

PARTICULATE MATTER AND METEOROLOGICAL CONDITIONS: A MULTIVARIATE SPATIO-TEMPORAL ANALYSIS

Cappello C.¹ & Palma M.² & Posa D.³

¹ *Department of Management, Economics, Mathematics and Statistics, University of Salento, Italy, claudia.cappello@unisalento.it*

² *Department of Management, Economics, Mathematics and Statistics, University of Salento, Italy, sandra.deiaco@unisalento.it*

³ *Department of Management, Economics, Mathematics and Statistics, University of Salento, Italy, donato.posa@unisalento.it*

Abstract. In multivariate Geostatistics, spatial direct and cross-correlation among the variables under study is measured by the matrix variogram or the matrix covariance function. The well known linear coregionalization model (LCM) introduced by Matheron in 1982, is a valid and simple model widely used to describe the matrix variogram (or covariance). Even in the spatio-temporal context, the space-time LCM (ST-LCM) based on admissible spatio-temporal models may successfully capture the spatio-temporal behaviour of the phenomena under study and can be used for prediction purposes. In this paper, the simultaneous diagonalization-based procedure has been applied to identify the basic hidden components which characterize the spatio-temporal correlation of three environmental variables; moreover the ST-LCM, using different basic models at the selected scales of spatio-temporal variability, has been proposed to obtain predictions of the primary variable. A modified GSLib routine for spatio-temporal cokriging has been used.

Keywords. spatio-temporal cross-correlation, product-sum model, non-separability model, spatio-temporal cokriging

1 Introduction

An environmental data set often concerns different correlated variables measured at some locations of the study area and for several time points. In this case, the data set presents a multivariate spatio-temporal structure; therefore appropriate modeling techniques which take into account the spatio-temporal relationships among the variables are needed.

In multivariate spatial analysis, spatial direct and cross-correlation of the variables is measured by the matrix variogram or matrix covariance function. The LCM, introduced by Matheron (1982), and several algorithms proposed in the literature for fitting the LCM (Babak and Deutsch, 2009; Emery, 2010; Gneiting et al., 2010) have been widely used in the applications. In the multivariate spatio-temporal context, the recent developments

of the LCM using the product-sum model and the new fitting procedure based on the simultaneous diagonalization of the sample matrix variogram (De Iaco et al., 2012, 2013), represent useful tools to analyze the spatio-temporal interrelationships among the variables under study and, consequently, to identify the basic hidden components in space and in time which characterize the same variables.

In this paper, a multivariate analysis of particle pollution in Southern Apulian region (Italy), is presented. The environmental data set involves particulate matter daily concentrations of aero-dynamic diameter less than 10 micrometer (PM_{10}) and two meteorological variables, such as rainfall and wind speed, measured during March 2012 at some monitoring stations located over the area of interest. It is well known that meteorological variables play a significant role on particle pollution. In general, low values of rainfall and wind speed are associated with high values of PM_{10} and vice-versa. Therefore, after a brief presentation of the spatio-temporal multivariate geostatistical framework, a case study is proposed and the following aspects are considered:

1. estimating the spatio-temporal interrelationships among the variables under study and, consequently, identifying the basic hidden components in space and in time which characterize the same variables;
2. modeling the spatio-temporal correlation among the variables under study by using the ST-LCM; in the fitting procedure, the simultaneous diagonalization-based method is applied to several matrix variograms in order to detect the basic independent components which contribute to define the multivariate correlation structure of the observed variables;
3. spatio-temporal cokriging, performed by a modified version of GSLib routine to predict PM_{10} pollution levels over the study area.

Note that the ST-LCM used in this paper is based on two structures of spatio-temporal correlation, related to different scales of spatio-temporal variability and the ST-LCM has been fitted by selecting two different classes of spatio-temporal correlation measures. It is important to point out that the suitable basic models have been selected through the inspection of the non-separability index (De Iaco and Posa, 2013).

