STATISTICAL AIR QUALITY PREDICTION FOR LISBON

Jorge Neto[(1)(2)(*)], Pedro Torres[(1)], Francisco Ferreira[(1)], and Filomena Boavida[(3)]

[(1)] *Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa*
[(2)] *Instituto de Meteorologia / Departamento de Observação e Redes*
[(3)] *Instituto do Ambiente*

[(*)] *Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa*
Departamento de Ciências e Engenharia do Ambiente
Quinta da Torre,
2829-516 Caparica, Portugal
Email: jorge.neto@meteo.pt, Phone : +351-21-2948374

ABSTRACT

Ozone and particulate matter levels in Southern European countries are particularly high, exceeding the established limit values and the information and alert thresholds (in the case of ozone), it is relevant to develop a good prediction methodology for the concentrations of these pollutants. Statistical models based on multiple regression analysis (MR) and classification and regression trees analysis (CART) were developed successfully applied in forecasting the average daily concentrations for particulate matter and average maximum hourly ozone levels, for next day, for the group of existing air quality monitoring stations in the Lisbon area in Portugal.

1. INTRODUCTION

An important commitment of Portugal in the area of air quality is the fulfillment of the Portuguese and European legislation. The forecasting of air pollutant concentrations is very important for areas with air quality problems. Predictions can be developed through the integration of physico-chemical relationships from both meteorology and pollutants behaviour, or by using stochastic methods based on the analysis of data series. A combination of standard statistical methods was the selected process described in this paper.

Using past information on studies to understand the variability of ozone (O_3) and particulate matter (PM_{10}) (Clapp and Jenkinb, 2001; Ferreira *et al*., 2000; Ferreira *et al*., 2004; Neto *et al*., 2004; Wang *et al*., 2003) was the first step to start to build a highly detailed data base using all the existing data. Some variables initially selected were substituted afterwards by better descriptors. One of the examples is the wind direction which was substituted by the pressure difference between stations) in order to have the synoptic wind, and not the local one. Another example was the case of maximum mixing layer height by the Holzworth method (1964) which substituted the inversion layer height and thickness since these variables are currently not available as predicted values.

2. METHODOLOGY

Models based on multiple regression analysis (MR) and classification and regression trees analysis (CART) with data from 2001 to 2002 were developed following some techniques already used for the case of Los Angeles by Casmassi (1987) presented also by EPA (2003). Air quality data was obtained from the Commission for Regional Development of Lisbon and the Tagus River Valley Region (CCDR-LVT) network, and from one station from the Institute for the Environment. Air quality monitoring stations for the Lisbon area were included in this analysis. Surface meteorological parameters such as temperature, relative humidity and atmospheric pressure were collected from six stations of the Meteorology Institute. Meteorological radio soundings data were obtained from the Lisboa Gago-Coutinho station. Synoptic situations at surface and 500 mb level from the forecast global model of European Centre for Medium-Range Weather Forecasting (ECMWF and IM, 2003) were classified in seven and five classes, respectively, for model input. Surface synoptic classes: front systems, deepening low pressure (instability), low pressure influence, surface calm, N/NW circulation, high pressure and thermal trough from the north of Africa and NE/E circulation. 500 mb synoptic classes: cut off low, low pressure trough, approaching trough or ridge breakdown, building high pressure ridge or zonal flow and high pressure ridge.

The variables used are:

- Daily maximum temperature and average relative humidity for Lisbon, Santarém, Évora and Beja (both absolute values and the difference between yesterday and today's values);
- Pressure difference between Lisbon and others cities (Porto, Portalegre, Évora and Faro) at 12UTC (the difference between yesterday and today's values);
- 12UTC geopotential height at 1000, 850, 700 e 500 mb. 12UTC temperature and relative humidity at 925, 850 and 700 mb;
- Daily average values of air quality data for all the stations in the area of study: particles matter, ozone maximum, ozone and carbon monoxide eight hour average, nitrogen dioxide, sulphur dioxide (the difference between today and yesterday values; all the air quality values are calculated as in the existing air quality index for Lisbon);
- Others variables like duration of the solar day, day of week and type of day (if week or week-end).

