

PROBLEMA 4: CARACTERIZAÇÃO DOS NÍVEIS DE NO_2 EM PORTUGAL

Adriana Matias^{1,2}; *Fábio Ferreira*^{1,3}; *Maria Pinto*^{1,4}; *Raquel Leiras*^{1,5};
Rui Machado^{1,6}; *Susana Cardoso*^{1,7}; *Coordenadora: Andreia Monteiro*^{8,9}

¹Faculdade de Ciências da Universidade do Porto - Departamento de Matemática

²up200907459@fc.up.pt, ³up201206807@fc.up.pt, ⁴up201507770@fc.up.pt,

⁵up201201441@fc.up.pt, ⁶up201204616@fc.up.pt, ⁷up200002915@fc.up.pt,

⁸ Universidade do Minho, CMAT & CIDMA, ⁹andreiaforte50@gmail.com

1. INTRODUÇÃO

No presente trabalho pretendemos caracterizar a evolução dos níveis de NO_2 em Portugal, através de abordagens geoestatísticas, indexadas no tempo e no espaço. Para desenvolver este projeto, foi realizada uma primeira análise exploratória dos valores diários da variável-alvo, medidos diariamente em Portugal, de 1 de outubro a 31 de dezembro de 2014, o que levou à identificação de três covariáveis (influência, ambiente e dia da semana). Em seguida, foram aplicadas ferramentas estatísticas apropriadas para modelar a tendência e a variabilidade espaço-temporal, permitindo-nos usar o método de kriging para fazer previsões. Esta metodologia tem importantes aplicações, uma vez que pode estimar as concentrações de dióxido de azoto quer em locais onde os dados foram perdidos, quer em locais onde não existem estações de monitorização.

2. POLUIÇÃO DO AR

A qualidade do ar é o termo geralmente usado para se referir ao grau de poluição no ar que respiramos. A poluição do ar pode ser provocada por uma mistura de elementos químicos ou por reações químicas que alteram a constituição natural da atmosfera. Entre os inúmeros problemas derivados da poluição do ar, podemos citar os que estão relacionados com a saúde (nomeadamente os problemas respiratórios) e os ambientais (como a deterioração da camada de ozono).

O índice de qualidade do ar é uma ferramenta que permite uma classificação simples e compreensível do estado da qualidade do ar. Este índice foi desenvolvido para poder traduzir a qualidade do ar, especialmente das aglomerações existentes no país, mas também de algumas áreas industriais e cidades.

A estimativa do índice de qualidade do ar envolve medições dos seguintes elementos químicos, dos quais se faz uma breve descrição em seguida. [1]

- **Monóxido de carbono (CO):**

O monóxido de carbono (CO) provém essencialmente da combustão de combustíveis fósseis ou de outras matérias orgânicas. As principais fontes naturais deste poluente são as erupções vulcânicas, os fogos florestais e a decomposição da clorofila. No meio urbano, o tráfego automóvel é a principal fonte de CO sendo as zonas de tráfego intenso as que apresentam concentrações mais elevadas deste poluente. A exposição a concentrações elevadas de CO está associada à diminuição da perceção visual, capacidade de trabalho, destreza manual, capacidade de aprendizagem e desempenho de tarefas complexas. Na atmosfera, transforma-se em dióxido de carbono, contribuindo assim para o efeito de estufa.

- **Dióxido de azoto (NO_2):**

O dióxido de azoto ou dióxido de nitrogénio é um gás inodoro, incolor e pouco tóxico, não sendo considerado um poluente nocivo para as concentrações normalmente presentes na atmosfera. As grandes fontes destes compostos são as centrais termoelétricas, os transportes rodoviários, os navios e alguns processos de fabrico como por exemplo a indústria química de produção de fertilizantes azotados. Em áreas urbanas a principal fonte de NO_2 são os veículos automóveis, pelo que as concentrações deste poluente acompanham geralmente as variações do tráfego automóvel. O NO_2 , em concentrações elevadas, causa efeitos que vão desde a irritação dos olhos e garganta, até afetação da membrana dos órgãos respiratórios, diminuição da capacidade respiratória, dores no peito, edema pulmonar e danos no sistema nervoso central e nos tecidos. Contribuem igualmente para o fenómeno das chuvas ácidas assim como para a eutrofização dos cursos de água e dos lagos, para a destruição da camada de ozono

estratosférico e para o efeito de estufa.

- **Dióxido de enxofre (SO_2):**

O dióxido de enxofre (SO_2) é emitido no momento da queima de combustíveis fósseis tais como o carvão e o fuelóleo. As principais fontes são as centrais térmicas, as grandes instalações de combustão industriais e as unidades de aquecimento doméstico. As emissões provenientes dos veículos têm vindo a baixar com a diminuição progressiva do enxofre nos combustíveis.

- **Ozono (O_3):**

O ozono (O_3), nas camadas altas da atmosfera, desempenha um papel vital ao filtrar a radiação solar ultravioleta, protegendo assim a vida sobre a Terra. Nas áreas urbanas, na proximidade das fontes emissoras, o monóxido de azoto emitido pelos veículos automóveis pode reagir com o O_3 , reduzindo-se assim localmente as concentrações deste poluente. Tem um efeito nocivo sobre a vegetação ao reduzir a atividade fotossintética e, ao nível da troposfera, é um gás com efeito de estufa, contribuindo para o aquecimento do planeta.