2 Multivariate spatio-temporal modeling and prediction

Let $\{\mathbf{Z}(\mathbf{s}, t), (\mathbf{s}, t) \in D \times T \subseteq \mathbb{R}^{d+1}\}$, be a second-order stationary multivariate space-time random function (STRF), with $\mathbf{Z}(\mathbf{s}, t) = [Z_1(\mathbf{s}, t), \dots, Z_p(\mathbf{s}, t)]^T, p \geq 2$, whose matrix variogram is

$$\mathbf{\Gamma}(\mathbf{h}) = [\gamma_{ij}(\mathbf{h})], \quad (1)$$

where

- $\mathbf{h} = (\mathbf{h}_s, h_t)$, with $\mathbf{h}_s = (\mathbf{s} - \mathbf{s}')$ and $h_t = (t' - t)$;
- $\gamma_{ij}(\mathbf{h}) = Cov[(Z_i(\mathbf{s} + \mathbf{h}_s, t + h_t) - Z_i(\mathbf{s}, t)), (Z_j(\mathbf{s} + \mathbf{h}_s, t + h_t) - Z_j(\mathbf{s}, t))]$, are the cross-variograms between the Z_i and Z_j , when $i \neq j$, and the variograms of the Z_i when $i = j$, with $i, j = 1, \dots, p$.

As it is well known, a ST-LCM is a straightforward extension of the spatial LCM in the spatio-temporal context, i.e. the ST-LCM for the matrix variogram $\mathbf{\Gamma}(\mathbf{h})$ can be written as follows:

$$\mathbf{\Gamma}(\mathbf{h}) = \sum_{l=1}^L \mathbf{B}_l g_l(\mathbf{h}), \quad (2)$$

where $\mathbf{B}_l = [b_{ij}^l]$, $i, j = 1, \dots, p$, are $(p \times p)$ positive-definite matrices, called *coregionalization matrices*, and $g_l(\mathbf{h})$, $l = 1, \dots, L$, ($L \geq 2$), are basic admissible spatio-temporal variograms at L variability scales.

As shown in De Iaco et al. (2013), the space and time correlation structures of the basic components, which contribute to determine the correlation structures of the p observed variables, can be easily obtained through the simultaneous diagonalization of the sample matrix variograms for a selection of spatio-temporal lags. From the spatio-temporal surfaces of the basic components, it is very simple to determine the number of the scales of variability, as well as the spatial and temporal ranges associated with each scale.

Moreover, by the inspection of the non-separability index (De Iaco and Posa, 2013) computed for all lags, in space and in time, of the basic components sample variograms, it is easy to choose the class of spatio-temporal models which is suitable for the basic components. Then the ST-LCM, proposed by De Iaco et al. (2005, 2012, 2013), can be implemented by using two or more different spatio-temporal variogram models for the independent basic components.

Finally the fitted model is used to predict the variable of interest (primary variable). In particular, for a second-order stationary multivariate random function \mathbf{Z} , a prediction of the primary variable, at an unsampled point $\mathbf{u} = \mathbf{s}, t$, can be obtained by the cokriging predictor:

$$\hat{\mathbf{Z}}(\mathbf{u}) = \sum_{\alpha=1}^n \Lambda_{\alpha}(\mathbf{u}) \mathbf{Z}(\mathbf{u}_{\alpha}), \quad (3)$$

where $\mathbf{u}_{\alpha} \in D \times T$, $\alpha = 1, \dots, n$, are the sampled points and $\Lambda_{\alpha}(\mathbf{u})$, $\alpha = 1, \dots, n$, are matrices of the weights which are determined so that predictor (3) is unbiased and efficient (Journel, 1981).

3 Case study

Particulate matter is a complex mixture of extremely small particles and liquid droplets, made up essentially of nitrates and sulphates, which could cause serious health problems,

such as respiratory illness, among others. Moreover meteorological variables, such as rainfall and wind speed, play a significant role on the particle pollution, since low values of rainfall and wind speed are associated with high values of particle pollution, and vice-versa. Hence, efficient monitoring systems which control PM_{10} concentration and the application of appropriate spatio-temporal models to predict particle pollution are useful tools for the environmental and human health protection.

