Modelling intra- and inter-day variability of NO_2 concentrations in Portugal

Andreia Monteiro[1]*, Raquel Menezes[1], Maria Eduarda Silva[2]

[1] *Department of Mathematics and Applications, University of Minho; andreiaforte50@gmail.com, rmenezes@math.uminho.pt*
[2] *CIDMA & Faculty of Economics, University of Porto; mesilva@fep.up.pt*
**CIDMA & Centre of Mathematics, University of Minho, Campus of Azurem, 4800-058 Guimarães, Portugal*

Abstract. *Nitrogen dioxide (NO_2), is a pollutant that is toxic by inhalation and there is evidence that long-term exposure to this, at high concentrations, has adverse health effects, namely in respiratory and cardiovascular systems. The goal of this study is to characterize the spatial and temporal evolution of NO_2 concentration levels, taking into account that environmental data often incorporate distinct recurring patterns in time, imposed by social habits. We aim at capturing the cyclic nature of these environmental indicators, identifying the intra and inter-day variability. Simultaneously, we aim at modelling the temporal and spatial correlation inherent to this type of data. This study focus on NO_2 hourly data collected in Portugal from October 1 to December 31, 2014. An initial exploratory study suggests that there are two main seasonal effects in the data and identifies variables such as the type of site, environment, and the day of the week as possible explanatory variables. Furthermore, the analysis of the correlation between meteorological parameters, as air temperature, wind speed and relative humidity and NO_2 levels identifies significant negative associations among them. After describing the trend function, 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. This methodology can be used to complement the current design sampling, where there are districts without monitoring stations or with many missing values. Moreover, as meteorological data are available earlier than NO_2 levels, we draw scenarios for NO_2 levels for 2015.*

Keywords. *Geostatistics; Spatio-temporal modelling; Time Series; Environment; NO_2.*

1 Motivating Example

This study analyzes hourly measurements of NO_2 from October 1 to December 31, 2014 corresponding to highest mean NO_2 concentrations along the year. The data set is available in QualAr site. The available data include information about the type of the site where the station is placed (background, industrial or traffic) and the environment of the zone (urban, suburban or rural). From the 49 stations analyzed, 33 are background, 10 traffic and 6 industrial, 29 are located in urban areas, 11 in rural areas and 9 in suburban areas. The mean of the hourly NO_2 concentrations is 20.6 $\mu g/m^3$ and standard deviation 21.9. Hourly mean values for both weekdays and weekends, represented in Figure 1 indicate that NO_2 levels show two daily peaks, one in the morning (8:00 A.M.) and one in the afternoon (6:00 P.M.) which coincide with rushhour traffic, with the second peak being more pronounced than the first. Moreover, the mean

NO_2 concentrations are much lower on weekends than on weekdays, displaying, also, smaller variations on weekends, which reflect reduced levels of vehicular emissions on non-working days. Thus, the two main seasonal effects in the data: intra-day as well as intra-week periodicities, may be, at least partially, explained by characteristics of the station. In fact, the stations located in traffic areas and urban zones present higher values for their NO_2 quartiles as well higher variability. This analysis indicates that the type of site and environment must be considered as explanatory variables.

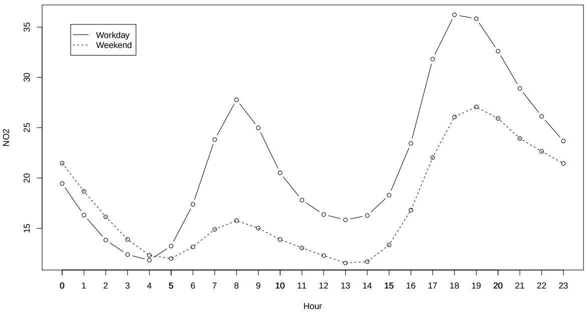

Figure 1: Mean hourly NO_2 concentrations, for workdays and weekends.

Furthermore, it is acknowledge that meteorological conditions influence NO_2 levels. Hourly data from the following meteorological variables were obtained from Weather Underground site: wind speed (km/h); air temperature (^oC) and relative humidity (%). The analysis of the correlation between these meteorological parameters and NO_2 levels identified significant negative associations among them with the strongest correlations occurring at a 9-hour lag with air temperature, a 1-hour lag with wind-speed and a 5-hour lag with relative humidity. Therefore, these meteorological variables are also be considered as explanatory variables.

