A new procedure for fitting a multivariate space-time linear coregionalization model [1]

Sandra De Iaco
Dip.to di Scienze Economiche e Matematico-Statistiche, Facolta' di Economia, Universita' del Salento, Italy, sandra.deiaco@unisalento.it

Monica Palma
Dip.to di Scienze Economiche e Matematico-Statistiche, Facolta' di Economia, Universita' del Salento, Italy.

Donato Posa
Dip.to di Scienze Economiche e Matematico-Statistiche, Facolta' di Economia, Universita' del Salento, Italy.

Abstract: New classes of cross-covariance functions have been recently proposed, nevertheless the linear coregionalization model (*LCM*) is still of interest and widely applied. In this paper, a new fitting procedure of the space-time *LCM* (*ST-LCM*) using the generalized product-sum model is proposed. This procedure is based on the well known algorithm of matrix simultaneous diagonalization, applied on the sample matrix variograms computed for multiple spatial-temporal lags.

Keywords: spatial-temporal correlation, product-sum variogram model, linear coreginalization model.

1 Introduction

The *LCM*, firstly introduced by Matheron in 1982, is still one of the most utilized models for multivariate spatial and spatial-temporal data analysis (Zhang, 2007; Babak and Deutsch, 2009; Emery, 2010). However, in the space-time context several theoretical and practical aspects must be considered, such as the fitting process. In geostatistics, there is a wide literature concerning the *LCM* fitting stage (Goulard and Voltz, 1989; Lark and Papritz, 2003). In this paper, a new fitting procedure of the *ST-LCM* using the generalized product-sum variogram model is proposed. It is shown that the simultaneous diagonalization of the sample matrix variograms is useful to identify the basic components of the coregionalization model.

[1] Supported by Fondazione Cassa di Risparmio di Puglia.

2 Multivariate space-time random field

Given a second-order stationary vector-valued space-time random function (*STRF*) $\{\mathbf{Z}(\mathbf{s},t),\ (\mathbf{s},t) \in D \times T \subseteq \mathbb{R}^{d+1}\}$, with $\mathbf{Z}(\mathbf{s},t) = [Z_1(\mathbf{s},t),\ldots,Z_p(\mathbf{s},t)]^T$, $p \geq 2$, where $\mathbf{s} = (s_1, s_2, \ldots, s_d) \in D$ (generally, $d \leq 3$), denotes the spatial coordinates and $t \in T$ is the temporal coordinate, the cross-variogram of two space-time random functions $Z(\mathbf{s},t)$ and $Z(\mathbf{s}',t')$ exists and depends on the space-time separation vector $\mathbf{h} = (\mathbf{h}_s, h_t)$, with $\mathbf{h}_s = (\mathbf{s} - \mathbf{s}')$ and $h_t = (t - t')$. As in the spatial context, a second-order stationary multivariate *STRF* can be modelled as a *ST-LCM*. Hence, the variogram matrix can be written as

$$\mathbf{\Gamma}(\mathbf{h}) = \mathbf{\Gamma}(\mathbf{h}_s, h_t) = \sum_{l=1}^{L} \mathbf{B}_l \, g_l(\mathbf{h}_s, h_t), \tag{1}$$

where $\mathbf{B}_l = [b^l_{\alpha\beta}],\ l = 1,\ldots,L,\ \alpha,\beta = 1,\ldots,p$, are positive definite $(p \times p)$ matrices, commonly known as *coregionalization matrices*, while $g_l(\mathbf{h}_s, h_t),\ l = 1,\ldots,L$, are basic space-time variograms associated with the L scales of variability.
In De Iaco et al. (2003, 2005), each space-time basic variogram is modelled as a generalized product-sum model (De Iaco et al., 2001):

$$g_l(\mathbf{h}_s, h_t) = \gamma_l(\mathbf{h}_s, 0) + \gamma_l(\mathbf{0}, h_t) - k_l \, \gamma_l(\mathbf{h}_s, 0) \, \gamma_l(\mathbf{0}, h_t) \ , \quad l = 1,\ldots,L, \tag{2}$$

where $\gamma_l(\mathbf{h}_s, 0)$ and $\gamma_l(\mathbf{0}, h_t)$ are the spatial and temporal marginal variogram models, respectively, while parameters $k_l, l = 1,\ldots,L$, are given by:

$$k_l = \frac{sill[\gamma_l(\mathbf{h}_s,0)] + sill[\gamma_l(\mathbf{0},h_t)] - sill[g_l(\mathbf{h}_s,h_t)]}{sill[\gamma_l(\mathbf{h}_s,0)] \cdot sill[\gamma_l(\mathbf{0},h_t)]}, \qquad l = 1,\ldots,L. \tag{3}$$

By substituting (2) in (1), the *ST-LCM* based on the generalized product-sum variogram models is determined by two marginal *LCM*'s:

$$\mathbf{\Gamma}(\mathbf{h}_s, 0) = \sum_{l=1}^{L} \mathbf{B}_l \, \gamma_l(\mathbf{h}_s, 0), \qquad \mathbf{\Gamma}(\mathbf{0}, h_t) = \sum_{l=1}^{L} \mathbf{B}_l \, \gamma_l(\mathbf{0}, h_t). \tag{4}$$

Note that other space-time variogram models (Gneiting, 2002; Ma, 2002; Stein, 2005; Porcu et al., 2008) can be used to describe the basic components of the *ST-LCM*. However, the flexibility of the product-sum variogram, in estimating and modeling the spatial-temporal variability, is often convenient (De Iaco et al. 2003, 2005).

