AIR QUALITY FORECASTING BY STATISTICAL METHODS: MODEL EVALUATION FOR LISBON AND OPORTO

Luísa Mendes[(1)], Jorge Neto[(2)], Filipa Marques[(3)], Ana Carvalho[(1)], Joana Monjardino[(1)], Sandra Mesquita[(4)], Francisco Ferreira[(1)].

[(1)] CENSE, Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa. DCEA-FCT/UNL – Campus da Caparica, 2829-516, Monte da Caparica, Almada, Portugal lc.mendes@fct.unl.pt, ac.carvalho@fct.unl.pt, jvm@fct.unl.pt, ff@fct.unl.pt

[(2)] Instituto de Meteorologia, Rua C do Aeroporto, 1749-077 Lisboa, Portugal, jorge.neto@meteo.pt

[(3)] Agência Portuguesa do Ambiente, Rua da Murgueira, 9/9ª, 2610-124 Amadora, Portugal, filipa.marques@apambiente.pt

[(4)] Comissão de Coordenação e Desenvolvimento Regional de Lisboa e Vale do Tejo, Rua Braamcamp, nº 7, 1250-048 Lisboa, Portugal, sandra.mesquita@ccdr-lvt.pt

Introduction

The work presented in this paper is the PREVQUALAR project, which aims to forecast the air quality index (AQI) on Portuguese zones and agglomerations for air quality assessment and management purposes. The Directive 2008/50/EC on ambient air quality and cleaner air for Europe, in the Annex XVI related to public information, states that "*Member States shall ensure that timely information about actual or predicted exceedances of alert thresholds, and any information threshold is provided to the public. Details supplied shall include at least the following information: (a) information on observed exceedance(s); (b) forecast for the following afternoon/day(s); (c) information on the type of population concerned, possible health effects and recommended behaviour; (d) information on preventive action to reduce pollution and/or exposure to it: indication of main source sectors; recommendations for action to reduce emissions; (e) in the case of predicted exceedances, Member State shall take steps to ensure that such details are supplied to the extent practicable*". This Directive was transposed into Portuguese law 102/2010 of September 22.

Along with high ozone and particulate matter levels registered in Portugal, sometimes exceeding the legal limit values established by referred European legislation, the motivation for the PREVQUALAR project was more than justified.

The AQI calculated is based on the statistical forecast of air pollutants concentrations, subsequently converted in indexes for each PM_{10} and O_3 pollutants. The global AQI is given by the contribution of the worst pollutant index. The project involves the collaboration between the Portuguese Environment Agency (APA), the Portuguese Meteorological Institute (IM) and the Faculty of Sciences and Technology of New University of Lisbon (FCT/UNL). The present work regards the evaluation of the statistical models, developed by FCT/UNL, operating since 2006 for the Lisbon and Oporto areas (represented in Figure 1).

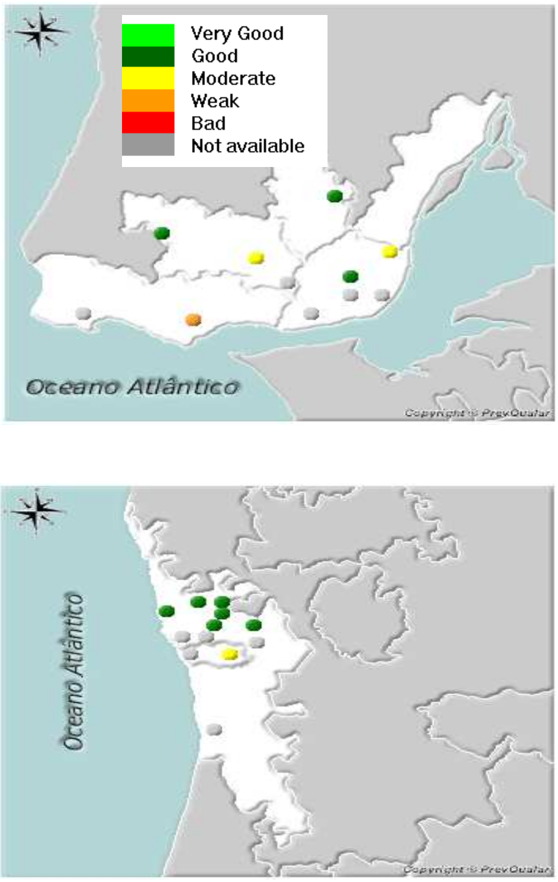

Figure 1 – Lisbon (top) and Oporto (bottom) air quality observation network.

Further information about the PREVQUALAR project can be found at http://www.prevqualar.org.

