Computers & Geosciences 36 (2010) 636–646

Contents lists available at ScienceDirect

Computers & Geosciences

journal homepage: www.elsevier.com/locate/cageo

FORTRAN programs for space–time multivariate modeling and prediction ☆

S. De Iaco [a,*], D.E. Myers [b], M. Palma [a], D. Posa [a]

[a] Università del Salento, Facoltà di Economia, Dip.to di Scienze Economiche e Matematico-Statistiche, Complesso Ecotekne, Via per Monteroni, 73100 Lecce, Italy
[b] University of Arizona, Department of Mathematics, Tucson, AZ, USA

ARTICLE INFO

Article history:
Received 27 March 2009
Received in revised form
6 October 2009
Accepted 9 October 2009

Keywords:
Multivariate space–time random field
Space–time linear coregionalization model
Product–sum model
Hourly carbon monoxide predictions

ABSTRACT

Environmental data is nearly always multivariate and often spatial–temporal. Thus to interpolate the data in space or to predict in space–time it is necessary to use a multivariate spatial–temporal method. Cokriging is easily extended to spatial–temporal data if there are valid space–time variograms or covariance functions. Various authors have proposed such models. In this paper, a generalized product–sum model is used with a linear coregionalization model for cokriging. The *GSLib* "COKB3D" program was modified to incorporate the space–time linear coregionalization model (*ST-LCM*), using the generalized product–sum variogram model. Hence, a new *GSLib* software, named "COK2ST", is proposed. To demonstrate the use of the software, hourly measurements of carbon monoxide and nitrogen dioxide from the Puglia region in Italy are used.

1. Introduction

Because environmental data is most often multivariate, i.e., there are data for several variables at many spatial locations, multivariate methods are necessary and in particular cokriging. One of the difficulties in using cokriging is how to model the cross-variograms (or cross-covariances). The common solution to this problem is using a linear coregionalization model. This approach is implemented in the *R* package *gstat*. However, environmental data, in particular air pollution data is usually not only multivariate but also spatial temporal. Thus there is the additional problem of how to model spatial–temporal variograms; this problem has been addressed by a number of authors including Bilonick (1986), Cressie and Huang (1999), De Iaco et al. (2001, 2002), Gneiting (2002), and De Iaco (in press). Several authors have given examples of the analysis of multivariate spatial–temporal data including De Iaco et al. (2003), De Iaco et al. (2005), Vanderlinden et al. (2006), Lark et al. (2006), and Sicard et al. (2002). Brown et al. (1994) use a Bayesian approach which requires a Multigaussian distribution assumption and uses a transformation on the space to account for the anisotropy. Some authors use multivariate methods to avoid the need to apply cokriging to spatial–temporal data, e.g., Rouhani and Myers (1990), De Iaco et al. (2000). Sen et al. (2006) as well as Kyriakidis et al. (2001a, b) used spatial time series. Koike et al. (2002) and Liu and Koike (2007) claim to have another method for fitting space–time variograms and cross-variograms, namely fitting the empirical variogram and cross-variograms with cubic splines. This does not ensure that the resulting variogram model is conditionally negative definite, nor does it ensure that the matrix variogram is a conditionally negative definite function. They also erroneously claim that it is necessary to have data for all variables at all locations in order to use cokriging.

The principal remaining problem when analyzing multivariate space–time data is the lack of software. De Cesare et al. (2002) modified *GSLib* (Deutsch and Journel, 1997) programs to obtain *FORTRAN* programs for computing empirical space-time variograms, marginal space variograms and marginal time variograms. They also modified the *GSLib* program for kriging to allow the use of spatial–temporal data but only included the product model variogram. Although De Iaco et al. (2003, 2005) showed that the general product–sum model could be used with an *LCM* for cokriging, they did not publish the software at that time. While various authors have used either a generalized product–sum model or Gneiting's model, the only other published geostatistics software for space–time modeling is the *R* package *random fields*. It has limited choices for the space–time variogram model, and does not provide for computation of the empirical marginal variograms nor incorporate cokriging.

In this paper we present a modified *GSLib* code for cokriging using spatial–temporal data. The new software is called "COK2ST".

Section 2 reviews the basics of using an *LCM* and generalized product–sum spatial temporal variograms for ordinary cokriging (De Iaco et al., 2005). Ordinary cokriging assumes that the mean of each component variable is constant.

☆ Code available from server at http://www.iamg.org/CGEditor/index.htm.

* Corresponding author. Tel./fax: +39 832298737.
E-mail addresses: sandra.deiaco@unisalento.it (S. De Iaco), myers@math.arizona.edu (D.E. Myers), monica.palma@unisalento.it (M. Palma), donato.posa@unisalento.it (D. Posa).

0098-3004/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cageo.2009.10.004

Section 3 describes the fitting of a diurnal component for each component variable. Air pollution data nearly always exhibits a diurnal pattern and the maximum magnitudes can vary with respect to the spatial location. This is somewhat analogous to a non-constant mean. Instead each of the original variables is viewed as the sum of two parts, a random component defined in space–time and a component corresponding to the diurnal behavior. The second component is not deterministic but instead incorporates a random component for the magnitude.

Section 4 is a case study using hourly measurements of carbon monoxide (CO) and nitrogen doxide (NO_2) in the Puglia region of Italy. The measurements were obtained during November 2006. This case study demonstrates the use of a modified *GSLib* for cokriging using a *ST-LCM* with the generalized product–sum variograms.

2. Vector-valued space–time random functions: modeling and prediction

Let $\{\mathbf{Z}(\mathbf{s},t),(\mathbf{s},t)\in D\times T\subseteq\mathbb{R}^{d+1}\}$ be a vector-valued space–time random function (*STRF*) with $p\geq 2$ scalar *STRF*s, i.e.

$$\mathbf{Z}(\mathbf{s},t)=[Z_1(\mathbf{s},t),Z_2(\mathbf{s},t),\ldots,Z_p(\mathbf{s},t)]^T, \tag{1}$$

where $\mathbf{s}\in D\subseteq\mathbb{R}^d$ (generally, $d\leq 3$) denotes the spatial point and $t\in T$ is the temporal coordinate.

The components of $\mathbf{Z}(\mathbf{s},\ t)$ are assumed to be second-order stationary which implies that the matrix variogram

$$\boldsymbol{\Gamma}(\mathbf{h})=[\gamma_{\alpha\beta}(\mathbf{h})], \tag{2}$$

exists and does not depend on ($\mathbf{s}$, t), where

- $\mathbf{h}$=($\mathbf{h}_s$, h_t), with $\mathbf{h}_s=(\mathbf{s}-\mathbf{s}')$ and $h_t=(t-t')$;
- $\gamma_{\alpha\beta}(\mathbf{h})=Cov[(Z_\alpha(\mathbf{s}+\mathbf{h}_s,t+h_t)-Z_\alpha(\mathbf{s},t)),(Z_\beta(\mathbf{s}+\mathbf{h}_s,t+h_t)-Z_\beta(\mathbf{s},t))]$, $\alpha,\ \beta=1,2,\ldots,p$, are the cross-variograms between the Z_α and Z_β *STRF*s, when $\alpha\neq\beta$, and the direct variograms of the Z_α *STRF*, when $\alpha=\beta$.

2.1. ST-LCM using a product–sum model

Recall that for an *LCM*, each component of the vector-valued random function $\mathbf{Z}(\mathbf{s},\ t)$ is assumed to be represented as a linear combination of uncorrelated second-order stationary random functions, e.g.

$$\mathbf{Y}_l(\mathbf{s},t)=[Y_1^l(\mathbf{s},t),Y_2^l(\mathbf{s},t),\ldots,Y_p^l(\mathbf{s},t)]^T,\quad l=1,2,\ldots,L. \tag{3}$$

Then

$$\mathbf{Z}(\mathbf{s},t)=\sum_{l=1}^{L}\mathbf{A}_l\mathbf{Y}_l(\mathbf{s},t), \tag{4}$$

where $\mathbf{A}_l$ is a ($p\ \times p$) coefficient matrix for each $l=1,2,\ldots,L$.

