MODELLING ENVIRONMENTAL MONITORING DATA COMING FROM DIFFERENT SURVEYS

Luís Margalho [1]
Coimbra Polytechnic – ISEC, Coimbra, Portugal
e-mail: *lmelo@isec.pt*

Raquel Menezes
Dep. Mathematics and Applications, Centre of Molecular and Environmental Biology,
University of Minho, Braga, Portugal

Inês Sousa
Dep. Mathematics and Applications, Centre of Molecular and Environmental Biology,
University of Minho, Braga, Portugal

With this work we propose a spatio-temporal model for Gaussian data collected in a small number of surveys. We assume the spatial correlation structure to be the same in all surveys. In the application concerning heavy metal concentrations in mosses, the data set is dense in the spatial dimension but sparse in the temporal one, thus our model-based approach corresponds to a correlation model depending on survey orders. One advantage of this approach is its computational simplicity. An interpretation for the space-time covariance function, decomposing the overall variance of the process as the product of the spatial component variance by the temporal component variance, is introduced. A simulation study, aiming to validate the model, provided better results in terms of accuracy with the novel covariance function. Maps of predicted heavy metal concentrations and of interpolation error, for the most recent survey, are presented.

Data of this kind is recurrent in environmental sciences, which is why we argue that this will be a practical tool to be used very often.

Key words: Environmental pollution monitoring, Separable covariance structure, Space-time modelling, Sparse time dimension.

1. Introduction

Nowadays, due to technological 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. Some references on the subject of spatio-temporal modelling include the following: Kyriakidis and Journel (1999), reviewing stochastic models involving the extension of spatial analysis tools to include the time dimension; de Cesare, Myers and Posa (2001), with a discussion on some classes of models and the introduction of the so-called product-sum model; Sahu and Mardia (2005), with a review on methods for modelling spatio-temporal point referenced data; Sherman (2011), with a

---

brief survey of several types of spatio-temporal covariance models; Huang, Martinez, Mateu and Montes (2007), comparing four spatio-temporal models covering the situations of separability and non-separability; and Cameletti, Ignaccolo and Bande (2011), with a comparison of six alternative spatio-temporal models belonging to the class of Bayesian hierarchical models and providing criteria to choose among them.

In environmental sciences, data are typically collected through monitoring stations. Shaddick and Wakefield (2002) use daily pollution data collected at eight monitoring sites within London, measuring the pollutants particulate matter PM_{10}, carbon monoxide, nitrogen oxide and sulphur dioxide over the period from 1994 to 1997. Fanshawe et al. (2008) models PM_4 levels in the United Kingdom, using data routinely collected from 20 monitoring stations between 1961 and 1992. Lindstrom et al. (2011) consider data from 20 sites in and around Los Angeles from an Air Quality System network measuring ambient concentrations of air pollutants collected at a 2-week timescale from the beginning of 1999 until September 2009. Cameletti et al. (2011) models PM_{10} concentration in the Piemonte region, analyzing daily data collected during 182 day from 24 sites.

However, data 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, Real, Fernández and Carballeira (2006) and Diggle, Menezes and Su (2010) with data from Galicia, northern Spain; Harmens et al. (2010) with data from several countries across Europe; Steinnes, Berg and Sjobakk (2003) and Steinnes, Berg and Uggerud (2011) with data from Norway; Zechmeister et al. (2008) with data from Austria.

It is common to have studies involving environmental spatio-temporal data containing a dense time dimension, but only a sparse spatial one, which is a result of the ease of data gathering enabled by modern technologies. However, 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 spatio-temporal framework which incorporates into the model both time and space correlations, capable of fitting 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 one has a small number of parameters to estimate in the temporal correlation function.

Margalho, Menezes and Sousa (2014), using the same data set, proposed an extension of an existing spatio-temporal prediction model accommodating the existence of explanatory variables to the process under observation. This extension aggregates temporal data at each location by considering the mean value, thus not taking into account the specificity of each particular survey. Here, inversely, we aim to include the temporal contribution of any particular survey in the spatial prediction process.

This work is organized as follows. Section 2 contains an exploratory analysis of the data set motivating this study. Details on the proposed spatio-temporal model are in Section 3. A simulation study, for model validation purposes, is conducted in Section 4. In Section 5 an application of the model, with the aim of predicting heavy metal concentration at non-sampled locations over mainland Portugal, is given, along with a comparison of results obtained by Margalho et al. (2014). Some conclusions are formulated in Section 6.

2. 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 mapping spatial and temporal patterns of accumulation in ecosystems (Holy et al., 2009). Later, the international mapping project *Atmospheric Heavy Metal Deposition in Europe* was established, which surveys the atmospheric deposition of heavy metals using moss species as biomonitors, with the aim of investigating the existence of correlations between heavy metal concentrations in mosses.

Uyar, Ören, Yildirim and Ince (2007) mention several advantages of using moss samples: the vast geographical distribution and the abundant growth in various natural habitats; the non-existence of epidermis or cuticle enabling their cell walls to be easily penetrable for metal ions; the fact that mosses obtain minerals mainly from air and precipitation as they have no root systems; the effect and contamination of soil by heavy metals is negligible for most mosses, so they show the concentrations of the most metals correlated to the amount of atmospheric deposition. The simple procedure of sampling and of cheap chemical analysis also makes mosses especially suitable organisms for the purpose of monitoring.

Some examples of studies related to the use of moss samples are in Diggle et al. (2010), where moss data obtained from Galicia, northern Spain, were used in the context of analyzing the effect of preferential sampling on prediction, but not taking into account the temporal representativeness of data; Harmens et al. (2010) who considered data from several countries across Europe; Steinnes et al. (2003) and Steinnes et al. (2011) with data from Norway; Zechmeister et al. (2008) with data from Austria.

Portugal was one of the participating countries in the *Atmospheric Heavy Metal Deposition in Europe* project, performing surveys every 5 years since its beginning in 1990. Moss samples of the species *Hypnum cupressiforme* and *Scleropodium touretti* were collected in three nationwide surveys across mainland Portugal, referred to as the 1992, 1996 and 2002 surveys.

Although the number of sampling locations was not the same throughout the surveys, 146 of those were common to the three surveys. Sampling locations were selected in order to be representative of background areas, collected in a 30×30 km^2 grid, although near large urban or industrial areas the sampling design was intensified, where the grid was adjusted to 10×10 km^2. Chemical analysis yielded concentration measurements of cadmium, chromium, copper, iron, manganese, nickel, lead and zinc. Further details on the sampling and analysis procedure can be found in Figueira, Sérgio and Sousa (2002) or Martins, Figueira, Sousa and Sérgio (2012).

As stated before, the sampling design was not uniform throughout the whole region under study, in contrast to the examples presented in Boquete, Fernández, Aboal and Carballeira (2011) and Steinnes et al. (2011). There are cases where sampling is intensified in subregions where a large gradient of the measured variable is expected (Diggle et al., 2010). In fact, one possible aim of an air monitoring network might be to identify larger values of pollution, so sampling locations could be selected based on high values rather than randomly (Guttorp and Loperfido, 2008).

Specifically, in the Portuguese case, sampling locations were in a larger number at regions with high industrial or urban density. This motivates the use of a function of the spatial sampling intensity as an explanatory variable when modelling air pollution data from Portugal.

The issue of using covariates in spatio-temporal models and testing for its significance is addressed in Díaz-Avalos, Juan and Mateu (2014) in the context of point processes. Particularly in the

applications that follow, the inclusion of sampling intensity as an explanatory variable will be obtained by a spatial kernel smoothing of the sampling locations density. As should be expected, the larger values of this smooth function, ranging from 3.14 to 27.48 and with mean value equal to 11.62, occur in regions with higher sampling intensity where the industrial or urban areas are located.