Methodology and Data

To develop local forecast regression models for the last years, a statistical approach, based on Classification and Regression Trees (CART) and Multiple Regression (MR) analysis, first used by Casmassi (1987), has been systematically used and improved since 2006. These models use daily measured concentrations of the atmospheric pollutants, local meteorological conditions, and the ECMWF forecasts.

Concerning air quality inputs, they are, if available, daily averages of particulate matter (PM_{10}), nitrogen dioxide (NO_2) and sulphur dioxide (SO_2) for the previous 24 hours of the forecast time deliver; daily ozone (O_3) maximum of the hourly data and octo-hourly moving averages; and maximum octo-hourly moving averages of carbon monoxide (CO). Concerning meteorological data both measured and deterministic forecast are considered. Measured meteorological information in different altitudes, at noon, gives the atmospheric stability to the statistical models, such as maximum value of the boundary layer mixing height, base and thickness of the inversion layer, temperature and relative humidity at pressure levels of 925, 850 and 700 hPa, and geopotential height at 1000, 850, 700 and 500 hPa. Additional meteorological information regarding surface values of extremes, daily averages, of temperature, relative humidity, pressure, and pressure gradients between locations will be included, or not, as predictors depending on the statistical significance at each location. Finally, a flag for week and weekend day is introduced.

The time frame for the model development was 2001-2002, years for which the data base was more robust and had less missing values.

The models allow forecasting the average daily concentrations for PM_{10} and average maximum hourly O_3 levels. Based on these concentrations, an AQI is established and the forecast is daily delivered around 16H00 UTC. The described methodology was used to develop regression models for 11 air quality stations in the metropolitan area of Lisbon and 13 in the metropolitan area of Oporto (Table 1).

To better understand the model development methodology several steps where taken:

- Validation of air quality data
- Validation of meteorological data
- Computing daily values of observed hourly PM_{10} and O_3 concentrations
- Establishment of derived meteorological parameters
- Classification and Regression Trees (CART) Analysis
- Multiple Regression (MR) model development
- Model validation

Table 1 - PM_{10} and O_3 developed models by agglomeration and air quality station.

Agglomeration	Air Quality Station	PM_{10}	O_3
Lisbon	Beato	•	•
	Olivais	•	•
	Entrecampos	•	•
	Avenida da Liberdade	•	
	Alfragide/Amadora		•
	Reboleira	•	•
	Loures	•	•
	Restelo		•
	Mem-Martins	•	•
	Cascais-Mercado	•	
	Oeiras-Quinta do Marquês	•	•
Oporto	Espinho	•	
	Antas	•	•
	Custoias	•	•
	Leça do Balio	•	•
	Vermoim	•	•
	Baguim		•
	Vila Nova da Telha	•	•
	Matosinhos	•	
	Perafita	•	•
	Senhora da Hora	•	
	Boavista	•	
	Ermesinde	•	•

Statistical measurements of agreement, between model results and measured air pollutant concentrations, used in this evaluation were Bias, the Root Mean Square Error (RMSE) and the Absolute Error (ABSE):

$$BIAS = \frac{1}{n}\sum_{i=1}^{n}(f_i - o_i) \qquad ABSE = \frac{1}{n}\sum_{i=1}^{n}|f_i - o_i|$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(f_i - o_i)^2}$$

f = forecast

o = observation

n = forecast/observation pairs

Results and discussion

Observed and predicted daily mean concentrations, for each pollutant, were obtained as the average of the metropolitan area sets of regression models, for the 2006-2010 period (Figure 2 and Figure 3 shows an example of daily estimates during the year 2007).

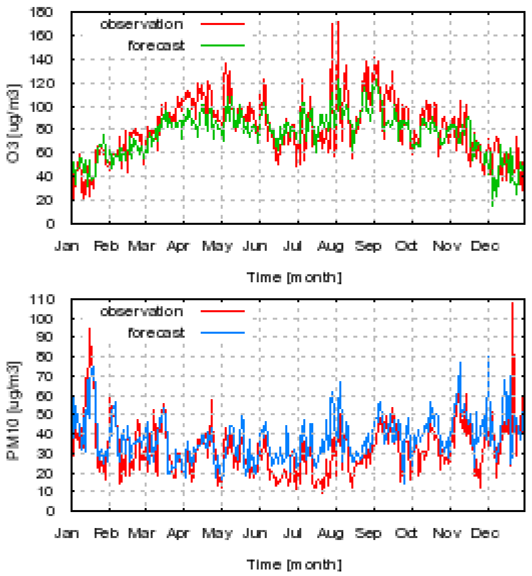

Figure 2 - Observed and predicted O_3 (top) and PM_{10} (bottom) mean concentrations in Lisbon 2007.

