UNIVERSITÀ DEGLI STUDI DI BERGAMO
DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE
E METODI MATEMATICI°

QUADERNI DEL DIPARTIMENTO

Department of Information Technology and Mathematical Methods

Working Paper

Series "*Mathematics and Statistics*"

n. 15/MS – 2009

Maximum likelihood estimation of the dynamic coregionalization model with heterotopic data

by

A. Fassò, F. Finazzi, C. D'Ariano

° Viale Marconi. 5, I – 24044 Dalmine (BG), ITALY, Tel. +39-035-2052339; Fax. +39-035-562779

[§] L'accesso alle *Series* è approvato dal Comitato di Redazione. I *Working Papers* della Collana dei Quaderni del Dipartimento di Ingegneria dell'Informazione e Metodi Matematici costituiscono un servizio atto a fornire la tempestiva divulgazione dei risultati dell'attività di ricerca, siano essi in forma provvisoria o definitiva.

Maximum likelihood estimation of the dynamic coregionalization model with heterotopic data

Alessandro Fassò, Francesco Finazzi, Cinzia D'Ariano
DIIMM, University of Bergamo
alessandro.fasso@unibg.it

November 30, 2009

Abstract

The information content of multivariable spatio-temporal data depends on the underlying spatial sampling scheme. The most informative case is represented by the isotopic configuration where all variables are measured at all sites. The opposite case is the completely heterotopic case where different variables are observed only at different locations. A well known approach to multivariate spatio-temporal modelling is based on the linear coregionalization model (LCM).

In this paper, the maximum likelihood estimation of the heterotopic spatio-temporal model with spatial LCM components and temporal dynamics is developed. In particular, the computation of the estimates is based on the EM algorithm and two solutions are proposed: one is based on the more cumbersome exact maximization of the a posteriori expected log likelihood and the other is an approximate closed-form solution, whose properties are assessed in terms of bias and effi ciency through an extensive Monte Carlo simulation.

KEY WORDS: EM algorithm; multivariate spatio-temporal models; dynamic mapping; particulate matters; Aerosol Optical Thickness.

1 Introduction

Multivariate environmental data collected over space and time may arise from two basically different spatial sampling schemes. Considering a monitoring system based on a network with instruments which do not move over time, the simplest case is represented by the isotopic or collocated configuration where each variable is measured at each network site. For example, this is the case in modelling the joint distribution of air pollution concentrations and meteorological variables coming from the same ground level monitoring network with every stations equipped with all the instruments.

The opposite case is represented by the completely heterotopic or non-collocated configuration, within which two different variables are never observed at the same site. An example arises when modelling the joint distribution of pointwise satellite data and pollution concentrations from a ground level monitoring network.

The approach to multivariate spatial modelling based on the linear coregionalization model (LCM) has been extended to the multivariate spatio-temporal modelling in classical papers (e.g. Rouhani and Wackernagel, 1990) and still deserves special attention in literature. For example, De Iaco et al. (2005) used a bivariate LCM model for spatio-temporal cokriging of isotopic temperature and humidity in the Lombardy region in northern Italy. Similarly, Jost et al. (2005) used the same cross-product model to analyse water storage in a forest ecosystem in Austria. Moreover, Liu and Koike (2007) used the spatio-temporal coregionalization model for partially heterotopic data on water quality in the Arike Sea, in Japan.

Identification and estimation is a key issue in the coregionalization model as most of applications are based on heuristic variogram fitting. For example, Lark and Papritz (2003) used simulated annealing for fitting the covariograms to soil data of Swiss Jura, and Bishop and Lark (2008) considered nonparametric coregionalization. Recently, maximum likelihood estimation for the purely spatial LCM has been considered by Zhang (2007), who showed that the EM algorithm gives an iterative procedure based on quasi-closed-form formulas, at least in the isotopic case. In particular, at each EM step the coregionalization matrices are computed by closed-form while only the low dimensional geostatistical parameters require numerical optimization.

In this paper, the maximum likelihood estimation is considered for the heterotopic spatio-temporal model with spatial LCM component and temporal dynamics proposed by Fassò et al. (2009) for satellite data calibra-

tion. The problem of spatio-temporal calibration has been attached using the Kalman filter approach (Brown et al., 2001) and the cokriging (Orasi et al., 2009). Our result is based on the EM algorithm and generalizes Zhang (2007) as it allows for time correlation, covariates and heterotopic data. Moreover, it generalizes Fassò and Cameletti (2009a, b) who gave the EM algorithm for the univariate case which has been extended to the locally weighted setup by Bodnar and Schmid (2009). Two solutions are proposed, one is based on the more cumbersome exact maximization of the a posteriori expected log likelihood, while the other is an approximate solution with closed forms for the coregionalization matrices.

The rest of the paper is organized as follows. Section 2 introduces the multivariate spatio-temporal model based on coregionalization and autoregressive temporal dynamics. Section 3 discusses the structure of both data and variance-covariance matrices of the various model components. Section 4 defines the EM algorithm and proposes both an exact procedure and an approximate solution, which has the advantage of reducing the dimension of the subset of parameters which are not computed by closed-form formulas.

Section 5 illustrates the algorithm capabilities thanks to an extensive Monte Carlo simulation campaign based on realistic data related to satellite AOT data calibration. In particular, the algorithm is tested over large data sets that give rise to covariance matrices up to millions of elements. Bias and effi ciency of the approximate solution are shown to be satisfactory. After the conclusions of Section 6, the Appendix gives technical details on both the exact and approximate EM algorithms.

2 Model

For any time $t \in \mathbb{N}$ and any site $s \in D \subset \mathbb{R}^2$, let $Y(s,t) = (Y_1(s,t), ..., Y_q(s,t))'$ be a $q-$dimensional spatio-temporal process. The first stage measurement equation is

$$Y(s,t) = U(s,t) + \varepsilon(s,t) \tag{1}$$

where $\varepsilon(s,t)$ is a Gaussian instrumental error which is white noise in space and time with $q \times q$ covariance diagonal matrix $\Gamma_0(s)$.

At the second stage, the underlying true local variable $U(s,t)$ has the following structure:

$$U(s,t) = X(s,t)\beta + KZ(t) + \bar{W}(s,t) \tag{2}$$

where $X(s,t)$ is a b-dimensional spatio-temporal field of known covariates, $Z(t)$ is the d-dimensional latent temporal state, which is constant in space and has Markovian temporal dynamics, while $\bar{W}(s,t)$ is the latent random spatial effect at time t. In particular, the latent state at time t is defined as

$$Z(t) = GZ(t-1) + \eta(t)$$

where G is a $d \times d$ transition matrix with eigenvalues λ_i such that $|\lambda_i| < 1$, $\eta(t) \sim N_d(0, \Sigma_\eta)$ and $Z(0) \sim N_d(\mu_0, \Sigma_{Z_0})$. The $q \times d$ matrix K is fixed in time and it accounts for the weights of the d components of $Z(t)$ for each spatial location $s \in D$.