2.1 Exploratory data analysis

Among the various heavy metals found in nature, manganese (Mn) is one of the most abundant and widely distributed, being found in water, rocks and soil (Pinsino, Roccheri and Matranga, 2012). The presence of this metal has not only a natural cause as a result of mineral weathering and atmospheric deposition, but can also have anthropogenic origins, such as municipal wastewater discharges, mining and mineral processing, combustion of fossil fuels or emissions from the combustion of fuel additives (Howe, Malcolm and Dobson, 2004). Although essential for humans, at higher levels of contamination manganese can become toxic. Several studies relate chronic manganese excess with disturbances in the central nervous system, with symptoms resembling those of Parkinson's disease (Perl and Olanow, 2007; Rocks and Levy, 2008).

Table 1 summarizes a descriptive analysis of Mn concentration data for the three considered surveys. Values are expressed in units of mg(metal)/kg(moss). Mn concentrations in the original scale ranged from a minimum of 4.0 mg/kg, recorded in the second survey, and a maximum value of 970.0 mg/kg, which occurred in the first survey. One can observe that the mean value has increased from around 160 mg/kg in the first survey to almost 180 mg/kg in the second survey, and decreased to less than 150 mg/kg in the third survey. A similar behavior occurred with the median, and one can also observe that the variability of Mn concentration values was always decreasing. The maximum recorded value for each survey should be noted, particularly in the first one when the maximum was about six times the mean value. This difference between the mean and the maximum value was not so marked in the second and third surveys, although it was also pronounced. All surveys show the presence of outlier values, with the first survey being the one with the strongest right asymmetric distribution, where over time the asymmetry becomes not so marked. A Box-Cox transformation of data was carried out in order to reduce the effect that these values could cause in the estimation process. The particular value of the Box-Cox transformation parameter was 0.15. After performing the transformation of data, the asymmetry fades away and the resulting distribution behaves like a Gaussian one (p-values in a Shapiro-Wilk test for normality of 0.435, 0.199 and 0.0798 for the first, second and third surveys, respectively).

2.2 Spatial analysis restricted to 2002

To better understand the spatial behaviour of data, transformed concentration values were predicted by means of Ordinary Kriging over a fine 300×100 grid covering mainland Portugal, allowing grid points at a distance of around 2 km from each other. Interpolation errors, defined as the squared root of the interpolation variance, were computed. The two leftmost columns in Table 2 show a summary of the predicted Mn transformed concentrations for the 2002 survey and the associated interpolation errors.

A spatial analysis was also performed by considering the sampling location intensity as an explanatory variable. Here, the expected value of the process under observation is a function of the observed location. This way, the interpolation procedure corresponds, according to Cressie

Table 1. Data Summary of Mn concentration (observed and Box-Cox transformed values).

| | Survey | | | | | |
| | 1992 | | 1996 | | 2002 | |
Concentration	Observ.	Transf.	Observ.	Transf.	Observ.	Transf.
Min	16.00	3.39	4.03	1.54	23.14	3.96
Median	123.50	6.90	149.18	7.28	123.20	6.89
Max	970.00	11.63	685.55	10.73	503.11	9.97
Mean	161.62	6.88	178.62	7.13	147.12	6.85
St. dev.	147.97	1.69	136.01	1.65	99.28	1.41

Table 2. Predicted Mn transformed concentration for the 2002 survey and associated interpolation error, obtained by Ordinary Kriging (left) and Universal Kriging (right).

| | OK | | UK | |
	Predicted	Error	Predicted	Error
Min	4.88	1.04	4.39	1.03
Median	6.99	1.26	7.57	1.25
Max	8.59	1.47	8.58	1.37
Mean	6.97	1.26	7.40	1.24
St. dev.	0.63	0.12	0.63	0.09

(1993), to the Universal Kriging.

Predicted transformed concentration at unobserved locations, placed over the same grid as in the previous application, and the associated interpolation error are summarized in the two rightmost columns of Table 2. It is worthwhile noticing that the inclusion of the mentioned covariate in the interpolation procedure has the effect of reducing the amount of interpolation error.

Maps of predicted values and of the associated interpolation error were created, although not presented here for space reasons. It is known that contamination by Mn is mainly associated with the soil typology (Figueira et al., 2002). In accordance with this, higher predicted values were found in eastern and south-western Portugal, i.e. regions with less forestry and hence more soil erosion.

Comparison of spatial prediction results

By comparing the prediction results presented in Table 2 for the most recent survey, one can register that the predicted Mn values are of larger magnitude when considering the covariate. This means that by incorporating this information in the prediction spatial model, the effect of ignoring a sampling design not evenly representative of the area under observation was mitigated. In the case of Portuguese moss data, where areas with more sampled locations correspond to lower values of Mn, this approach allows one to solve the issue of underestimating Mn concentrations.

3. The model

Aiming to consider data available from different surveys, we propose a spatio-temporal model for Gaussian data collected at location $\mathbf{s} \in \mathbb{R}^2$ and time $t \in \mathbb{N}$, defined as

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

Considering N locations observed at T surveys, the mean component $\mu(\mathbf{s},t)$, depending on possibly observed covariates $f(\mathbf{s},t)$, indexed in space or in time, will be defined as

$$\mu(\mathbf{s}_i,t_k) = \sum_{j=1}^{p} \beta_j f_j(\mathbf{s}_i,t_k), \quad i = 1,\ldots,N, \quad k = 1,\ldots,T. \tag{2}$$

Under matrix notation, one has

$$\boldsymbol{Y} = \boldsymbol{\mu} + \boldsymbol{Z} + \boldsymbol{\varepsilon},$$

with

$$\boldsymbol{\mu} = M \cdot \boldsymbol{\beta},$$

the product of the $(NT) \times (p+1)$ design matrix M (where p is the number of considered covariates) and $\boldsymbol{\beta}$, the vector of regression coefficients in the mean component of (2). The columns of M, with exception of the first, represent each covariate $f_j(\mathbf{s},t)$ in (2).

The unobserved spatio-temporal process $Z(\mathbf{s},t)$ is such that

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

and $\varepsilon(\mathbf{s},t)$ represents Gaussian space-time measurements errors,

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

with I_{NT} the identity matrix of order NT.

In the spatio-temporal process (3), Σ is a $(NT) \times (NT)$ symmetric matrix, which can be interpreted as a $T \times T$ block matrix such that, for $k, l = 1,\ldots,T$, the element on line i and column j is

$$\Sigma^{i,j,k,l} = \text{Cov}_{ST}\big[Z(\mathbf{s}_i,t_k), Z(\mathbf{s}_j,t_l)\big], \quad i, j = 1,\ldots,N. \tag{5}$$

Furthermore, the proposed model assumes second order stationarity and an isotropic and separable covariance structure, so we define purely spatial and purely temporal covariance functions Cov_S and Cov_T (Sherman, 2011) resulting in

$$\begin{aligned}
\Sigma^{i,j,k,l} &= \text{Cov}_S\big(\|\mathbf{s}_i - \mathbf{s}_j\|\big) \times \text{Cov}_T\big(|t_k - t_l|\big) \\
&= \text{Cov}_S\big(h_S\big) \times \text{Cov}_T\big(h_T\big).
\end{aligned}$$

The reason to propose this decomposition of the covariance structure is that one could consider, for each sampling location, an autoregressive process (along time) and find, simultaneously, spatial correlation among locations. In case the real data set considered here had more time observations, this would be expected to happen.

In spite of the difficulty of identifiability that this decomposition brings, it seems an interesting point of comparison in terms of accuracy of prediction.