- **Partículas finas (PM_{10}):**

As partículas são um conjunto complexo de substâncias, minerais ou orgânicas, que se encontram em suspensão na atmosfera, sob a forma líquida ou sólida. Em zonas urbanas os transportes rodoviários são considerados a maior fonte emissora de partículas, observando-se as maiores concentrações na proximidade de vias de tráfego intenso. Quanto mais pequenas as partículas, maior a probabilidade de penetrarem profundamente no aparelho respiratório e maior o risco de induzirem efeitos negativos. Os efeitos de sujidade nos edifícios e monumentos são os efeitos mais evidentes das partículas no ambiente.

Em relação aos poluentes acima mencionados, pretendemos estudar as concentrações de NO_2 . Note-se que se regista uma forte presença de dióxido de azoto nas zonas urbanas e áreas industriais, pelo que é um bom marcador da exposição à poluição relacionada com o tráfego.

Um estudo recente de investigadores da *University of California*, nos EUA, sugere uma ligação entre os níveis de NO_2 e a incidência da síndrome da morte súbita em recém-nascidos, há também estudos que relacionam o NO_2 à incidência de autismo. Pessoas que possuam doenças que possam ser agravadas devido à exposição ao dióxido de azoto não devem ser autorizadas para trabalhos com este produto, pois, em altas concentrações, ele pode provocar pneumonia química retardada, edema pulmonar, irritação da mucosa do nariz, danos severos aos pulmões, semelhantes aos provocados pelo enfisema pulmonar e a exposição contínua pode

causar diminuição permanente das funções pulmonares.

Quanto aos efeitos no meio ambiente, ao ser oxidado na atmosfera, o óxido de azoto produz o ácido nítrico (HNO_3), um dos componentes que aumenta a acidez da chuva, e causa vários danos à natureza por ser corrosivo. A chuva ácida, ao cair na superfície, altera a composição química do solo e das águas, atinge as cadeias alimentares, destrói florestas e lavouras, corrói estruturas metálicas, monumentos históricos e edificações. Muitas vezes, o facto do céu ter um tom cinzento em cidades com tantos veículos, deve-se à formação do NO_2 na atmosfera.

As medições de NO_2 indicam a localização da sua fonte, pois o tempo de residência deste poluente é de aproximadamente um dia, ficando concentrado próximo das fontes poluidoras.

3. BASE DE DADOS

A qualidade do ar tem vindo a ser objeto de um vasto trabalho ao nível do Ministério do Ambiente, do Ordenamento do Território e do Desenvolvimento Regional no quadro da Agência Portuguesa do Ambiente, em coordenação com as Comissões de Coordenação e Desenvolvimento Regional no território de Portugal Continental e com as Direções Regionais do Ambiente das Regiões Autónomas.

No âmbito do sistema de monitorização de qualidade do ar em Portugal, a Agência Portuguesa do Ambiente promoveu o desenvolvimento da base de dados sobre qualidade do ar QualAr, com a possibilidade de consulta via Internet. Esta aplicação visa centralizar todos os dados de qualidade do ar medidos em Portugal e disponibilizar informação ao público sobre qualidade do ar. [2]

A base de dados em estudo, diz respeito a dados horários obtidos em 49 estações de monitorização, instaladas em diferentes tipos de zonas - rurais, suburbanas e urbanas - e apresentam tipologias distintas, dependentes das emissões dos poluentes dominantes nas zonas onde se encontram instaladas, sendo classificadas como estações de fundo, industriais e de tráfego, representando, assim, diferentes tipos de exposição da população à poluição atmosférica. Estas apresentam as seguintes características:

- As **estações de fundo** (background) não se encontram sob a influência direta de vias de tráfego ou de qualquer fonte próxima de poluição. Permitem conhecer a exposição média da população aos fenómenos de poluição de fundo;

- As **estações industriais** (industrial) encontram-se situadas na proximidade de zonas industriais. Permitem conhecer as concentrações máximas de certos poluentes de origem industrial aos quais a população pode estar pontualmente exposta;
- As **estações de tráfego** (traffic) situam-se na proximidade de vias de tráfego intenso e permitem avaliar o risco máximo de exposição da população às emissões do tráfego automóvel. Esta exposição é, regra geral, de curta duração mas os níveis de poluição observados são normalmente elevados.

As estações de monitorização estão localizadas predominantemente em zonas urbanas e na costa de Portugal, como se constata na figura 1.

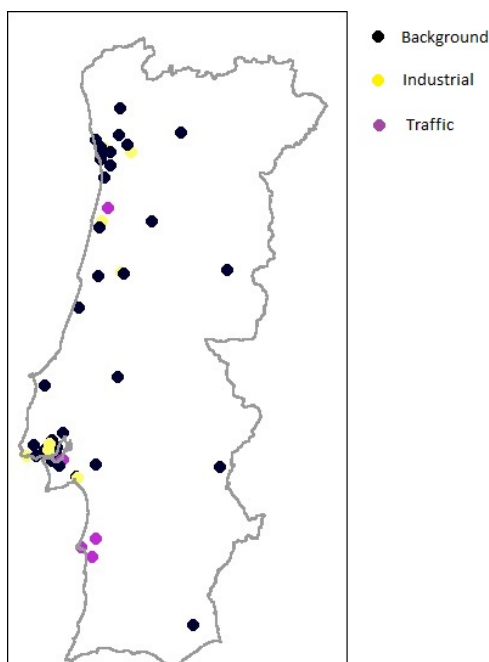


Fig. 1. Localização e tipo de sítio onde se inserem as estações na área em estudo (Portugal).