3.1 Exploratory data analysis

The analyzed data set consists of PM_{10} daily concentrations (expressed in $\mu g/m^3$), total daily rainfall (mm) and wind speed daily averages (m/sec), collected during March 2012 at several monitoring stations located in Southern Apulian region (Lecce, Brindisi, and Taranto districts), as shown in Fig. 1. In particular, there are 63 meteorological stations for monitoring rainfall and wind speed and 31 PM_{10} survey stations, which are either traffic or industrial stations, depending on the area where they are located. Note that,

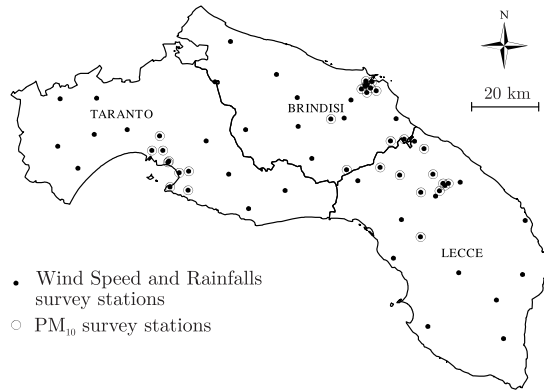


Figure 1: Posting map of the survey stations in Southern Apulian region.

according to National Laws concerning the human health protection, PM_{10} daily average concentrations, cannot be greater than $50 \mu g/m^3$ for more than 35 times per year. By analyzing the PM_{10} daily averages measured during March 2012 in the study area, the threshold value has been overcome 47 times, especially on the 3rd, 4th, 5th and 20th of March, as shown in Fig. 2-a).

It is well known that rainfall and wind speed are negatively correlated: generally low values of rainfall are associated to high values of wind speed and vice-versa (Fig. 2-b)). Moreover, the dispersion of particulate matter is strongly influenced by meteorological variables, since high values of rainfall and wind speed favour the dispersion of the pollutant, however this effect is not immediate and a delay effect can be noted.

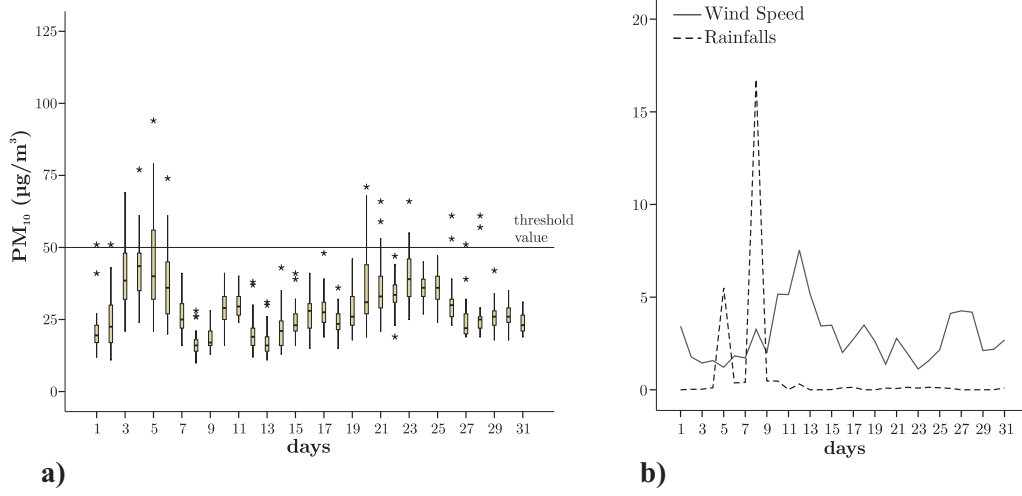


Figure 2: a) boxplot of PM_{10} daily concentrations, classified by days, b) time series plot of total daily rainfall and wind speed daily averages

3.2 Modeling ST-LCM

The space-time variogram surfaces of the variables under study (Fig. 3) have been computed for 7 spatial lags and 6 temporal lags, resulting in 42 symmetric (3×3) matrices of sample direct and cross-variograms. Then, by using the simultaneous diagonalization-based procedure, these matrices have been simultaneously nearly diagonalized and 42 diagonal matrices have been obtained to generate the sample spatio-temporal variogram surfaces of the latent basic components. Since the variograms of the basic components have highlighted two scales of variability (at 10 km and 4 days, the first scale; at 24 km and 7 days, the second scale) solely two basic components have been retained.