The *ST-LCM* for the variogram matrix $\boldsymbol{\Gamma}(\mathbf{h})$, can be written as

$$\boldsymbol{\Gamma}(\mathbf{h})=\boldsymbol{\Gamma}(\mathbf{h}_s,h_t)=\sum_{l=1}^{L}\mathbf{B}_l g_l(\mathbf{h}_s,h_t), \tag{5}$$

where $\mathbf{B}_l=\mathbf{A}_l\mathbf{A}_l^T=[b_{\alpha\beta}^l]$, $l=1,2,\ldots,L$, $\alpha,\beta=1,2,\ldots,p$, are positive definite ($p\ \times p$) matrices and $g_l(\mathbf{h}_s,\ h_t)$, $l=1,2,\ldots,L$, are basic space–time variograms which might correspond to different scales of variability.

In the extension of the *LCM* given by De Iaco et al. (2005), each basic space–time variogram is a generalized product–sum model

$$g_l(\mathbf{h}_s,h_t)=\gamma_l(\mathbf{h}_s,0)+\gamma_l(\mathbf{0},h_t)-k_l\gamma_l(\mathbf{h}_s,0)\gamma_l(\mathbf{0},h_t),\quad l=1,2,\ldots,L, \tag{6}$$

where $\gamma_l(\mathbf{h}_s,0)$ and $\gamma_l(\mathbf{0},h_t)$, $l=1,2,\ldots,L$, are spatial and temporal marginal variogram models, while $k_l, l=1,2,\ldots,L$, are parameters given by

$$k_l=\frac{sill[\gamma_l(\mathbf{h}_s,0)]+sill[\gamma_l(\mathbf{0},h_t)]-sill[g_l(\mathbf{h}_s,h_t)]}{sill[\gamma_l(\mathbf{h}_s,0)]\cdot sill[\gamma_l(\mathbf{0},h_t)]},\quad l=1,2,\ldots,L. \tag{7}$$

The parameters given in (7) must satisfy a necessary and sufficient condition (De Iaco et al., 2001) to ensure that $g_l(\mathbf{h}_s,\ h_t)$ is strictly conditionally negative definite.

By substituting (6) in (5), the *ST-LCM* with basic generalized product–sum variogram models is determined by two marginal *LCM*, one in space:

$$\boldsymbol{\Gamma}(\mathbf{h}_s,0)=\sum_{l=1}^{L}\mathbf{B}_l\gamma_l(\mathbf{h}_s,0), \tag{8}$$

and the other one in time:

$$\boldsymbol{\Gamma}(\mathbf{0},h_t)=\sum_{l=1}^{L}\mathbf{B}_l\gamma_l(\mathbf{0},h_t). \tag{9}$$

Moreover, the diagonal elements of $\mathbf{B}_l$ are easily determined after modeling marginal direct variograms, while the off-diagonal elements are obtained by marginal cross-variogram models in such a way to ensure positive definiteness of the matrices $\mathbf{B}_l$.

Since the only difference between cokriging in space–time and cokriging in space is in the modeling of the variograms and cross-variograms, as shown in Myers (1991), the ordinary space–time cokriging predictor can be written as

$$\hat{\mathbf{Z}}(\mathbf{u})=\sum_{i=1}^{n}\boldsymbol{\Lambda}_i(\mathbf{u})\mathbf{Z}(\mathbf{u}_i), \tag{10}$$

where $\mathbf{u}=(\mathbf{s},t)\in D\times T$ is any point in the space–time domain, $\mathbf{u}_i=(\mathbf{s}_i,t_i)\in D\times T$, $i=1,2,\ldots,n$, are the data points and $\boldsymbol{\Lambda}_i(\mathbf{u})$, $i=1,2,\ldots,n$, are ($p\times p$) matrices of weights whose elements $\lambda_i^{\alpha\beta}(\mathbf{u})$ are the weights assigned to the value of the β-th variable, $\beta=1,2,\ldots,p$, at the i-th data point to predict the α-th variable, $\alpha=1,2,\ldots,p$, at the point $\mathbf{u}\in D\times T$.

2.2. COK2ST for space–time prediction

In this section we will describe the modification of the *GSLib* program "COKB3D" (Deutsch and Journel, 1997).

To determine the space–time variogram matrix for $\mathbf{Z}(\mathbf{s},\ t)$ and to make multivariate predictions in space–time, various parameters must be specified in the new *GSLib* software "COK2ST". As shown in the Appendix B, these are:

(a) the option for making prediction over a grid or at specified points, or for making cross-validation (`0=grid, 1=jack-knife, 2=cross`);
(b) the number of nested product or product–sum variogram models (`np/sp`);
(c) the option for using the product or the product–sum model (`0 = prod, 1 =sum-prod`);
(d) the number of spatial and temporal variograms for each structure (`nst`);
(e) the spatial and temporal nugget effects;
(f) the type of spatial and temporal basic structures (`it`);
(g) the spatial and temporal contributions of the basic structures (`cc`);
(h) the spatial and temporal angles (`ang1,ang2,ang3`) and ranges or scales of variability (`a_hmax,a_hmin,a_vert`);
(i) the global sills.

The (f), (g), and (h) parameters must be specified for each direct and cross-variogram and for each scale of variability, both in space and also in time.

Note that

- in contrast with the original program which works on a three-dimensional space, this modified version works on $D \times T \subseteq \mathbb{R}^{2+1}$, where the vertical dimension now is the temporal dimension. Hence, spatial angles and ranges pertain to two-dimensional space, the temporal angle and range pertain to one-dimensional space. The options for three angles and ranges could be useful for a metric model in two-dimensional space and time;
- variogram models with zero type, range and sill parameters are used to denote the nugget effect;
- the same number of spatial and temporal structures (`nst`) for each nested product or product–sum model has to be considered.

The type of spatial and temporal basic structures is specified by an integer code, as shown in Table 1.

3. Modeling the diurnal component

Air pollution data nearly always exhibits a diurnal pattern and the maximum magnitude can vary with respect to the spatial location. This is somewhat analogous to a non-constant mean. In this case, each of the original variables is viewed as the sum of two parts, a space–time random component corresponding to the residual and a component corresponding to the diurnal behavior, which incorporates a random component for the magnitude. Hence, the fitting of a diurnal component for each component variable is proposed.

Each of the original variables $X_\alpha(\mathbf{u})$, $\alpha = 1,2,\ldots,p$, with $\mathbf{u} = (\mathbf{s},t) \in D \times T \subseteq \mathbb{R}^{d+1} (d=2)$, is treated as a partial realization of a non-stationary *STRF*, decomposed as follows:

$$X_\alpha(\mathbf{u}) = Z_\alpha(\mathbf{u}) + M_\alpha(\mathbf{u}), \quad \mathbf{u} = (\mathbf{s},t) \in D \times T \subseteq \mathbb{R}^{d+1}, \tag{11}$$

where

- $Z_\alpha(\mathbf{u})$ is, as in Section 2, a space–time random component which is assumed to be second-order stationary;
- $M_\alpha(\mathbf{u})$ is the diurnal component.

The components of $M_\alpha(\mathbf{u})$ are assumed to be of the form:

$$M_\alpha(\mathbf{u}) = \mu_\alpha + \beta_\alpha(t) \cdot V_\alpha(\mathbf{s}), \quad \alpha = 1,2,\ldots,p, \tag{12}$$

where

- μ_α is constant in the space–time domain;
- $\beta_\alpha(t)$ is a periodic function representing the standardized diurnal cycle of X_α;
- $V_\alpha(\mathbf{s})$ is a second-order stationary spatial random field representing the magnitude of the diurnal cycle at location $\mathbf{s}$ of X_α.

Table 1
Integer codes used to specify type of spatial and temporal basic structure in "COK2ST" parameter file.

Type of variogram model	Integer code	
	In space	In time
Spherical	1	6
Exponential	2	7
Gaussian	3	8
Power	4	9
Hole effect	5	10

The components $\beta_\alpha(t)$ and $V_\alpha(\mathbf{s})$ of $M_\alpha(\mathbf{u})$ are computed as follows:

1. the diurnal component is estimated for each monitoring station by the *MAE* method (Brockwell and Davis, 1987), as described hereafter;
2. the 24 h diurnal components are standardized for each monitoring station;
3. a periodic function $\beta_\alpha(t)$ is fitted to the diurnal standardized components for all the survey stations;
4. the standard deviations of the diurnal components, estimated for each survey station, are considered as a realization of the random field $V_\alpha(\mathbf{s})$; hence, ordinary cokriging can be used in order to estimate V_α over the area of interest.

