

PREVISÃO DA QUALIDADE DO AR PARA LISBOA POR MÉTODOS ESTATÍSTICOS

AIR QUALITY FORECAST FOR LISBOA USING STATISTICAL METHODS

Jorge Neto⁽¹⁾⁽²⁾, Pedro Torres⁽¹⁾, Francisco Ferreira⁽¹⁾ e Filomena Boavida⁽³⁾

⁽¹⁾Departamento de Ciências e Engenharia do ambiente, Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa, Quinta da Torre, 2829-516 Caparica, Portugal

⁽²⁾Departamento de Observação e Redes, Instituto de Meteorologia, Rua C ao Aeroporto, 1749-077 Lisboa, Portugal

⁽³⁾Instituto do Ambiente, Rua da Murgueira, 9/9^a, 2610-124 Amadora, Portugal
jorge.neto@meteo.pt; pmt@fct.unl.pt; ff@fct.unl.pt; filomena.boavida@iambiente.pt

SUMMARY

One of the important commitments of Portugal in the area of air quality is the fulfillment of the National 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. In this context, this presentation represents the work that has been developed using a statistical data base 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 statistic methods for the Lisbon region.

Resumo

O cumprimento por parte de Portugal da legislação nacional e europeia no domínio da qualidade do ar é um dos objectivos fundamentais da política de ambiente, nomeadamente como salvaguarda da saúde pública e dos ecossistemas. Nas últimas épocas de Verão, Lisboa tem revelado elevados valores de diversos poluentes atmosféricos, ultrapassando o estabelecido na legislação vigente. Considerou-se assim a necessidade da elaboração diária da previsão para o dia seguinte da qualidade do ar em Lisboa mais precisamente do ozono (O_3) e das partículas de diâmetro inferior a $10 \mu m$ (PM_{10}), tendo-se concretizado esse objectivo através da realização de protocolos entre o Instituto do Ambiente (IA), a Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa (FCT/UNL) e o Instituto de Meteorologia (IM). Inicialmente fez-se um estudo dos dados da qualidade do ar e algumas comparações com dados meteorológicos com o objectivo de conhecer melhor os comportamentos horários, diários e mensais dos mesmos, bem como sobre as excedências dos limiares horários fixados de acordo com a legislação em vigor.

Desenvolveram-se modelos estatísticos para a determinação das concentrações de O_3 e PM_{10} usando dados da qualidade do ar da rede da Comissão de Coordenação e Desenvolvimento

Regional de Lisboa e Vale do Tejo (CCDR-LVT) e de uma estação do IA (Alfragide), e dados meteorológicos (IM) à superfície de diversas estações da Rede de Estações Automáticas e em altitude das radiosondagens efectuadas na estação aerológica de Lisboa Gago Coutinho. Foi também efectuada a classificação subjectiva das situações sinópticas à superfície e aos 500hPa.

Metodologia

A primeira parte do trabalho contemplou a análise do comportamento horário, diário e mensal do O_3 na cidade de Lisboa e área envolvente, bem como sobre as excedências aos limiares horários fixados de acordo com a legislação em vigor. As estações da qualidade do ar usadas para tal são nove (Figura 1): oito da rede de qualidade do ar da Comissão de Coordenação e Desenvolvimento Regional de Lisboa e Vale do Tejo (CCDR-LVT) (Beato, Entrecampos, Loures, Olivais e Reboleira situadas na cidade de Lisboa e Loures, Odivelas e Oeiras-Marquês localizadas na área suburbana de Lisboa) e uma estação do IA em Alfragide.

O período de dados usado para este fim foi de 3 anos (2001 a 2003). A eficiência de medição em todas as estações foi superior a 75%, excepto para o caso dos Olivais a qual teve uma eficiência de 55% (a medição de ozono em Olivais começou em Fevereiro de 2002).

A legislação nacional que regula os níveis de O₃, Decreto-Lei n.º 320/2003 que resulta da transposição da Directiva Europeia 2002/3/CE, estabelece limiares que não deverão ser excedidos:

180 µ.g.m⁻³ (médias horárias): designado como limiar de informação ao público;

120 µ.g.m⁻³ (médias octohorárias): designado limiar de protecção da saúde humana.