The notion of separability, according to Bruno, Guttorp, Sampson and Cocchi (2003), satisfies the two following statements:

(i) the spatial covariance function is constant in time, so that

$$\text{Cov}\big[Z(\mathbf{s}_i, t_k), Z(\mathbf{s}_j, t_k)\big] = \text{Cov}\big[Z(\mathbf{s}_i, t_l), Z(\mathbf{s}_j, t_l)\big],$$

for all (t_k, t_l), $\mathbf{s}_i$, $\mathbf{s}_j \in \mathbb{R}^2$, and

(ii) the temporal covariance function is the same at all monitoring locations, regardless of the displacement between locations,

$$\text{Cov}\big[Z(\mathbf{s}_i, t_k), Z(\mathbf{s}_i, t_l)\big] = \text{Cov}\big[Z(\mathbf{s}_j, t_k), Z(\mathbf{s}_j, t_l)\big],$$

for all $(\mathbf{s}_i, \mathbf{s}_j)$, $t_k, t_l = 1, \ldots, T$.

3.1 Inference on model parameters

Under the assumption of second order stationarity, we allow two different interpretations for the covariance function. The most common one (e.g. Rodriguez-Iturbe and Mejia, 1974; Sherman, 2011) considers a scale parameter σ_{total}^2 representing the overall variance, being the covariance matrix given by

$$\Sigma = \sigma_{total}^2 R_S \otimes R_T, \tag{6}$$

where $\otimes$ is the Kronecker product of matrices, and R_S and R_T are, respectively, the $N \times N$ spatial correlation matrix and the $T \times T$ temporal correlation matrix. As an alternative we propose to take into account different scale parameters for the spatial and the temporal components,

$$\Sigma = \sigma_S^2 R_S \otimes \sigma_T^2 R_T, \tag{7}$$

where σ_S^2 is the spatial variance and σ_T^2 is the temporal variance. Basically, this corresponds to performing a reparametrization of the overall variance and then decomposing it into the product of a spatial variance and a temporal one. To the best of our knowledge, such a decomposition of the overall variance has not yet been proposed.

If only a reduced number of time observations are available, R_T is a $T \times T$ matrix whose elements, denoted by $\rho_T(|t_k - t_l|)$, $k, l = 1, \ldots, T$, measure the association between observations at time t_k and t_l.

Denoting by $\mathbf{C}_Y$ the covariance matrix of $Y(\mathbf{s}, t)$, which depends on the parameters τ^2 and $\boldsymbol{\theta} = (\sigma_{total}^2, \phi_S, \rho_T(|t_k - t_l|))$ or $\boldsymbol{\theta} = (\sigma_S^2, \sigma_T^2, \phi_S, \rho_T(|t_k - t_l|))$, we have

$$\mathbf{C}_Y(\tau^2, \boldsymbol{\theta}) = \Sigma(\boldsymbol{\theta}) + \tau^2 I_{NT}.$$

Following standard results from Gaussian distribution theory, the log-likelihood function is

$$\log L(\tau^2, \boldsymbol{\theta}) = -\frac{NT}{2}\log(2\pi) - \frac{1}{2}\log\Big(det\big(\mathbf{C}_Y(\tau^2, \boldsymbol{\theta})\big)\Big) - \frac{1}{2}\Big(Y - M\boldsymbol{\beta}\Big)^t \mathbf{C}_Y^{-1}(\tau^2, \boldsymbol{\theta})\Big(Y - M\boldsymbol{\beta}\Big), \tag{8}$$

where the profile Likelihood Estimator for $\boldsymbol{\beta}$ is given by

$$\widehat{\boldsymbol{\beta}} = \Big(M^t \mathbf{C}_Y^{-1}(\tau^2, \boldsymbol{\theta})M\Big)^{-1} M^t \mathbf{C}_Y^{-1}(\tau^2, \boldsymbol{\theta})\,Y. \tag{9}$$

The log-likelihood function (8) depends on $3 + (T(T-1)/2 - 1) + (p+1)$ parameters, if one considers (6), or $4 + (T(T-1)/2 - 1) + (p+1)$ in the case of (7). The advantage resulting from the knowledge

of an analytic estimator for the parameter vector $\boldsymbol{\beta}$ is a reduction on the computational effort required for the parameter estimation.

The difference in the number of parameters in the log-likelihood function results from the reparametrization of the overall variance σ_{total}^2 in (6) by $(\sigma_S \cdot \sigma_T)^2$. Hence, one should expect to find an estimate $\widehat{\sigma}_{total}^2$ close to the product $\widehat{\sigma}_S^2 \cdot \widehat{\sigma}_T^2$.

3.2   The theoretical semi-variogram

The space-time semi-variogram $\gamma(h_S, h_T)$ of the unobserved spatio-temporal process $Z(\mathbf{s}, t)$, where $h_S = \|\mathbf{s}_i - \mathbf{s}_j\|$ and $h_T = |t_k - t_l|$ are the spatial distances between locations and temporal distances between surveys orders, respectively, is

$$\gamma(h_S, h_T) = \frac{1}{2}\mathrm{Var}\big[Z(\mathbf{s}, t) - Z(\mathbf{s} + h_S, t + h_T)\big].$$

The relationship between $\gamma(h_S, h_T)$ and the space-time covariance function

$$\gamma(h_S, h_T) = \mathrm{Cov}_{ST}(\mathbf{0}, 0) - \mathrm{Cov}_{ST}(h_S, h_T)$$

under second order stationarity conditions is well known (see, e.g., Cressie and Wikle, 2011; Sherman, 2011).

If the covariance structure in (6) is considered, then $\mathrm{Cov}_{ST}(\mathbf{0}, 0) = \sigma_{total}^2$ and $\mathrm{Cov}_{ST}(h_S, h_t) = \sigma_{total}^2 \cdot \rho_S(h_S) \cdot \rho_T(h_T)$, where $\rho_S(h_S)$ and $\rho_T(h_T)$ represent the respective elements of the spatial and temporal correlation matrices.

Equivalently, if (7) is assumed, one has $\mathrm{Cov}_{ST}(\mathbf{0}, 0) = \sigma_S^2 \sigma_T^2$ and $\mathrm{Cov}_{ST}(h_S, h_t) = \sigma_S^2 \sigma_T^2 \cdot \rho_S(h_S) \cdot \rho_T(h_T)$. These two versions of the semi-variogram are analogous, only differing in the reparametrization of the variance. The resulting advantage of considering this decomposition comes from the capacity to assign different magnitudes to the spatial and the temporal variability.

3.3   Prediction at unsampled locations

The model described in (1) assumes that the hidden process $Z(\mathbf{s}, t)$ and the measurement error $\varepsilon(\mathbf{s}, t)$ are Gaussian ((3) and (4)). It is well known (see, e.g., Cressie and Wikle, 2011), that for an unobserved location $\mathbf{s}_0$ and a time t_0, the joint distribution of $Y(\mathbf{s}_0, t_0)$ and $\boldsymbol{Y}$ is

$$\left[\begin{array}{c} Y(\mathbf{s}_0, t_0) \\ \boldsymbol{Y} \end{array}\right] \sim \mathrm{MVN}\left(\left[\begin{array}{c} \mu(\mathbf{s}_0, t_0) \\ \boldsymbol{\mu} \end{array}\right], \left[\begin{array}{cc} C_{0,0} & \mathbf{c}_0^T \\ \mathbf{c}_0 & \mathbf{C}_Y(\tau^2, \boldsymbol{\theta}) \end{array}\right]\right), \tag{10}$$

where $\mu(\mathbf{s}_0, t_0) = \sum_{i=1}^{p} \widehat{\beta}_i f_i(\mathbf{s}_0, t_0)$, $\boldsymbol{\mu}$ is defined by (2), $C_{0,0} = \mathrm{Var}\,[Y(\mathbf{s}_0, t_0)]$, $\mathbf{c}_0 = \mathrm{Cov}\,[Y(\mathbf{s}_0, t_0), \boldsymbol{Y}]$, and $\mathbf{C}_Y(\tau^2, \boldsymbol{\theta}) = \Sigma(\boldsymbol{\theta}) + \tau^2 I_{NT}$ with $\Sigma(\boldsymbol{\theta})$ as in (5).