3.1. Missing values and removal of the diurnal component

Using the *FORTRAN* program "REMOVEMULT", the diurnal components of the variables under study are estimated by moving averages (Brockwell and Davis, 1987) and removed simultaneously. In case of missing values, this program uses linear interpolation to replace a prescribed number of consecutive missing values.

"REMOVEMULT" is a modified version of the *FORTRAN* program "REMOVE" given in De Cesare et al. (2002), which allows deseasonalization of several variables simultaneously.

The input parameters for the program "REMOVEMULT" (see Appendix A) are, for each variable of interest:

- the maximum number of missing consecutive data;
- the period of the seasonal component;
- the minimum number of valid consecutive data for deseasonalization;
- the time unit;
- the number of time intervals;
- the interval definition, i.e. the first instant and the last instant +1 of each time interval.

After removing the diurnal component at each station, the residuals of each variable are used for modeling and prediction purposes.

4. A case study

In this paper, a multivariate approach was applied to analyze the direct and cross correlation for two space–time air pollutants,

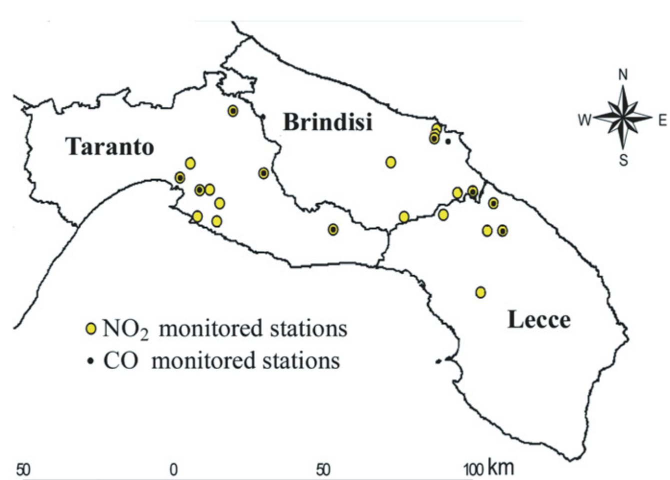

Fig. 1. Location map of survey stations.

carbon monoxide (CO) and nitrogen dioxide (NO_2), measured during November 2006 at monitoring stations in Puglia region (Italy).

As previously pointed out, the aim of this case study is to show the flexibility of the space–time linear coregionalization model, based on the generalized product–sum variogram, in modeling and interpolation a multivariate space–time random field; the study proceeds as follows:

1. exploratory analysis for each of the two air pollutants;
2. structural analysis for the two air pollutants;
3. cross-validation and space–time prediction of CO residuals (primary variable) using data for CO and NO_2 residuals;
4. prediction maps of CO concentrations for the 1st of December 2006.

4.1. Exploratory data analysis

The data set, provided by the Environmental Protection Agency of the Puglia region, Italy, consists of CO (mg/m^3) and NO_2 (μg/m^3) hourly averages measured during November 2006 at 10 and 22 monitoring stations, respectively (Fig. 1). NO_2 is a secondary pollutant caused, mainly in winter, by heating systems, both civil and industrial ones, and also by the traffic; therefore it reaches very high values in those urban areas characterized by high-density population. On the other hand, CO is caused by the motor vehicles emissions and it reaches very high values in those areas with heavy traffic and poor ventilation.

Fig. 2 shows the box-plots of CO and NO_2 hourly averages where it is evident that the temporal behavior of each pollutant is characterized by a diurnal cycle.

4.2. Estimating and modeling the diurnal component

The observations for CO and NO_2 are treated as a partial realization of a non-stationary bivariate *STRF*

$$\mathbf{X}(\mathbf{u}) = [X_1(\mathbf{u}), X_2(\mathbf{u})]^T, \tag{13}$$

with $\mathbf{u} = (\mathbf{s}, t) \in D \times T \subseteq \mathbb{R}^{d+1} (d=2)$, decomposed as follows:

$$\mathbf{X}(\mathbf{u}) = \mathbf{Z}(\mathbf{u}) + \mathbf{M}(\mathbf{u}), \quad \mathbf{u} = (\mathbf{s}, t) \in D \times T \subseteq \mathbb{R}^{d+1}, \tag{14}$$

where

- $\mathbf{Z}(\mathbf{u})$ is as in Section 2 but with only two components;
- $\mathbf{M}(\mathbf{u})$ is the diurnal component.

The components of $\mathbf{M}(\mathbf{u})$ are assumed to be of the form:

$$M_\alpha(\mathbf{u}) = \mu_\alpha + \beta_\alpha(t) \cdot V_\alpha(\mathbf{s}), \quad \alpha = 1, 2, \tag{15}$$

where each term is as in Section 3.

4.2.1. Missing values and removal of CO and NO_2 diurnal components

Since at each monitoring station, hourly measurements of both pollutants exhibit a diurnal behavior, before performing structural analysis, it was necessary to deseasonalize the CO and NO_2 values.

Using the *FORTRAN* program "REMOVEMULT" described in Section 3.1 and the parameter file shown in Appendix A, the diurnal components of the two variables under study were estimated and removed simultaneously. In case of missing values, linear interpolation was used.

After removing the diurnal component at each station, the residuals of CO and NO_2 were used for:

1. modeling the spatial-temporal correlation of the two variables by the *ST-LCM*, based on the generalized product–sum variogram model,
2. predicting CO hourly concentrations in the domain under study, for the 1st of December 2006, using the multivariate space–time model, and the CO and NO_2 data.

4.3. Modeling the ST-LCM

The *ST-LCM* for CO and NO_2 residuals was constructed as follows.

Step 1: Using the residuals obtained by subtracting the estimated diurnal components, the sample marginal variograms were computed. These were fitted using nested models as follows:

$$\tilde{\gamma}_{11}(\mathbf{h}_s, 0) = 0.062\ \gamma_1(\mathbf{h}_s, 0) + 0.107\ \gamma_2(\mathbf{h}_s, 0) + 0.032\ \gamma_3(\mathbf{h}_s, 0), \tag{16}$$

$$\tilde{\gamma}_{11}(\mathbf{0}, h_t) = 0.062\ \gamma_1(\mathbf{0}, h_t) + 0.107\ \gamma_2(\mathbf{0}, h_t) + 0.032\ \gamma_3(\mathbf{0}, h_t), \tag{17}$$

for CO residuals, and

$$\tilde{\gamma}_{22}(\mathbf{h}_s, 0) = 180\ \gamma_1(\mathbf{h}_s, 0) + 100\ \gamma_2(\mathbf{h}_s, 0) + 73.5\ \gamma_3(\mathbf{h}_s, 0), \tag{18}$$

$$\tilde{\gamma}_{22}(\mathbf{0}, h_t) = 180\ \gamma_1(\mathbf{0}, h_t) + 100\ \gamma_2(\mathbf{0}, h_t) + 73.5\ \gamma_3(\mathbf{0}, h_t), \tag{19}$$

for NO_2 residuals, where the three marginal basic structures in space and time are given below:

$$\gamma_1(\mathbf{h}_s, 0) = \begin{cases} 0, & \|\mathbf{h}_s\| = 0, \\ 1, & \|\mathbf{h}_s\| > 0, \end{cases} \tag{20}$$

$$\gamma_2(\mathbf{h}_s, 0) = 0.81\, Sph(\|\mathbf{h}_s\|/28), \tag{21}$$

$$\gamma_3(\mathbf{h}_s, 0) = 1.61\, Exp(\|\mathbf{h}_s\|/45), \tag{22}$$

$$\gamma_1(\mathbf{0}, h_t) = \begin{cases} 0, & |h_t| = 0, \\ 0.35, & |h_t| > 0, \end{cases} \tag{23}$$

$$\gamma_2(\mathbf{0}, h_t) = 0.87\ Exp(|h_t|/12), \tag{24}$$

$$\gamma_3(\mathbf{0}, h_t) = 1.07\ Exp(|h_t|/36). \tag{25}$$

Fig. 3 shows empirical spatial and temporal marginal variograms for CO and NO_2 residuals and their corresponding models.