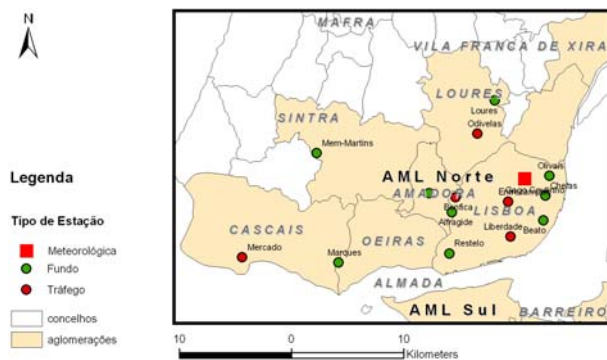


Figura 1 – Estações de monitorização da qualidade do ar na região de Lisboa

Numa primeira fase foram efectuadas correlações entre as concentrações de ozono e algumas variáveis meteorológicas de superfície da estação de Lisboa Gago Coutinho (assinalada pela estrela azul na Figura 1) do IM. Posteriormente foram desenvolvidos modelos estatísticos para a previsão da qualidade do ar no que se refere ao O₃ e PM₁₀.

As simulações iniciais foram realizadas com base em cinco anos de dados (1998-2002) sobre algumas variáveis meteorológicas à superfície da estação de Lisboa Gago Coutinho e concentrações de poluentes medidos nas estações da rede de qualidade do ar. Obtiveram-se modelos de regressão linear múltipla para algumas estações em diferentes períodos do ano.

Os modelos obtidos foram melhorados de acordo com as metodologias desenvolvidas para Los Angeles, Estados Unidos da América (Casmassi, 1987 e outros), recorrendo-se a um conjunto de variáveis mais alargado e diversificado. O período de dados utilizado foi diminuído para apenas dois anos (2001-2002) de modo criar uma base de dados mais consistente, sem as falhas existentes no intervalo mais longo, fazendo diminuir a capacidade de previsão dos modelos desenvolvidos. As variáveis utilizadas para a nova base de dados são apresentadas na Tabela 1. Das variáveis acrescentadas em relação ao ensaio inicial, destaca-se a diferença de pressão entre estações da rede de estações automáticas do IM que veio substituir a intensidade e rumo do vento usadas anteriormente. As variáveis em altitude são obtidas através das observações aerológicas (radiosondagens), das 12 horas UTC. Calcularam-se ainda variáveis retiradas indirectamente do perfil vertical de temperatura: a altura máxima da camada de mistura, segundo o

método de Holzworth, 1964, a altura da camada de inversão e a espessura da mesma. Foi ainda efectuada uma análise qualitativa das situações sinópticas observadas à superfície e aos 500 hPa para os dois anos acima mencionados (2001-2002), com base nas análises do modelo global de previsão do European Centre for Medium-Range Weather Forecasting (ECMWF). A análise permitiu identificar sete classes de três tipos de situações sinópticas distintas verificadas à superfície (Tabela 2). Em altitude, aos 500 hPa, obtiveram-se cinco classes (Tabela 3).

Tabela 1 – Variáveis utilizadas para a previsão do modelo

Variáveis meteorológicas à superfície	
Situação sinóptica às 12UTC	Temperatura máxima e humidade relativa média para Lisboa, Santarém, Évora e Beja e sua variação diária
	Diferença de pressão entre Lisboa-Porto, Lisboa-Portalegre, Lisboa-Évora e Lisboa-Faro (12UTC) e sua variação diária
Variáveis meteorológicas em altitude	
Situação sinóptica às 12UTC	Altura máxima da camada de mistura (Holzworth, 1964)
	Base e espessura da camada de inversão
	Altura do geopotencial (1000, 850, 700 e 500 hPa) – 12UTC
	Temperatura e humidade relativa (925, 850 e 700 hPa) – 12UTC
Variáveis da qualidade do ar	
PM ₁₀ , O ₃ , CO, NO ₂ e SO ₂	Os dados são tratados da mesma forma que o índice da Qualidade do Ar. Para cada poluente, cada valor diário é determinado para cada estação sendo depois feita a média entre todas as estações:
	- PM ₁₀ , NO ₂ e SO ₂ : médias diárias
	- O ₃ : máximo diário para valores horários e médias octo-horárias
	- CO: máximo diário para médias octo-horárias
	Variação diária de cada valor
Outras	
	Dia da semana
	“flag” tipo de dia (semana/fim semana)
	Duarção do dia solar

Tabela 2 – Classificação das situações sinópticas à superfície

Sistemas frontais
1 Sistemas frontais
Sistemas depressionários
2 Depressão/instabilidade
3 Influência de depressão
Sistemas de altas pressões
4 Calma à superfície
5 Circulação de N/NW
6 Anticiclone e vale invertido desde o Norte de África
7 Circulação de NE/E