The predicted value $\widehat{Y}(\mathbf{s}_0, t_0)$ at an unsampled location can be obtained from (10), and is given by

$$\widehat{Y}(\mathbf{s}_0, t_0) = \mathrm{E}\big[Y(\mathbf{s}_0, t_0)|\boldsymbol{Y}\big] = \mu(\mathbf{s}_0, t_0) + \mathbf{c}_0^T\, \mathbf{C}_Y^{-1}(\tau^2, \boldsymbol{\theta})\,(\boldsymbol{Y} - \boldsymbol{\mu}). \tag{11}$$

The variance of the prediction, also resulting from (10), is

$$\sigma^2(\mathbf{s}_0, t_0) = \mathrm{E}\left[Y(\mathbf{s}_0, t_0) - \widehat{Y}(\mathbf{s}_0, t_0)\right]^2 = C_{0,0} - \mathbf{c}_0^T\, \mathbf{C}_Y^{-1}(\tau^2, \boldsymbol{\theta})\, \mathbf{c}_0. \tag{12}$$

In particular, if an exponential spatial model is assumed, $\mathbf{c}_0$ is the concatenation of T vectors $c_1, c_2, \ldots, c_T$, each with dimension $N \times 1$, where

$$c_i = \widehat{\sigma}_{total}^2 \left(\exp\left(-\widehat{\phi}_S^{-1}\|\mathbf{s}_0 - \mathbf{s}_1\|\right)\widehat{\rho}_T(|t_i - t_0|), \ldots, \exp\left(-\widehat{\phi}_S^{-1}\|\mathbf{s}_0 - \mathbf{s}_N\|\right)\widehat{\rho}_T(|t_i - t_0|)\right)^t$$

or

$$c_i = \widehat{\sigma}_S^2 \widehat{\sigma}_T^2 \left(\exp\left(-\widehat{\phi}_S^{-1}\|\mathbf{s}_0 - \mathbf{s}_1\|\right)\widehat{\rho}_T(|t_i - t_0|), \cdots, \exp\left(-\widehat{\phi}_S^{-1}\|\mathbf{s}_0 - \mathbf{s}_N\|\right)\widehat{\rho}_T(|t_i - t_0|)\right)^t,$$

with $i = 1, 2, \ldots, T$, depending on whether one assumes the covariance structure in (6) or in (7).

4. Simulation study

For model validation purposes a simulation study was conducted. Gaussian data spatially correlated with zero mean was generated on a set of $N = 50$ randomly chosen locations in the square $[0, 10]^2$ at time points $T = 3$, $T = 4$ and $T = 6$ according to AR(2), AR(3) and AR(5) models, respectively, in order to be representative of a strong temporal correlation among the time points. We should note that each time point represents a specific survey, assuming that they occur approximately regularly over time. To replicate the behaviour of the real data set to be introduced in Section 5, which exhibits a region with more intensified sampling density, 15 of those locations belong to the square $[4.5, 5.5]^2$.

The mean component in (2) includes the covariates *intensity of sampling locations*, $int(\mathbf{s})$, and a specific contribution for survey $i = 2, \ldots, T$, assuming survey 1 as the reference. For example, under the simulation study with $T = 3$, this results in

$$\mu(\mathbf{s}, t) = \beta_0 + \beta_1 int(\mathbf{s}) + \beta_2 v_2(t) + \beta_3 v_3(t), \tag{13}$$

where, for $i = 2, 3$,

$$v_i(t) = \begin{cases} 1 & \text{if } t = i, \\ 0 & \text{otherwise.} \end{cases}$$

The reason to consider the covariate sampling intensity is that, as stated before, the data set to be used later in the application of model (1) shows that the sampling design is not uniform throughout the whole region. The covariates v_1 and v_2 are intended to accommodate situations in which, for instance, a change in policies is expected to modify some outcome. Margalho et al. (2014), who also consider data of heavy metal concentration in mosses as in the present work, identified a reduction on the scale of Pb measurements at the 2002 survey, which could probably be attributed to new Portuguese legislation forcing the use of unleaded fuel.

The two different space-time covariance functions described in (6) and (7) were compared. The particular choice of the parameter values for the mean component (13) were as follows: $\beta_0 = 0$, which indicates that the expected value for the first survey is zero; $\beta_1 = 1$, once the inclusion of the covariate *sampling intensity* in the design matrix M in (2) was made by subtracting the mean intensity of all locations to the intensity of each location, thus an unitary coefficient turns the contribution of the covariate to the mean dependent on the sign of that difference; $\beta_2 = \cdots = \beta_T = 0$, which means that the difference between the expected values of, respectively, the second and the first survey and each new survey, is zero.

The variance parameters were $\sigma_S^2 = 5$, representing the mean of spatial variances along time, and $\sigma_T^2 = 0.8$, representing the mean of temporal variances at each sample location; $\sigma^2 = 4$, representing

Table 3. Estimates for the model parameters when considering the space-time covariance function in (6).

True Parameter	AR(2)		AR(3)		AR(5)	
	Mean	**Std. Error**	**Mean**	**Std. Error**	**Mean**	**Std. Error**
$\beta_0 = 0$	−0.0341	0.0566	−0.1016	0.0534	−0.0844	0.0583
$\beta_1 = 1$	1.0496	0.1591	1.0340	0.1329	0.9454	0.1162
$\beta_2 = 0$	−0.0007	0.0239	0.0303	0.0237	0.0291	0.0252
$\beta_3 = 0$	−0.0177	0.0339	0.0336	0.0265	−0.0017	0.0278
$\beta_4 = 0$	-	-	0.0408	0.0304	0.0112	0.0271
$\beta_5 = 0$	-	-	-	-	0.0036	0.0331
$\beta_6 = 0$	-	-	-	-	0.0159	0.0341
$\rho_T(|t_1 - t_2|) = 0.850$	0.8349	0.0029	0.8311	0.0029	0.8329	0.0018
$\rho_T(|t_1 - t_3|) = 0.810$	0.8007	0.0034	0.8032	0.0032	0.8054	0.0019
$\rho_T(|t_1 - t_4|) = 0.750$	-	-	0.7519	0.0036	0.7490	0.0021
$\rho_T(|t_1 - t_5|) = 0.725$	-	-	-	-	0.7356	0.0025
$\rho_T(|t_1 - t_6|) = 0.700$	-	-	-	-	0.7063	0.0027
$\sigma^2 = 4$	3.5666	0.0841	3.7167	0.0552	3.7647	0.0382
$\tau^2 = 0.25$	0.2409	0.0043	0.2437	0.0035	0.2494	0.0024
$\phi = 3$	2.8459	0.0816	3.0456	0.0533	3.0366	0.0308

the variance of all observations along time and space (and also satisfying $\sigma^2 = \sigma_S^2 \cdot \sigma_T^2$); $\tau^2 = 0.25$; $\phi = 3$, being less than 25% of the maximum possible distance between two sampled locations.