The latent component $\bar{W}(s,t)$ is a Gaussian process $GP_q(0, \Gamma_W)$ and it is described by a q-dimensional linear coregionalization model (LCM) of c components:

$$\bar{W}(s,t) = \sum_{p=1}^{c} W_p(s,t)$$

where $W_p = \left(W_p^1, ..., W_p^q\right)$ is white noise in time but correlated over space with a $q \times q$ covariance and cross-covariance matrix function given by

$$\Gamma_p(h, \theta_p) = \left(\text{cov}(W_p^i(s), W_p^j(s'))\right)_{i,j=1,...,q} = V_p \rho_p(h, \theta_p)$$

The above covariance and cross-covariance functions are assumed to be isotropic and $h = \|s - s'\|$ is the Euclidean distance between two sites $s, s' \in D$. For each $p = 1, ..., c$, V_p is a positive semi-definite $q \times q$ matrix and $\rho_p(h, \theta_p)$ is a valid correlation function, for example the Matern function, characterized by the parameter vector θ_p. In addition, the processes W_p are uncorrelated in the sense that, for any $i \neq j$

$$\text{cov}(W_i(s), W_j(s')) = 0, \ \forall s, s' \in D$$

The multivariate $q \times q$ covariance matrix for W is then

$$\Gamma_W(h, \theta_1, ..., \theta_c) = \sum_{p=1}^{c} \Gamma_p(h, \theta_p) = \sum_{p=1}^{c} V_p \rho_p(h, \theta_p)$$

The model parameters are collected in the following $R-$ dimensional vector

$$\Psi = vec^*(\beta, \Gamma_0; \mu_0; G, \Sigma_\eta, ; V_1, \theta_1, ..., V_c, \theta_c) = (\Psi_Y, \Psi_{Z_0}, \Psi_Z, \Psi_W) \qquad (3)$$

where the operator vec^* vectorizes all the unique parameters contained in the covariance matrices excluding zeros. The matrices K and Σ_{Z_0} are assumed to be known and do not take part in model parametrization and estimation.

3 Data structure

In order to deal with the most general case, it is assumed that each variable is possibly observed over different sets of sites. If $S_i = \{s_{i,1}, ..., s_{i,n_i}\}$ is the set of sites for the variable Y_i, $i = 1, ..., q$, three cases can be distinguished: the isotopic case, the partially heterotopic case and the entirely heterotopic case (Wackernagel, 1998). The isotopic case is characterized by the fact that each variable is observed at each site, so that $S_1 = ... = S_q$. Within the entirely heterotopic case, two different variables are never observed at the same site and $\bigcap_{i=1}^{q} S_i = \emptyset$. A less narrow case is the partially heterotopic setting, in which only some variables share only some sites. In this work it is also assumed that the sets S_i do not change with time $t \in \mathcal{T} = \{1, ..., T\}$.

Now, let $Y_t = (Y_1(s_{1,1}, t), ..., Y_1(s_{1,n_1}, t), ..., Y_q(s_{q,1}, t), ..., Y_q(s_{q,n_q}, t))'$ be the $N = n_1 + ... + n_q$ dimensional observation vector at time t at the sampling sites $S = \{S_1, ..., S_q\}$ and let X_t be the $N \times b$ matrix of known regressors at time t. Let $Y = (Y_1, ..., Y_T)$ be the $N \times T$ full data matrix, $Z = (Z_0, Z_1, ..., Z_T)$ be the $T + 1$ vector of the latent temporal process and $W_p = (W_{p,1}, W_{p,2}, ..., W_{p,T})$, $p = 1, ..., c$, be the $N \times T$ matrices of the c spatial latent processes. The collection of all processes is $W = (W_1, ..., W_c)$.

The $N \times N$ symmetric matrix of distances between each pair of sampling locations in S is given by the block representation

$$H = \left(H^{i,j}\right)_{i,j=1,...,q}$$

with $n_i \times n_j$ dimensional blocks $H^{i,j}$. If $\rho_p(H, \theta_p)$ is the spatial correlation matrix for the W_p process observed at S, then

$$\Sigma_p(V_p, \theta_p, H) = \left(v_p^{i,j} \rho_p(H^{i,j}, \theta_p)\right)_{i,j=1,...,q} \tag{4}$$

is the $N \times N$ variance-covariance block matrix of the process W_p. In particular, the k, l-th element of the block $v_p^{i,j} \rho_p(H^{i,j}, \theta_p)$ is $\mathrm{cov}(W_p(s_{i,k}), W_p(s_{j,l}))$, $k = 1, ..., n_i$, $l = 1, ..., n_j$. It must be noted that, unless the isotopic setting is considered, only the diagonal blocks of Σ_p are square.

4 Maximum-likelihood estimation

The maximum likelihood (ML) estimation of the unknown parameter vector Ψ defined by (3) is performed by means of the EM algorithm (McLachlan

and Krishnan, 1997). This method is particularly useful in the presence of either latent variables or missing data. The former being the case of the model defined by (1), in which the processes $Z(t)$ and $\bar{W}(s,t)$ are not directly observed.

4.1 Likelihood function

The complete data likelihood $L(\Psi; Y, Z, W)$ can be written as

$$L(\Psi; Y, Z, W) = L(\Psi_Y; Y \mid Z, W) \cdot L(\Psi_Z; Z) \cdot L(\Psi_W; W)$$

Note that, since Z in uncorrelated with W and, also, $W_{p,t}$ is uncorrelated with $W_{p',\tau}$ for each $t \neq \tau$ and $p \neq p'$, $p, p' = 1, ..., c$ it follows that

$$\begin{aligned} L(\Psi_Z, \Psi_W; Z, W) &= L(\Psi_Z; Z) L(\Psi_W; W) \\ &= L(\Psi_Z; Z) \cdot \prod_{p=1}^{c} \prod_{t=1}^{T} L(\Psi_W; W_{p,t}) \end{aligned}$$

Moreover, since the observations Y_t are mutually independent given the latent components (Z, W), we have

$$L(\Psi_Y; Y \mid Z, W) = \prod_{t=1}^{T} L(\Psi_Y; Y_t \mid Z, W)$$

Finally the first-order Markovian assumption implies that

$$L(\Psi_Z; Z) = L\left(\Psi_Z; Z_0\right) \cdot \prod_{t=1}^{T} L(\Psi_Z; Z_t \mid Z_{t-1})$$

Now from model assumptions of Section 2, we have the following probability distributions

$$(Y_t \mid Z, W) \sim N_N(X_t\beta + KZ_t + \bar{W}_t, \Sigma_0) \tag{5}$$

$$Z_0 \sim N_d(\mu_0, \Sigma_{Z_0})$$

$$(Z_t \mid Z_{t-1}) \sim N_d(GZ_{t-1}, \Sigma_\eta)$$

$$W_{p,t} \sim N_N(0, \Sigma_p), \ p = 1, ..., c \tag{6}$$

These give the log-likelihood function which, apart from an additive constant, is based on the following summands

$$-2l(\Psi_Y; Y \mid Z, W) = T \log |\Sigma_0| + \\ + \sum_{t=1}^{T} \left(Y_t - X_t\beta - KZ_t - \bar{W}_t\right)' (\Sigma_0)^{-1} \left(Y_t - X_t\beta - KZ_t - \bar{W}_t\right)$$

$$-2l\left(\Psi_{Z_0}; Z_0\right) = \log |\Sigma_{Z_0}| + (Z_0 - \mu_0)' \left(\Sigma_{Z_0}\right)^{-1} (Z_0 - \mu_0)$$

$$-2l(\Psi_Z; Z) = T \log |\Sigma_\eta| + \sum_{t=1}^{T} (Z_t - GZ_{t-1})' (\Sigma_\eta)^{-1} (Z_t - GZ_{t-1})$$

$$-2l(\Psi_W; W) = \sum_{p=1}^{c} \left[T \log |\Sigma_p| + \sum_{t=1}^{T} (W_{p,t})' \Sigma_p^{-1} (W_{p,t}) \right]$$

