II Galician-Portuguese Meeting of Biometry
Santiago de Compostela, 30 June – 2 July 2016

MODELLING ENVIRONMENTAL MONITORING DATA COMING FROM DIFFERENT SURVEYS

L. Margalho [1], R. Menezes [2] and I. Sousa[2]

[1] Dep. of Physics and Mathematics-Coimbra Institute of Engineering & CMAT-Minho University
[2] Dep. of Mathematics and Applications-Minho University & CMAT-Minho University

ABSTRACT

Environmental monitoring networks are providing large amounts of spatio-temporal data. Air pollution data, as other environmental data, exhibit a spatial and a temporal correlated nature. To improve the accuracy of predictions at unmonitored locations, there is a growing need for models capturing those spatio-temporal correlations.
With this work, we propose a spatio-temporal model for gaussian data collected in a few number of surveys. We assume the spatial correlation structure to be the same in all surveys. In an application of this model to real data, concerning heavy metal concentrations in mosses collected from three surveys occurring between 1992 and 2002 in mainland Portugal, the data set is dense in the spatial dimension but sparse in the temporal one, thus our model-based approach corresponds to a saturated correlation model in the time dimension. A novel interpretation for the space-time covariance function is introduced. A simulation study, aiming to validate the model, provided better results in terms of accuracy with the novel covariance function.
Prediction maps of the observed variable for the most recent survey, and of the interpolation error as a measure of accuracy, are presented.

Keywords: Environmental pollution monitoring; Space-time modelling; Sparse time dimension; Separable covariance structure.

1. INTRODUCTION

Nowadays, due to technology developments and worldwide policies, environmental monitoring networks are providing large amounts of data exhibiting a spatial and a temporal correlated nature, and as a consequence a large number of models and techniques to analyze this sort of data has emerged.
Although in environmental sciences data are typically collected through monitoring stations, it may also be collected through biomonitoring surveys covering extensive areas. Some examples of studies involving moss samples as biomonitors of atmospheric heavy metal deposition are Aboal *et al.* (2006) and Diggle *et al.* (2010) with data from Galicia, northern Spain, Harmens *et al.* (2010) considering data from several countries across Europe, Steinnes *et al.* (2003) and Steinnes *et al.* (2011) concerning Norway data, Zechmeister *et al.* (2008) with data from Austria.
It is common to have studies, such as the ones mentioned before, involving environmental spatio-temporal data containing a dense time dimension but only a sparse spatial one, as a result of the

easiness of gathering data enabled by modern technologies. That is not the case of the biomonitoring data being used in this work, which are related to measurements of heavy metal concentrations made at 146 spatial locations in only 3 surveys.
Our aim is to propose a naive spatio-temporal framework which incorporates into the model both time and space correlations, capable to fit spatio-temporal data containing a reduced number of time observations. Due to this particular characteristic of having few temporal records, and under the hypothesis of separability of the correlation structure, it may be the case that the number of parameters to estimate in the temporal correlation function equals the number of temporal observations, which corresponds to have a saturated correlation model in the time dimension, *i.e.*, a model perfectly reproducing the data.

2. MODEL PROPOSAL

We propose a spatio-temporal model for Gaussian data, collected at location $\mathbf{s}$ and time t,

$$Y(\mathbf{s},t) = \mu(\mathbf{s},t) + \boldsymbol{Z}(\mathbf{s},t) + \boldsymbol{\varepsilon}(\mathbf{s},t) \tag{1}$$

The mean component $\mu(\mathbf{s},t)$, depending on possibly observed covariates $f_i(\mathbf{s},t)$, will be considered as

$$\mu(\mathbf{s},t) = \sum_{i=1}^{p} \beta_i f_i(\mathbf{s},t) \tag{2}$$

where $E[Y(\mathbf{s},t)] = \mu(\mathbf{s},t)$. The non-observed spatio-temporal process $\boldsymbol{Z}(\mathbf{s},t)$ is defined by

$$\boldsymbol{Z} \sim MVN\left(\mathbf{0}, \Sigma\right) \tag{3}$$

and $\boldsymbol{\varepsilon}$ represents Gaussian space-time measurement errors,

$$\boldsymbol{\varepsilon} \sim MVN\big(\mathbf{0}, \tau^2 I_{NT}\big) \tag{4}$$

We assume a separable covariance structure,

$$\begin{aligned}
\Sigma^{i,j,k,l} &= \mathrm{Cov}_{ST}\big[Z(\mathbf{s}_i,t_k), Z(\mathbf{s}_j,t_l)\big] \\
&= \mathrm{Cov}_S\left(\|\mathbf{s}_i - \mathbf{s}_j\|\right) \times \mathrm{Cov}_T\left(|t_k - t_l|\right) \\
&= \mathrm{Cov}_S\left(h_S\right) \times \mathrm{Cov}_T\left(h_T\right)
\end{aligned}$$

and propose two interpretations for Σ

- $\Sigma(h_S, h_T) = \sigma^2_{total}\, R_S(h_S) \otimes R_T(h_T)$ (Rodriguez-Iturbe & Mejia (1974))
- $\Sigma(h_S, h_T) = \sigma^2_S R_S(h_S) \otimes \sigma^2_T R_T(h_T)$

($\otimes$ represents the Kronecker product of matrices). In the application of the model, the covariates included in the model are indexed in space (the sampling locations intensity) and indexed in time (the specific contribution of a given survey).