The correlation parameters in the time dimension were chosen taking into account non-null correlations of a AR(2), AR(3), or AR(5) process, depending on whether the number of surveys T is 3, 4 or 6, respectively. In particular, under the simulation study with $T = 3$, $\rho_T(|t_1 - t_2|) = \rho_T(|t_2 - t_3|) = 0.85$ and $\rho_T(|t_1 - t_3|) = 0.81$.

For each scenario 250 replicates were computed. For each replicate, the β coefficients were estimated according to (9) and the optimization method L-BFGS-B of the function *optim* in *R* was used to obtain estimates for the other parameters. This method is a bound constrained variant of the BFGS algorithm which in turn approximates the Newton method for solving nonlinear optimization problems. The L-BFGS-B algorithm is an iterative algorithm that minimizes the objective function subject to some boundary constraints (Keskar and Waechter, 2016).

To assess the predictive performance of each scenario, the mean and the standard error of the estimates were computed. Results on the estimation of the model parameters may be found in Tables 3 and 4. As $\rho_T(|t_1 - t_2|) = \rho_T(|t_2 - t_3|) = \rho_T(|t_3 - t_4|) = \rho_T(|t_4 - t_5|) = \rho_T(|t_5 - t_6|)$, for the sake of simplicity these tables only display the value for $\rho_T(|t_1 - t_2|)$. The same display applies for lags of 2, 3, 4 or 5 temporal units.

Although there are no major differences, these estimates show lower values of standard errors for the majority of the parameters if one considers the spatial and the temporal variance contribution separately. These simulation results also highlight the fact that the proposed model produces better estimates as the number of time observations increases.

In this simulation study an iterative process to find parameter estimates could also be considered

Table 4. Estimates for the model parameters when considering the space-time covariance function in (7).

True Parameter	AR(2) Mean	AR(2) Std. Error	AR(3) Mean	AR(3) Std. Error	AR(5) Mean	AR(5) Std. Error
$\beta_0 = 0$	−0.0954	0.0541	−0.0710	0.0544	−0.0046	0.0536
$\beta_1 = 1$	1.0615	0.1469	1.0109	0.1478	0.9807	0.1153
$\beta_2 = 0$	0.0184	0.0215	0.0012	0.0222	−0.0036	0.0249
$\beta_3 = 0$	0.0465	0.0304	0.0339	0.0247	−0.0114	0.0268
$\beta_4 = 0$	-	-	0.0408	0.0178	−0.0075	0.0256
$\beta_5 = 0$	-	-	-	-	−0.0266	0.0321
$\beta_6 = 0$	-	-	-	-	−0.0311	0.0346
$\rho_T(|t_1 - t_2|) = 0.850$	0.8379	0.0031	0.8305	0.0029	0.8344	0.0018
$\rho_T(|t_1 - t_3|) = 0.810$	0.8097	0.0034	0.8066	0.0028	0.8030	0.0018
$\rho_T(|t_1 - t_4|) = 0.750$	-	-	0.7502	0.0029	0.7501	0.0021
$\rho_T(|t_1 - t_5|) = 0.725$	-	-	-	-	0.7242	0.0023
$\rho_T(|t_1 - t_6|) = 0.700$	-	-	-	-	0.7017	0.0021
$\sigma_S^2 = 5$	4.7047	0.0419	4.7299	0.0317	4.5086	0.0258
$\sigma_T^2 = 0.8$	0.7528	0.0067	0.7568	0.0051	0.7688	0.0041
$\tau^2 = 0.25$	0.2438	0.0037	0.2506	0.0032	0.2454	0.0021
$\phi = 3$	2.7996	0.0563	2.9487	0.0407	3.0664	0.0333

by following an algorithm where *(i)* an initial estimate $\widehat{\beta}_{(0)}$ for β is proposed, *(ii)* estimates $\tau_{(0)}^2$ and of $\theta_{(0)}$ are derived from (8) using the estimate $\widehat{\beta}_{(0)}$, *(iii)* the estimate $\widehat{\beta}_{(1)}$ is found by updating $\widehat{\beta}_{(0)}$ from the estimates $\tau_{(0)}^2$ and $\theta_{(0)}$, *(iv)* steps *(ii)* and *(iii)* are repeated until a given error tolerance is reached.

5. Application to environmental data

The model previously described in Section 3 will now be used to obtain predicted values of heavy metal concentration at each point of a fine 300×100 grid covering mainland Portugal, allowing grid points at a distance of around 2 km from each other.

5.1 Assessing the separability assumption

An important assumption in the definition of model (1) is the separability of the covariance structure. This is a simplifying assumption frequently taken into account in environmental modelling, as substantial computational benefits arise from this property. For the application under consideration and due to the reduced number of time observations, it is not possible to perform a formal test to ascertain the separability of the covariance structure.

Alternatively, we use the aid of the empirical spatio-temporal semivariogram to adjust a spatio-temporal separable model (Rodriguez-Iturbe and Mejia, 1974; de Cesare et al., 2001),

$$C_{ST}(h_S, h_T) = C_S(h_S) \cdot C_T(h_t),$$

Table 5. Parameter estimates for the adjusted variograms in Figure 1 ($\widehat{\phi}_S$ in meters and $\widehat{\phi}_T$ in years).

Model	Component	$\widehat{\tau}^2$	$\widehat{\sigma}^2$	$\widehat{\phi}$
Separable	Spatial	0.41	0.59	53 108.81
	Temporal	0.24	0.76	2.69
	Joint	—	2.45	—
Metric	Joint	1.13	0.91	—
ProductSum	Spatial	—	1.58	69 065.92
	Temporal	—	1.94	2.04
	Joint	0.99	3.49	—

a metric model (Dimitrakopoulos and Luo, 1994),

$$C_{ST}(h_S, h_T) = C\left(a_1 \|h_S\|^2 + a_2 |h_t|^2\right),$$

and a product-sum model (de Cesare et al., 2001),

$$C_{ST}(h_S, h_T) = k_1 \cdot C_S(h_S) \cdot C_T(h_t) + k_2 \cdot C_S(h_S) + k_3 \cdot C_T(h_t).$$

For each model, the resulting parameter estimates are in Table 5. We consider exponential models in the covariance functions to obtain these estimates.

There are some graphical differences between these several adjusted theoretical models, as represented in Figure 1. The separable model captures more precisely the behavior shown by the empirical semivariogram, detecting an increase of variances from the first to the second survey, and a decrease for the third. Cressie and Huang (1999) state that *separable models are often chosen for convenience rather than for their ability to fit the data well*, so the assumption of separability will be assumed for the Mn data.

As stated before, the definition of separability satisfies the statements *(i)* and *(ii)* on page 41. In our application case, *(i)* means that the spatial structure is the same for all surveys. The left side of Figure 2 shows the empirical semivariograms related with each survey and, on the right, cross semivariograms for time lags of 0, 1 and 2 surveys. The structure revealed by the cross semivariograms is similar independently of the lag being considered, which corroboraties the assumption of equality of spatial structure for the performed surveys.