3.3 Non-separability index

Given the spatio-temporal surfaces of the basic components, the problem of the choice of an appropriate class of models has been faced up and the non-separability index (De Iaco and Posa, 2013) has been computed.

In particular, the new function called `sep_index` (Cappello et al., 2014), defined for determining the non-separability index (De Iaco and Posa, 2013) in the R environment, has been applied in order to analyze the space-time correlation structure of either the basic components at small and large scale of variability. The function `sep_index` returns the purely spatial and purely temporal sample covariance matrices, empirical non-separability index ratio and box-plots of sample non-separability ratios, classified for spatial lags and temporal lags. In the present case study, the non-separability index computed for the first (small scale) basic component sample variogram, is greater than 1 for all lags in space and

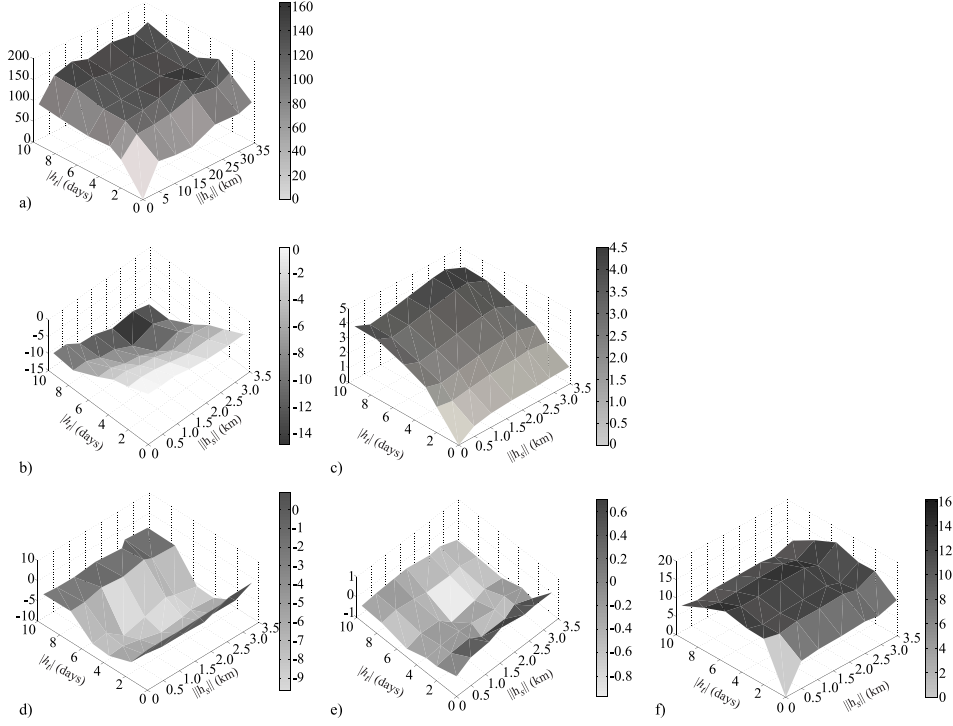


Figure 3: Spatio-temporal variogram and cross-variogram surfaces of a) PM_{10} , b) wind speed vs PM_{10} c) wind speed d) rainfall vs PM_{10} , e) rainfall vs wind speed, f) rainfall

