasting for Ozone episodes in Northwest Europe (TWG-DFO).
- Yi J. and Prybutok V. R. (1996), A neural network model forecasting for prediction of daily maximum ozone concentration in an industrialized urban area, *Environmental Pollution*, **92**, 349-357.
- Wilks S.D. (1995), *Statistical Methods in the Atmospheric Science, An Introduction*. Academic Press, New York.
- Wilmott C.J. (1982), Some comments on the evaluation of model performance, *Bulletin of American Meteorological Society*, **63**, 1309-1313.
- Wilmott C.J., Ackleson S. G., Davis R.E., Feddema J.J., Klink K.M., Legates D. R., O'Donnell J. and Rowe C.M. (1985), Statistics for the evaluation and comparison of models, *Journal of Geophysics Research*, **90**, 8995-9005.
- Zannetti P. (1990), Air pollution modelling, Van Nostrand Reinhold, New York.