Applying 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: The near simultaneous diagonalization of the sample space-time matrix covariances or variograms makes the fitting procedure of a space-time linear coregionalization model (*ST-LCM*) easier. The method is illustrated by a case study involving data on three environmental variables measured at some monitoring stations of the Puglia region, Italy. It is shown that the near diagonalization works very well for this data set and the cross validation results show that the fitted matrix variogram is appropriate for the data.

Keywords: space-time linear coregionalization model, simultaneous diagonalization, environmental data.

1 Introduction

In this paper, the new fitting procedure of a *ST-LCM* (De Iaco et al., 2011) based on the generalized product-sum variogram model, is illustrated through an application to a multivariate space-time data set concerning three environmental variables. This method, based on the simultaneous diagonalization of the matrix variograms computed for several spatial-temporal lags, makes the identification of the parameters of the *ST-LCM* very simple and flexible.

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

2 The case study

The data set consists of ozone, O_3 ($\mu g/m^3$), Temperature (°C) and Relative Humidity (%) daily maximum values, collected during June 2009 at some monitoring stations of the Puglia region, Italy (Fig.1). The space-time correlation structure of

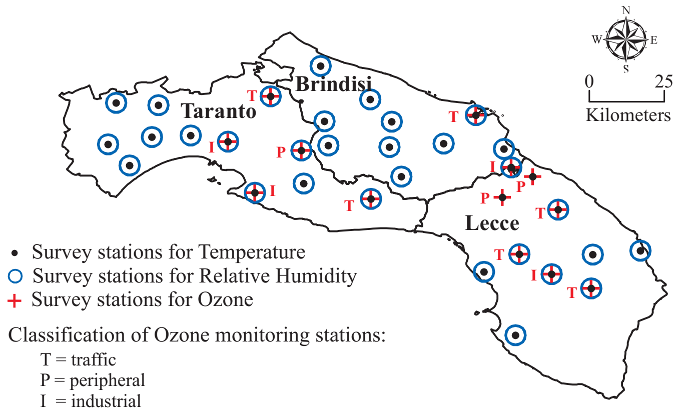

Figure 1: Posting map of survey stations in the South of Puglia region, Italy.

the variables under study has been modelled by a *ST-LCM*, whose basis components are generalized product-sum variograms.

2.1 Fitting process of a *ST-LCM*

The first step of the fitting process consists of computing the space-time direct and cross-variogram surfaces for the variables under study. Fig. 2 shows the variogram surfaces computed for 5 spatial lags and 10 temporal lags. Hence, 150 symmetric (3×3) matrices of sample direct and cross-variograms have been obtained. Afterwards, the 150 symmetric (3×3) matrices of sample direct and cross-variograms have been simultaneously diagonalized by using the matlab code "joint_diag_r.m" (Cardoso, 1996). Hence, the orthogonal (3×3) matrix $\mathbf{\Psi}$ which simultaneously diagonalizes all these matrices is given below:

$$\mathbf{\Psi} = \begin{bmatrix} 0.9725 & -0.0719 & 0.2213 \\ 0.0969 & 0.9898 & -0.1041 \\ -0.2116 & 0.1227 & 0.9696 \end{bmatrix}. \quad (1)$$

Successively, by extracting the diagonal elements from the 150 diagonal matrices, the sample spatial-temporal variograms of the independent basic components have been obtained. Since the spatial and temporal marginal variograms of the second and third basic component show the same behavior, meaning that the spatial and temporal ranges are almost equal for the second and the third basic component, solely the first and the second basic component have been retained. Two different scales of spatial-temporal variability have been considered: 21 kilometers in space, and 3 days in time, at the first scale of variability; 35 kilometers in space, and 8 days in time, at the second scale of variability. Hence, spatial and temporal marginal basic variograms, fitted to the empirical basic components have been the following:

$$\gamma_1(\mathbf{h}_s, 0) = 206\, Exp(||\mathbf{h}_s||/21), \qquad \gamma_1(\mathbf{0}, h_t) = 185\, Sph(|h_t|/3), \tag{2}$$

$$\gamma_2(\mathbf{h}_s, 0) = 3.1\, Sph(||\mathbf{h}_s||/35), \qquad \gamma_2(\mathbf{0}, h_t) = 8.3\, Exp(|h_t|/8), \tag{3}$$

where $Sph(\cdot)$ and $Exp(\cdot)$ are the abbreviated forms for the spherical and the exponential models, respectively. The contributions associated to the first and the