3. SIMULATION STUDY

For a model validation purpose, a simulation study was conducted. A set of 50 randomly chosen space locations was considered in the square $[0,1]^2$. In order to have a region with more intensified sampling density, mimicking the behavior of the real data set used in the application, 15 of those locations belong to the square $[0.45, 0.55]^2$. Observations are assumed to be collected at 3 different moments, according to an AR(2) model. Using the absolute value of the coefficient of variation as an accuracy measure, simulations with $\Sigma(h_S, h_T) = \sigma^2_S R_S(h_S) \otimes \sigma^2_T R_T(h_T)$ provided better results.

4. CASE STUDY. HEAVY METAL CONCENTRATION IN MOSSES

Mosses are widely used as biomonitors of atmospheric heavy metal deposition. In Europe, they have been used since 1990, with the aim of map spatial and temporal patterns of accumulation in ecosystems, after the establishment of the international mapping project *Atmospheric Heavy Metal Deposition in Europe*, which is surveying the atmospheric deposition of heavy metals using moss species as biomonitors, with the aim of investigate the existence of correlations between heavy metal concentrations in mosses. Portugal was one of the participating countries in the mentioned project, performing surveys every 5 years since its beginning in 1990. Moss samples of species *Hypnum cupressiforme* and *Scleropodium touretti* were collected in three nationwide surveys across mainland Portugal, referred to as the 1992, 1996 and 2002 surveys.

The model described previously assumes that the hidden process $\boldsymbol{Z}(\mathbf{s},t)$ and the measurement error $\boldsymbol{\varepsilon}(\mathbf{s},t)$ are Gaussian. It is well known that for a non-observed location $\mathbf{s}_0$ and a time t_0, the joint distribution of $Y(\mathbf{s}_0,t_0)$ and $Y(\mathbf{s},t)$ is

$$\begin{bmatrix} Y(\mathbf{s}_0,t_0) \\ \mathbf{Y}(\mathbf{s},t) \end{bmatrix} \sim \text{MVN}\left(\begin{bmatrix} \mu(\mathbf{s}_0,t_0) \\ \mu(\mathbf{s},t) \end{bmatrix}, \begin{bmatrix} C_{0,0} & \mathbf{c}_0^T \\ \mathbf{c}_0 & \mathbf{C}_Y \end{bmatrix} \right) \tag{5}$$

where $\mu(\mathbf{s}_0,t_0)$ and $\mu(\mathbf{s},t)$ are defined by (2), $C_{0,0} = \text{Var}(Y(\mathbf{s}_0,t_0))$, $\mathbf{c}_0 = \text{Cov}(Y(\mathbf{s}_0,t_0), Y(\mathbf{s},t))$, and $\mathbf{C}_Y = \Sigma + \tau^2 I$.

Under the assumption (5), the predicted value at an unsampled location $Y^*(\mathbf{s}_0,t_0)$ is given (Cressie and Wikle (2011)) by

$$Y^*(\mathbf{s}_0,t_0) = \text{E}\left[Y^*(\mathbf{s}_0,t_0)|\mathbf{Y}(\mathbf{s},t)\right] = \mu(\mathbf{s}_0,t_0) + \mathbf{c}_0^T\mathbf{C}_Y^{-1}\left(\mathbf{Y}(\mathbf{s},t) - \mu(\mathbf{s},t)\right) \tag{6}$$

and the variance of the prediction is

$$\sigma^2(\mathbf{s}_0,t_0) = \text{E}\left(Y(\mathbf{s}_0,t_0) - Y^*(\mathbf{s}_0,t_0)\right)^2 = C_{0,0} - \mathbf{c}_0^T\mathbf{C}_Y^{-1}\mathbf{c}_0 \tag{7}$$

The estimates of the model parameters are in Table 1. The computation of the standard errors was made via Monte-Carlo simulation.

Param.	β_0	β_1	β_2	β_3	ρ_{12}	ρ_{13}	ρ_{23}	σ_S^2	σ_T^2	τ^2	ϕ
Estim.	7.45	0.01	0.15	-0.24	0.97	0.91	0.96	0.98	1.48	1.02	58624.08
St. Error	0.04	<0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.02	0.01	1470.20

Table 1: Model parameter estimates with standard errors

Figure 1 shows the predicted Mn transformed concentration map for the most recent survey and the corresponding interpolation error map, over mainland Portugal.