Models were first built using a CART analysis. Figure 1 represents the CART analysis for both pollutants predicted. In both cases the variable most determinant was the value of the pollutant obtained in the day before (PM10_1 and O3_1). For the highest values of PM_{10} a high relationship with the relative humidity in the 850mb level (HR850) was found. In the O_3 model, for low values, a good relationship exists with the duration of the solar day (DD); for higher values, the same happens with the maximum temperature observed in Santarém (TX734). The last result may show that the air masses that come from East in the summer are related with high values O_3.

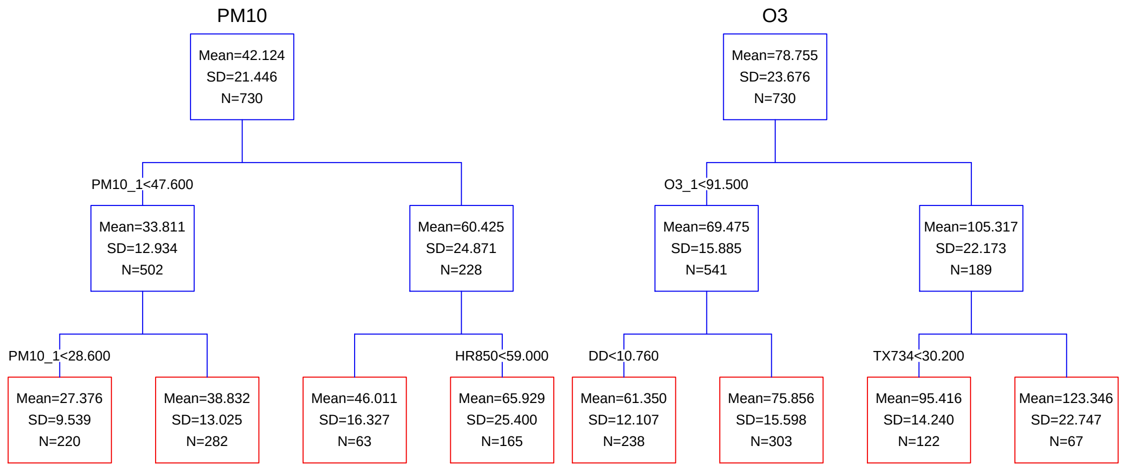

Figure 1 – CART models for O_3 e PM_{10} for the period 2001-2002

The MR analysis was performed using each one of the final four groups obtained in the CART analysis for both pollutants. The MR used backward stepwise analysis with a significance level of 0.15.

3. RESULTS AND DISCUSSION

One can observe from Tables 1 and 2 the results from the development of the different models for both pollutants. In the case of O_3 the correlation values are highest than for PM_{10} (0.99 for O_3 and 0.97-0.99 for PM_{10}).

Table 1 – Results from the several multiple regression models for PM_{10}

PM_{10}	PM10_1<47.6		PM10_1>=47.6	
	PM10_1<28.6	PM10_1>=28.6	HR850>=59	HR850<59
r	0.976	0.973	0.997	0.981
r^2	0.952	0.946	0.994	0.962
Std. error	6.769	10.054	6.055	15.338
N	218	280	61	164

Table 2 – Results from the several multiple regression models for O_3

O_3	O3_1<91.5		O3_1>=91.5	
	DD<10.76	DD>=10.76	T734<30.2	T734>=30.2
r	0.994	0.992	0.997	0.999
r^2	0.988	0.984	0.995	0.997
Std. error	7.472	10.469	8.383	9.926
N	238	303	121	67

Then models were tested with other periods (winter and summer data from year 2000). The results show a high correlation, statistically significant at a 95% confidence level. Variance explained between predicted and measured concentrations were up to 78% for ozone and 72% for particulate matter. The results are showed in **Erro! A origem da referência não foi encontrada.** and **Erro! A origem da referência não foi encontrada.**.