Um total de 49 estações foram utilizadas na análise (33 fundo, 6 industriais e 10 tráfego), onde 29 delas foram localizadas em áreas urbanas, 9 em áreas suburbanas e 11 em áreas rurais.

A Base de dados foi retirada do site QualAr, the Environment Portuguese Agency (<http://qualar.apambiente.pt/index.php?page=1&day=15&month=4&year=2015&x=10&y=4>).

4. PROBLEMA

De um modo genérico os objetivos deste trabalho são compreender como diferentes fatores afetam as concentrações de NO_2 detetadas, modelar o efeito sazonal em conjunto com a estimação da variabilidade espaço-temporal e caracterizar a evolução espacial e temporal dos níveis de concentração de NO_2 utilizando ferramentas geostatísticas.

5. O PROCESSO ESPAÇO-TEMPORAL

A variabilidade espacial e temporal é usada para caracterizar o desempenho de processos ambientais, tais como concentrações de poluentes atmosféricos, precipitação ou ventos de superfície.

Considere-se $\{Z(\mathbf{s}, t) : (\mathbf{s}, t) \in \mathbb{R}^d \times \mathbb{R}\}$, um processo aleatório, indexado no espaço por $\mathbf{s} \in \mathbb{R}^d$ e no tempo por $t \in \mathbb{R}$ que se pode decompor da seguinte forma:

$$Z(\mathbf{s}, t) = \mu(\mathbf{s}, t) + \delta(\mathbf{s}, t)$$

onde $\mu(\mathbf{s}, t)$ denota a média e $\delta(\mathbf{s}, t)$ os resíduos.

6. IMPLEMENTAÇÃO

De modo a estudar o proposto, criaram-se rotinas em R . Estas rotinas implicaram a instalação de alguns pacotes, assim como a utilização das funcionalidades de geração de gráficos da linguagem para produzir os resultados presentes na próxima Secção.

Mostra-se ao longo das secções seguintes os resultados obtidos através destas rotinas, começando por uma análise exploratória dos dados e estimação da variação de larga escala (correspondente à média), seguida da caracterização dos resíduos (variação de pequena escala). Por fim, faz-se a previsão das concentrações de NO_2 através de técnicas de *kri-ting*.

7. ANÁLISE EXPLORATÓRIA

Os dados utilizados são referentes à leitura dos níveis de concentração de NO_2 , nas 49 estações, de 1 de outubro de 2014 a 31 de dezembro de 2014, estudou-se estes três meses devido ao facto de os valores da concentração de NO_2 serem mais significativos em comparação com os restantes meses do ano de 2014. [2]

As 11 variáveis presentes na base de dados são:

- **NO2:** concentração de NO_2 ;
- **x:** latitude corresponde à localização de cada estação;
- **y:** longitude corresponde à localização de cada estação;

- **hour**: hora da medição (0,1,...,23);
- **month**: mês da medição (outubro, novembro e dezembro);
- **year**: ano da medição (2014);
- **influence**: tipo de estação (fundo, tráfego e industrial);
- **environment**: tipo de ambiente onde a estação se insere (rural, suburbano e urbano).

A matriz dos dados contém 18% de observações em falta relativamente às medições da variável NO_2 , o que se deve, em parte, ao facto das estações passarem por períodos de manutenção.

A média da concentração de NO_2 é de $20.6\mu g/m^3$ e o seu desvio de padrão é de $21.9\mu g/m^3$.

Nos três meses em estudo (outubro, novembro e dezembro), o limite de $200\mu g/m^3$ na variável em estudo foi excedido 20 vezes enquanto que uma diretiva europeia indica que o limite não deverá ser excedido mais de 18 vezes ao longo de um ano. Podemos concluir assim que é importante controlar as concentrações de NO_2 , para garantir uma melhoria da qualidade do ar.

Foi calculada a periodicidade para as estações sem dados em falta e obteve-se periodicidade de 12:00h e 24:00h, esta informação é depois importante para a definição do modelo de larga escala.

7.1. Boxplots

Apresentam-se os boxplots referentes às concentrações de NO_2 consoante o tipo de estação (fundo, tráfego e industrial) e o tipo de ambiente em que a estação se insere (rural, suburbano e urbano), respetivamente nas figuras 2 e 3.

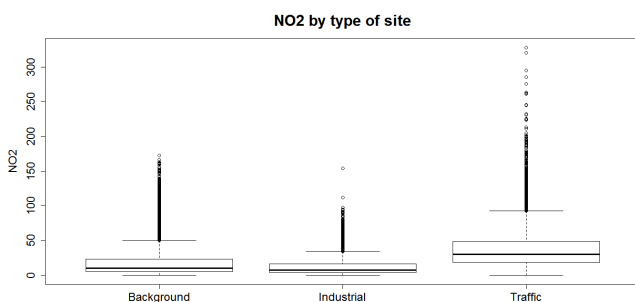


Fig. 2. Boxplots das concentrações de NO_2 consoante o tipo de estação (fundo, industrial e tráfego).

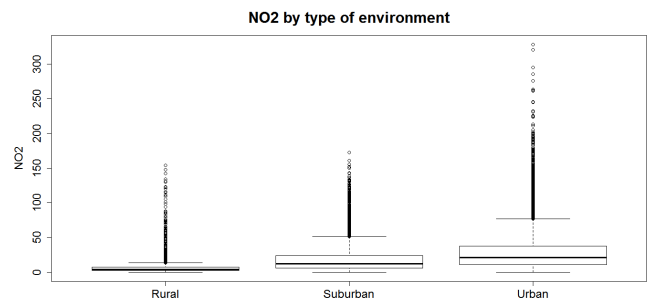


Fig. 3. Boxplots das concentrações de NO_2 consoante o tipo de ambiente em que a estação se insere (rural, suburbano e urbano).