4.2 The E-step

The $(k+1)-$iteration of the EM algorithm is composed of two steps, the E-step and the M-step. During the E-step, the algorithm computes the expectation of the complete data log-likelihood conditionally on the observed data Y under the parameter $\Psi^{(k)}$, that is

$$Q\left(\Psi, \Psi^{(k)}\right) = E_{\Psi^{(k)}} \left[-2l(\Psi; Y, Z, W) \mid Y\right] \\ = Q\left(\Psi_Y, \Psi^{(k)}\right) + Q\left(\Psi_{Z_0}, \Psi^{(k)}\right) + Q\left(\Psi_Z, \Psi^{(k)}\right) + Q\left(\Psi_W, \Psi^{(k)}\right) \qquad (7)$$

The first term in (7) is

$$Q\left(\Psi_Y, \Psi^{(k)}\right) = T \log |\Sigma_0| + \sum_{t=1}^{T} \operatorname{tr}\left[\Sigma_0^{-1} \left(\bar{e}_t \bar{e}_t' + \bar{A}_t^T + K P_t^T K'\right)\right]$$

where

$$\bar{e}_t = E(Y_t - X_t\beta - KZ_t - \bar{W}_t \mid Y) = Y_t - X_t\beta - KZ_t^T - \bar{W}_t^T$$
$$\bar{W}_t^T = E(\bar{W}_t \mid Y) = \sum_{p=1}^{c} W_{p,t}^T = \sum_{p=1}^{c} E(W_{p,t} \mid Y)$$
$$\bar{A}_t^T = \mathrm{Var}(\bar{W}_t \mid Y) = \sum_{p=1}^{c} A_{p,t}^T = \sum_{p=1}^{c} \mathrm{Var}(W_{p,t} \mid Y)$$
$$Z_t^T = E(Z_t \mid Y)$$
$$P_t^T = \mathrm{Var}(Z_t \mid Y)$$

Where Z_t^T, P_t^T are the Kalman Smoother outputs with formulas given in Appendix (A.1), while $\bar{W}_t^T$ and $\bar{A}_t^T$ are Kriging quantities given in Appendix (A.2).

The second term in (7) is given by

$$Q\left(\Psi_{Z_0}, \Psi^{(k)}\right) = \log|\Sigma_{Z_0}| + \mathrm{tr}\left\{\Sigma_{Z_0}^{-1}\left[\left(Z_0^T - \mu_0\right)\left(Z_0^T - \mu_0\right)' + P_0^T\right]\right\}$$

The third term in (7) can be written as

$$Q\left(\Psi_Z, \Psi^{(k)}\right) = T\log|\Sigma_\eta| + \mathrm{tr}\left[\Sigma_\eta^{-1}\left(S_{11} - S_{10}G' - GS_{10}' + GS_{00}G'\right)\right]$$

where

$$S_{11} = \sum_{t=1}^{T} Z_t^T\left(Z_t^T\right)' + P_t^T$$
$$S_{10} = \sum_{t=1}^{T} Z_t^T\left(Z_{t-1}^T\right)' + P_{t,t-1}^T$$
$$S_{00} = \sum_{t=1}^{T} Z_{t-1}^T\left(Z_{t-1}^T\right)' + P_{t-1}^T$$

The last term in (7) is given by

$$Q\left(\Psi_W, \Psi^{(k)}\right) = \sum_{p=1}^{c} T\log|\Sigma_p| + \sum_{t=1}^{T} \mathrm{tr}\left\{\Sigma_p^{-1}\left[A_{p,t}^T + W_{p,t}^T\left(W_{p,t}^T\right)'\right]\right\}$$

4.3 The M-step

At the M-step, $Q\left(\Psi, \Psi^{(k)}\right)$ is maximized with respect to Ψ, and $\Psi^{(k+1)}$ is chosen so that

$$\Psi^{(k+1)} = \arg\max Q\left(\Psi, \Psi^{(k)}\right)$$

In this case the maximization can mostly be solved in closed form and the solution of $\frac{\partial Q}{\partial \mathbf{\Psi}} = 0$ is obtained by partitioning the parameter vector $\Psi = \left\{\mathring{\Psi}, \breve{\Psi}\right\}$, where $\mathring{\Psi} = \{\Psi_Y, \Psi_{Z_0}, \Psi_Z\}$, $\breve{\Psi} = \Psi_W$. The matrix Σ_{Z_0} is assumed constant.

Extending the results of Fassò & Cameletti (2008), the closed-form estimation formulas for the parameters in $\mathring{\Psi}$ are given by

$$\begin{aligned}
\hat{\beta}^{(k+1)} &= \left[\sum_{t=1}^{T} X_t' \left(\Sigma_0^{(k)}\right)^{-1} X_t\right]^{-1} \cdot \\
&\quad \cdot \left[\sum_{t=1}^{T} X_t' \left(\Sigma_0^{(k)}\right)^{-1} \left(Y_t^{(k)} - K Z_t^{T,(k)} - \bar{W}_t^{T,(k)}\right)\right] \\
\hat{\mu}_0^{(k+1)} &= Z_0^{T,(k)} \\
\hat{G}^{(k+1)} &= S_{10}^{(k)} \left(S_{00}^{(k)}\right)^{-1} \\
\hat{\Sigma}_\eta^{(k+1)} &= \frac{1}{T}\left[S_{11}^{(k)} - S_{10}^{(k)} G' - G\left(S_{10}^{(k)}\right)' + G S_{00}^{(k)} G'\right]
\end{aligned}$$

Moreover, the closed form for the $q \times q$ diagonal variance matrix of the instrumental error ε is given by:

$$\hat{\Gamma}_0^{(k+1)} = diag\left(\frac{1}{n_1} \operatorname{tr} \bar{\Omega}^{1,1,(k)}, ..., \frac{1}{n_q} \operatorname{tr} \bar{\Omega}^{q,q,(k)}\right) \tag{8}$$

where

$$\bar{\Omega}^{(k)} = \left(\bar{\Omega}^{i,j,(k)}\right)_{i,j=1,\ldots,q} = \frac{1}{T}\sum_{t=1}^{T} \Omega_t^{(k)} = \frac{1}{T}\sum_{t=1}^{T} \bar{e}_t \bar{e}_t' + \bar{A}_t^T + K P_t^T K'$$

and $\Omega_t^{(k)}$ is the $N \times N$ dimensional sample variance matrix of ε at time t and $k-th$ step. The temporal average is a consequence of the time-invariance property assumed for the distribution of ε. The proof of (8) is given in Appendix (A.3.1).

For what concerns V_p, a closed form estimation formula, obtained from the minimization of $Q\left(\Psi_{W_p}, \Psi^{(k)}\right)$, cannot be easily derived since, unless this is an isotopic case, it is diffi cult to write the$q \times q$ matrix V_p as a function of the $N \times N$ matrix $\Sigma_p(V_p, \theta_p, H)$ in closed form. For this reason, a closed-form estimation formula is proposed here for V_p based on unconstrained minimization of $Q\left(\Psi_{W_p}, \Psi^{(k)}\right)$ and least square matrix approximation. According to the proof given in Appendix $(A.3.2)$, the M-step can be approximated by

$$\tilde{V}_p^{(k+1)} = \left(\frac{\operatorname{tr}\left[\rho_p\left(H^{i,j}, \theta_p^{(k)}\right) \bar{U}_p^{i,j,(k)}\right]}{\operatorname{tr}\left[\rho_p\left(H^{i,j}, \theta_p^{(k)}\right)^2\right]} \right)_{i,j=1,...,q} \qquad p = 1, ..., c \tag{9}$$

where

$$\bar{U}_p^{(k)} = \left(\bar{U}_p^{i,j,(k)}\right)_{i,j=1,...,q} \tag{10}$$

$$= \frac{1}{T}\sum_{t=1}^{T} U_{p,t}^{(k)} = \frac{1}{T}\sum_{t=1}^{T}\left[A_{p,t}^{T,(k)} + W_{p,t}^{T,(k)}\left(W_{p,t}^{T,(k)}\right)'\right] \tag{11}$$

and $U_{p,t}^{(k)}$ is the sample $N \times N$ cross-covariance matrix of W_p at time t and step $k - th$. Note that the temporal average is a consequence of the time-invariance property assumed for the distribution of $W_{p,t}$.