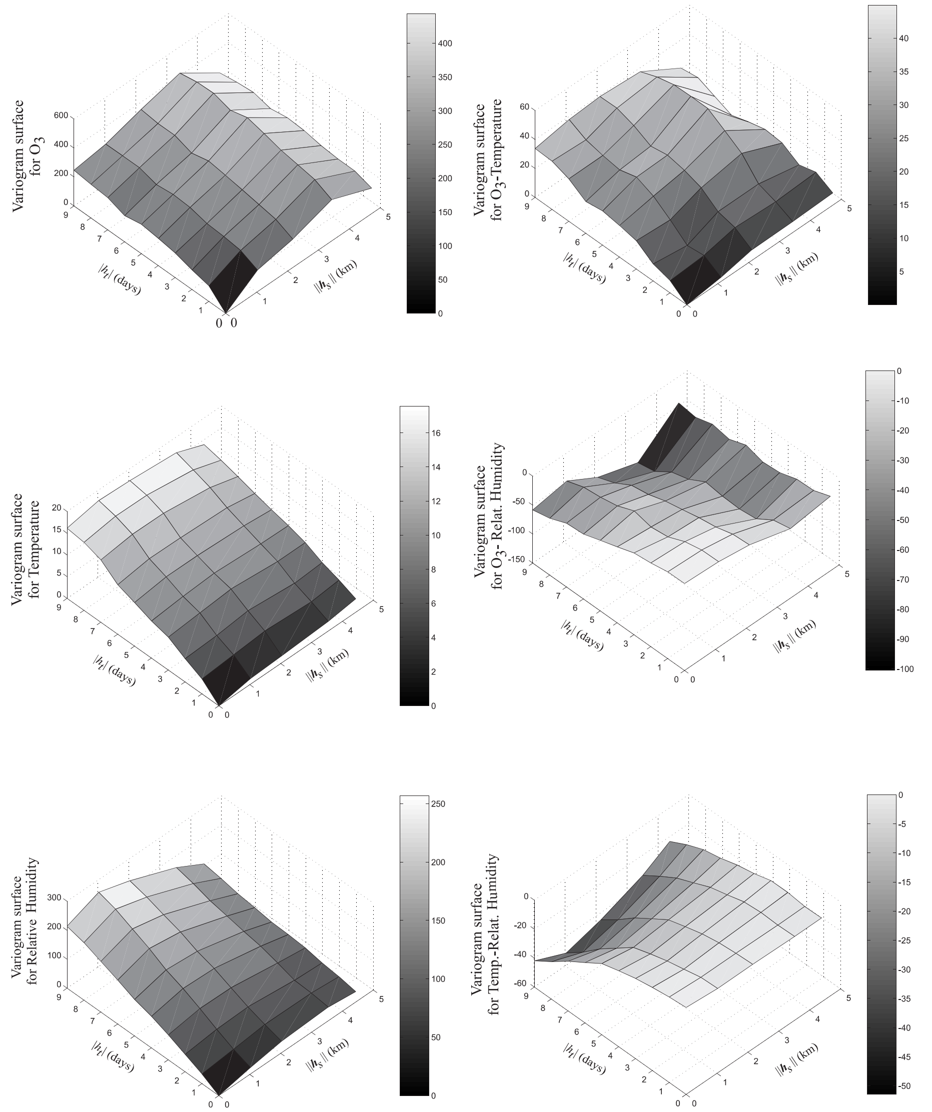

Figure 2: Space-time variogram surfaces of O_3, Temperature and Relative Humidity daily maximum values.