Observando a figura 2, podemos ver que passando de uma influência de fundo para uma de tráfego verifica-se um aumento significativo da concentração de NO_2 e um aumento da variabilidade, tal como previsto uma vez que as estações de fundo não se encontram sob a influência direta de vias de tráfego nem de fontes diretas de poluição, enquanto que as estações de tráfego situam-se na proximidade de vias de tráfego intenso. Para além disso, verifica-se também que os valores da concentração de NO_2 são mais baixos para as estações do tipo industrial, o que pode ser explicado pelo facto de a queima de combustíveis fósseis (nos veículos automóveis, por exemplo) ser a fonte mais relevante na origem do NO_2 .

A figura 3 mostra a variação das concentrações de NO_2 para os ambientes rural, suburbano e urbano. Como já foi mencionado, ambientes urbanos referem-se a ambientes com elevadas concentrações populacionais, ao contrário de ambientes rurais onde a atividade agrícola predomina. Assim verifica-se que, na figura 3, os valores das concentrações de NO_2 vão sendo superiores passando do ambiente rural para suburbano e para o urbano, assim como se verifica um aumento da variabilidade nos valores das concentrações. Estes resultados eram esperados, uma vez que também o tráfego automóvel e a atividade industrial é superior em regiões urbanas.

7.2. Concentrações de NO_2 durante Semana e Fim de Semana

Estudou-se a concentração de NO_2 ao longo de todos os dias da semana, para verificar se existia alguma diferença significativa nas concentrações de NO_2 medidas. Observou-se que as concentrações de NO_2 durante os dias da semana não apresentavam diferenças significativas, no entanto à sexta-feira eram mais elevadas, como se pode observar na figura 4. Já para sábado e domingo verificaram-se menores concentrações. Deste modo, optou-se por apenas fazer a

distinção entre dias da semana e fim de semana.

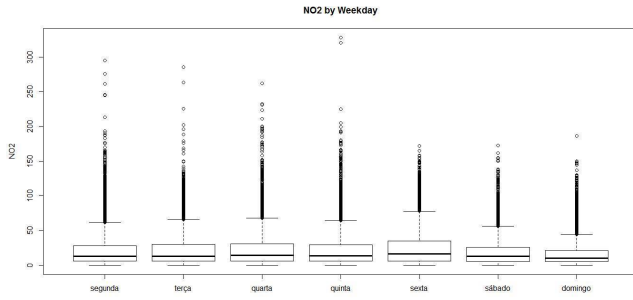


Fig. 4. Boxplot referente aos níveis de NO_2 em cada dia da semana.

Os valores médios horários para os dias úteis e de fim de semana estão representados na figura 5, revelando que os níveis de NO_2 apresentam 2 picos máximos às 08:00h e 18:00h, sendo o segundo mais pronunciado que o primeiro, provavelmente relacionados com as horas de maior tráfego automóvel. Também é possível denotar que às 13:00h há uma diminuição de concentração de NO_2 , possivelmente devido ao facto de ser uma hora em que há menos tráfego.

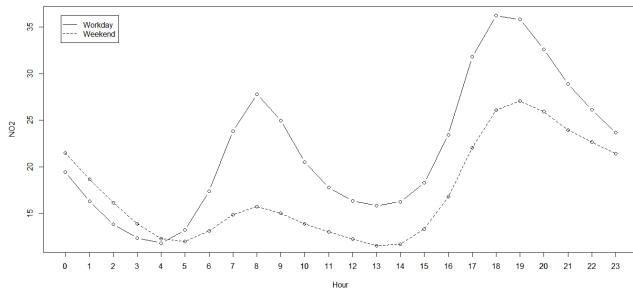


Fig. 5. Comparação da distribuição de NO_2 em dias úteis e fins de semana por hora.

7.3. Histograma de NO_2

As concentrações horárias de NO_2 estão representadas na figura 6 e revelam uma distribuição assimétrica. O facto do histograma ser assimétrico será importante para a modelação da média.

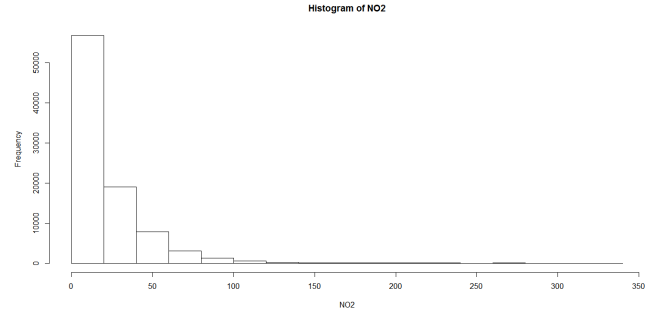


Fig. 6. Histograma da concentração do NO_2 .

8. ESTIMAÇÃO DA MÉDIA

Para se estimar a variação da média $\mu(s, t)$ usamos o modelo linear generalizado (GLM):

$$\eta(s, t) = \alpha + \beta_1 X_1(s, t) + \dots + \beta_k X_k(s, t)$$

com $\alpha, \beta_i, \in \mathbb{R}$, $g(\mu(s, t)) = \eta(s, t)$ e $g(\cdot)$ uma função de ligação suave e invertível.