5.2 Maximum likelihood estimates of the parameters

The parameter estimates of model (1), considering the covariance functions defined by (6) and by (7), and derived according to (8) and (9), are given in Table 6. In particular, one can observe that, in either case, the baseline effect estimate β_0 is above seven units, being of similar magnitude to the mean of the transformed Mn concentration. The inclusion of the covariate *sampling intensity*, as detailed in the simulation study, was done considering the difference between the intensity of each location and the mean of the intensity for all locations. This way, the estimate of β_1 resulting from (7) reflects more precisely the fact that larger values of Mn concentration occur in regions with
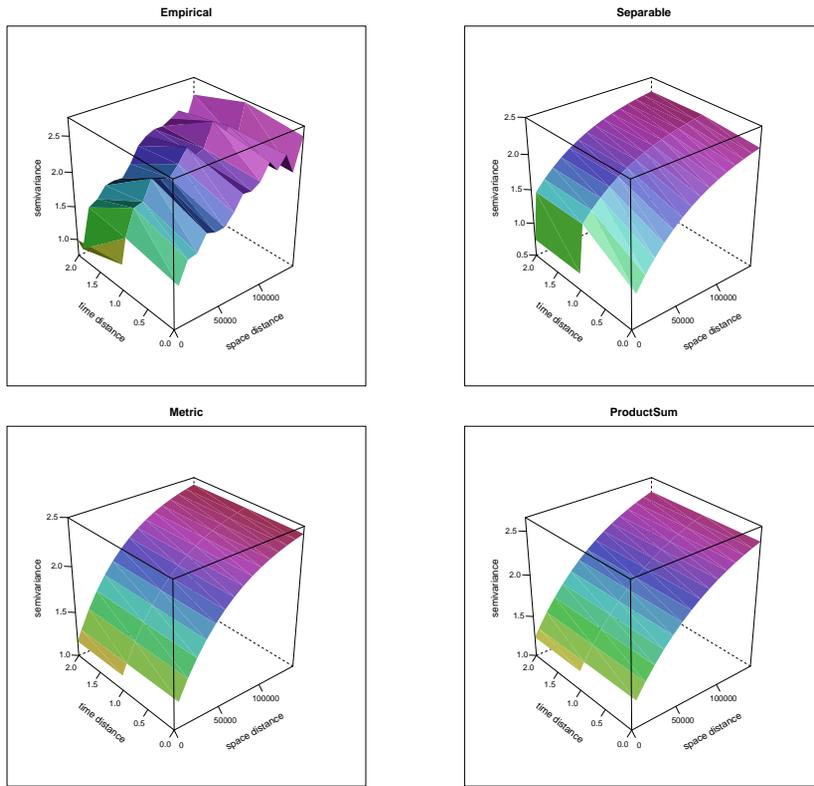
Figure 1. Empirical (top left), Separable (top right), Metric (bottom left) and ProductSum (bottom right) spatio-temporal semivariograms for Mn transformed data.

low urban or industrial density, which corresponds also to regions with lower sampling locations intensity. We note that these less industrialized areas are the ones where soil typology favours high Mn concentration values. The contribution of the second and of the third surveys are estimated to be of different signs, namely the second survey with a positive contribution and the third with a negative one. However, the estimate of β_3 for the covariance function described in (7) reinforces the contribution of the third survey to the decrease of predicted concentrations.

As was to be expected, the parameters related to the temporal correlation are equal in either approach, indicating data that strongly time correlated. Similarly, the nugget effect is estimated equally by either approach. The covariance function given by (7) assumes different spatial and temporal variances, contributing as a product to the covariance of the process. It is worthwhile to note that the product of the estimates of these parameters equals the overall variance estimate for (6). The estimates for the radius of influence ϕ are also of similar magnitude for either approach, being approximately 58.5 km. The computation of the standard errors was done via Monte-Carlo simulation.
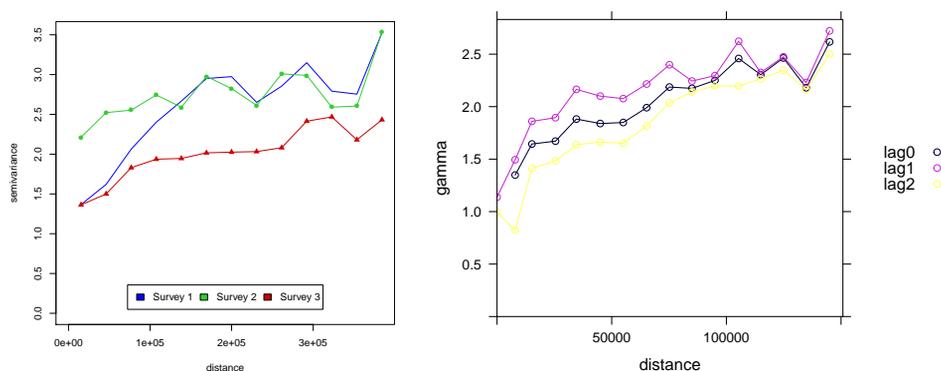
Figure 2. Empirical semivariograms (left) and cross semivariograms for lags 0, 1 and 2 (right) for Mn transformed data.

5.3 Prediction of Mn concentration for the most recent survey

These parameter estimates, together with (11) and (12) (setting t_0 equal to 2002 in these equations), allows one to predict, for the most recent survey, the Box-Cox transformed Mn concentration values and the associated interpolation errors. This corresponds to adopting plug-in estimates of the parameters that define the mean and the covariance structure of the Mn variable to proceed with prediction. A brief summary of those predicted values is in Table 7. One can observe that the range of predicted values, in either situation, is similar, although with larger predicted values when considering spatial and temporal scale parameters. In fact, all the measures of the predicted values according to (7), with exception of the standard deviation, are larger than the ones obtained considering (6). This might be explained by taking into account that, according to Table 6, when applying the adjustment of sampling design, the expected value for Mn under (7) becomes larger than the one under (6).

The left panel of Figure 3 represents the Mn predicted concentration map, while the associated prediction error is in the right panel. As was to be expected, once contamination by Mn is more associated to parameters describing soil typology and sampling site conditions, and less related to anthropogenic contamination sources, higher predicted contamination values occur in the eastern part of mainland Portugal. This spatial structure is captured by either approach for the covariance function.

5.4 Cross validation study

Cross-validation of results was developed to assess the accuracy of the predicted concentrations obtained considering the two approaches for the covariance structure in (6) and in (7). The technique in use for this cross-validation study corresponds to leave out at each time one sampled location and interpolate the value at that location as a function of all other locations.

The absolute prediction error (APE) was considered to measure the discrepancy between the observed and the interpolated value at each location. Results revealed the value 1.02 for the mean

Table 6. Model (1) parameter estimates with standard errors.

Parameter	Covariance function in (6)		Covariance function in (7)	
	Estimate	**St. Error**	**Estimate**	**St. Error**
β_0	7.39	0.04	7.45	0.04
β_1	-0.01	<0.01	0.01	<0.01
β_2	0.21	0.01	0.15	0.01
β_3	-0.14	0.02	-0.24	0.02
$\rho_T(|t_1 - t_2|)$	0.97	0.01	0.97	0.01
$\rho_T(|t_1 - t_3|)$	0.91	0.01	0.91	0.01
$\rho_T(|t_2 - t_3|)$	0.96	0.01	0.96	0.01
σ^2_{total}	1.45	0.03	—	—
σ^2_S	—	—	0.98	0.01
σ^2_T	—	—	1.48	0.02
τ^2	1.02	0.01	1.02	0.01
$\phi\,(m)$	58 596.89	1421.10	58 624.08	1470.20

Table 7. Predicted transformed Mn concentration for the 2002 survey and interpolation error values, obtained from the spatio-temporal model in (1).

	Covariance function in (6)		Covariance function in (7)	
	Predicted	**Error**	**Predicted**	**Error**
Min	4.05	0.23	4.53	0.41
Median	7.10	0.72	7.21	0.79
Max	9.05	1.19	9.16	1.19
Mean	6.92	0.78	7.09	0.85
St. dev.	0.92	0.23	0.74	0.19

APE, with a standard deviation of 0.81, no matter which approach for the covariance function is considered.