second basic components, i.e. the first and the second scales, are 293 and 8.7, respectively. Then, the coefficients $k_l, l = 1, 2$, are obtained as follows:

$$k_1 = \frac{206 + 185 - 293}{206 \cdot 185} \Rightarrow k_1 = 0.00257, \qquad k_2 = \frac{3.1 + 8.3 - 8.7}{3.1 \cdot 8.3} \Rightarrow k_2 = 0.10494.$$

Next, the entries in the matrices $\mathbf{B}_l, l = 1, 2$, have been computed:

$$\mathbf{B}_1 = \begin{bmatrix} 0.98635 & 0.08532 & -0.08360 \\ 0.08532 & 0.03038 & -0.01980 \\ -0.08360 & -0.01980 & 0.33447 \end{bmatrix}, \mathbf{B}_2 = \begin{bmatrix} 17.55814 & 2.09302 & -7.61628 \\ 2.09302 & 0.70930 & -2.45349 \\ -7.61628 & -2.45349 & 4.53488 \end{bmatrix}. \qquad (4)$$

Hence, the *ST-LCM* for the analyzed variables is given below:

$$\boldsymbol{\Gamma}(\mathbf{h}_s, h_t) = \mathbf{B}_1\, g_1(\mathbf{h}_s, h_t) + \mathbf{B}_2\, g_2(\mathbf{h}_s, h_t), \qquad (5)$$

where the matrices $\mathbf{B}_l, l = 1, 2$, are as in (4) and the space-time variograms $g_l(\mathbf{h}_s, h_t)$, $l = 1, 2$, are modelled as a generalized product-sum model as follows:

$$g_1(\mathbf{h}_s, h_t) = \gamma_1(\mathbf{h}_s, 0) + \gamma_1(\mathbf{0}, h_t) - k_1\gamma_1(\mathbf{h}_s, 0) \cdot \gamma_1(\mathbf{0}, h_t),$$
$$g_2(\mathbf{h}_s, h_t) = \gamma_2(\mathbf{h}_s, 0) + \gamma_2(\mathbf{0}, h_t) - k_2\gamma_2(\mathbf{h}_s, 0) \cdot \gamma_2(\mathbf{0}, h_t),$$

where $\gamma_1(\mathbf{h}_s, 0)$, $\gamma_2(\mathbf{h}_s, 0)$, and $\gamma_1(\mathbf{0}, h_t)$, $\gamma_2(\mathbf{0}, h_t)$, are the spatial and temporal marginal basic variogram models respectively, as indicated in (2) and (3), while k_1 and k_2 are the coefficients of the model.

2.2 Validation of the fitted coregionalization model

Using the modified *GSLib* routine "COK2ST" proposed in De Iaco et al. (2010), cross-validation has been performed in order to evaluate the goodness of the fitted *ST-LCM*. O_3, Temperature and Relative Humidity daily maximum values have been estimated at all data points by space-time cokriging using model (5). The proportion of absolute normalized errors (normalized by the cokriging standard deviations) exceeding 2.5, is very small (less than 1.5%) for each variable: from Chebyshev's inequality this proportion should be less than 1/6.25 (i.e, 16%). Hence, the fitted matrix variogram can be considered appropriate for the observed data.

References

Cardoso, J.F., Souloumiac, A. (1996) Jacobi angles for simultaneous diagonalization, *SIAM J. Mat. Anal. Appl.*, 17, 161-164.

De Iaco, S., Myers, D.E., Palma, M., Posa, D. (2010) FORTRAN programs for space-time multivariate modeling and prediction, *Comput. & Geosc.*, 36, 5, 636-646.

De Iaco, S., Palma, M., Posa, D. (2011) A new procedure for fitting a multivariate space-time linear coregionalization model, *Proceedings of 2011 European Regional Conference in Spatial Data Methods for Environmental and Ecological Processes, 1-2 September 2011, Foggia (Italy).*