5. CONCLUSIONS

The new model proposal provides better results in terms of interpolation prediction error, than the ones obtained in Margalho *et al.* (2014) with the same data set using a different geostatistical model.

References

[1] Aboal, J.; Real, C.; Fernández, J. and Carballeira, A. (2006). Mapping the results of extensive surveys: The case of atmospheric biomonitoring and terrestrial mosses, *Science of The Total Environment*, **356**, 256–274.

[2] Cressie, N. and Wikle, C. (2011). *Statistics for spatio-temporal data*. John Wiley & Sons Inc, New Jersey.

[3] Diggle, P.; Menezes, R. and Su, T. (2010). Geostatistical inference under preferential sampling, *Applied Statistics*, **59**, 2, 191–232.

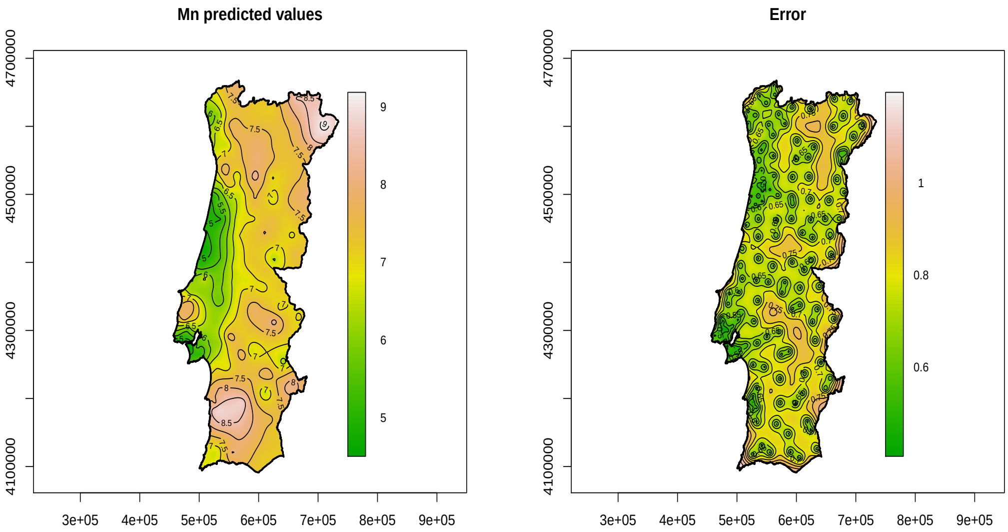

Figure 1: Prediction map of the Box-Cox transformed Mn data (left) and interpolation error map (right).

[4] Harmensm H.; Norris, D.; Steinnes, E.; Kubin, E.; Piispanen, J.; Alber, R.; Aleksiayenak, Y.; Blum, O.; Coskun, M.; Dam, M.; De Temmerman, L.; Fernández, J.; Frolova, M.; Frontasyeva, M.; González-Miqueo, L.; Grodzińska, K.; Jeran, Z.; Korzekwa, S.; Krmar, M.; Kubin, E.; Kvietkus, K.; Leblond, S.; Liiv, S.; Magnússon, S.. Maňkovská, B.; Pesch, R.; Rüling, A.; Santamaria, J.; Schöder, W.; Spiric, Z.; Suchara, I.; Thöni, L.; Urumov, L.; Yurukova, L. and Zechmeister, H. (2010). Mosses as biomonitors of atmospheric heavy metal deposition: Spatial patterns and temporal trends in Europe, *Environmental Pollution*, **158**, 3144–3156.

[5] Margalho, L.; Menezes, R. and Sousa, I. (2014). Assessing interpolation error for space-time monitoring data, *Journal of Stochastic Environmental Research and Risk Assessment*, **28**, 5, 1307–1321.

[6] Rodriguez-Iturbe, I. and Mejia, J. (1974). The design of networks in time and space, *Water Resources Research*, **10**, 4, 713–728.

[7] Steinnes, E.; Berg, T. and Sjøbakk, T. (2003). Temporal and spatial trends in Hg deposition monitored by moss analysis, *The Science of the Total Environment*, **304**, 215–219.

[8] Steinnes, E.; Berg, T. and Uggerud, H. (2011). Three decades of atmospheric metal deposition in Norway as evident from analysis of moss samples, *The Science of the Total Environment*, **412-413**, 351–358.

[9] Zechmeister, H.; Hohenwallner, D.; Hanus-Illnar, A.; Hagendorfer, H.; Roder, I. and Riss, A. (2008). Temporal patterns of metal deposition at various scales during the last two decades, *Atmospheric Environment*, **42**, 1301–1309.