Tabela 3 – Classificação das situações sinópticas aos 500 hPa

1 Depressão
2 Vale
3 Aproximação de vale
4 Crista em desenvolvimento ou fluxo zonal

Os dados serviram para testar estatisticamente novos modelos usando a análise de regressão múltipla (RM) e análise de regressão em árvore (CART - "Classification and Regression Tree"). Depois de feita a análise CART para ambos os poluentes, desenvolveu-se para cada um dos grupos finais um modelo de RM (Casmassi, 1987). Deve referir-se que a RM foi efectuada passo a passo, em modo de incrementação de variáveis, eliminando aquelas que apresentam um nível de significância de superior a $\alpha=0,10$.

Após a obtenção dos modelos foram efectuadas uma validação e uma análise de sensibilidade utilizando os dados referentes ao período de Janeiro, Fevereiro, Julho e Agosto de 2000. Este período de dados não foi utilizado no desenvolvimento do modelo. Pode então comparar-se os diversos resultados e fazer uma análise minuciosa da correlação dos modelos obtidos.

Resultados

A análise da Figura 2 mostra que o maior número de excedências de O_3 ocorre durante a tarde, tal como era esperado, uma vez que é o período mais quente do dia (86% das excedências do limiar de alerta do público ocorrem entre as 13h00 e 17h00). O tráfego rodoviário é de facto a maior fonte de poluição atmosférica na região de Lisboa (Ferreira et al., 2000). Nesta figura visualizam-se os perfis do O_3 e NO_2 destacando-se o seu comportamento inverso.

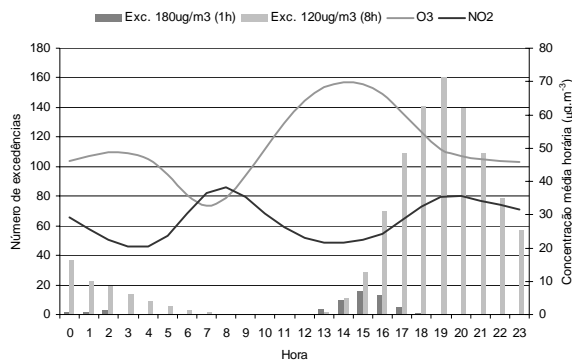


Figura 2 – Concentrações médias de O_3 e NO_2 e número de excedências horárias de O_3 para Lisboa no período de 2001-2003 (adaptado de Ferreira et al., 2004)

Na Figura 3 observa-se que o número de excedências de O_3 é superior nos meses mais quentes. No entanto, os valores médios horários mais elevados ocorrem na Primavera. Isto deve-se ao facto de amplitude diária do valor do O_3 ser menor nesta estação do ano.

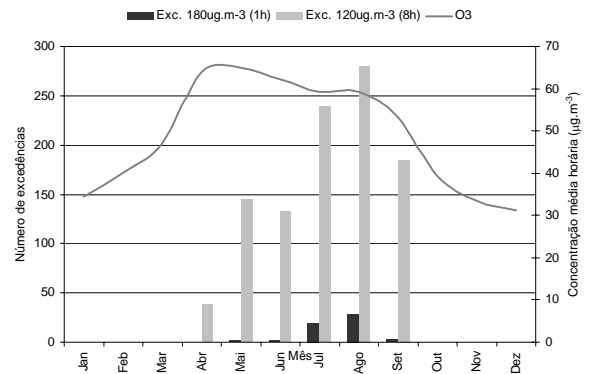


Figura 3 – Análise mensal de excedências e concentrações de O_3 para Lisboa no período de 2001-2003 (adaptado de Ferreira et al., 2004)

Na Figura 4 está representada a altura máxima da camada de mistura, constatando claramente a relação entre este parâmetro e as concentrações de O_3 .

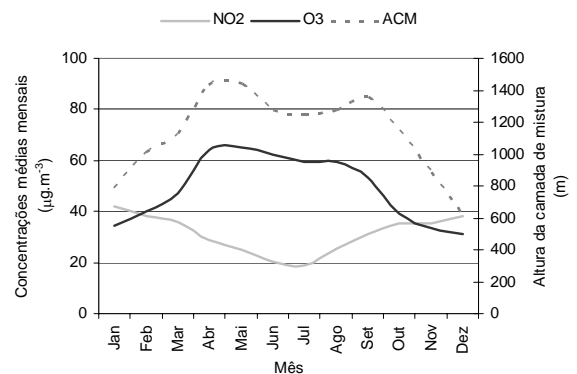


Figura 4 – Concentrações médias mensais de O_3 e NO_2 e altura máxima da camada de mistura em Lisboa (adaptado de Ferreira et al., 2004)

O padrão semanal do O_3 mostra uma boa correlação com o volume de tráfego, medido indirectamente através das concentrações de CO (Figura 5) (Ferreira, 2002). As concentrações mais elevadas de O_3 ocorrem ao domingo e as mais reduzidas à sexta-feira.