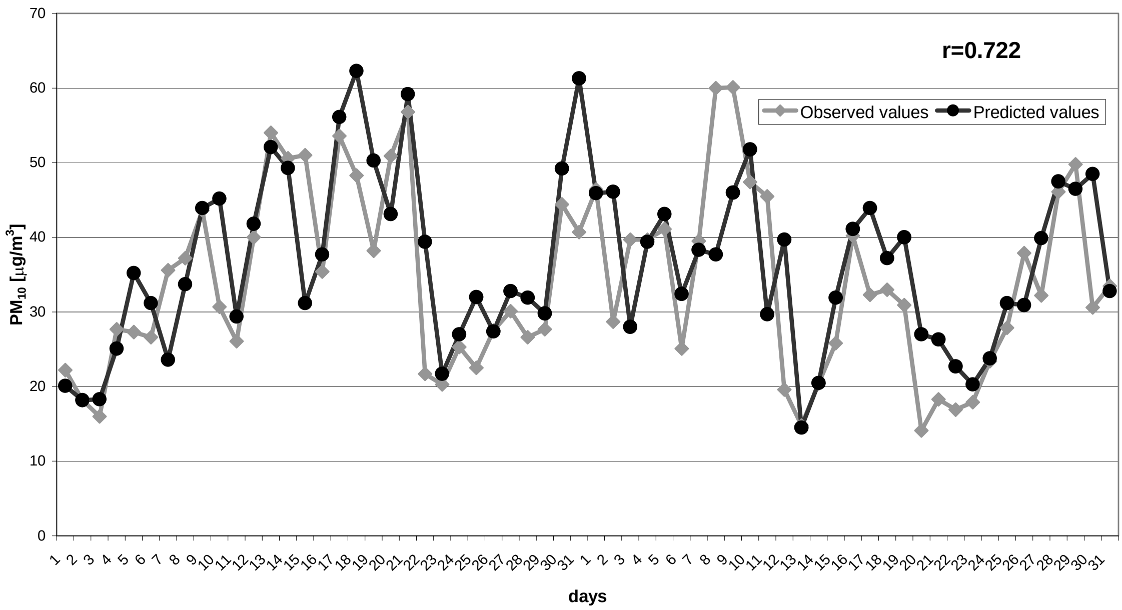

Figure 2 - Observed and predicted PM_{10} concentrations values using CART+MR models (July 1st to August 31st 2000)