The coefficients in Eqs. (16) and (18) are the diagonal entries in the matrices $\mathbf{B}_l$, l = 1,2,3.

Step 2: The sample space–time variogram surfaces for the residuals of each variable were obtained by using the modified version of the program "GAMV" in De Cesare et al. (2002). Then the variogram surfaces were fitted to product–sum nested models, by appropriate choices of the k_l, l=1,2,3, parameters defined in (7).

Fig. 4 shows the sample space–time variogram surfaces and the fitted product–sum nested models of the residuals of the two variables.

Step 3: The sample marginal cross-variograms in space and time for the residuals of the two variables (Fig. 5) were computed and the fitted nested models were the following:

$$\tilde{\gamma}_{12}(\mathbf{h}_s, 0) = \tilde{\gamma}_{21}(\mathbf{h}_s, 0) = 0.1\ \gamma_1(\mathbf{h}_s, 0) + 2.229\ \gamma_2(\mathbf{h}_s, 0) + 1.398\ \gamma_3(\mathbf{h}_s, 0), \tag{26}$$

$$\tilde{\gamma}_{12}(\mathbf{0}, h_t) = \tilde{\gamma}_{21}(\mathbf{0}, h_t) = 0.1\ \gamma_1(\mathbf{0}, h_t) + 2.229\ \gamma_2(\mathbf{0}, h_t) + 1.398\ \gamma_3(\mathbf{0}, h_t), \tag{27}$$

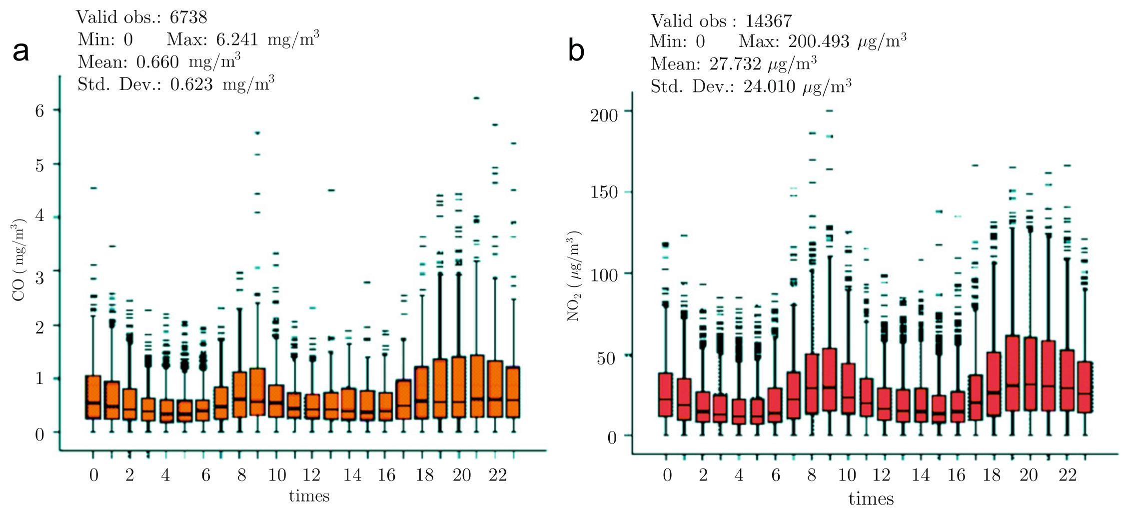

Fig. 2. Box-plots of (a) CO hourly averages; (b) NO_2 hourly averages.

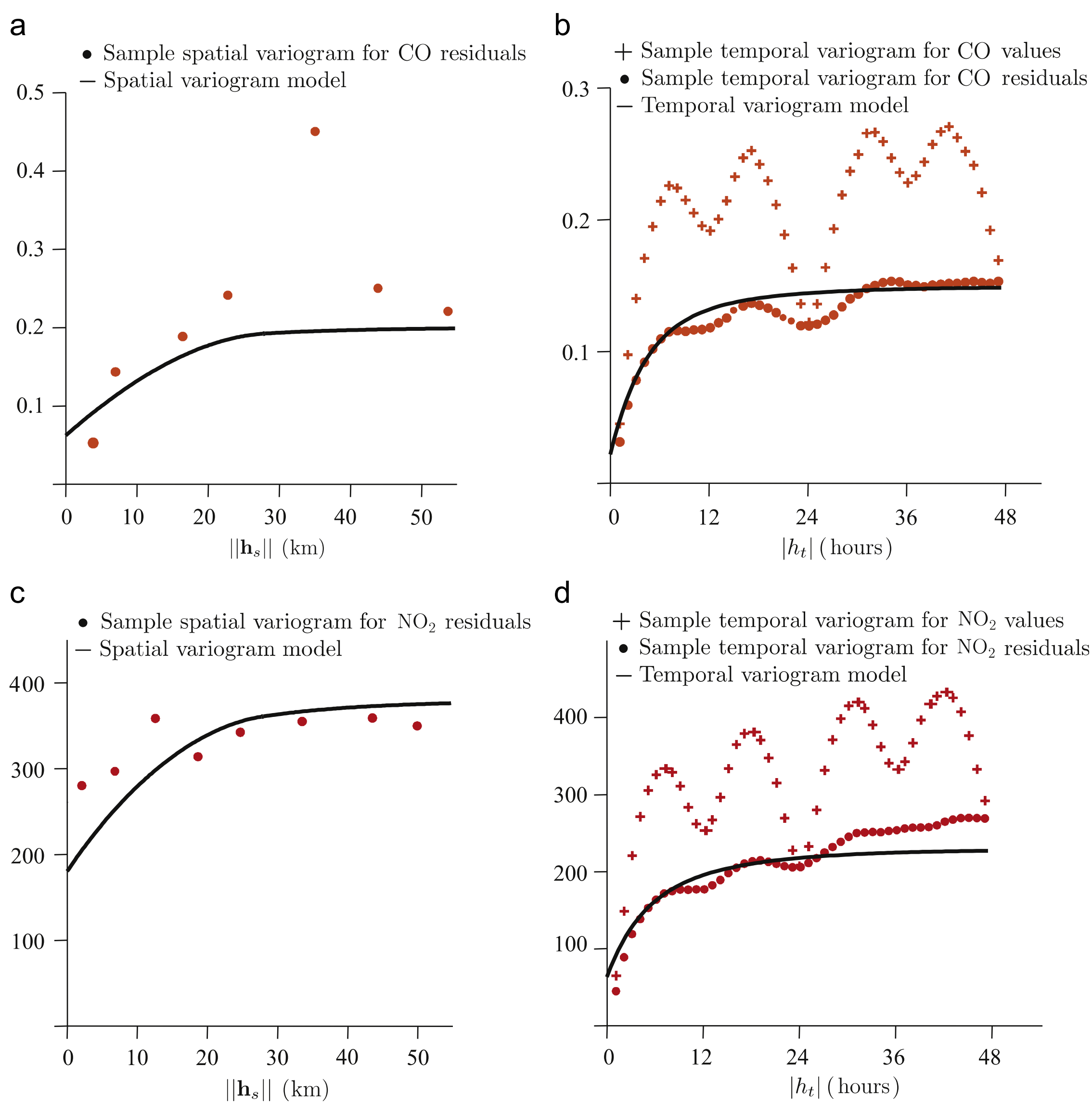

Fig. 3. Spatial and temporal marginal variograms and models for (a,b) CO and (c,d) NO_2 residuals.