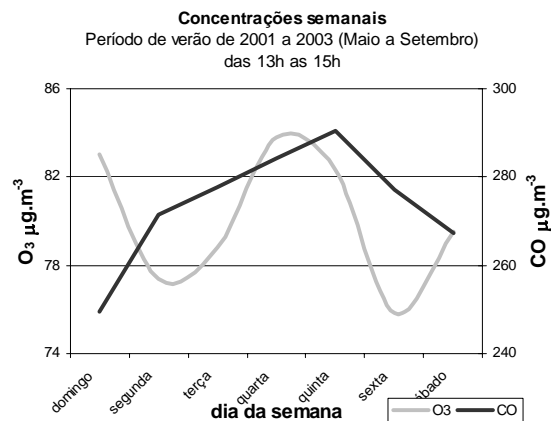


Figura 5 – Padrão semanal das concentrações do O₃ e CO em Lisboa no período de verão de 2001-2003 entre as 13 e 15 h

A Figura 6 realça uma boa relação entre a temperatura do ar à superfície (2 metros) observada e os valores de O₃, tal como era esperado.

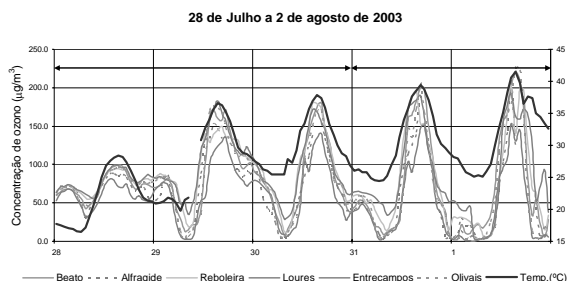


Figura 6 – Relação dos valores da temperatura do ar com concentrações de ozono em diferentes estações de Lisboa (adaptado de Ferreira et al., 2004)

Ao desenvolver modelos estatísticos e tendo construído a base de dados referente aos anos de 2001-2002, foram obtidos diversos modelos de RM para o O₃ e para PM₁₀ directamente dos diversos conjuntos finais obtidos na análise CART, tal como já foi referido na metodologia (Figura 7). Realça-se o facto de que, para a estimação da concentração de O₃, a temperatura máxima em Santarém acaba por ter um papel mais importante que a temperatura máxima de Lisboa. Este resultado parece ser um indicador de que a massa de ar, quando proveniente do quadrante leste, e portanto quente no Verão, está associado a valores altos de O₃.

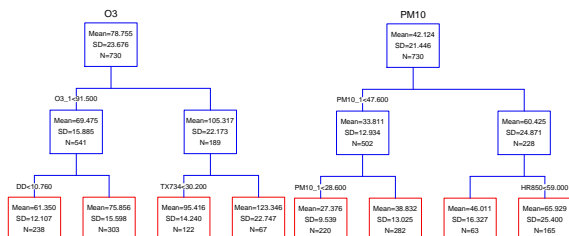


Figura 7 – Modelos de regressão em árvore obtidos para o O₃ e PM₁₀ do período de 2001-2002

Para o modelo do O₃, outra variável encontrada é o valor do dia anterior. Relativamente aos resultados obtidos com PM₁₀, a variável encontrada com maior importância é o valor do dia anterior. Não se encontrou discriminação de dados com base em variáveis meteorológicas.

A variância explicada pelos modelos obtidos (Tabela 4 e Tabela 5), variam entre 97.6 e 99.9% para o O₃ e 97.3 e 99.7% para as PM₁₀.

Os resultados dos modelos obtidos são considerados bons, não sendo apresentadas as equações correspondentes aos seis modelos seleccionados

(três para o ozono e três para partículas, de acordo com a discriminação em árvore apresentada), porque tal foi considerado demasiado extenso em termos de apresentação.