2 Modelling the large and small scale variation

Consider a random function, $\left\{Z(\mathbf{s},t) : (\mathbf{s},t) \in \mathbb{R}^d \times \mathbb{R}\right\}$ index in space by $s \in \mathbb{R}^d$ and in time by $t \in \mathbb{R}$. The random process $Z(s,t)$ can be decomposed into, $Z(\mathbf{s},t) = \mu(\mathbf{s},t) + \delta(\mathbf{s},t)$, where $\mu(\mathbf{s},t)$ denotes the trend and $\delta(\mathbf{s},t)$ is a zero-mean stationary residual. For NO_2 concentrations, exploratory analysis revealed that it is a continuous variable with an asymmetric distribution. To estimate $\mu(\mathbf{s},t)$ we use a Generalized Linear Model (GLM) and assume that the response variable follows a Gamma distribution with log link. The seasonal effects in the data can be modeled by adding some components of a mixed model [1], leading to:

$$\begin{aligned}\eta(\mathbf{s},t) &= \alpha + \beta_1 X_1(\mathbf{s},t) + \beta_2 X_2(\mathbf{s},t) + \ldots + \beta_k X_k(\mathbf{s},t) + \phi_{1,1}\cos(2\pi t f_j) + \phi_{1,2}\sin(2\pi t f_j) + \ldots + \\ &+ \phi_{l,1}\cos(l\, 2\pi t f_j) + \phi_{l,2}\sin(l\, 2\pi t f_j) \qquad (1)\end{aligned}$$

where X_i are the regressors, function of the explanatory variables, α, β_i, $\phi_{l,1}$, $\phi_{l,2} \in \mathbb{R}$ the regression parameters estimated by the model and f_j is the frequency of the jth harmonic.

After estimation of the large-scale variation (trend), the next step is to analyze the dependence structure of the small-scale variation (residuals) with the aim of characterizing the space-time variability, following a similar approach to that described in [2].

3 Characterization of $\mu(\mathbf{s},t)$ and $\delta(\mathbf{s},t)$

As described in Section 2, we first model the trend of NO_2 data using a GLM. We consider six explanatory variables: type of site, type of environment, if weekend, air temperature (9-hour lag effect), wind speed (1-hour lag effect) and relative humidity (5-hour lag effect). For modeling the seasonal effects in the data set we use the model given in (1).

To estimate the frequency of the data, the periodicity for stations without missing values was calculated and we obtain periodicities equal to 12 and 24 hours. The results of the gamma regression of the hourly NO_2 concentrations show that the values of NO_2 concentrations are greater during the week and in monitoring stations where the environment is urban or suburban and the type of site is traffic. Regarding the meteorological variables, these variables have significant negative associations with NO_2 levels, wind speed has a stronger influence on NO_2 concentrations than humidity and air temperature. The acquired coefficient of determination shows that 41% of the large-scale variation of NO_2 concentrations is explained under this trend model.

After estimation of the large-scale variation, we fit a valid variogram. We start by analyzing the marginal spatial and marginal temporal correlation structures. To decide which non-separable covariance structure to adopt we use a cross-validation study. We select the sum-metric model with a Exponential function for the temporal variogram and Gaussian functions for the spatial and the spatio-temporal components. Under this selection, the fitted final model is represented in Figure 2 (right). According to the results, we conclude that the majority of the total variation is explained by the spatial component. The temporal and spatio-temporal components have a smaller contribution. Furthermore, NO_2 concentrations have a significative spatial correlation up to 40 km and a temporal correlation up to 100 hours (4 days).

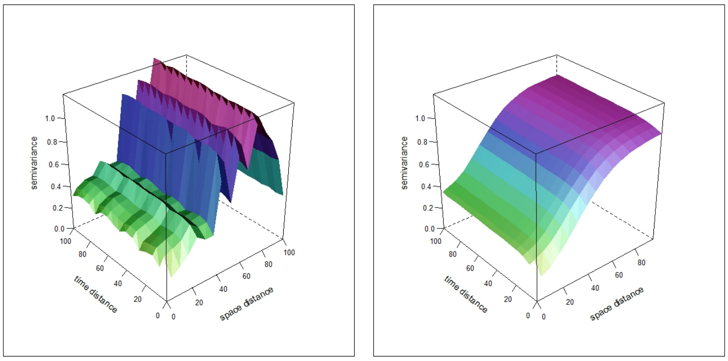

Figure 2: Plots of the experimental estimator (left) and the fitted model (right) for the space-time variogram.