Numerical optimization methods are used in order to update the parameter vector θ_p when a closed-form solution cannot be found. This gives

$$\hat{\theta}_p^{(k+1)} = \arg\min_{\theta_p}\left\{\log\left|\Sigma_p^{(k)}\right| + \operatorname{tr}\left[\left(\Sigma_p^{(k)}\right)^{-1}\bar{U}_p^{(k)}\right]\right\},\ p = 1, ..., c$$

where $\Sigma_p^{(k)} = \Sigma_p^{(k)}(V_p, \theta_p, H)$. Of course, it is possible to estimate V_p by numerical optimization too, giving

$$\left[\hat{V}_p^{(k+1)}, \hat{\theta}_p^{(k+1)}\right] = \arg\min_{V_p, \theta_p}\left\{\log\left|\Sigma_p^{(k)}\right| + \operatorname{tr}\left[\left(\Sigma_p^{(k)}\right)^{-1}\bar{U}_p^{(k)}\right]\right\},\ p = 1, ..., c \tag{12}$$

It must be noted, however, that the matrix V_p is composed by $\left(q^2 + q\right)/2$ independent elements and numerical optimization becomes cumbersome when q is high.

5 Monte Carlo simulation study

In order to evaluate the EM algorithm of Section 4.1, a large simulation study is implemented. Two aspects are taken into consideration: how well the parameter vector Ψ can be estimated from data and how starting values may influence the estimation procedure. The first question is related to the spatial and temporal structure of data, namely the length T of the temporal frame and the number of sites in S. The second question is considered by forcing different perturbations on the starting values. In particular we consider both random perturbations and systematic offsets.

Moreover we compare the exact and approximate formulas given by equations (12) and (19) in terms of bias and effi ciency.

5.1 Simulation of realistic data

Although different spatio-temporal structures could be simulated, we test the method of the previous section in connection to the realistic case of air quality modelling and mapping discussed in more detail by Fassò et al. (2009). This is interesting not only because of the plausible parameters and correlations involved but also because the size of the covariance matrices involved is rather large and computational feasibility may become a relevant issue.

In particular, the application concerns dynamical mapping of the concentration of particulate matter (PM_{10}) over the north-west part of Italy depicted in Figure 1. The PM_{10} concentration is collected every day by a ground level monitoring network, S_{PM}, composed of 84 sites in area D_3, 360×220 km in size. Also see Bodnar et al. (2008).

The second variable considered is Aerosol Optical Thickness (AOT) which is collected by remote sensing and is disposed on a 10×10 km regular grid over area D_3 resulting in a grid of $|S_{AOT}| = 858$ sites. The two variables, PM_{10} concentration and AOT, are never observed at the same site, giving an entirely heterotopic configuration.

Moreover, two known covariates are used, namely the meteorological variable known as mixing height, acronymized by MH (see Dandou et al., 2002) and land elevation, LE. With these data, the daily temporal frame considered is given by the first $T = 240$ days of year 2006.

Now, in order to define the model to be simulated with equations of Section 2, we use the results of a preliminary analysis. This gives a model setup characterized by a univariate latent temporal state Z, that is $d = 1$.

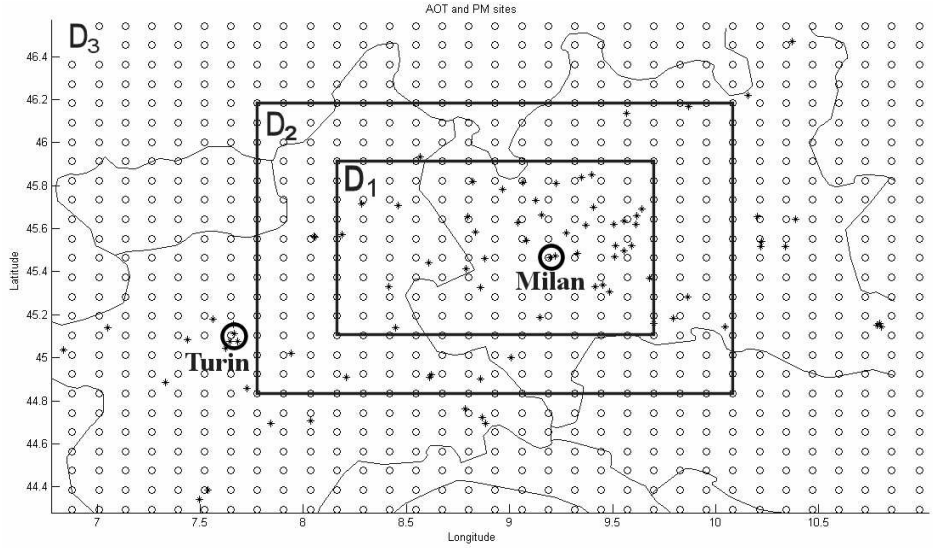

Figure 1: Study area with AOT grid (circles) and PM_{10} monitoring network (stars).

The related transition matrix G is represented by a scalar g, satisfying the condition $|g| < 1$, and matrix K is set to scalar 1.

Moreover, the linear coregionalization model for the latent random spatial effect consists of one component ($c = 1$) and the correlation structure between different sites is based on the exponential model, that is

$$\rho(h, \theta_1) = \exp\left(-\frac{h}{\theta_1}\right)$$

Finally, the parameter vector denoted by Ψ^0 and appropriate for standardized data is reported in Table 1.

Using this realistic setup, the subsequent simulations, which are conditional on both the covariates and the site configuration, are obtained by running model of Section 2 with conditional distributions given by $(5) - (6)$ and parameter vector given by Ψ^0.

5.1.1 Design of experiment

In order to study the EM algorithm effi cacy in estimating the parameter vector Ψ^0, the simulation campaign is based on 25 different simulation setups,

denoted by π_i, and is reported in Table 3, where each simulation setup, π_i, has $M = 100$ Monte Carlo replications.

The main aspects considered are the dimension of area D, the number of days T and the variability of the starting values $\Psi_{i,j,0}$, where $i = 1, ..., 25$ is the index of the simulation setup π_i and $j = 1, ..., M$ is the index of each replication. Moreover, the exact solution and the approximated closed-form solution are compared by evaluating their relative effi ciency defined in (13).

We consider three different areas, which are reported in Table 2, characterized by a different dimension and a different number of sites (also see Figure 1). Moreover, we examine three temporal frames given by $T = 60, 120$ and 240.

For what concerns starting values, four cases are considered: starting values equal to the simulation vector parameter ($\Psi_{i,j,0} = \Psi^0$), starting values with random perturbation centred on the simulation vector parameter ($\Psi_{i,j,0} = \Psi^0 + \delta_j$), starting values with offset ($\Psi_{i,j,0} = \Psi^0 + \phi$) and both starting values with offset and random perturbation ($\Psi_{i,j,0} = \Psi^0 + \delta_j + \phi$).

