Spatial-Temporal modelization of the NO2 concentration data in mainland Portugal

R.Menezes[1,*], H.Piairo[1], P.Garcia-Soidán[2] and I.Sousa[1]

[1] *Dept. Mathematics and Applications, University of Minho, Guimarães (Portugal); rmenezes@math.uminho.pt, helena.piairo@gmail.com, isousa@math.uminho.pt*
[2] *Dept. Statistics and Operations Research, University of Vigo, Pontevedra (Spain); pgarcia@uvigo.es*
[*] *Corresponding author*

Abstract. *In this work, we propose characterizing the evolution of the NO_2 levels in Portugal, by using geostatistical approaches that deal with both the space and time coordinates and by establishing different factors affecting the concentrations detected. We also discuss appropriate tools for modelling the presence of a trend or a seasonal effect, together with specific approaches for estimation of the space-time variability. The correlation structure of the NO_2 levels can be approximated, enabling to make use of the kriging techniques for prediction, without requiring data from a dense monitoring network. Proceeding in this way, the spatio-temporal patterns of the NO_2 data can be derived, as well as the corresponding deviation errors. Among the potentialities of the information provided, compliance with the regulations about the NO_2 concentrations can be checked.*

Keywords. *NO_2; Geostatistics;Time series analysis; Space-time analysis.*

1 Introduction

The NO_2 is considered a primary pollutant, regarded for the estimation of the air quality index, whose excessive presence may cause significant environmental and health problems. The EU legislation regarding environmental pollution was recently revised to incorporate the latest scientific and technical developments, yielding the publication of the Directive 2008/50/EC on ambient air quality and cleaner air for Europe. The EU First Daughter Directive (99/30/EC) sets an annual mean limit of 40 $\mu g/m^3$, together with an hourly limit of 200 $\mu g/m^3$ that must not be exceeded on more than 18 occasions a year [4].

The current research is focused on providing a procedure to characterize the spatial and temporal evolution of the NO_2 concentration levels through the use of geostatistical approaches. In particular, our proposal is applied to data collected in Portugal from 2004 to 2012 in 88 monitoring sites (see Figure 1).

Figure 1: Location and type of site of the stations in the area studied (Portugal).

At the same time, we aim to deal with the main environmental risks associated to the presence of NO_2, which are signalled for each of the monitored locations. The values observed show dependence on the type of site (background, industrial and traffic) and the environment of the zone (urban, suburban and rural) where each station is placed.

For the aforementioned purposes, we first model the trend of the NO_2 data. Then, geostatistical approaches are applied to the resulting residuals with the aim of characterizing the space-time variability and deriving the predicted values through the kriging tools. The former issue asks for the approximation of spatio-temporal variograms, which are addressed by using various valid models, for both the detrended and deseasonalised data obtained for the period 2004-2012. The results achieved for the different variogram models have been compared through a cross-validation approach, showing that the estimation of the seasonality into the space-time variogram through a sum-metric model presents better values than those derived for the other alternatives, based on the incorporation of the seasonal effect into the trend estimation and/or on the use of different parametric variograms. The subsequent application of the kriging approach enables the reconstruction of the space-time pattern followed by the NO_2 concentrations as well as the associated prediction error.

2 Methodology

Consider a random function $\{Z(\mathbf{s},t) : (\mathbf{s},t) \in \mathbb{R}^d \times \mathbb{R}\}$, indexed in space by $\mathbf{s} \in \mathbb{R}^d$ and in time by $t \in \mathbb{R}$ [3]. Typically, the goal of the space-time analysis is to obtain a prediction of $Z(\cdot)$ at an unsampled point $(\mathbf{s}_0,t_0)$, namely, $Z(\mathbf{s}_0,t_0)$, which requires the specification of the space-time dependence structure. However, sometimes, knowledge of the space-time dependence of Z itself is the aim of a study. For instance, quantified space-time statistical dependencies may be instructive for the comparison or interpretation of the magnitude of variation in the spatial and/or temporal components. In addition, the information provided by the spatio-temporal correlation can be essential to optimize the monitoring design itself [2].

