LISBON AIR QUALITY FORECAST USING STATISTICAL METHODS

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

[1] Instituto de Meteorologia / Departamento de Observação e Redes
[2] Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa
[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). Therefore, it is relevant to develop a good prediction methodology for the concentrations of these pollutants. Statistical models based on multiple regression analysis and classification and regression trees analysis were developed successfully. The models were 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 Metropolitan Area of North Lisbon in Portugal.

Key Words: Statistical forecast, particles, ozone

1. INTRODUCTION

An important commitment of Portugal in the area of air quality is the fulfillment of the Portuguese and European legislation. Since 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. 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. 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 both particulate matter (PM_{10}) and average maximum hourly ozone (O_3) levels for next day, for the existing air quality monitoring stations in the Metropolitan Area of North Lisbon in Portugal.

2. METHODOLOGY

Using past information on studies to understand the variability of O_3 and 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. The data base is based on a detailed analysis of both historical and expert knowledge involving meteorology and air quality aspects with data from the years 2000 to 2002. 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 of the Institute for the Environment. Surface meteorological parameters such as maximum temperature, relative humidity, and stations pressure differences were collected from seven stations of the Meteorology Institute. Geopotential heights, temperature, relative humidity, maximum mixing layer height by the Holzworth method (1964), and the inversion layer height, and thickness were also collected from the daily radio soundings of Lisbon 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, 2003; IM, 2003) were classified in seven and five classes (Tables 1 and 2). A future objective is to use ten automatic weather types using pressure points centered in Portugal from the ECMWF forecast (Mendes et al., 2002; Trigo and DaCamara, 2000). The weather types are the eight typical circulations (NE, E, SE, S, SW, W, NW and N) and the pure high and low pressure systems.

Table 1 – Classification of surface synoptic situations

Front systems	1 Front systems
Low pressure systems	2 Deepening low pressure (instability) 3 Low pressure influence
High pressure systems	4 Surface calm 5 N/NW circulation 6 Building high pressure ridge or zonal flow 7 NE/E circulation

Table 2 – Classification of 500 hPa synoptic situations

1 Cut off low
2 Low pressure trough
3 Approaching trough or ridge breakdown
4 Building high pressure ridge or zonal flow
5 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: PM_{10}, O_3 maximum, O_3 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-day or week-end).

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.

Using the techniques already used for the case of Los Angeles by Casmassi (1987) presented also by EPA (2003), the statistical models were first built using CART analysis, and then MR analysis. The next step was to perform the MR analysis 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.

For daily prediction, each pollutant value is transformed in terms of an index from very good to very bad (five levels). The worst index calculated from the two pollutants is the one chosen to represent our predicted index value for the next day.

3. RESULTS AND DISCUSSION

Figure 1 represents the CART analysis for both pollutants predicted. In both cases persistence is very important. The most determinant variable was the value of the pollutant obtained in the day before (PM10_1 and O3_1). In the O_3 model for higher values a good relationship exists with the maximum temperature observed in Santarém (TX734). This result may indicate that the air masses that come from East in the summer are related with high values O_3.