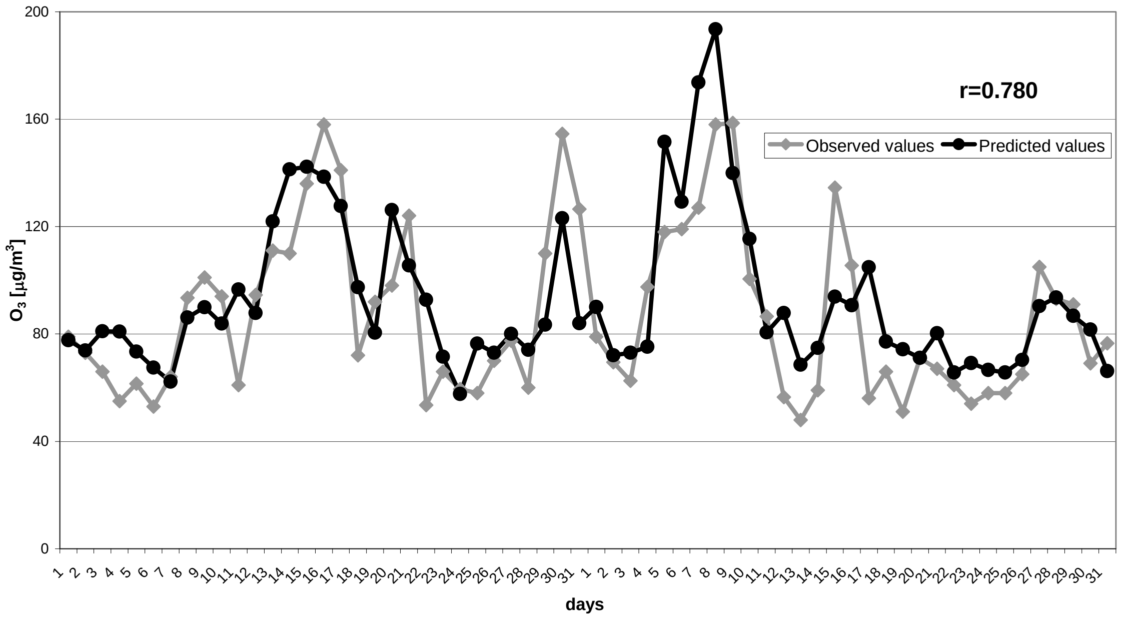

Figure 3 - Observed and predicted O_3 concentrations values using CART+MR models (July 1st to August 31st 2000)

The results are always better for O_3 because this pollutant has a very good relationship with temperature, mostly in summer, when both variables are very high. For PM_{10} no meteorological variable have the same explanatory impact.

4. CONCLUSIONS

The work presented is a statistical attempt based on a detailed analysis of both historical and expert knowledge involving meteorology and air quality aspects concerning ozone and particulate matter. The final goal was to develop a daily air quality forecast using statistical methods for the Lisbon region. A two-year period (2001-2002) was selected as the fitness period for the models, while other periods were select for the validation of the model. The use of statistical models based on multiple regression analysis (MR) and classification and regression trees analysis (CART) was very successful in forecasting the average daily concentrations for both particulate matter and average maximum hourly ozone levels for next day in the Lisbon area in Portugal. The developed models are currently being tested with current meteorology and air quality variables being measured with good results that will be further evaluated. The models developed also enabled a better understanding of the role of the different variables involved and their relationships.

5. ACKNOWLEDGEMENTS

The work developed was supported by Instituto do Ambiente and by Instituto de Meteorologia. The authors also wish to acknowledge the Comissão de Coordenação e Desenvolvimento Regional de Lisboa e Vale do Tejo for the data provided.

6. REFERENCES

Casmassi, J., 1987 Development of an objective ozone forecast model for the south coast air basin. 80th Annual Meeting of APCA (The Association Dedicated to Air Pollution Control and Hazardous Waste Management). New York – June 21-26, 1987.

Clapp, L.J. and Jenkinb, M.E., 2001 Analysis of the relationship between ambient levels of O_3, NO_2 and NO as a function of NO_x in the UK, Atmospheric Environment, 35, 6391-6405.

ECMWF 2003 Analyses Charts, European Centre for Medium-Range Weather Forecasts, 2000 - 2002.

EPA 2003 Guidelines for Developing an Air Quality (Ozone and PM2.5) Forecasting Program, U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, North Carolina, EPA-456/R-03-002.

Ferreira, F., Tente, H., Torres, P., Cardoso, S., Palma-Oliveira, J.M., 2000 Air Quality Monitoring and Management in Lisbon, Environmental Monitoring and Assessment 2000, 65, 443-450.

Ferreira, F., Torres, P., Neto, J., Tente, H., 2004 Ozone Levels in Portugal: the Lisbon Region Assessment. Em Proceedings of Air & Waste Management's 97th Annual Conference & Exhibition. June 22-25, 2004, Indianapolis, Indiana. CD-ROM, pp. 18.

Holzworth, G.C., 1967 Mixing depths, wind speeds and air pollution potential for selected locations in the United States, J. Appl. Meteor., 1967, 5, 1039-1044.

Instituto de Meteorologia 2003 Boletim diário, Instituto de Meteorologia. 2000 a 2002

Neto, J., Torres, P. & Ferreira, F., 2004 Previsão da qualidade do ar para Lisboa – a abordagem estatística. In 8ª Conferência Nacional de Ambiente. 27-29 Outubro de 2004, Lisboa. CD-ROM, pp. 11.

Wang, W., Weizhen, L., Wang, X., Leung, A., 2003 Prediction of maximum daily ozone level using combined neural network and statistical characteristics. Environment International 29, 555–562.