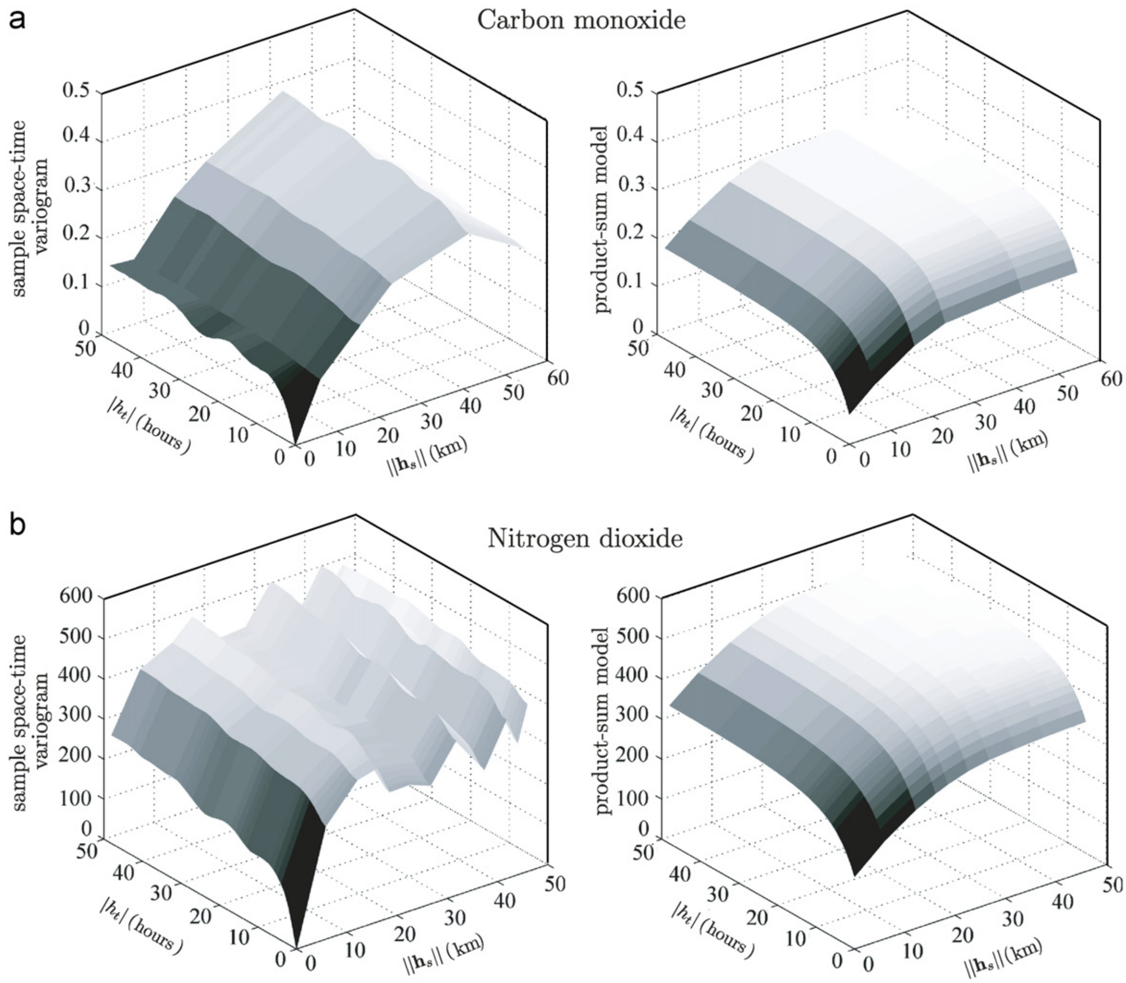

Fig. 4. Sample space–time variogram surfaces and their product–sum models for (a) CO and (b) NO_2 residuals (k_1=0.99; k_2=0.02; k_3=0.1).

where $\gamma_1(\mathbf{h}_s,0)$, $\gamma_2(\mathbf{h}_s,0), \gamma_3(\mathbf{h}_s,0)$, $\gamma_1(\mathbf{0},h_t)$, $\gamma_2(\mathbf{0},h_t)$ and $\gamma_3(\mathbf{0},h_t)$ had been previously defined.

The marginal cross-variograms (26) and (27) were chosen to ensure that the matrices $\mathbf{B}_l$, l=1,2,3, are positive definite and the *ST-LCM* is strictly conditionally negative definite.

Finally, the *ST-LCM* for CO and NO_2 residuals is of the following form:

$$\boldsymbol{\Gamma}(\mathbf{h}_s,h_t)=\sum_{l=1}^{3}\mathbf{B}_l g_l(\mathbf{h}_s,h_t)=\mathbf{B}_1 g_1(\mathbf{h}_s,h_t)+\mathbf{B}_2 g_2(\mathbf{h}_s,h_t)+\mathbf{B}_3 g_3(\mathbf{h}_s,h_t), \tag{28}$$

where the matrices $\mathbf{B}_l$, l=1, 2, 3, are, respectively:

$$\mathbf{B}_1=\begin{vmatrix}0.062 & 0.1\\ 0.1 & 180\end{vmatrix},\quad \mathbf{B}_2=\begin{vmatrix}0.107 & 2.229\\ 2.229 & 100\end{vmatrix},\quad \mathbf{B}_3=\begin{vmatrix}0.032 & 1.398\\ 1.398 & 73.5\end{vmatrix}, \tag{29}$$

and the basic space–time variograms $g_l(\mathbf{h}_s,\ h_t)$, l=1, 2, 3, are modeled as a generalized product–sum model as follows:

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

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

$$g_3(\mathbf{h}_s,h_t)=\gamma_3(\mathbf{h}_s,0)+\gamma_3(\mathbf{0},h_t)-k_3\gamma_3(\mathbf{h}_s,0)\gamma_3(\mathbf{0},h_t),$$

with

- k_1 = 0.99, k_2=0.02, k_3=0.1;
- $\gamma_1(\mathbf{h}_s,0), \gamma_2(\mathbf{h}_s,0), \gamma_3(\mathbf{h}_s,0)$, are marginal basic structures in space;
- $\gamma_1(\mathbf{0},h_t), \gamma_2(\mathbf{0},h_t)$ and $\gamma_3(\mathbf{0},h_t)$, are marginal basic structures in time.

To ensure a fit for the global sill, the following equations must be satisfied:

$$0.062217=0.062[1+0.35-k_1(1\cdot 0.35)],$$

$$0.1783=0.107[0.81+0.87-k_2(0.81\cdot 0.87)],$$

$$0.0802=0.032[1.61+1.07-k_3(1.61\cdot 1.07)],$$

for CO residuals;

$$180.63=180[1+0.35-k_1(1\cdot 0.35)],$$

$$166.6=100[0.81+0.87-k_2(0.81\cdot 0.87)],$$

$$184.32=73.5[1.61+1.07-k_3(1.61\cdot 1.07)],$$

for NO_2 residuals;

$$0.10035=0.1[1+0.35-k_1(1\cdot 0.35)],$$

$$3.7135=2.229[0.81+0.87-k_2(0.81\cdot 0.87)],$$

$$3.5058=1.398[1.61+1.07-k_3(1.61\cdot 1.07)],$$

for the cross-correlation of CO and NO_2 residuals, where k_1=0.99, k_2=0.02, and k_3=0.1.

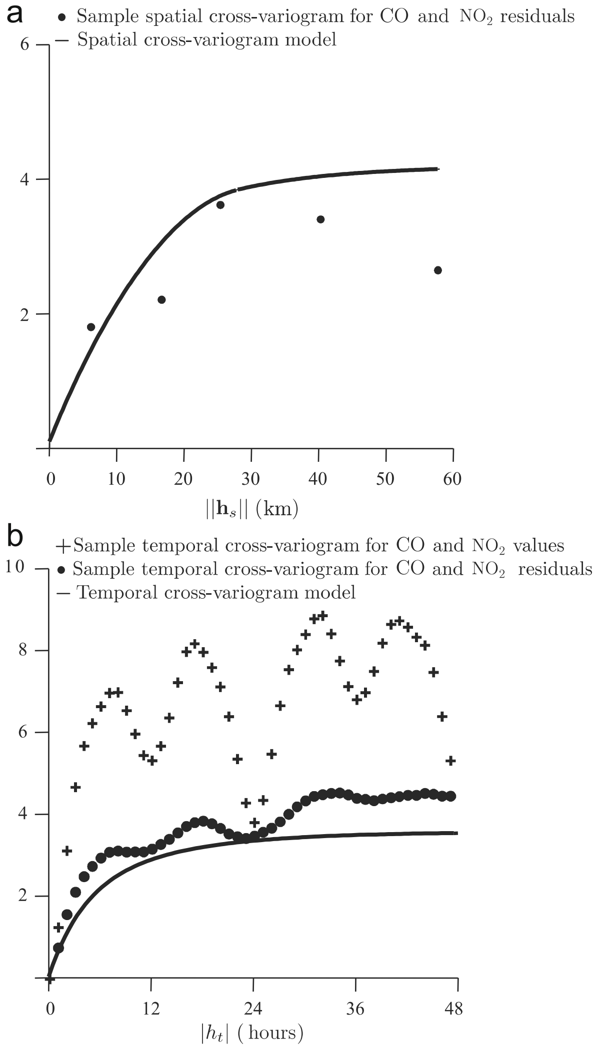

Fig. 5. Sample marginal cross-variograms and their models (a) in space and (b) in time.