This is not a surprising result as the two sets of predicted values were obtained from the same model, with the only difference being the parametrization of the separable spatio-temporal covariance structure. In fact, this finding reinforces the advantage of being able to identify two distinct parameters for variability, one for the spatial dimension and the other for the temporal dimension, without compromising the accuracy of predicted values.

5.5 Comparison with other models

Margalho et al. (2014), with the same data set, proposed an extension of an existing spatio-temporal model introduced in Høst, Omre and Switzer (1995), allowing for considering explanatory covariates relevant to the process under observation. As in the present work, the separability of the spatial and the temporal contribution was taken into account.
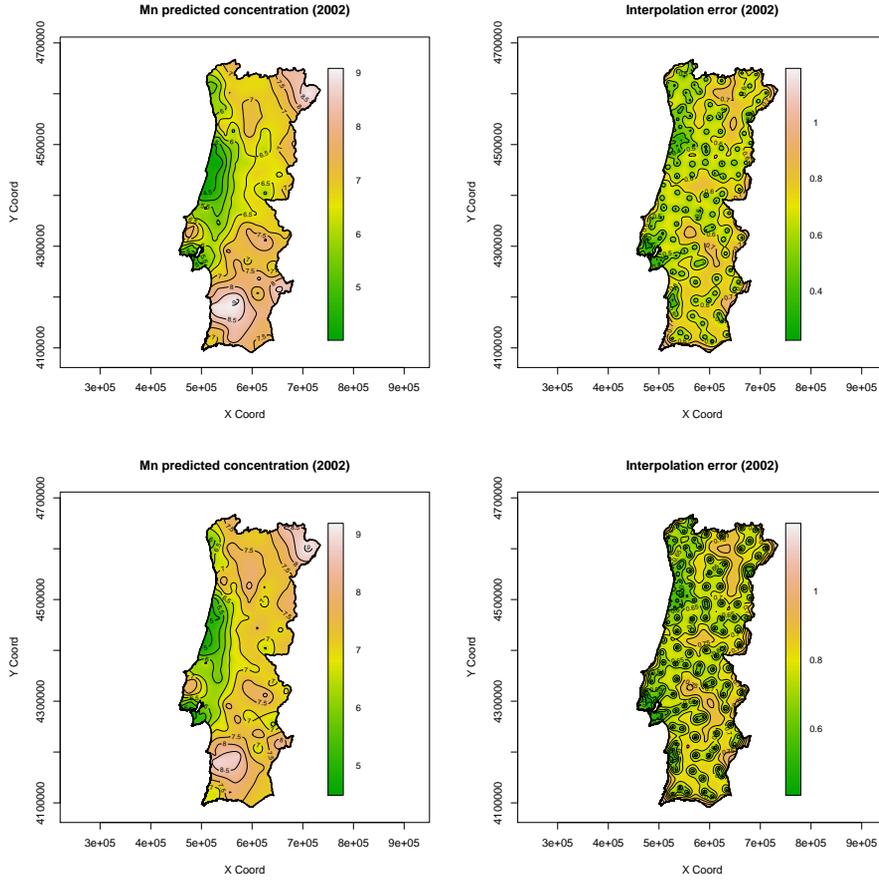
Figure 3. Prediction map for the 2002 survey (left) and the associated interpolation error map (right) for Mn transformed data, considering the covariance function given in (6) (top row) or (7) (bottom row).

The response variable value $Y(\mathbf{s}, t)$ at location $\mathbf{s}$ and time t was written as

$$Y(\mathbf{s}, t) = \mu(\mathbf{s}, t) + \omega(\mathbf{s}, t)\varepsilon(\mathbf{s}, t),$$

where $\mu(\mathbf{s}, t)$, $\omega(\mathbf{s}, t)$ and $\varepsilon(\mathbf{s}, t)$ stand for space-time mutually independent random fields representing mean, scale and residuals.

The mean and the scale components were decomposed between spatial and temporal effects. Specifically, the mean component was decomposed as

$$\mu(\mathbf{s}, t) = M_1(\mathbf{s}) + m_2(t),$$

that is, the sum of a purely spatial component $M_1(\mathbf{s})$, modelling the spatial mean variation, and $m_2(t)$, a temporal modulation at discrete times, which corresponds to consider an additive separability in

Table 8. Comparison of predicted Box-Cox transformed Mn concentration for the 2002 survey and interpolation error values under two different spatio-temporal models.

	Model (1)		**Model in Margalho et al. (2014)**	
	Concentration	**Error**	**Concentration**	**Error**
Min	4.53	0.41	4.30	0.80
Median	7.21	0.79	7.53	1.07
Max	9.16	1.19	8.93	1.18
Mean	7.09	0.85	7.37	1.06
St. dev.	0.74	0.19	0.66	0.12

the mean field $\mu(\mathbf{s}, t)$. Regarding the scale field, a multiplicative separability was considered,

$$\omega(\mathbf{s}, t) = S_1(\mathbf{s}) s_2(t),$$

that is, the product of a purely spatial component, $S_1(\mathbf{s})$, and a purely temporal component, $s_2(t)$. $M_1(\mathbf{s})$ represents a mean random effect in space and $m_2(t)$ the associated time correction, being approximately equal to zero. In a similar way $S_1(\mathbf{s})$ represents a scale random effect in space and $s_2(t)$ the associated time correction being approximately one. The remaining space-time interactions not captured by the foregoing components were identified by the random component $\varepsilon(\mathbf{s}, t)$.

The model aimed to be simple and, mainly, to cover a wide range of practical situations, where space and time separability is reasonable to be assumed.

Although both the prediction and the interpolation error maps obtained by each methodology are similar, the interpolation error values are of smaller magnitude with the new model proposal (1). This statement can be confirmed from Table 8, which summarizes a comparison of the predicted values and the interpolation errors obtained when applying model (1) or the model in Margalho et al. (2014). Moreover, the comparison of predictions accuracy, presented in Table 2 and Table 8, highlights the importance of including in the prediction procedure not only explanatory covariates for the process under observation, but also the information from the past surveys. One can register for spatio-temporal prediction models lower values of interpolation error.

6. Conclusions

We propose a spatio-temporal model-based approach to accommodate the possible existence of explanatory covariates to the process under observation and, under the assumption of the separability of the Gaussian process under observation, construct the model in a way that the reduced number of time observations lead to a temporal correlation model with a small number of parameters. The proposed model was derived in order to accommodate not exclusively geo-referenced covariates, but also covariates associated with the temporal behavior of the process. Complementing the most common interpretation of the covariance structure, different spatial and temporal sources of variability were allowed.

The model validation was performed via a simulation study where better results in terms of precision were achieved when assuming different spatial and temporal variances, in contrast to an

approach an overall variance.

We have shown how to use this model via an application to a real data set, predicting heavy metal concentrations, namely Mn concentration, along with the associated uncertainty maps. The separability of the covariance structure in the spatio-temporal context was not possible to test, due to the reduced number of temporal observations. Instead, we have used the aid of a graphical comparison of the empirical spatio-temporal semivariogram with the ones produced by different models to observe that the separable model fitted the data better. The model's good performance in predicting concentration values of contaminant was compared with results obtained by using a different spatio-temporal model which also assumes spatial and temporal separability and enables the use of covariates. In terms of interpolation error, values of lower magnitude were found with this new proposal.

Acknowledgments

References

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.

BOQUETE, M., FERNÁNDEZ, J., ABOAL, J., AND CARBALLEIRA, A. (2011). Analysis of temporal variability in the concentrations of some elements in the terrestrial moss *Pseudoscleropodium purum*. *Environmental and Experimental Botany*, **72**, 210–217.