Each element of the vector δ_j is randomly generated from a uniform distribution $U(0.7\psi_r^0, 1.3\psi_r^0)$ where ψ_r^0 is the $r-th$ element of Ψ^0. On the other hand, elements of the vector ϕ are constant for each $j = 1, ..., M$ and their values represent an either positive or negative offset having a magnitude of 20% of the value of the respective element of the vector Ψ^0.

5.1.2 EM estimation

Once the data set is simulated, the EM algorithm is triggered with particular starting values depending on the simulation π_i considered. The maximum iteration number of the EM algorithm is set to 250 while the exit condition is based on the convergence criterion $\frac{\left\|\Psi^{(k+1)}-\Psi^{(k)}\right\|}{\left\|\Psi^{(k)}\right\|} < 5 \cdot 10^{-4}$. The solution $\hat{\Psi}_{i,j}$ is the $j-th$ replication of the estimate of Ψ^0 within simulation π_i. In particular, $\hat{\Psi}_{i,j}$ denotes the result of the exact algorithm while $\tilde{\Psi}_{i,j}$ denotes the result of the approximate solution.

5.2 Simulation results

For each simulation π_i considered, the sets $\hat{\Psi}_i = \left\{\hat{\Psi}_{i,1}, ..., \hat{\Psi}_{i,M}\right\}$ and $\tilde{\Psi}_i = \left\{\tilde{\Psi}_{i,1}, ..., \tilde{\Psi}_{i,M}\right\}$ are analysed in order to evaluate bias and variability. These

are induced, on the one hand, from the estimation procedure of the EM algorithm, and on the other, from the natural variability of the data sets considered.

By considering $\bar{\Psi}_i$ as the arithmetic mean vector of $\hat{\Psi}_i$ ($\tilde{\Psi}_i$ for the approximate solution), the bias is evaluated by the metric $b_i = \left\| \bar{\Psi}_i - \Psi^0 \right\| / \left\| \Psi^0 \right\|$. If $s_{i,r}$ is the sample standard deviation of $\hat{\Psi}_i^r = \left\{ \hat{\Psi}_{i,1}^r, ..., \hat{\Psi}_{i,M}^r \right\}$, with $\hat{\Psi}_{i,j}^r$ the $r-th$ element of the vector $\hat{\Psi}_{i,j}$, then the average coeffi cient of variation is $CV_i = \frac{1}{R} \sum_{r=1}^{R} s_{i,r} / \left| \bar{\Psi}_i^r \right|$, where $\bar{\Psi}_i^r$ is the $r-th$ element of the vector $\bar{\Psi}_i$ and R is the dimension of the parameter Ψ^0.

The relative effi ciency between the exact and the approximate solutions for estimating the matrix V_1 is defined as

$$eff_i = \frac{1}{R} \sum_{r=1}^{R} \frac{RMSE\left(\hat{\Psi}_i^r\right)}{RMSE\left(\tilde{\Psi}_i^r\right)} \tag{13}$$

Table 4 reports detailed simulation results for π_{17}. Note that the bias is small if compared to the value of the respective element of the vector Ψ^0. Also note that standard deviations and RMSEs are higher for parameters referring to the PM variable. This is a consequence of the different number of sites over which the two variables are observed ($|S_{AOT}| \gg |S_{PM}|$).

Concerning the DOE of Tab 3, aggregated results are reported in Table 5, 6 and 7 for $T = 60$, 120 and 240 respectively.

The first point worth to be pointing out is the overall stability of both EM algorithm solutions proposed, which is confirmed by the convergence for all $3'800 = M \times 38$ Monte Carlo simulations, irrespective on the initial values.

Secondly, as expected, the estimation error, as assessed by the total RMSE, decreases with both geographic area and temporal frame sizes, $|S|$ and T respectively. This convergence result seems not to be influenced by the starting value perturbation. Hence the proposal results to be robust with respect to initial values.

Finally considering the approximate solution of equation (19), it is interesting to note that bias and RMSE are generally speaking rather small in all simulation setups. Moreover, the effi ciency with respect to the exact solution of equation (12) is stable, which confirms the idea that the approximate solution gives a consistent estimate. The effi ciency value, for our realistic simulation setup, has a lower bound close to 0.7.

6 Conclusions

This paper develops the EM algorithm for the linear coregionalization model (LCM) with dynamic components, which is useful for spatio-temporal multivariate data observed on non-collocated sites.

The extensive Monte Carlo simulations, based on a large realistic data set, show that the method has good stability properties even with rather large covariance matrices. Moreover it is quite robust with respect to initial values.

In order to improve computational performance for large LCMs, the exact formulas are coupled with approximate closed-form formulas for the coregionalization matrices. According to our simulations, the approximate EM gives estimates which are consistent though sub-optimal.

We then suggest to use the previous two algorithms in series. According to this, the approximate approach is used at first and then the exact algorithm is fed with the approximate results as initial values.

A Appendix

A.1 Kalman smoother

At the generic $k-th$ iteration of the EM algorithm, the Kalman smoothed Z_t^T is used instead of the unknown state Z_t, following the forward and backward Kalman recursions whose formulas are given in this section.

Let Z_t^τ, P_t^τ and $P_{t,t-1}^\tau$ denote the mean, the variance and the lag-one covariance of Z_t, $t = 1, ..., T$ conditional on the observation matrix Y up to time τ, respectively. The Kalman filter equations for the predicted values are given by

$$\begin{cases} Z_t^{t-1} = GZ_{t-1}^{t-1} \\ P_t^{t-1} = \Sigma_\eta + GP_{t-1}^{t-1}G' \end{cases}$$

Moreover, the updating equations are given by

$$\begin{cases} Z_t^t = Z_t^{t-1} + J_t\left(Y_t - X_t\beta - KZ_t^{t-1}\right) \\ P_t^t = (I - J_tK)P_t^{t-1} \end{cases} \tag{14}$$

where the Kalman gain J_t is

$$J_t = P_t^{t-1}K'(KP_t^{t-1}K' + \Sigma_{\bar{W}})^{-1}$$

and $\Sigma_{\bar{W}} = \sum_{p=1}^{c} \Sigma_p$. Above recursions are based on the initial values given by $Z_0^0 = \mu_0$ and $P_0^0 = \Sigma_{Z_0}$.

With these results, the smoothed values are computed from the following backward recursion formulas for $t = T, ..., 1$:

$$Z_{t-1}^T = Z_{t-1}^{t-1} + P_{t,t-1}^T G \left(P_t^{t-1}\right)^{-1} \left[Z_t^T - Z_t^{t-1}\right] \tag{15}$$

$$P_{t-1}^T = P_{t-1}^{t-1} + P_{t-1}^{t-1} G \left(P_t^{t-1}\right)^{-1} \left[P_t^T - P_t^{t-1}\right] \left(P_{t-1}^{t-1} G \left(P_t^{t-1}\right)^{-1}\right)' \tag{16}$$

where the initial values Z_T^T and P_T^T are the outputs of the previously defined Kalman updating equations.