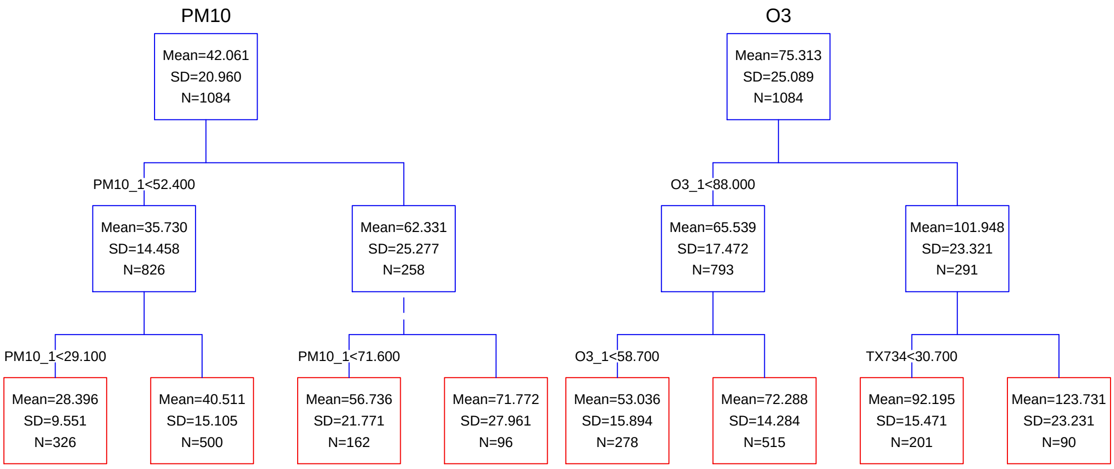

Figure 1 – CART models for O_3 e PM_{10} for the period 2000-2002

The results obtained RM models constructions for both pollutants perform correlations between 0.964 and 0.996. Tables 3 and 4 represent the results from the development of the different models for both pollutants.

Table 3 – Results from the several MR models for PM_{10}

	PM10_1<52.4		PM10_1>=52.4	
	PM10_1<29.1	PM10_1>=29.1	PM10_1<71.6	PM10_1>=71.6
r	0.976	0.964	0.982	0.975
r^2	0.953	0.930	0.965	0.950
Std.error	6.731	11.715	12.359	21.845
N	322	498	159	96

Table 4 – Results from the several MR models for O_3

	O3_1<88		O3_1>=88	
	O3_1<58.7	O3_1>=58.7	T734<30.7	T734>=30.7
r	0.985	0.992	0.994	0.996
r^2	0.970	0.984	0.989	0.993
Std.error	10.223	9.725	11.018	13.645
N	278	512	201	90

Then models were tested with other periods (winter and summer data from year 2003). The results show a high correlation, statistically significant at a 95% confidence level. Variance explained between predicted and measured concentrations were up to 88% for O_3, and 74% for PM_{10}, in best cases. The selected models allow a better understanding of each synoptic situation, and its relation with air quality leading to a forecast with considerable certainty for the majority of the identified scenarios. Figures 2 and 3 show an example of the test for the 2003 summer (June to August). 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 has the same explanatory impact. In the figure 2 two significant differences exists, due the existence of intense fires in Portugal in this days.