4.4. Cross-validation of the ST-LCM

As previously discussed, the new program "COK2ST" incorporates cross-validation; hence, it was possible to cross-validate the *ST-LCM* constructed for the CO and NO_2 residuals. In particular, CO hourly residuals were estimated at all data points by using space–time cokriging. Fig. 6 shows the scatter plot of CO hourly residuals towards the estimated ones. It is evident that the linear correlation coefficient between CO hourly residuals and estimated values is very high. Hence, the *ST-LCM* (28) could be considered valid for modeling the spatial–temporal correlation of the two variables.

4.5. Prediction maps of CO concentrations

"COK2ST" and the parameter file shown in Appendix B, were used to predict CO hourly residuals for the 1st of December 2006

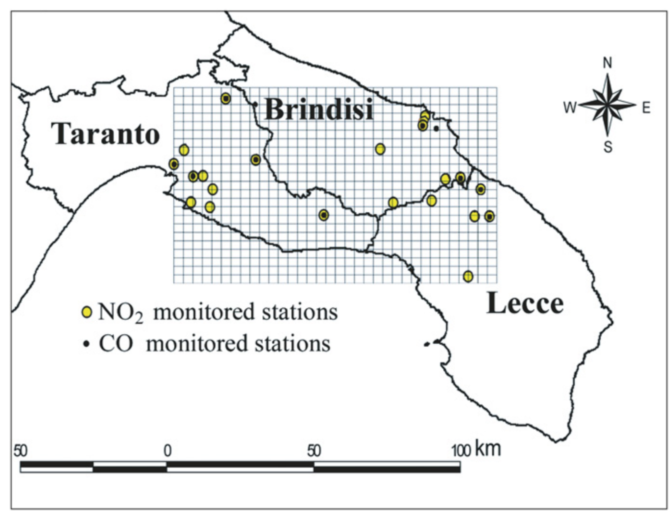

Fig. 7. Grid of 23 × 34 nodes which covers spatial domain.

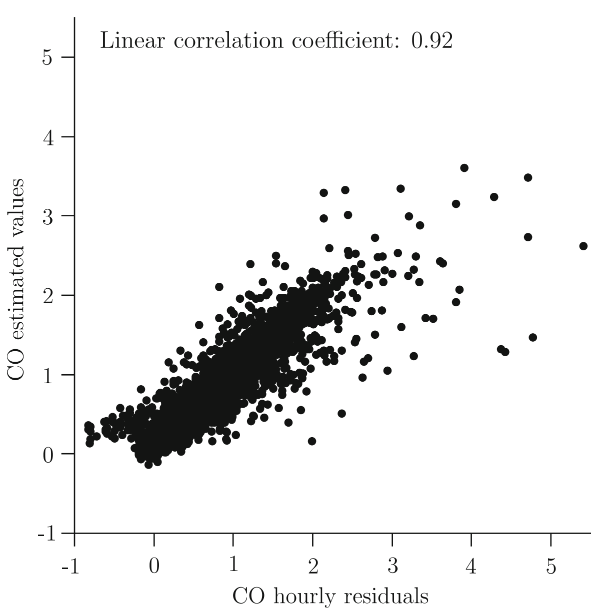

Fig. 6. Scatter plot of CO hourly residuals towards estimated ones.

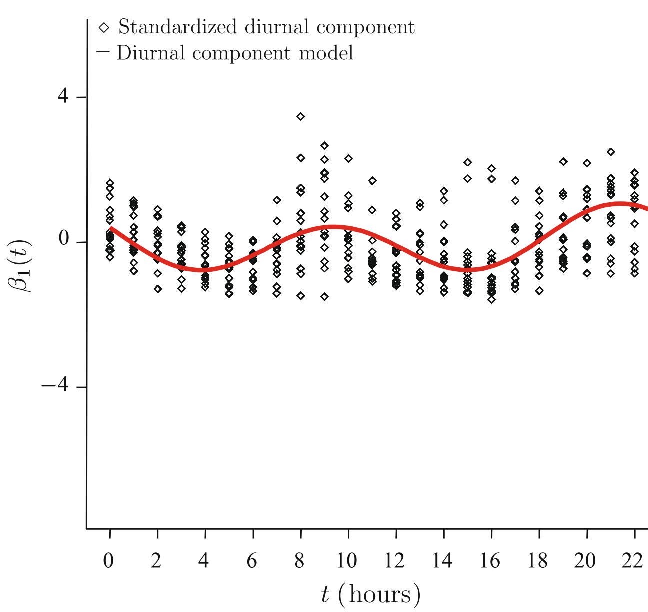

Fig. 8. Standardized diurnal component for CO values and its model.

over a 23×34 grid nodes which covers the area of interest (Fig. 7).

In order to generate the prediction maps of CO for the 1st of December 2006 in the study area, it is necessary to model $\beta_1(t)$ and $V_1(\mathbf{s})$.

- Standardized diurnal values for CO (Fig. 8) were used to fit the following periodic function

$$\beta_1(t)=0.75\cos\left[\frac{2\pi}{12}(t+2.6)\right]+0.32\cos\left[\frac{2\pi}{24}(t+2.6)\right]. \qquad (30)$$

- The standard deviations for the diurnal values for CO and NO_2 were computed, these were used as data to compute the sample variograms and cross-variograms for the $V_\alpha(\mathbf{s}),\alpha=1,2,$ as shown in Fig. 9. The fitted models were

$$\gamma_{V_1}(\mathbf{h}_s)=0.003+0.016\,Sph(\|\mathbf{h}_s\|/28)+0.0247\,Exp(\|\mathbf{h}_s\|/45),$$

$$\gamma_{V_2}(\mathbf{h}_s)=14+6\,Sph(\|\mathbf{h}_s\|/28)+15\,Exp(\|\mathbf{h}_s\|/45),$$

$$\gamma_{V_{12}}(\mathbf{h}_s)=\gamma_{V_{21}}(\mathbf{h}_s)=0.2+0.016\,Sph(\|\mathbf{h}_s\|/28)+0.33\,Exp(\|\mathbf{h}_s\|/45).$$

Then, using spatial cokriging, $V_1(\mathbf{s})$ was estimated at 23×34 grid nodes.

Finally, the predicted CO hourly residuals were added to the estimated CO diurnal component.

Fig. 10 shows the spatial-temporal behavior of CO concentrations over the study area for the 1st of December 2006.

It is evident that the highest CO concentrations occurred in the southern area of Brindisi adjoining the northern area of Lecce, which is an industrial area situated close to high population density towns.

5. Summary

The *GSLib* cokriging program "COKB3D" was modified to incorporate space–time data and also generalized product–sum space variograms in an *LCM*. "COK2ST" is the new *GSLib* program proposed in this paper. To demonstrate the use of the software, data from a case study on air pollution in the Puglia region of Italy were used, hourly measurements were obtained for CO and NO_2.