4 Prediction and Forecasting in the Portuguese case

This methodology is important because it can be used to complement the current design sampling, where there are districts without monitoring stations or with many missing values. To illustrate this idea, we obtain the predicted maps for two days, 21 and 23 of November, Friday and Sunday which correspond, respectively, to the days of the week with higher and lower levels of NO_2 and select three times of the day, 8:00 A.M. and 6:00 P.M. which are the daily peaks and 13:00 P.M. a daily minimum. The results are displayed in Figure 3, and confirm that the general spatial pattern of the NO_2 concentrations are not persistent during the day and over the days, achieving higher values on the coast, namely in zones of Braga, Porto, Aveiro, Lisbon and Sines.

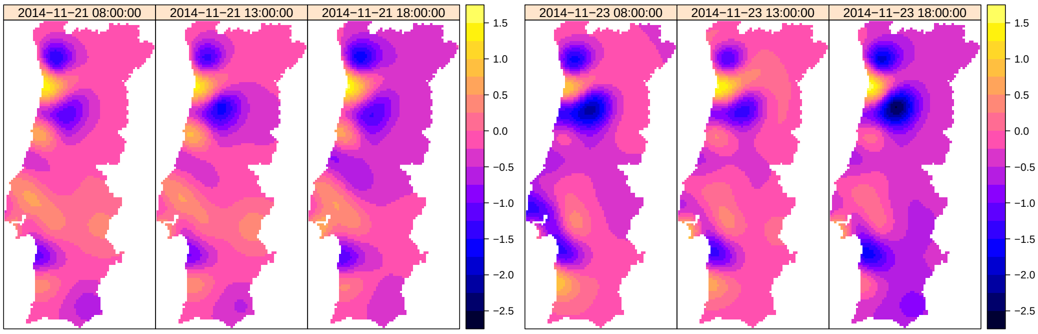

Figure 3: Space kriging maps for 2014-11-21 (Friday) and 2014-11-23 (Sunday).

Knowledge of the space-time dependence allows to predict missing values in a specific station. These missing values may occur occasionally at some time points or when the station becomes inactive. The estimation of small-scale and large-scale variation from 2014-10-06 00:00 to 2014-10-11 00:00 for Vila Nova da Telha, a suburban and background station from Maia county are displayed in Figure 4.

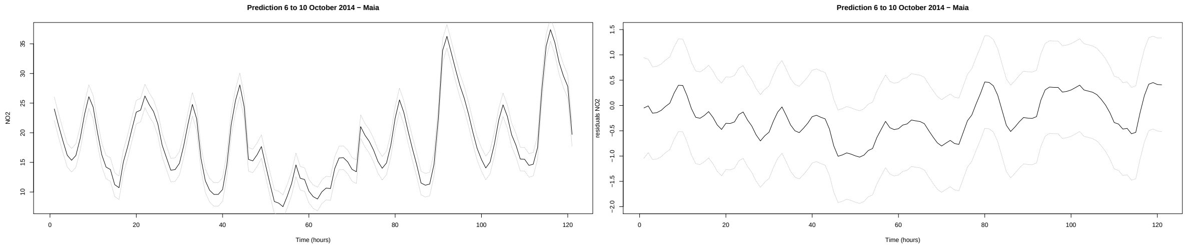

Figure 4: Estimation of the small-scale (left) and large-scale (right) variation of NO_2 concentrations in Maia Station. The dashed lines identify the 95% confidence intervals.

Acknowledgments. The authors acknowledge Foundation FCT (Fundação para a Ciência e Tecnologia) for funding through Individual Scholarship PhD PD/BD/ 105743/2014, Centre of Mathematics of Minho University and Center for Research & Development in Mathematics and Applications of Aveiro University within project UID/MAT/04106/2013.

References

[1] Kyriakidis, P. and Journel, A. (1999). Geostatistical space-time models: A review. *Mathematical Geology* **31**, 651–684.

[2] Menezes, R., Piairo, H., García-Sóidan, P. and Sousa, I. (2015). Spatialtemporal modellization of the NO_2 concentration data through geostatistical tools. *Statistical Methods Applications* **25**, 107–124.