Tabela 4 – Coeficientes de correlação entre os dados observados e os diferentes resultados dos modelos para o O₃

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
Erro padrão	7.472	10.469	8.383	9.926
N	238	303	121	67

Tabela 5 – Coeficientes de correlação entre os dados observados e os diferentes resultados dos modelos para o PM₁₀

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
Erro padrão	6.769	10.054	6.055	15.338
N	218	280	61	164

Deve salientar-se que, para ambos os poluentes, a variável que representa a situação sinóptica à superfície (G) é apenas usada no modelos de previsão para ambos os poluentes quando as concentrações são mais elevadas. Ao contrário do que era esperado, os modelos parecem não atribuir muito peso aos efeitos dessa variável. Outra característica obtida para o estudo de O₃ como variável dependente no último modelo é que, ao contrário de outras, as variáveis meteorológicas em altitude não são utilizadas. Já para as PM₁₀, refira-se que as variáveis meteorológicas em altitude são muito mais utilizadas do que para os modelos de O₃.

A Figura 8 e a Figura 9 mostram os resultados obtidos através da utilização dos modelos no período de 2 de Julho a 31 de Agosto de 2000, para ambos os poluentes estudados.

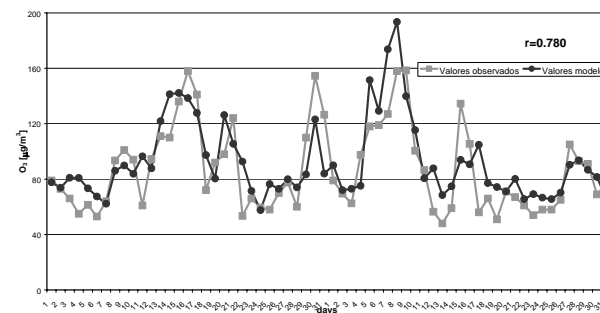


Figura 8 - Valores observados e simulados de concentrações máximas de O₃ usando os modelos de regressão múltipla e regressão em árvore+regressão múltipla (2 de Julho a 31 de Agosto de 2000)

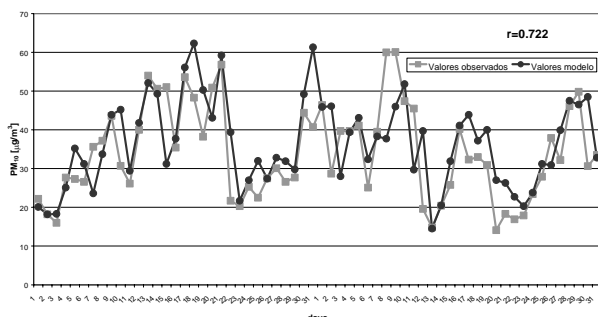


Figura 9 - Valores observados e simulados de concentrações médias de PM₁₀ usando os modelos de regressão múltipla e regressão em árvore+regressão múltipla (2 de Julho a 31 de Agosto de 2000)

O coeficiente de correlação mais elevado para o O₃ foi de 0.780. Para o PM₁₀ o coeficiente de correlação mais elevado é de apenas 0.722. O modelo para PM₁₀ apresenta uma maior dificuldade em prever grandes alterações de concentrações ao contrário do O₃. Isto pode dever-se ao facto da concentração de O₃ estar muito condicionada pela temperatura, enquanto as PM₁₀ não têm uma relação forte com nenhuma variável meteorológica. Daí que o estudo realizado das situações meteorológicas de forma mais integrada possa vir a ser fundamental para o desenvolvimento de modelos mais eficazes na previsão da concentração de PM₁₀.

Desenvolvimentos futuros, como a utilização de redes neuronais, poderão revelar-se métodos eficazes na determinação de melhores modelos de previsão (Wang et al., 2003).

Conclusões

O presente estudo mostra que a utilização de modelos de regressão linear múltipla desenvolvidos no quadro duma estrutura em árvore, como forma de melhor perceber as relações entre variáveis de qualidade do ar e meteorológicas e permitir a previsão de concentrações de O₃ e PM₁₀ para a área de Lisboa, tem resultados muito satisfatórios. A variância explicada pelos modelos é bastante grande e sempre superior a 72%, permitindo a sua aplicação em termos de previsão. A aplicação dos modelos num período de tempo diferente do usado para o estabelecimento das regressões apresenta correlações inferiores, cujos valores são no entanto suficientemente significativos para demonstrar a robustez da metodologia aplicada.

Referências

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.

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

EPA (2003): Guidelines for Developing an Air Quality (Ozone and PM_{2.5}) 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, pp. 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.

IM (2003): Boletim diário, Instituto de Meteorologia. 2000 a 2002

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, pp. 555-562.