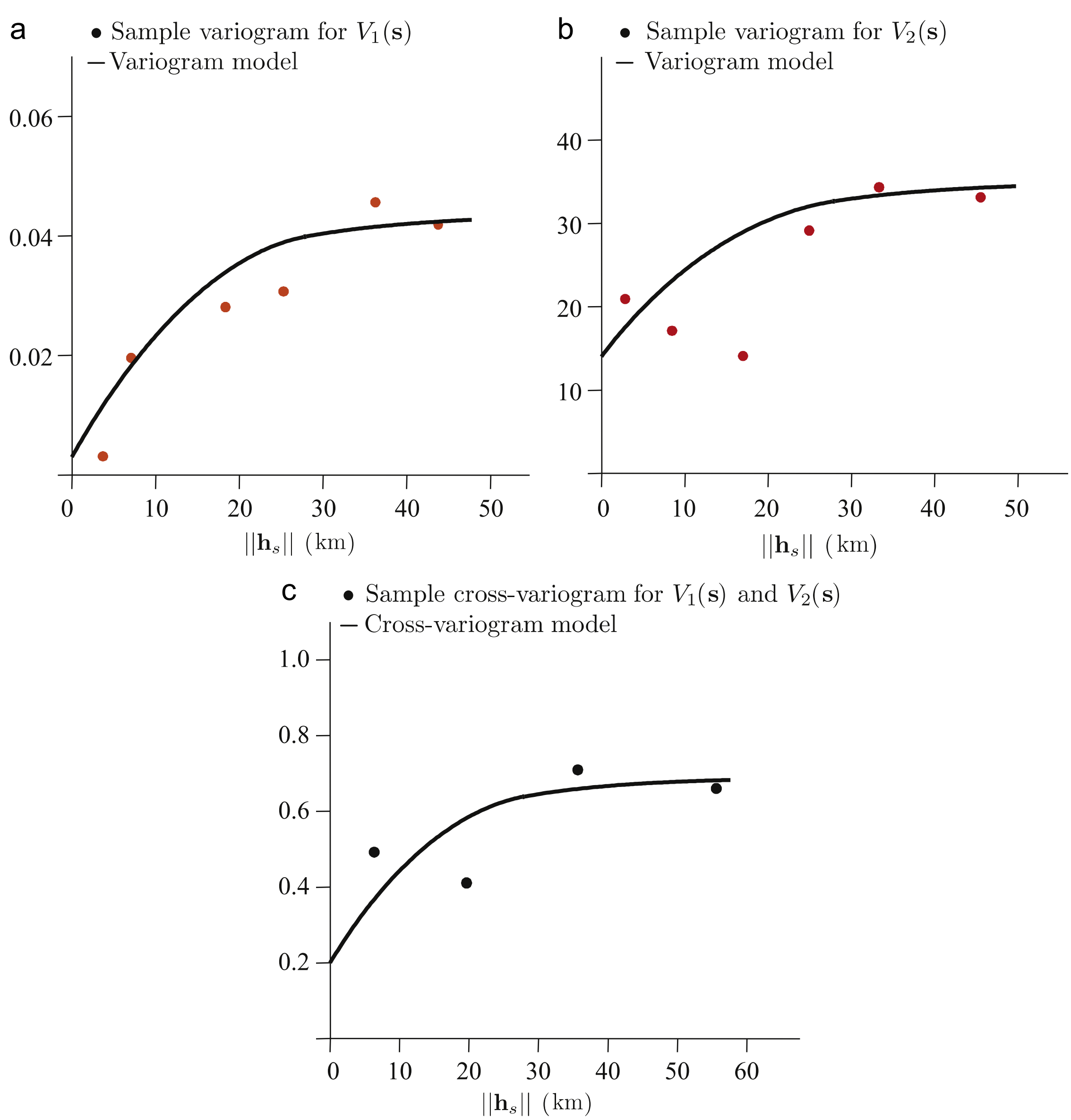

Fig. 9. (a) Sample variogram for $V_1(\mathbf{s})$; (b) sample variogram for $V_2(\mathbf{s})$; (c) sample cross-variogram for $V_1(\mathbf{s})$ and $V_2(\mathbf{s})$ and their corresponding models.

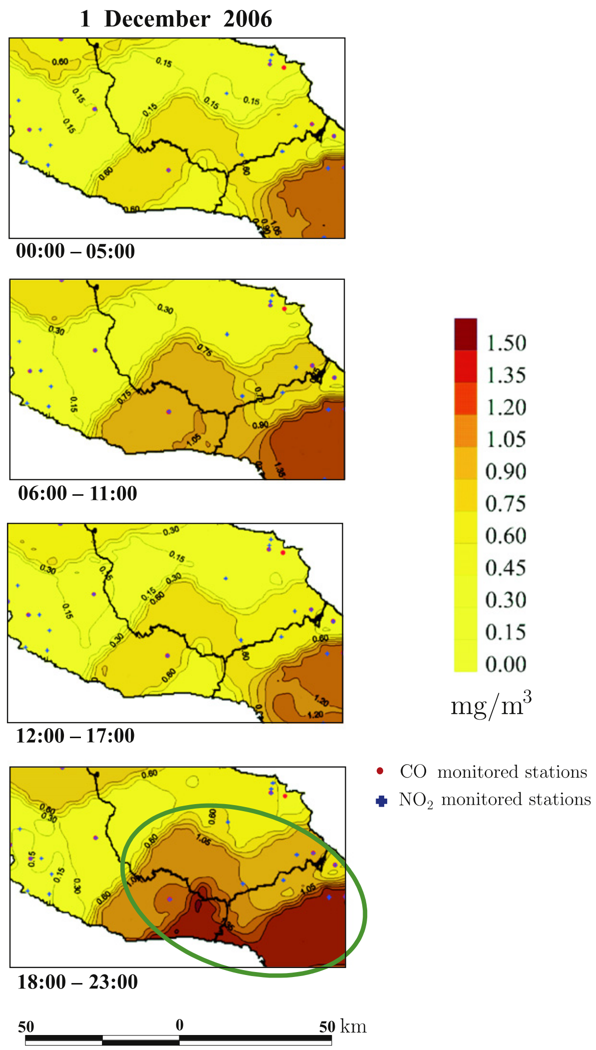

Fig. 10. Contour maps for predicted CO values averaged for each hour for 1st of December 2006.

Because there is a diurnal effect in this type of data, the values are modeled as the sum of two components: a diurnal component and a space–time random component. The diurnal components were estimated using a modified version of the "REMOVE" software in De Cesare et al. (2002) and spatial cokriging. Sample marginal space and temporal variograms were computed for the residuals using the modified "GAMV" program (De Cesare et al., 2002). These were used to model the *ST-LCM* which was then used with the "COK2ST" to predict values of CO for the first day after the dates for which data was available.

Acknowledgements

The authors wish to thank the Editor and the Referees for their suggestions that improved the final version of the paper.

Appendix A. REMOVEMULT.PAR

The parameter file required by the *FORTRAN* program "REMOVEMULT" is shown.

```
Parameters for REMOVEMULT
*************************

START OF PARAMETERS:
nov2006-          \data file
 NO2CO.dat
1 2 3 4           \column for cod,x,y, t
                  coordinates
nov2006-          \output file
 NO2COdst.dat
2                 \number of variable to consider
5                 \var ith column number
-9. 1.0e21        \tmin, tmax (trimming limits)
5                 \Max number of missing
                  consecutive values
24                \length
60                \Min number of data for
                  deseasonalization
1                 \time unit
1                 \number of time intervals (ti)
7297              \ti definition ti+1 values
8017
6                 \var ith column number
-9. 1.0e21        \tmin, tmax (trimming limits)
5                 \Max number of missing
                  consecutive values
24                \period
60                \Min number of data for
                  deseasonalization
1                 \time unit
1                 \number of time intervals (ti)
7297              \ti definition ti+1 values
8017
```

Appendix B. COK2ST.PAR

The parameter file required by the *GSLib* program "COK2ST", in order to provide space–time predictions for CO hourly residuals, is shown. This program and the related parameter file have been properly modified to model the space–time variogram matrix, involved in the prediction system, by the *ST-LCM* based on the generalized product–sum variogram model.