Some random processes can be considered non-stationary in the mean, thus requiring a preliminary characterization of the trend function involved. That is the case of the random variable here studied, the NO_2 concentrations, for which an exploratory analysis shows that is a continuous variable, with an asymmetric distribution and a conditional variance that grows with its mean. So, we decide to use a generalized linear model, whose response variable is gamma distributed, to estimate the expectation of the response.

Denote the residuals by $\varepsilon(\mathbf{s},t) = Z(\mathbf{s},t) - \mu(\mathbf{s},t)$. We assume that the residuals are normally distributed with zero mean. With regard to the dependence structure of the residuals, it can be characterized in terms of the the variogram γ, as $\gamma_{st}(\mathbf{h}_s,h_t) = \frac{1}{2}Var\left[\varepsilon(\mathbf{s}+\mathbf{h}_s,t+h_t) - \varepsilon(\mathbf{s},t)\right]$, where $\mathbf{h}_s,h_t$ takes values in $\mathbb{R}^d \times \mathbb{R}$.

The theoretical variogram γ_{st} satisfies the conditionally negative definiteness condition, namely

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \gamma_{st}\left(\mathbf{s}_i - \mathbf{s}_j, t_i - t_j\right) \leq 0, \quad \forall n \in \mathbb{N}, \quad \text{and for } i = 1,\cdots,n, (\mathbf{s}_i,t_i) \in \mathbb{R}^d \times \mathbb{R}, \;\; a_i \in \mathbb{R}$$

with $\sum_{i=1}^{n} a_i = 0$. This property must also be required from the variogram estimator in order to be valid for its application to prediction through the kriging tools. An example of a valid model is the sum-metric model:

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

where $\alpha \in \mathbb{R}$ and γ_s and γ_t are the corresponding variogram functions in space and time.

As the NO_2 data reveals a cyclical or periodic component, this issue can be dealt with by using a mean function model such as:

$$\eta(\mathbf{s},t) = \alpha + \beta_1 X_1(\mathbf{s},t) + \beta_2 X_2(\mathbf{s},t) + \cdots + \beta_k X_k(\mathbf{s},t) + \beta_{k+1}\cos\left(\frac{2\pi t}{frequency}\right) + \beta_{k+2}\sin\left(\frac{2\pi t}{frequency}\right)$$

or by using a periodic autocorrelation function for the time component in the space-time variogram. In this work, the latter two options for modelling the seasonality are considered as well as different space-time variogram models.

When the trend function and the variogram of the residuals have been specified, the space-time interpolation can be done in the usual way [1]. Prediction in both the spatial and the spatio-temporal settings is addressed through the kriging approaches. In particular, a linear kriging technique is a regression procedure, which yields the best linear unbiased predictor at an unsampled point by computing a weighted linear combination of the surrounding observations, under the basis that the prediction error is minimized.

Acknowledgments. This research was financed by FEDER Founds through "Programa Operacional Factores de Competitividade - COMPETE" and by Portuguese Founds through FCT - "Fundação para a Ciência e a Tecnologia", within project PTDC/MAT/112338/2009 and Est-C/MAT/UI0013/2011.

References

[1] Bivand, R., Pebesma, E. and Gómez-Rubio, V. (2008). *Applied Spatial Data Analysis with R*. Springer,NewYork.

[2] De Gruijter, J., Brus, D., Bierkens, M. and Knotters, M. (2006). *Sampling for Natural Resource Monitoring*. Springer, Germany.

[3] Gneiting, T., Genton, M.G. and Guttorp, P. (2007). *Statistical Methods for Spatio-Temporal Systems*. Chapman and Hall, Cambridge.

[4] WHO (2003). *Health Aspects of Air Pollution with Particulate Matter, Ozone and Nitrogen Dioxide*. World Health Organization, Germany.