Moreover, the smoothed lag-one covariance is computed using the following equation for $t = T - 1, ..., 1$

$$\begin{aligned} P_{t,t-1}^T = {} & P_t^t \left(P_{t-1}^{t-1} G \left(P_t^{t-1}\right)^{-1}\right)' + \\ & + \left(P_t^t G \left(P_{t+1}^t\right)^{-1}\right) \left(P_{t+1,t}^T - G P_t^t\right) \left(P_{t-1}^{t-1} G \left(P_t^{t-1}\right)^{-1}\right)' \end{aligned}$$

and, for $t = T$, we have

$$P_{T,T-1}^T = \left(I - J_T K\right) G P_{T-1}^{T-1}$$

A.2 Latent spatial effect estimation

In this section, using the joint multivariate normality of W and Y, closed-form formulas for W are given which generalize the corresponding formulas of Zhang (2007) . To see this, note that, for a fixed $t \in \mathcal{T} = \{1, ..., T\}$ and for each $p = 1, ..., c$, we have that

$$\operatorname{cov}\left(W_{p,t}, Y_t\right) = \operatorname{cov}(W_{p,t}, W_{p,t}) = \Sigma_p$$

Moreover note that the expectation of $W_{p,t}$ conditionally on the observed data Y may be written as

$$W_{p,t}^T = E(W_{p,t} \mid Y) = E\left[E(W_{p,t} \mid Y, Z) \mid Y\right]$$

If Z_t is observed for each $t \in \mathcal{T}$, then

$$E(W_{p,t} \mid Y, Z) = E(W_{p,t} \mid Y_t, Z_t) \tag{17}$$

From the well known properties of the multivariate normal distribution, it now follows that

$$E(W_{p,t} \mid Y_t, Z_t) = \Sigma_p \Sigma^{-1} \left(Y_t - X_t\beta - KZ_t\right)$$

Conditioning on Y gives

$$\begin{aligned} W_{p,t}^T &= E\left[\Sigma_p \Sigma^{-1} \left(Y_t - X_t\beta - KZ_t\right) \mid Y\right] \\ &= \Sigma_p \Sigma^{-1} \left(Y_t - X_t\beta - KZ_t^T\right) \end{aligned}$$

In a similar way, given the observed data Y, the conditional variance of $W_{p,t}$ is given by

$$\begin{aligned} A_{p,t}^T &= \mathrm{Var}(W_{p,t} \mid Y) \\ &= \mathrm{Var}\left[E\left(W_{p,t} \mid Y, Z\right) \mid Y\right] + E\left[\mathrm{Var}\left(W_{p,t} \mid Y, Z\right)|Y\right] \end{aligned}$$

and

$$\mathrm{Var}\left(W_{p,t} \mid Y, Z\right) = \mathrm{Var}\left(W_{p,t} \mid Y_t, Z_t\right) = \Sigma_p - \Sigma_p \Sigma^{-1} \Sigma_p$$

$$\begin{aligned} A_{p,t}^T &= \mathrm{Var}\left[\Sigma_p \Sigma^{-1} \left(Y_t - X_t\beta - KZ_t\right) \mid Y\right] + \left[\Sigma_p - \Sigma_p \Sigma^{-1} \Sigma_p\right] \\ &= \Sigma_p \Sigma^{-1} K P_t^T K' \Sigma^{-1} \Sigma_p + \left[\Sigma_p - \Sigma_p \Sigma^{-1} \Sigma_p\right] \end{aligned}$$

where

$$\Sigma = \mathrm{Var}\left(Y_t\right) = \sum_{p=0}^{c} \Sigma_p$$

A.3 Expected log-likelihood derivatives

This section provides details about the expected log-likelihood derivatives implemented in order to obtain closed-form estimation formulas for $\hat{\Gamma}_0$ and $\tilde{V}_p$, $p = 1, ..., c$.

A.3.1 Derivation of $Q\left(\Psi_Y, \Psi^{(k)}\right)$

The expected log-likelihood $Q\left(\Psi_Y, \Psi^{(k)}\right)$ can be restated as follows

$$
\begin{aligned}
Q\left(\Psi_Y, \Psi^{(k)}\right) &= T \log |\Sigma_0| + \sum_{t=1}^{T} \operatorname{tr}\left[\Sigma_0^{-1}\left(\bar{e}_t \bar{e}_t' + \bar{A}_t^T + K P_t^T K'\right)\right] \\
&= T \log |\Sigma_0| + \operatorname{tr}\left[\Sigma_0^{-1} \sum_{t=1}^{T}\left(\bar{e}_t \bar{e}_t' + \bar{A}_t^T + K P_t^T K'\right)\right] \\
&= T \log |\Sigma_0| + T \operatorname{tr}\left(\Sigma_0^{-1} \bar{\Omega}\right)
\end{aligned}
$$

By assuming that ε has uncorrelated components, the $N \times N$ matrix Σ_0 happens to be block diagonal and

$$
\begin{aligned}
|\Sigma_0| &= \prod_{i=1}^{q} \left(\gamma_0^{i,i}\right)^{n_i} \\
\operatorname{tr}\left(\Sigma_0^{-1} \bar{\Omega}\right) &= \sum_{i=1}^{q} \frac{1}{\gamma_0^{i,i}} \operatorname{tr}\left(\bar{\Omega}^{i,i}\right)
\end{aligned}
$$

Then

$$
Q\left(\Psi_Y, \Psi^{(k)}\right) = T \sum_{i=1}^{q} n_i \log \gamma_0^{i,i} + T\left[\sum_{i=1}^{q} \frac{1}{\gamma_0^{i,i}} \operatorname{tr}\left(\bar{\Omega}^{i,i}\right)\right]
$$

Deriving $Q\left(\Psi_Y, \Psi^{(k)}\right)$ with respect to $\gamma_0^{i,i}$, $i = 1, ..., q$, yields to

$$
\frac{\partial Q\left(\Psi_Y, \Psi^{(k)}\right)}{\gamma_0^{i,i}} = \frac{T n_i}{\gamma_0^{i,i}} - \frac{T \operatorname{tr}\left(\bar{\Omega}^{i,i}\right)}{\left(\gamma_0^{i,i}\right)^2}
$$

and $Q\left(\Psi_Y, \Psi^{(k)}\right)$ is minimized by the $q \times q$ diagonal matrix $\hat{\Gamma}_0$ whose diagonal elements are given by

$$
\hat{\gamma}_0^{i,i} = \frac{1}{n_i} \operatorname{tr}\left(\bar{\Omega}^{i,i}\right)
$$

A.3.2 Derivation of $Q\left(\Psi_{W_p}, \Psi^{(k)}\right)$

As mentioned in Section 2, when partially or completely heterotopic settings are considered, closed-form estimation formulas for V_p cannot be easily obtained since it is diffi cult to write the$q \times q$ matrix V_p in closed-form from the $N \times N$ matrix $\Sigma_p(V_p, \theta_p, H)$. This section proposes a closed-form formula

for V_p based on unconstrained minimization of the expected log-likelihood $Q\left(\Psi_{W_p}, \Psi^{(k)}\right)$ and least square matrix approximation.

The $Q\left(\Psi_{W_p}, \Psi^{(k)}\right)$ can be re-written in a more compact way as follows

$$\begin{aligned} Q\left(\Psi_{W_p}, \Psi^{(k)}\right) &= T \log\left|\tilde{\Sigma}_p\right| + \sum_{t=1}^{T} \operatorname{tr}\left\{\tilde{\Sigma}_p^{-1}\left[A_{p,t}^T + W_{p,t}^T\left(W_{p,t}^T\right)'\right]\right\} \\ &= T \log\left|\tilde{\Sigma}_p\right| + \operatorname{tr}\left[\tilde{\Sigma}_p^{-1}\sum_{t=1}^{T} A_{p,t}^T + W_{p,t}^T\left(W_{p,t}^T\right)'\right] \\ &= -T \log\left|\tilde{\Sigma}_p^{-1}\right| + T \operatorname{tr}\left\{\tilde{\Sigma}_p^{-1}\bar{U}_p\right\} \end{aligned}$$

where $\tilde{\Sigma}_p$ is a generic symmetric semi-positive definite matrix. The derivative of $Q\left(\Psi_{W_p}, \Psi^{(k)}\right)$ with respect to $\tilde{\Sigma}_p^{-1}$ is given by

$$\frac{\partial Q\left(\Psi_{W_p}, \Psi^{(k)}\right)}{\partial \tilde{\Sigma}_p^{-1}} = -T\tilde{\Sigma}_p + T\bar{U}_p$$

and $Q\left(\Psi_{W_p}, \Psi^{(k)}\right)$ is minimized at

$$\tilde{\Sigma}_p = \bar{U}_p \tag{18}$$

The solution in (18) is not constrained, namely it does not reflect the coregionalization structure described in Section 2. In particular, the matrix $\Sigma_p(V_p, \theta_p, H)$ must respect the block structure of (4), in which case there is, in general, no V_p that satisfies the equation $\Sigma_p(V_p, \theta_p, H) = \bar{U}_p$.