Para as concentrações de NO_2 a resposta assume uma distribuição gamma e $g(\cdot)$ é a função logarítmica.

Com vista a modelar a periodicidade das estações considera-se um modelo misto onde incluímos as funções seno e coseno:

$$\eta(s, t) = \alpha + \beta_1 X_1(s, t) + \dots + \beta_k X_k(s, t) + \phi_{1,1} \cos(\pi t \omega) + \phi_{1,2} \sin(\pi t \omega) + \dots + \phi_{l,1} \cos(l \pi t \omega) + \phi_{l,2} \cos(l \pi t \omega)$$

com $\phi_{l,1}, \phi_{l,2}, \omega \in \mathbb{R}$, sendo ω o inverso da frequência.

Foram analisados diferentes modelos, o *Modelo 1* apenas incluía as covariáveis ambiente, influência e dia da semana, no *Modelo 2* utilizou-se o modelo misto para modelar a sazonalidade e no *Modelo 3* considerou-se o modelo com todas as variáveis significativas.

Usou-se o critério AIC (Critério de Informação de Akaike) e a percentagem de variabilidade explicada por cada modelo para escolher qual o melhor a ser usado.

Modelos	AIC	% Variabilidade
Modelo 1	675862	32%
Modelo 2	667302	37.6389%
Modelo 3	667281	37.6432%

Table 1. Comparação de Modelos.

Analisando a tabela 1, concluímos que o melhor modelo é o *Modelo 3* pois apresenta menor AIC e maior percenta-

gem de variabilidade. Os resultados para a regressão, considerando o Modelo 3, são apresentados na tabela 2.

Parâmetro	Estimativa	Erro Padrão
(Intercepto)	1.674796	0.007689
Tipo de estação (baseline: Fundo)- Industrial	-0.500388	0.008781
Tipo de estação (baseline: Fundo)- Tráfego	0.443375	0.008418
Dia (baseline: Fim de Semana)- Semana	0.215237	0.006663
Ambiente (baseline: Rural)- Suburbano	1.135722	0.009552
Ambiente (baseline: Rural)- Urbano	1.297806	0.007796

Table 2. Estimação dos coeficientes da regressão gama para concentrações de NO_2 horárias.

Conclui-se então que o pior cenário para os níveis de NO_2 é encontrado no ambiente urbano e sob influência de tráfego. Note-se também que a concentração de NO_2 aumenta com um fator de $exp(1.3) = 4$ ao mudar de ambiente rural para ambiente urbano e com um fator de $exp(0.44) = 2$ de uma estação de tipo fundo para tipo tráfego.

9. ESTIMAÇÃO DOS RESÍDUOS

Depois da estimação de larga escala, os resíduos podem ser calculados e o próximo passo é a caracterização da estrutura de dependência dos resíduos. Para a caracterização dos resíduos $\delta(s, t)$ usou-se a função variograma, que é dada por:

$$\gamma_{st}(h_s, h_t) = \frac{Var[\delta(s + h_s, t + h_t) - \delta(s, t)]}{2}$$

onde (h_s, h_t) toma valores em $\mathbb{R}^d \times \mathbb{R}$.

Um exemplo da função variograma é demonstrado na figura 7 onde o parâmetro *sill* representa a variância total do processo, o parâmetro *nugget* corresponde à variância associada aos erros de medição, o *partial sill* a diferença entre o *sill* e o *nugget*. Por fim, o parâmetro *range* associa-se ao ponto a partir do qual as observações deixam de estar correlacionadas.

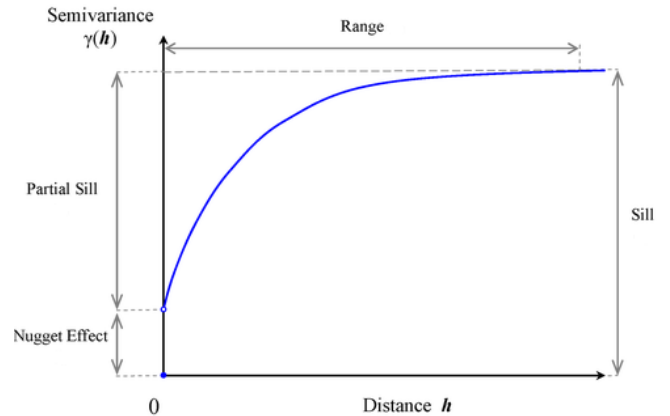


Fig. 7. Exemplo de uma função variograma, com os respectivos parâmetros.

Note-se também que numa parte mais avançada do projeto, nomeadamente na aplicação de técnicas de Kriging, devemos garantir que a função variograma, γ_{st} , é definida negativa. Como o modelo empírico pode não satisfazer essa condição, aproximou-se a um modelo teórico, como vamos ver de seguida.

Podemos considerar dois tipos de modelos teóricos para o estudo da dependência dos resíduos - os modelos separáveis e os modelos não separáveis. Nos modelos separáveis consideramos as variáveis do tempo e espaço separadamente e não temos em conta a sua interação. Por outro lado, os modelos não separáveis têm em conta a interação espaço-temporal. Os primeiros têm a vantagem de ser de implementação computacional mais simples e rápida, contudo os segundos representam de uma forma mais fidedigna os dados.

Para o estudo dos resíduos vamos considerar o modelo *product-sum* e o modelo *sum-metric*, ambos modelos não separáveis.