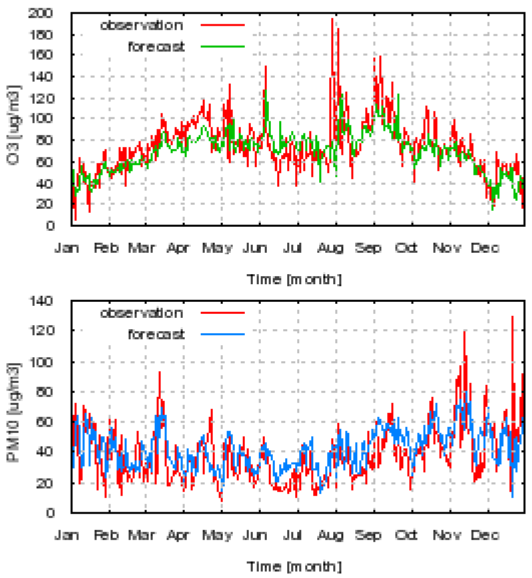

Figure 3 - Observed and predicted O_3 (top) and PM_{10} (bottom) mean concentrations in Oporto 2007.

The results obtained for the year 2007 over Lisbon and Oporto show a good agreement between the forecasted and measured concentrations, as shown in previous studies Ferreira *et al* (2000) and (2004), Marques *et al* (2006).

A first analysis of the Figures 2 and 3 seems to indicate that the persistence of chemical species concentrations has a major contribution on the statistical model of both regions. However, it is of paramount importance to understand the reason why the statistical model does not perform well, to act on and modify it, in order to get better results. Hence, an evaluation of the individual air quality station contributing to each pollutant index was attempted for the period between 2006-2010, over both metropolitan areas.

As expected, results show a local variability over each agglomeration region regarding ozone (Figure 4). Except for Loures air quality station (located in Lisbon), AQI hit rate over the Lisbon stations are higher than 75 %.

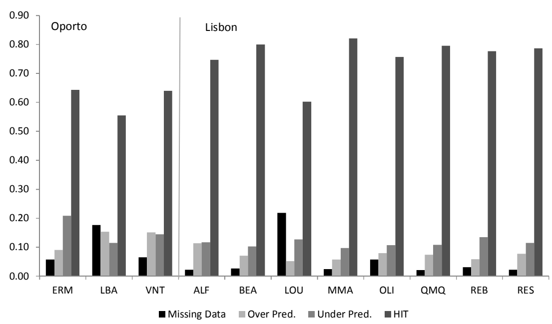

Figure 4 - Hit rates, under and overestimation of the AQI for ozone over Lisbon and Oporto air quality stations considered in the statistical forecast models.

Lower hit rates are associated with air quality stations with high missing data. When the AQI between the models forecast and the observed are not paired, the Lisbon region tends to be underestimated by the AQI, while for Oporto the opposite trend occurs.

Table 2 shows that the AQI forecasted for Lisbon tends to give Moderate values when Good are observed. Concerning the AQI expected for Oporto (Table 3), when severe AQI are observed the model gives better air quality indication, PM_{10} seems to be the driver to fail the observed AQI.

Table 2 - Contingency table for AQI model results, for Lisbon, from July 2006 to December 2010.

	AQI	Observed					
		Bad	Weak	Moderate	Good	Very Good	Total
Forecast	Bad	**0**	0	0	0	0	0
	Weak	0	**53**	92	31	1	177
	Moderate	0	40	**202**	307	13	562
	Good	0	13	59	**629**	40	741
	Very Good	0	0	0	1	**0**	1
	Total	0	106	353	0	54	**1481**

Table 3 - Contingency table for AQI model results, Oporto, from July 2006 to December 2010.

AQI		Observed					
		Bad	Weak	Moderate	Good	Very Good	Total
Forecast	Bad	**0**	0	0	0	0	0
	Weak	2	**123**	91	44	5	265
	Moderate	1	51	**166**	317	26	561
	Good	0	12	45	**377**	63	497
	Very Good	1	0	0	6	**0**	7
	Total	4	186	302	744	94	**1330**

Final Remarks

Considering the previous presented results, the statistical models need to be tuned in Oporto on rural/industrial stations and in the traffic Antas station. Lisbon agglomeration has tuned statistical models.

Figure 5 and Figure 6 show the existence of overestimation for PM_{10} and underestimation for O_3, in both agglomerations of Lisbon and Oporto.