BRUNO, F., GUTTORP, P., SAMPSON, P., AND COCCHI, D. (2003). Non-separability of space-time covariance models in environmental studies. *In* MATEU, J., HOLLAND, D., AND MANTEIGA, W. (Editors) *The ISI International Conference on Environmental Statistics and Health*. Universidade de Santiago de Compostela.

CAMELETTI, M., IGNACCOLO, R., AND BANDE, S. (2011). Comparing spatio-temporal models for particulate matter in Piemonte. *Environmetrics*, **22**, 985–996.

CRESSIE, N. (1993). *Statistics for Spatial Data*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. Wiley & Sons, New York.

CRESSIE, N. AND HUANG, H. (1999). Classes of nonseparable, spatio-temporal stationary covariance functions. *Journal of the American Statistical Association*, **94**, 1330–1340.

CRESSIE, N. AND WIKLE, C. (2011). *Statistics for Spatio-Temporal Data*. Revised edition. Wiley & Sons, Hoboken, New Jersey.

DE CESARE, L., MYERS, D., AND POSA, D. (2001). Estimating and modeling space-time correlation structures. *Statistics and Probability Letters*, **51**, 9–14.

DÍAZ-AVALOS, C., JUAN, P., AND MATEU, J. (2014). Significance tests for covariate-dependent trends in inhomogeneous spatio-temporal point processes. *Journal of Stochastic Environmental Research and Risk Assessment*, **28**, 593–609.

DIGGLE, P., MENEZES, R., AND SU, T. (2010). Geostatistical inference under preferential sampling. *Applied Statistics*, **59**, 191–232.

DIMITRAKOPOULOS, R. AND LUO, X. (1994). Spatiotemporal modelling: covariances and ordinary kriging systems. *In Geostatistics for the Next Century*. Springer, 88–93.

FANSHAWE, T., DIGGLE, P., RUSHTON, S., SANDERSON, R., LURZ, P., GLINIANAIA, S., PEARCE, M., PARKER, L., CHARLTON, M., AND PLESS-MULLOLI, T. (2008). Modelling spatio-temporal variation in exposure to particulate matter: a two-stage approach. *Environmetrics*, **19**, 549–566.

FIGUEIRA, R., SÉRGIO, C., AND SOUSA, A. (2002). Distribution of trace metals in moss biomonitors and assessment of contamination sources in Portugal. *Environmental Pollution*, **118**, 153–163.

GUTTORP, P. AND LOPERFIDO, N. (2008). Network bias in air quality monitoring design. *Environmetrics*, **19**, 661–671.

HARMENS, H., NORRIS, D., STEINNES, E., KUBIN, E., PIISPANEN, J., ALBER, R., ALEKSIAYENAK, Y., BLUM, O., COSKUN, M., DAM, M., TEMMERMAN, L. D., FERNÁNDEZ, J., FROLOVA, M., FRONTASYEVA, M., GONZÁLEZ-MIQUEO, L., GRODZINSKA, K., JERAN, Z., KORZEKWA, S., KRMAR, M., KUBIN, E., KVIETKUS, K., LEBLOND, S., LIIV, S., MAGNÚSSON, S., MANKOVSKÁ, B., PESCH, R., RULING, A., SANTAMARIA, J., SCHODER, W., SPIRIC, Z., SUCHARA, I., THONI, L., URUMOV, V., 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.

HOLY, M., PESCH, R., SCHODER, W., HARMENS, H., ILYIN, I., ALBER, R., ALEKSIAYENAK, Y., BLUM, O., COSKUN, M., DAM, M., TEMMERMAN, L. D., FEDORETS, N., FIGUEIRA, R., FROLOVA, M., FRONTASYEVA, M., GOLTSOVA, N., MIQUEO, L., GRODZINSKA, K., JERAN, Z., KORZEKWA, S., KRMAR, M., KUBIN, E., KVIETKUS, K., LARSEN, M., LEBLOND, S., LIIV, S., MAGNÚSSON, S., MANKOVSKÁ, B., MOCANU, R., PIISPANEN, J., RUHLING, A., SANTAMARIA, J., STEINNES, E., SUCHARA, I., THONI, L., TURCSÁNYI, G., URUMOV, V., WOLTERBEEK, B., YURUKOVA, L., AND ZECHMEISTER, H. (2009). First thorough identification of factors associated with Cd, Hg and Pb concentrations in mosses sampled in the European surveys 1990, 1995, 2000 and 2005. *Journal of Atmospheric Chemistry*, **63**, 109–124.

HØST, G., OMRE, H., AND SWITZER, P. (1995). Spatial interpolation errors for monitoring data. *Journal of the American Statistical Association*, **90**, 853–861.

HOWE, P., MALCOLM, H., AND DOBSON, S. (2004). Manganese and its compounds: environmental aspects. Concise International Chemical Assessment Document 63, World Health Organization, Geneva.

HUANG, H., MARTINEZ, F., MATEU, J., AND MONTES, F. (2007). Model comparison and selection for stationary space-time models. *Computational Statistics and Data Analysis*, **51**, 4577–4596.

KESKAR, N. S. AND WAECHTER, A. (2016). A limited-memory quasi-newton algorithm for bound-constrained nonsmooth optimization. *arXiv preprint arXiv:1612.07350*.

KYRIAKIDIS, P. AND JOURNEL, A. (1999). Geostatistical space-time models: A review. *Mathematical Geology*, **31**, 651–684.

LINDSTROM, J., SZPIRO, A., SAMPSON, P., SHEPPARD, L., ORON, A., RICAHRDS, M., AND LARSON, T. (2011). A flexible spatio-temporal model for air pollution: Allowing for spatio-temporal covariates.

UW Biostatistics Working Paper Series, working paper 370.

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**, 1307–1321.

MARTINS, A., FIGUEIRA, R., SOUSA, A., AND SÉRGIO, C. (2012). Spatio-temporal patterns of Cu contamination in mosses using geostatistical estimation. *Environmental Pollution*, **170**, 276–284.

PERL, D. AND OLANOW, C. (2007). The neuropathology of manganese-induced parkinsonism. *Journal of Neuropathology and Experimental Neurology*, **66**, 675.

PINSINO, A., ROCCHERI, M., AND MATRANGA, V. (2012). *Manganese: a new emerging contaminant in the environment*. INTECH Open Access Publisher.

ROCKS, S. AND LEVY, L. (2008). Manganese Health Research Program: Overview of research into the health effects of manganese (2007–2008). Report, Institute of Environment and Health, Cranfield University, Silsoe, Bedfordshire.

RODRIGUEZ-ITURBE, I. AND MEJIA, J. (1974). The design of networks in time and space. *Water Resources Research*, **10**, 713–728.

SAHU, S. AND MARDIA, K. (2005). Recent trends in modelling spatio-temporal data. *In Proceedings of the Special Meeting on Statistics and Environment*. 69–83.

SHADDICK, G. AND WAKEFIELD, J. (2002). Modelling daily multivariate pollutant data at multiple sites. *Journal of The Royal Statistical Society Series C – Applied Statistics*, **51**, 351–372.

SHERMAN, M. (2011). *Spatial Statistics and Spatio-Temporal Data: Covariance Functions and Directional Properties*. John Wiley & Sons Inc, Chichester.

STEINNES, E., BERG, T., AND SJOBAKK, T. (2003). Temporal and spatial trends in Hg deposition monitored by moss analysis. *The Science of the Total Environment*, **304**, 215–219.

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.

UYAR, G., ÖREN, M., YILDIRIM, Y., AND INCE, M. (2007). Mosses as indicators of atmospheric heavy metal deposition around a coal-fired power plant in Turkey. *Fresenius Environmental Bulletin*, **16**, 182–192.

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.