We ask for $\tilde{V}_p$ that minimizes the difference between the matrix $\Sigma_p(V_p, \theta_p, H)$ and the matrix $\bar{U}_p$. A possible solution is to choose matrix $\tilde{V}_p$ which minimizes the Frobenius norm, namely

$$\tilde{V}_p = \arg\min_{V_p} \left\|\Sigma_p(V_p, \theta_p, H) - \bar{U}_p\right\|_F \tag{19}$$

Exploiting the block structure of $\Sigma_p(V_p, \theta_p, H)$ and $\bar{U}_p$ (see eq. 10), the minimization in (19) can be implemented by considering each element $v_p^{i,j}$, $i, j = 1, ..., q$, of the matrix V_p separately:

$$\tilde{v}_p^{i,j} = \arg\min_{v_p^{i,j}} \left\|v_p^{i,j}\rho_p(H^{i,j}, \theta_p) - \bar{U}_p^{i,j}\right\|_F$$

With the following substitutions $v = v_p^{i,j}$, $\rho = vec(\rho_p(H^{i,j}, \theta_p))$, $u = vec(\bar{U}_p^{i,j})$, it is easily seen that

$$\tilde{v} = \frac{\rho' u}{\rho' \rho}$$

Now, from $vec(A')'vec(B) = \mathrm{tr}(AB)$, it follows that

$$\tilde{v}_p^{i,j} = \frac{\mathrm{tr}\left[\rho_p(H^{i,j}, \theta_p)\bar{U}_p^{i,j}\right]}{\mathrm{tr}\left[\rho_p(H^{i,j}, \theta_p)^2\right]} = \frac{\mathrm{tr}\left[\rho_p(H^{i,j}, \theta_p)\bar{U}_p^{i,j}\right]}{\|\rho_p(H^{i,j}, \theta_p)\|_F^2}$$

and, by collecting each $\tilde{v}_p^{i,j}$ within the matrix $\tilde{V}_p$, the solution in (9) is obtained.

Acknowledgements

The work is partially supported by PRIN 2006 project n.2006131039 *Statistical analysis of spatial and temporal dynamics and health impact of particulate matter*s and Regione Piemonte project CIPE 2004 *Statistical methods and spatio-temporal models for atmospheric pollution monitoring.*

References

Bishop TFA, Lark RM. 2008. A comparison of parametric and non-parametric methods for modelling a coregionalization. Geoderma **148**: 13-24.

Bodnar O, Cameletti M, Fassò A, Schmid W. 2008. Comparing air quality in Italy, Germany and Poland using BC indexes. Atmospheric Environment **42**: 8412-8421.

Bodnar O, Schmid W. 2009. Nonlinear locally weighted kriging prediction for spatio-temporal environmental processes. Environmetrics doi:10.1002/env.1005

Brown PE, Diggle PJ, Lord ME, Young P. 2001. Space-time calibration of radar rainfall data. Journal of the Royal Statistical Society, Series C **50**: 221-241.

Dandou A, Bossioli E, Tombrou M, Sifakis N, Paronis D, Soulakellis N, Sarigiannis D. 2002. The Importance of Mixing Height in Characterising Pollution Levels from Aerosol Optical Thickness Derived by Satellite. International Journal of Water, Air and Soil Pollution **2**: 17-28

De Iaco S, Palma M, Posa D. 2005. Modeling and prediction of multivariate space-time random fields. Computational Statistics and Data Analysis **48**: 525-547.

Di Nicolantonio W, Cacciari A, Bolzacchini F, Ferrero L, Volta L, Pisoni E. 2007. Modis aerosol optical properties over North Italy for estimating surface-level PM2.5. Proc. of the Envisat Symposium 2007, ESA SP-636

Fassò A, Cameletti M. 2009a. A unified statistical approach for simulation, modelling, analysis and mapping of environmental data. Simulation: Transactions of the Society for Modeling and Simulation International, accepted. OnlineFirst, doi:10.1177/0037549709102150

Fassò A, Cameletti M. 2009b. The EM algorithm in a distributed computing environment for modelling environmental space-time data. Environmental Modelling & Software **24**: 1027-1035.

Fassò A, Finazzi F, D'Ariano C. 2009. Integrating satellite and ground level data for air quality monitoring and dynamical mapping. GRASPA Working Paper n.34. (www.graspa.org).

Lark RM, Papritz A. 2003. Fitting a linear model of coregionalization for soil properties using simulated annealing. Geoderma **115**: 245-260

Le DN, Zidek JV. 2006. Statistical Analysis of Environmental Space-Time Processes. Springer, New York.

Liu C, Koike K. 2007. Extending Multivariate Space-Time Geostatistics for Environmental Data Analysis. Mathematical Geology **39**: 289-305

McLachlan GJ, Krishnan T. 1997. The EM Algorithm and Extensions. Wiley, New York

Orasi A, Jona Lasinio G., Ferrari C. 2009. Comparison of calibration methods for the reconstruction of space-time rainfall fields during a rain enhancement experiment in Southern Italy. Environmetrics **20**: 812-834

Rouhani S, Wackernagel H. 1990. Multivariate geostatistical approach to space-time data analysis: Water Resources Res. **36**: 585-591.

Wackernagel H. 1998. Multivariate Geostatistics: An Introduction with Applications. Springer-Verlag Inc. New York

Wang J, Christopher SA. 2003. Intercomparison between satellite-derived aerosol optical thickness and PM2.5 mass: Implications for air quality studies. Geophys. Res. Lett., 30, doi:10.1029/2003GL018174.

Zhang H. 2007. Maximum-likelihood estimation for multivariate spatial linear coregionalization models. Environmetrics **18**: 125-139

$\beta_{AOT,MH}$	$\beta_{AOT,LE}$	$\beta_{PM,MH}$	$\beta_{PM,LE}$	$\gamma_0^{AOT,AOT}$	$\gamma_0^{PM,PM}$
-0.3	-0.5	-0.4	-0.6	0.6	0.6
g	σ_η^2	$v_1^{AOT,AOT}$	$v_1^{AOT,PM}$	$v_1^{PM,PM}$	θ_1
0.8	0.2	1.0	0.6	1.0	60.0

Table 1: Value of the parameter vector Ψ^0.

$Area$	$Width\ (km)$	$Height\ (km)$	$\lvert S_{AOT}\rvert$	$\lvert S_{PM}\rvert$
D_1	130	100	130	42
D_2	190	160	304	60
D_3	330	260	858	84

Table 2: Configuration of the areas considered.