9.1. Modelo Product-Sum

O modelo *product-sum* é dado pela expressão:

$$\gamma_{st}(h_s, h_t) = \gamma_s(h_s) + \gamma_t(h_t) - k\gamma_s(h_s)\gamma_t(h_t)$$

com

$$k = \frac{sill_s + sill_t - sill_{st}}{sill_s sill_t}$$

onde $sill_s$ e $sill_t$ representam respetivamente a variância total dos semivariogramas marginais do espaço e tempo e $sill_{st}$ representa a variância total do variograma γ_{st} , que relaciona ambas as variáveis.

9.2. Modelo Sum-Metric

O modelo *sum-metric* é dado pela expressão:

$$\gamma_{st}(h_s, h_t) = \gamma_s(h_s) + \gamma_t(h_t) + \gamma(|h_s| + \alpha|h_t|)$$

onde γ representa a função variograma e $\alpha \in \mathbb{R}$.

Este modelo tem a vantagem de considerar o parâmetro α que permite efetuar um ajuste entre as escalas temporal e espacial.

9.3. Variograma Espacial

Começamos por analisar os variogramas separadamente como forma exploratória e também como apoio para uma futura modelação conjunta das variáveis. Começamos por apresentar o variograma espacial na figura 8. Neste variograma foi usada uma curva gaussiana para ajustar aos dados e foram obtidos os valores para os parâmetros da função variograma. O parâmetro *nugget*=0.16, o *partial sill*=0.6 e o *range*=34. O parâmetro *range* ter um valor de 34, indica que a partir de uma distância de 34 km deixa de haver uma correlação significativa entre as estações.

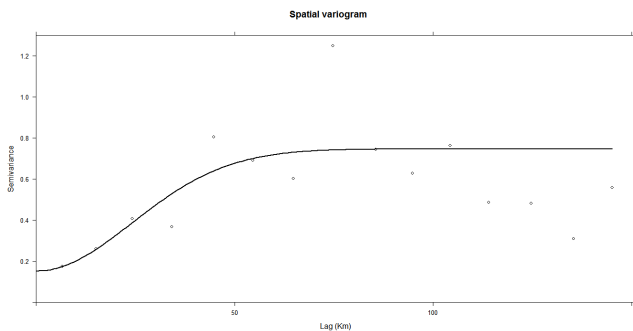


Fig. 8. Variograma Espacial.

9.4. Variograma Temporal

De seguida, estudou-se o variograma temporal, cujo gráfico está representado na figura 9, para qual se considerou uma curva exponencial para ajuste aos dados. Neste variograma o parâmetro *nugget* tem um valor de 0.6, o *partial sill*=0.1 e o *range*=57. Analogamente ao gráfico anterior, o parâmetro *range* indica que deixa de haver correlação significativa a partir de 57 horas, aproximadamente 2 dias.

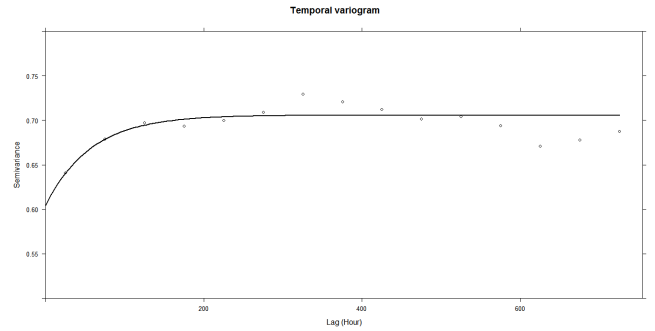


Fig. 9. Variograma Temporal.

9.5. Validação cruzada

Com o objetivo de decidir qual o modelo a adotar para caracterizar a estrutura da dependência dos resíduos, se o modelo *sum-metric* ou se o modelo *product-sum*, aplicamos o método da validação cruzada.

A validação cruzada é uma técnica que consiste na eliminação de uma observação e na sua predição a partir da restante base de dados.

Através dos valores previstos e dos valores reais de cada observação foram computados o erro médio (ME) e o erro quadrático médio (MSE). Os resultados obtidos são apresentados na tabela 3.

Modelo	Conjunto	Temporal	Espacial	ME	MSE
Modelo Product-sum	-	Exp	Gau	- 0.018	0.220
Modelo Sum-metric	Gau	Exp	Gau	-0.007	0.200

Table 3. Estimativas de ME e MSE do estudo de validação cruzada.

Analisando a tabela 3 conclui-se que foram obtidos resultados similares. Neste contexto optou-se pela escolha do modelo *sum-metric* pois permite-nos variogramas específicos para cada uma das componentes: componente espacial, temporal e espaço-temporal e para além disso contém o parâmetro α que nos permite o ajuste entre a escala temporal e espacial.

Para a componente temporal optou-se por escolher a função exponencial e para as componentes espacial e espaço-temporal a função gaussiana. Na tabela 4 são apresentados os resultados obtidos.

Variograma	Modelo	τ^2	σ^2	ϕ	α
Espacial	Gaussiana	0.007	0.552	37.784	
Temporal	Exponencial	0.010	0.071	100.000	
Conjunto	Gaussiana	0.173	0.148	69.978	13.065

Table 4. Parâmetros estimados para o variograma espacial, temporal e espaço-temporal.