3 Fitting a *ST-LCM*

After a brief review of the usual fitting process of the *ST-LCM* using the generalized product-sum model, the new, more flexible, fitting procedure is discussed.

The usual fitting procedure

In De Iaco et al. (2003) the process of fitting a ST-LCM using a generalized product-sum variogram model, was developed as follows.

1. Compute the empirical marginal direct variograms, in space and in time, for all the p variables under study and then fit nested variogram models. At this step, the diagonal elements of each matrix $\mathbf{B}_l$, $l = 1, \ldots, L$, are determined as well as the marginal basic structures $\gamma_l(\mathbf{h}_s, 0)$ and $\gamma_l(\mathbf{0}, h_t)$, $l = 1, \ldots, L$.

2. Determine the marginal cross-variograms and the off-diagonal elements of the matrices (4), ensuring that each matrix $\mathbf{B}_l$ is positive definite.

3. In order to complete the modeling of $g_l(\mathbf{h}_s, h_t)$, $l = 1, \ldots, L$, the k_l parameters must be determined. Hence, the space-time variogram surfaces are computed and fitted to product-sum nested models.

Using this procedure, different practical problems have to be faced: a) the identification of the $b^l_{ij}, i, j = 1, \ldots, p$, elements of the matrices $\mathbf{B}_l$, $l = 1, \ldots, L$, since for a fixed l, these coefficients must be the same for the marginal space and time variograms; b) the estimation of parameters k_l, with $l = 1, \ldots, L$.

The new fitting procedure

Given the multivariate space-time data set concerning the p variables (with $p \geq 2$) and the $p(p+1)/2$ spatio-temporal direct and cross-variograms, computed for a selection of H spatial-temporal lags, the new fitting algorithm goes on running 4 sub-procedures sequentially, as follows.

Sub-procedure I: identify the basic structures.

A simultaneous diagonalization technique is applied on the set of H square, symmetric and real-valued matrices $\hat{\mathbf{\Gamma}}(\mathbf{h}_s, h_t)_k, k = 1, \ldots, H$, of sample direct and cross-variograms, in order to find a $(p \times p)$ orthogonal matrix which diagonalizes or "nearly" diagonalizes these matrices. At this step, the l-th empirical basic spatial-temporal component are detected by extracting the l-th diagonal element from all the diagonal matrices.

Sub-procedure II: fit the basic structures.

Given the space-time surfaces of the basic components, the spatial and temporal ranges of the basic surfaces are determined so that the scales of space-time variability are identified. The number L ($L \leq p$) of scales depends on the number of different spatial and temporal ranges the basic components exhibit. Successively, the product-sum model $g_l(\mathbf{h}_s, h_t)$ in (2) is fitted to each empirical basic component, with $l = 1, \ldots, L$. Hence marginal variogram models, $\gamma_l(\mathbf{h}_s, 0)$ and $\gamma_l(\mathbf{0}, h_t)$ are fitted to the empirical basic marginals.

Sub-procedure III: compute the coregionalization matrices.

Given the direct and cross-variograms surfaces of the variables under study, estimated in step I, the global sill values at the L scales of spatial-temporal variability are detected. Successively, the elements $b^l_{\alpha\beta}$ of matrices $\mathbf{B}_l, l = 1, \ldots, L$, are determined by dividing the contributions of the direct and cross-variogram surfaces at the l-th scale of variability by $sill[g_l(\mathbf{h}_s, h_t)]$.

Sub-procedure IV: check the admissibility of the model.
Given the coregionalization matrices $\mathbf{B}_l, l = 1, \ldots, L$, the admissibility of the *ST-LCM* is checked. If the matrix $\mathbf{B}_l$, with $l = 1, \ldots, L$, presents some negative eigenvalues, they are replaced by zeros, such that the new coregionalization matrix $\mathbf{B}_l^+$, at the l-th scale of variability, is positive definite.

References

Babak, O., Deutsch, C.V. (2009) An intrinsic model of coregionalization that solves variance inflation in collocated cokriging, *Comput. & Geosc.* 35, 3, 603-614.

De Iaco, S., Myers, D.E., Posa, D. (2001) Space-time analysis using a general product-sum model, *Statist. and Probab. Lett.*, 52, 1, 21-28.

De Iaco, S., Myers, D.E., Posa, D. (2003) The linear coregionalization model and the product-sum space-time variogram, *Mathematical Geolology*, 35, 1, 25-38.

De Iaco, S., Palma, M., Posa, D. (2005) Modeling and prediction of multivariate space-time random fields, *Comput. Statist. and Data Anal.* 48, 525-547.

Emery, X. (2010) Iterative algorithms for fitting a linear model of coregionalization, *Comput. & Geosc.*, 36, 9, 836-846.

Gneiting, T. (2002) Nonseparable, stationary covariance functions for space-time data, *J. of the Am. Statist. Assoc.*, 97, 458, 590-600.

Goulard, M., Voltz, M. (1989) Linear coregionalization model: tools for estimaton and choice of cross-variogram matrix, *Math. Geol.*, 24, 269-286.

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

Ma, C. (2002) Spatio-temporal covariance functions generated by mixtures, *Math. Geol.*, 34, 965-975.

Porcu, E., Mateu, J., Saura, F. (2008) New classes of covariance and spectral density functions for spatio-temporal modelling, *Stoch. Environ. Res. and Risk Assess.*, 22, Supplement 1, 65-79.

Stein, M. (2005) Space-time covariance functions, *J. of the Am. Statist. Assoc.*, 100, 310-321.

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