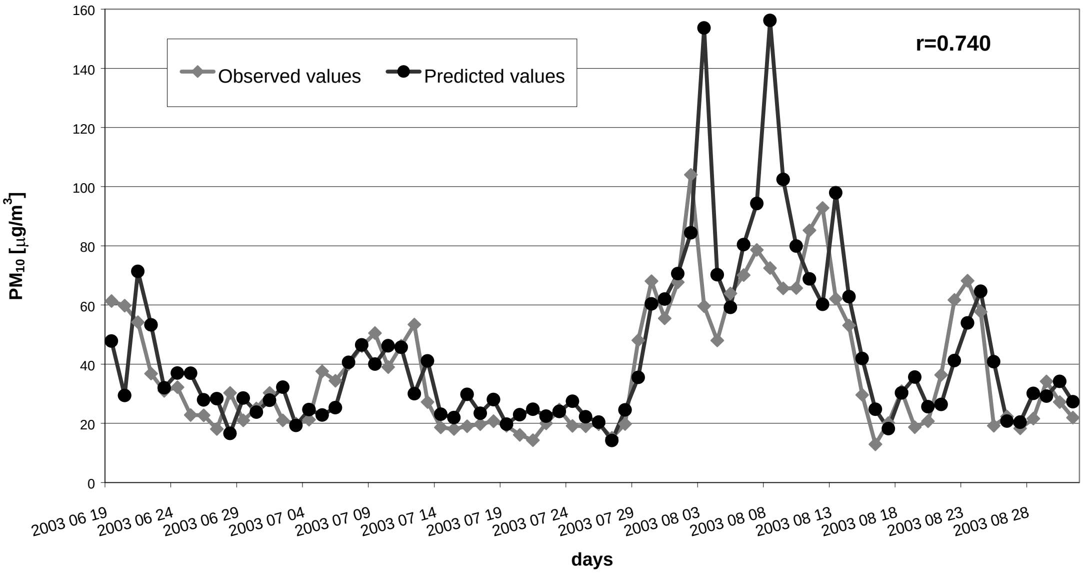

Figure 2 - Observed and predicted PM_{10} concentrations values using CART+MR models (July 1st to August 31st 2000)

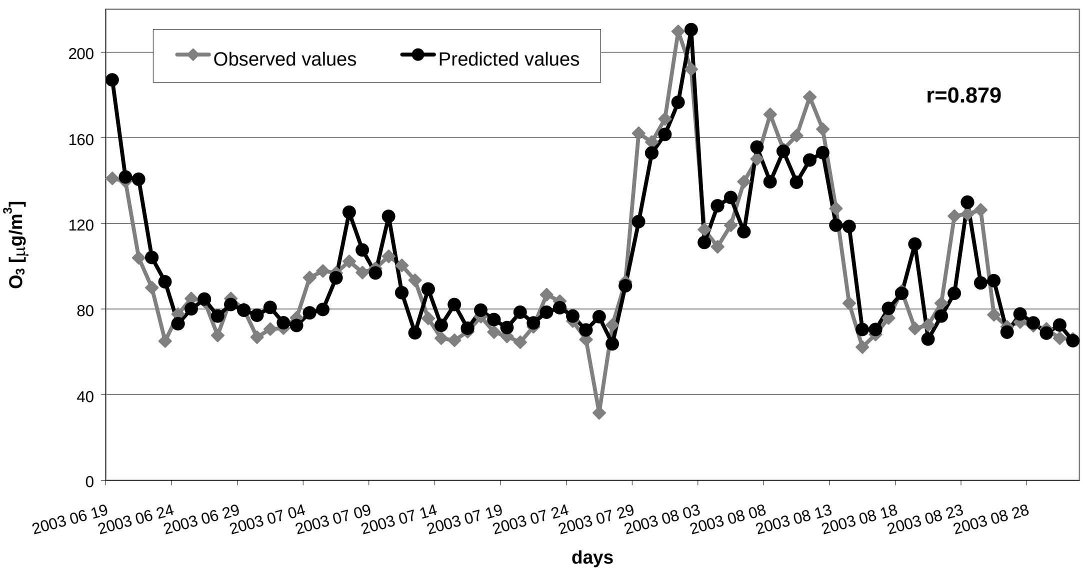

Figure 3 - Observed and predicted O_3 concentrations values using CART+MR models (July 1st to August 31st 2000)

The models are already in daily use since January 2005. Since then in 67% of days the final predicted index corresponds to the observed one. In all the other situations the predicted index is only one level worst above the observed one.

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 3-year period (2000-2002) was selected as the fitness period for the models, while another period was select for the validation of the model. The use of statistical models based on MR and CART analysis 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 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.

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-03-002.2

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

Mendes, M.T., Trigo, R.M. & DaCamara, C.C., 2002 Padrões de circulação atmosférica para Portugal Continental (1881-1995) in: Garcia, F.G. & Valero, J.L.B. (Ed.), 3ª Assembbleia Luso Espanhola de Geodesia e Geofísica – Valencia 2002. UPV, Valência. pp. 961-964.

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.

Trigo R.M. and DaCamara C.C., 2000 Circulation Weather Types and their influence on the Precipitation Regime in Portugal. Int. J. Climatol, 20, 1559-1581.

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.