Na tabela 4 são apresentados os seguintes parâmetros:

- τ^2 que corresponde ao parâmetro *nugget* (erros associados);
- σ^2 que corresponde ao parâmetro *partial sill* de onde podemos concluir que a componente espacial é a que mais contribui para o modelo;
- ϕ que corresponde, por exemplo, no caso da componente espacial à distância para além da qual deixa de haver correlação (neste caso $\approx 38\text{km}$) e no caso da componente temporal às horas para além da qual deixa de haver correlação (neste caso $\approx 100\text{h}$, isto é, ≈ 4 dias);
- α parâmetro que ajusta a escala temporal e espacial.

A figura 10 e a figura 11 correspondem, respetivamente, ao variograma empírico e ao variograma ajustado pelo modelo *sum-metric*. Como o variograma empírico não satisfaz a condição de ser definido negativo fez-se um ajuste por um modelo teórico.

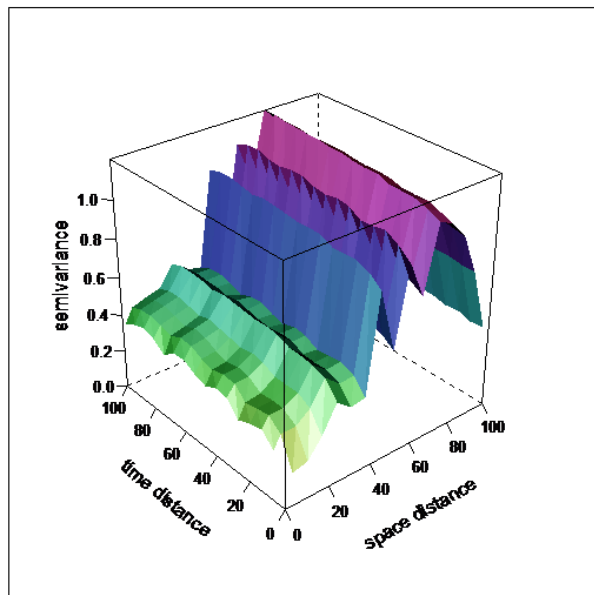


Fig. 10. Estimativas experimentais para o variograma espaço-temporal.

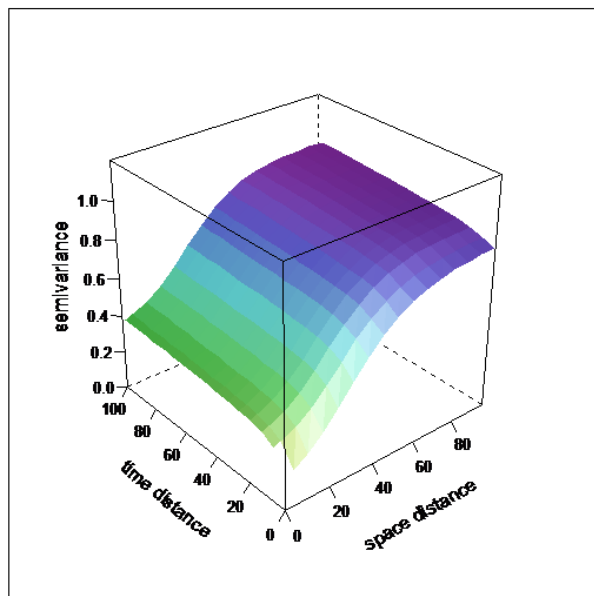


Fig. 11. Variograma espaço-temporal ajustado pelo modelo *sum-metric*.

O variograma ajustado pelo modelo *sum-metric* exige a estimativa dos parâmetros desconhecidos do modelo empírico,

ou seja, a estimativa de τ^2 , σ^2 , ϕ e α .

10. KRIGING PARA OS RESÍDUOS

Com o objetivo de modelizarmos os resíduos do nosso modelo para conseguirmos tirar conclusões sobre o nível de NO_2 em Portugal, utilizamos um método chamado Kriging.

Kriging é um método de interpolação que inclui relações estatísticas entre os pontos conhecidos, tal como a autocorrelação entre eles e utilização de uma média ponderada para a covariância espacial. Devido a isto, as técnicas estatísticas não só têm a capacidade de produzir uma superfície de previsão, mas também fornecer alguma medida de segurança das previsões.

O método inclui várias etapas, nomeadamente uma exploração dos dados e a criação de um variograma e é frequentemente usado em geologia.

Existe mais do que um modelo de Kriging: Simples, Ordinário, Universal, entre outros. O modelo mais adequado aos nossos dados foi o Kriging Ordinário, que assume que a média é constante na vizinhança de cada ponto de avaliação.

Posto isto, fomos estudar a variabilidade dos níveis de NO_2 em Portugal para uma sexta-feira (visto anteriormente como o dia da semana com maiores níveis de NO_2) e para um domingo (visto anteriormente como o dia da semana com menores níveis de NO_2), sendo que em cada um dos dias analisamos as horas: 08:00h e 18:00h (horas que se destacam por altos níveis de NO_2) e as 13:00h (hora que se destaca pelos baixos níveis de NO_2). Os resultados são apresentados na figura 12.

É realmente notório que os níveis de NO_2 são efetivamente mais altos para sexta-feira em comparação com domingo, e que as 08:00h e as 18:00h se destacam pela mesma razão em relação às 13:00h. Também podemos observar que as zonas do litoral são as mais afetadas.