π	D	T	Random	Offset	π	D	T	Random	Offset
1	D_1	60	*no*	*no*	13	D_2	60	*no*	*no*
2	D_1	60	*yes*	*no*	14	D_2	60	*yes*	*no*
3	D_1	60	*no*	*yes*	15	D_2	60	*no*	*yes*
4	D_1	60	*yes*	*yes*	16	D_2	60	*yes*	*yes*
5	D_1	120	*no*	*no*	17	D_2	120	*no*	*no*
6	D_1	120	*yes*	*no*	18	D_2	120	*yes*	*no*
7	D_1	120	*no*	*yes*	19	D_2	120	*no*	*yes*
8	D_1	120	*yes*	*yes*	20	D_2	120	*yes*	*yes*
9	D_1	240	*no*	*no*	21	D_2	240	*no*	*no*
10	D_1	240	*yes*	*no*	22	D_2	240	*yes*	*no*
11	D_1	240	*no*	*yes*	23	D_2	240	*no*	*yes*
12	D_1	240	*yes*	*yes*	24	D_2	240	*yes*	*yes*
					25	D_3	60	*no*	*no*

Table 3: Simulation campaign. Legend: π: simulation index; D: considered area; T: considered temporal frame; Random: randomized starting values; Offset: starting values with offset.

	Ψ^0	$\bar{\Psi}$	$bias$	std	$RMSE$	$q_{0.025}$	$q_{0.975}$
$\beta_{AOT,MH}$	-0.3	-0.301	-0.001	0.010	0.010	-0.327	-0.278
$\beta_{AOT,LE}$	-0.5	-0.500	$+0.000$	0.012	0.012	-0.526	-0.465
$\beta_{PM,MH}$	-0.4	-0.403	-0.003	0.027	0.028	-0.464	-0.346
$\beta_{PM,LE}$	-0.6	-0.604	-0.004	0.037	0.038	-0.675	-0.532
$\gamma_0^{AOT,AOT}$	$+0.6$	$+0.597$	-0.003	0.006	0.007	$+0.583$	$+0.609$
$\gamma_0^{PM,PM}$	$+0.6$	$+0.593$	-0.007	0.015	0.017	$+0.560$	$+0.620$
g	$+0.8$	$+0.776$	-0.024	0.077	0.080	$+0.598$	$+0.887$
σ_η^2	$+0.2$	$+0.209$	$+0.009$	0.047	0.048	$+0.132$	$+0.322$
$v_1^{AOT,AOT}$	$+1.0$	$+1.000$	$+0.000$	0.043	0.043	$+0.927$	$+1.082$
$v_1^{AOT,PM}$	$+0.6$	$+0.596$	-0.004	0.034	0.035	$+0.531$	$+0.654$
$v_1^{PM,PM}$	$+1.0$	$+1.001$	$+0.001$	0.048	0.048	$+0.915$	$+1.103$
θ_1	$+60.0$	$+59.429$	-0.571	3.228	3.278	$+53.203$	$+66.141$

Table 4: Simulation results for π_{17} ($D = D_2, T = 120$).

				D_1				D_2	
				$Offset$				$Offset$	
				absent	*present*			*absent*	*present*
			bias	0.048	0.010		*bias*	0.005	0.005
	a	Ex	CV	0.127	0.149	Ex	CV	0.095	0.110
	b		$RMSE$	7.909	9.743		$RMSE$	5.334	5.605
	s		*iter*	37	68		*iter*	22	77
	e		*bias*	0.003			*bias*	0.022	
R	n		CV	0.168			CV	0.136	
a	t	Ap	$RMSE$	9.910		Ap	$RMSE$	9.288	
n			eff	0.792			eff	0.704	
d			*iter*	44			*iter*	40	
o			*bias*	0.019	0.017		*bias*	0.012	0.007
m	p	Ex	CV	0.145	0.159	Ex	CV	0.119	0.130
	r		$RMSE$	7.502	8.042		$RMSE$	4.943	6.546
p	e		*iter*	57	72		*iter*	56	67
e	s		*bias*	0.005			*bias*	0.012	
r	e		CV	0.194			CV	0.154	
t	n	Ap	$RMSE$	11.461		Ap	$RMSE$	7.878	
u	t		eff	0.771			eff	0.767	
r			*iter*	51			*iter*	45	
b				D_3					
a				$Offset$					
t				*absent*	*present*				
i			*bias*	0.004					
o	a	Ex	CV	0.065					
n	b		$RMSE$	3.278					
	s		*iter*	16					
	e		*bias*	0.014					
	n		CV	0.102					
	t	Ap	$RMSE$	5.241					
			eff	0.697					
			iter	27					

Table 5: Aggregated simulation results for $T = 60$ and different perturbations on the starting values. Legend: Ex: exact solution; Ap: closed-form approximated solution; *bias*: average bias; CV: average coeffi cient of variation; $RMSE$: total RMSE; *iter*: average iteration number to convergence; eff: relative effi ciency between exact and approximated solution.

				D_1				D_2	
				$Offset$				$Offset$	
R				$absent$	$present$			$absent$	$present$
a			$bias$	0.026	0.027		$bias$	0.009	0.006
n	a	Ex	CV	0.085	0.102	Ex	CV	0.062	0.083
d	b		$RMSE$	5.262	6.250		$RMSE$	3.641	3.773
o	s		$iter$	29	80		$iter$	21	80
m	e		$bias$	0.033			$bias$	0.001	
	n		CV	0.116			CV	$0,087$	
p	t	Ap	$RMSE$	7.502		Ap	$RMSE$	4.895	
e			eff	0.771			eff	0.753	
r			$iter$	41			$iter$	34	
t			$bias$	0.029	0.015		$bias$	0.015	0.002
u	p	Ex	CV	0.110	0.112	Ex	CV	0.087	0.091
r	r		$RMSE$	6.017	6.980		$RMSE$	4.191	4.498
b	e		$iter$	62	76		$iter$	52	72
a	s		$bias$	0.027			$bias$	0.001	
t	e		CV	0.135			CV	0.110	
i	n	Ap	$RMSE$	8.201		Ap	$RMSE$	5.588	
o	t		eff	0.851			eff	0.824	
n			$iter$	50			$iter$	44	

Table 6: Aggregated simulation results for $T = 120$ and different perturbations on the starting values. Legend: see Table 5.

				D_1				D_2	
				$Offset$				$Offset$	
R				$absent$	$present$			$absent$	$present$
a			$bias$	0.018	0.019		$bias$	0.005	0.008
n	a	Ex	CV	0.056	0.067	Ex	CV	0.042	0.052
d	b		$RMSE$	3.404	3.727		$RMSE$	2.431	2.626
o	s		$iter$	24	83		$iter$	15	81
m	e		$bias$	0.013			$bias$	0.003	
	n		CV	0.087			CV	0.062	
p	t	Ap	$RMSE$	6.125		Ap	$RMSE$	3.704	
e			eff	0.703			eff	0.701	
r			$iter$	38			$iter$	29	
t			$bias$	0.032	0.015		$bias$	0.014	0.005
u	p	Ex	CV	0.087	0.088	Ex	CV	0.073	0.067
r	r		$RMSE$	4.538	4.297		$RMSE$	3.014	2.932
b	e		$iter$	62	75		$iter$	53	77
a	s		$bias$	0.005			$bias$	0.008	
t	e		CV	0.104			CV	0.089	
i	n	Ap	$RMSE$	5.710		Ap	$RMSE$	4.220	
o	t		eff	0.877			eff	0.844	
n			$iter$	50			$iter$	47	

Table 7: Aggregated simulation results for $T = 240$ and different perturbations on the starting values. Legend: see Table 5.