Note that the suffix "dst" to the data file name means "deseasonalized", since the values saved in the data file are the residuals of the variable under study, computed after removing the diurnal component at each station.

```
Parameters for COK2ST
*********************

START OF
 PARAMETERS:
nov2006-           \data file
 NO2COdst.dat
2                  \number of variables
                   primary+other
2 3 4 7 5          \columns for X,Y,T and variables
-0.01 1.0e21       \trimming limits
0                  \option: 0=grid, 1=jackknife,
                   2=cross
xxx.dat            \file with jackknife data
2 3 4 7 0          \columns for X,Y,T,primary vr
0                  \co-located cokriging? (0=no,
                   1=yes)
somedata.dat       \file with gridded covariate
4                  \column for covariate
3                  \debugging level: 0,1,2,3
cok2ST.dbg         \file for debugging output
```

```
1dic2006-COdst.dat       \file for output
34 684000 2500           \nx,xmn,xsiz
23 4455000 2500          \ny,ymn,ysiz
24 8017 1.0              \nt,tmn,tsiz
1 1 1                    \x, y, and z block discretization
2 10 6                   \min primary,max primary,max all sec
20000.0 20000.0 180.0    \maximum search radii: primary
20000.0 20000.0 180.0    \maximum search radii: all secondary
0.0 0.0 0.0              \angles for search ellipsoid
2                        \kriging type (0=SK, 1=OK, 2=OK-trad)
3.38 2.32 0.00 0.00      \mean(i),i=1,nvar
3                        \np/sp=num str. product or sum-prod
1                        \option: 0 = prod, 1 =sum-prod
1 1                      \variogram for "i" and "j"
2 0.062                  \ nst, spatial nugget effect
0 0.0 0.0 0.0 0.0        \ it,cc,ang1,ang2,ang3
0.0 0.0 0.0              \ a_hmax, a_hmin, a_vert
0 0.0 0.0 0.0 0.0        \ it,cc,ang1,ang2,ang3
0.0 0.0 0.0              \ a_hmax, a_hmin, a_vert
0.0217 0.062217          \ temporal nugget, global sill
2 0.0                    \ nst, spatial nugget effect
1 0.0867 0.0 0.0 0.0     \ it,cc,ang1,ang2,ang3
28000.0 28000.0 28000.0  \ a_hmax, a_hmin, a_vert
7 0.0931 0.0 0.0 0.0     \ it,cc,ang1,ang2,ang3
12.0 12.0 12.0           \ a_hmax, a_hmin, a_vert
0 0.1783                 \ temporal nugget, global sill
2 0.0                    \ nst, spatial nugget effect
2 0.05152 0.0 0.0 0.0    \ it,cc,ang1,ang2,ang3
45000.0 45000.0 45000.0  \ a_hmax, a_hmin, a_vert
7 0.03424 0.0 0.0 0.0    \ it,cc,ang1,ang2,ang3
36.0 36.0 36.0           \ a_hmax, a_hmin, a_vert
0 0.0802                 \ temporal nugget, global sill
1 2                      \variogram for "i" and "j"
2 0.1                    \ nst, spatial nugget effect
1 0.0 0.0 0.0 0.0        \ it,cc,ang1,ang2,ang3
0.0 0.0 0.0              \ a_hmax, a_hmin, a_vert
7 0.0 0.0 0.0 0.0        \ it,cc,ang1,ang2,ang3
0.0 0.0 0.0              \ a_hmax, a_hmin, a_vert
0.035 0.10035            \ temporal nugget, global sill
2 0.0                    \ nst, spatial nugget effect
1 1.8055 0.0 0.0 0.0     \ it,cc,ang1,ang2,ang3
28000.0 28000.0 28000.0  \a_hmax, a_hmin, a_vert
7 1.93923 0.0 0.0 0.0    \ it,cc,ang1,ang2,ang3
12.0 12.0 12.0           \ a_hmax, a_hmin, a_vert
0 3.7135                 \ temporal nugget, global sill
2 0.0                    \ nst, spatial nugget effect
2 2.25078 0.0 0.0 0.0    \ it,cc,ang1,ang2,ang3
45000.0 45000.0 45000.0  \ a_hmax, a_hmin, a_vert
7 1.49586 0.0 0.0 0.0    \ it,cc,ang1,ang2,ang3
36.0 36.0 36.0           \ a_hmax, a_hmin, a_vert
0 3.5058                 \ temporal nugget, global sill
2 2                      \variogram for "i" and "j"
2 180                    \ nst, spatial nugget effect
1 0.0 0.0 0.0 0.0        \ it,cc,ang1,ang2,ang3
0.0 0.0 0.0              \ a_hmax, a_hmin, a_vert
7 0.0 0.0 0.0 0.0        \ it,cc,ang1,ang2,ang3
0.0 0.0 0.0              \ a_hmax, a_hmin, a_vert
63.0 180.63              \ temporal nugget, global sill
2 0.0                    \ nst, spatial nugget effect
1 81.0 0.0 0.0 0.0       \ it,cc,ang1,ang2,ang3
28000.0 28000.0 28000.0  \ a_hmax, a_hmin, a_vert
7 87.0 0.0 0.0 0.0       \ it,cc,ang1,ang2,ang3
12.0 12.0 12.0           \ a_hmax, a_hmin, a_vert
0 166.6                  \ temporal nugget, global sill
2 0.0                    \ nst, spatial nugget effect
2 118.335 0.0 0.0 0.0    \ it,cc,ang1,ang2,ang3
45000.0 45000.0 45000.0  \ a_hmax, a_hmin, a_vert
7 78.645 0.0 0.0 0.0     \ it,cc,ang1,ang2,ang3
36.0 36.0 36.0           \ a_hmax, a_hmin, a_vert
0 184.32                 \ temporal nugget, global sill
```

References

Bilonick, R.A., 1986. Monthly hydrogen ion deposition maps for the Northeastern U.S. from July 1982 to September 1984. Atmospheric Environment 22, 1909–1924.

Brockwell, P.J., Davis, R.A., 1987. Time Series: Theory and Methods. Springer, New York, p. 577.

Brown, P.J., Le, N.D., Zidek, J.V., 1994. Multivariate spatial interpolation and exposure to air pollutants. The Canadian Journal of Statistics 22, 489–509.

Cressie, N., Huang, H., 1999. Classes of nonseparable, spatio-temporal stationary covariance functions. Journal of the American Statistical Association 94, 1330–1340.

De Cesare, L., Myers, D.E., Posa, D., 2002. FORTRAN programs for space–time modeling. Computers & Geosciences 28, 205–212.

De Iaco, S., Space–time correlation analysis: a comparative study. Journal of Applied Statistics, in press.

De Iaco, S., Myers, D.E., Posa, D., 2000. Total air pollution and space–time modeling. In: Monestiez, P., Allard, D., Froidevaux, R. (Eds.), GeoENV III Geostatistics for Environmental Applications. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 45–56.

De Iaco, S., Myers, D.E., Posa, D., 2001. Space–time analysis using a general product–sum model. Statistics and Probability Letters 52 (1), 21–28.

De Iaco, S., Myers, D.E., Posa, D., 2002. Nonseparable space–time covariance models: some parametric families. Mathematical Geology 34, 23–42.

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

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.

Deutsch, C.V., Journel, A.G., 1997. *GSLib*: Geostatistical Software Library and User's Guide, second ed. Oxford University Press, New York, p. 369.

Gneiting, T., 2002. Nonseparable, stationary covariance functions for space–time data. Journal of the American Statistical Association 97, 590–600.

Koike, K., Narikiyo, K., Matsuda, S., 2002. Spatio-temporal distribution modeling of iron component concentrations in groundwater using geostatistics. Japanese Journal of Applied Statistics 31, 41–58.

Kyriakidis, P.C., Journel, A.G., 2001a. Stochastic modeling of atmospheric pollution: a spatial time-series framework. Part I: methodology. Atmospheric Environment 35, 2331–2337.

Kyriakidis, P.C., Journel, A.G., 2001b. Stochastic modeling of atmospheric pollution: a spatial time-series framework. Part II: application to monitoring monthly sulfate deposition over Europe. Atmospheric Environment 35, 2339–2348.

Lark, R.M., Bellamy, P.H., Rawlins, B.G., 2006. Spatio-temporal variability of some metal concentrations in the soil of Eastern England and implications for soil monitoring. Geoderma 133, 363–379.

Liu, C., Koike, K., 2007. Extending multivariate space–time geostatistics for environmental data analysis. Mathematical Geology 39 (3), 289–305.

Myers, D.E., 1991. Matrix formulation of cokriging. Mathematical Geology 14 (3), 249–257.
Rouhani, S., Myers, D.E., 1990. Problems in space–time kriging of hydrogeological data. Mathematical Geology 22, 611–623.
Sen, Z., Altunkaynak, A., Özger, M., 2006. Space–time interpolation by combining air pollution and meteorologic variables. Applied Geophysics 163, 1435–1451.
Sicard, E., Sabatier, R., Niel, H., Cadier, E., 2002. A new approach in space–time analysis of multivariate hydrological data: application to Brazil's Nordeste region rainfall. Water Resources Research 38, 1319–1329.
Vanderlinden, K., Ordóñez, R., Polo, M.J., Giráldez, J.V., 2006. Mapping residual pyrite after a mine spill using non co-located spatiotemporal observations. Journal of Environmental Quality 35, 21–36.