10.1. Kriging espaço-temporal dos resíduos estocásticos para a estação da Maia

Tal como foi apresentado anteriormente, este método fornece-nos previsões para os níveis de NO_2 , isto pode nos ser muito útil em vários aspetos, nomeadamente por razões económicas, pois podemos dispensar de um uso excessivo de aparelhos de medição das concentrações de NO_2 ou para prever medições de um determinado aparelho que, durante um dado período de tempo, não registou os dados (devido a uma avaria ou manutenção, por exemplo).

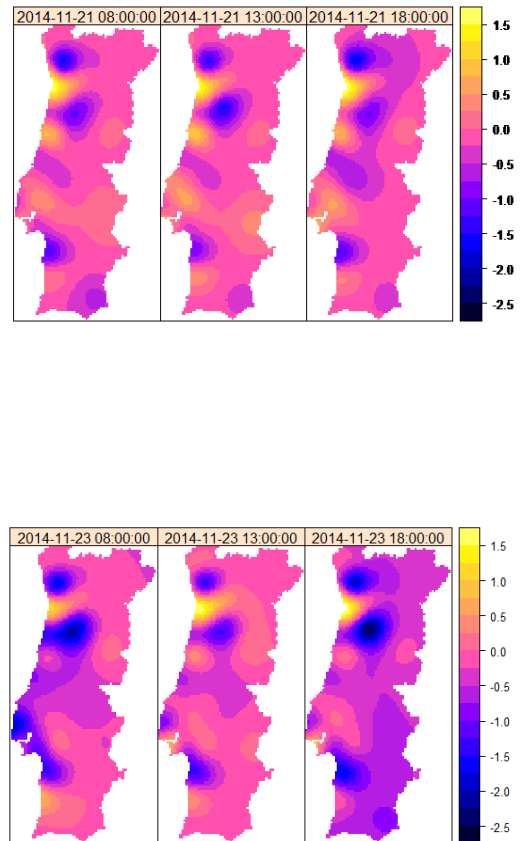


Fig. 12. Mapas obtidos com uso do Kriging espacial nas datas de 2014-11-21 (sexta-feira) e 2014-11-23 (domingo), sendo que quando mais a escala está afastada do 0 maiores são os níveis de NO_2

Vamos agora apresentar uma ilustração do segundo caso. A estação da Maia era uma estação com vários dados em falta e através deste método apresentamos um gráfico com os níveis previstos para esta estação durante uma dada semana, na figura 13.

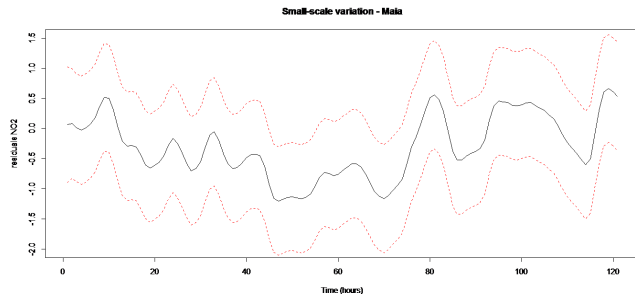


Fig. 13. Estimação dos resíduos para a estação da Maia em 2014-10-06 00:00 (Segunda) até 2014-10-11 00:00 (Sábado). As linhas a vermelho identificam um intervalo de confiança a 95% para a previsão.

11. CONCLUSÕES

Este trabalho mostrou ser uma mais-valia na nossa aprendizagem, com qual conseguimos adquirir vários conceitos e conteúdos novos, assim como nos permitiu a aplicação e manipulação de técnicas já conhecidas.

Mais especificamente, este projeto mostrou-nos que podemos fazer uso das técnicas de kriging para a previsão de dados, sem a necessidade de uma densa rede de controlo, assim como quantificar as dependências estatísticas espaço-temporais pode ser instrutivo para a explicação da grandeza de variação das componentes espaciais e/ou temporais.

Entre as potencialidades das informações fornecidas, a conformidade com os regulamentos sobre as concentrações de NO_2 pode ser verificado, com um certo grau de confiança, mesmo que existam dados omissos. Para além disso, as informações fornecidas pela correlação espaço-temporal podem ser essenciais para otimizar o projeto de monitorização.

Fica a informação de que os níveis de NO_2 não são constantes ao longo do tempo e que predominam nas grandes cidades, sendo estas o principal foco então, para repensar em medidas que melhoravam os níveis consideravelmente.

12. REFERÊNCIAS

- [1] Agência Portuguesa do Ambiente, disponível em <http://www.apambiente.pt/>
- [2] Base de Dados Online sobre a Qualidade do Ar, Agência Portuguesa do Ambiente, disponível em <http://qualar.apambiente.pt/>

- [3] Bivand R., Pebesma E., Gómez-Rubio V. (2008). Applied Spatial Data Analysis with R. Springer, New York. (**gstat** and **spacetime** R libraries).
- [4] Gneiting T., Genton M. G., Guttorp P. (2007). Statistical Methods for Spatio-Temporal Systems. Chapman and Hall, Cambridge.
- [5] Grice S., Stedman J., Kent A., Hobson M., Norris J., Abbott J., Cooke S. (2009). Recent trends and projections of primary NO_2 emissions in europe. Atmospheric Environment 43, 2154-2167.
- [6] Heuvelink G., Griffith D. (2010). Space-time geostatistics for geography: A case study of radiation monitoring across parts of germany. Geographical Analysis 42, 161-179.
- [7] Menezes R., Piairol, H., García-Sóidan, P. and Sousa, I. (2015). Spatial-temporal modellization of the NO_2 concentration data through geostatistical tools. Statistical Methods & Applications, 25, 107-124.