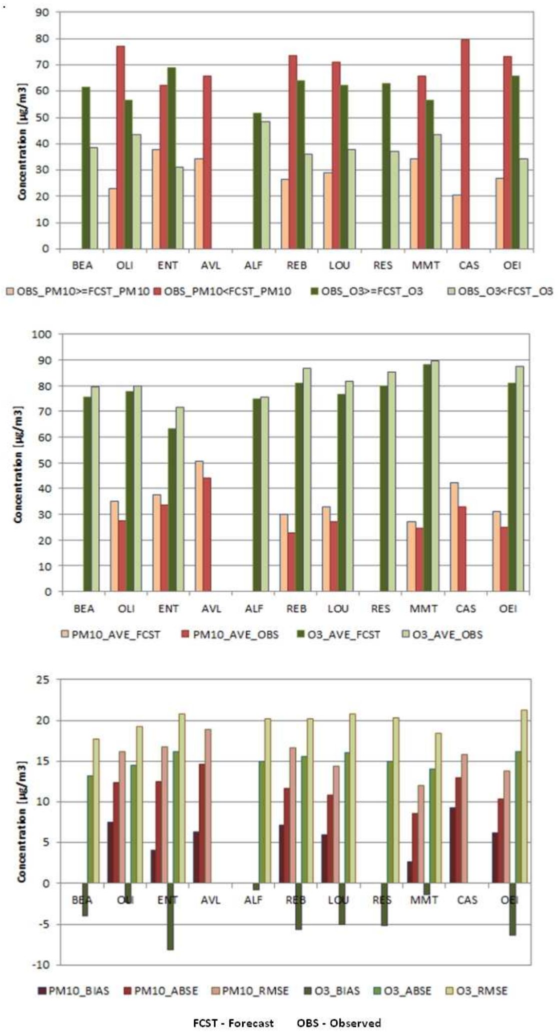

Figure 5 - Statistical measurements of agreement results and underestimation/overestimation analysis or for each air quality station, for Lisbon.

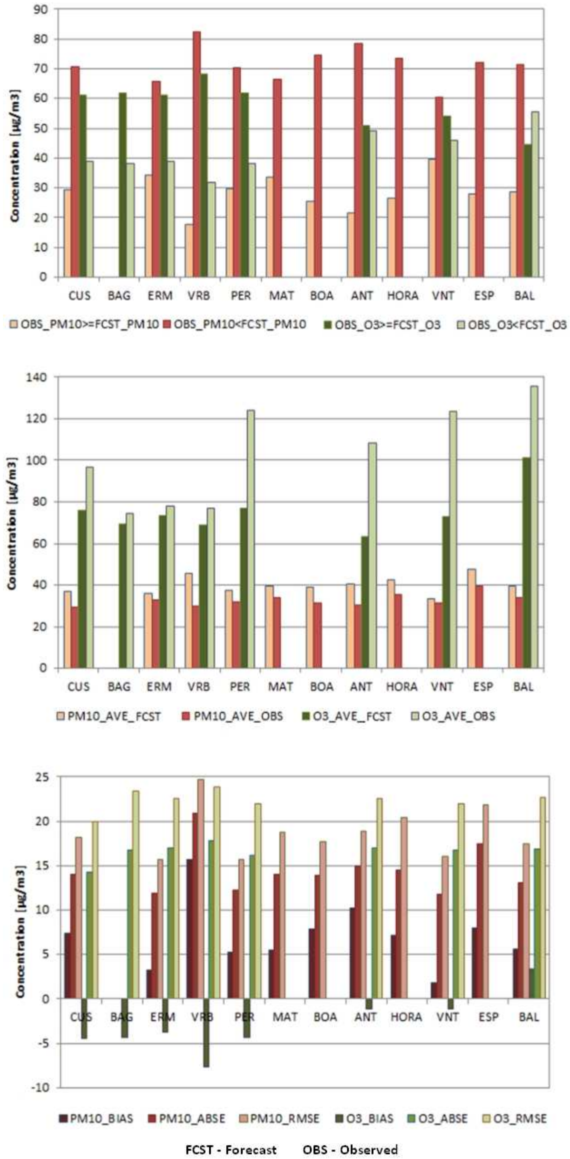

Figure 6 - Statistical measurements of agreement results and underestimation/overestimation analysis or for each air quality station, for Oporto.

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

Directive 2008/50/EC of the European Parliament and of The Council of 21 May 2008 on ambient air quality and cleaner air for Europe.

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

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

Marques, F., Cassmassi, J., Ferreira, F., Mesquita, S. (2006). Air quality forecast for Lisbon, Portugal: from data analysis to outreach., 2006 National Air Quality Conferences: Air Quality Forecasting, Mapping, and Monitoring and Communicating Air Quality and Communities in Motion "Quality of Air Means Quality of Life", February 5-8, 2006, San Antonio, Texas.