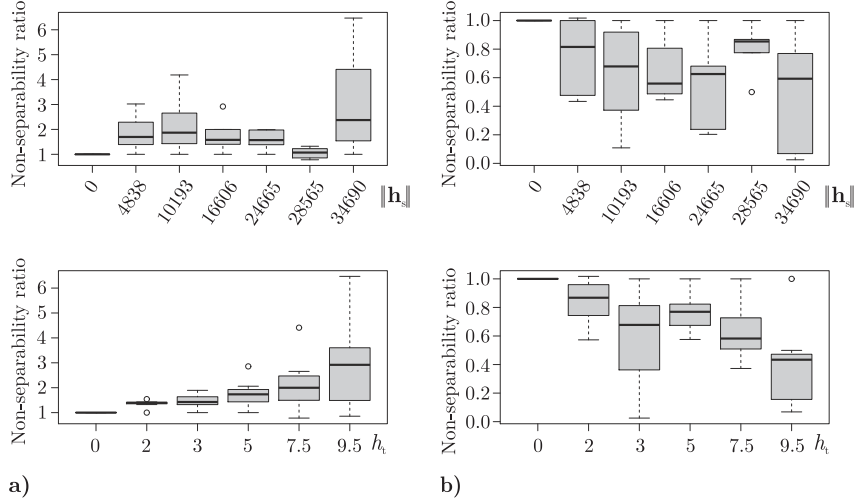


Figure 4: Box-plots of sample non-separability ratio, computed for basic components, classified for spatial lags and temporal lags, concerning a) first scale, b) second scale of spatio-temporal variability

in time (Fig. 4-a)). Hence, it should be suitable to select a correlation model characterized by uniformly positive non-separability, such as the Gneiting model. In particular, at the first scale of spatio-temporal variability, the following basic variogram model belonging to Gneiting class (Gneiting, 2002), has been considered:

$$g_1(\mathbf{h}_s, h_t) = 2.68 \left[1 - \frac{1}{(2.5|h_t| + 1)} \exp \left\{ -\frac{3\|\mathbf{h}_s\|^{2\cdot\lambda}}{10(2.5|h_t| + 1)^{\beta\cdot\lambda}} \right\} \right], \quad (4)$$

with $\lambda = 0.39$ and $\beta = 0.5$.

On the other hand, the non-separability index is less than 1 for all spatial and temporal lags (Fig. 4-b)) of the sample variogram computed for the second (large scale) basic component; therefore, it should be suitable to select a correlation model characterized by uniformly negative non-separability, such as the product-sum model. In this case the following generalized product-sum model (De Iaco et al., 2001, 2011) has been fitted to the sample spatio-temporal variogram of the second basic component:

$$g_2(\mathbf{h}_s, h_t) = \gamma_2(\mathbf{h}_s, 0) + \gamma_2(\mathbf{0}, h_t) - k_2 \gamma_2(\mathbf{h}_s, 0) \gamma_2(\mathbf{0}, h_t), \quad (5)$$

where $\gamma_2(\mathbf{h}_s, 0) = 93 \text{Exp}(\|\mathbf{h}_s\|; 24)$ and $\gamma_2(\mathbf{0}, h_t) = 97 \text{Exp}(h_t; 7)$, are, respectively, the spatial and temporal marginal variograms, while $k_2 = 0.0049$ is the spatio-temporal interaction parameter. Note that $\text{Exp}(\cdot; a)$ denotes the well known exponential model with practical range a (Deutsch and Journel, 1998).

3.4 The fitted ST-LCM

As described in De Iaco et al. (2012), by dividing the contribution (sill) of the direct and cross-variograms surfaces at the 2 scales of variability, by the corresponding global sills (2.689 at the first scale and 145.97 at the second one), the following coregionalization matrices have been determined:

$$\mathbf{B}_1 = \begin{bmatrix} 49.4236 & -1.8111 & -2.0937 \\ -1.8111 & 0.9520 & -0.0595 \\ -2.0937 & -0.0595 & 4.0907 \end{bmatrix}, \quad \mathbf{B}_2 = \begin{bmatrix} 0.0987 & -0.0536 & 0.0341 \\ -0.0536 & 0.0300 & -0.0195 \\ 0.0341 & -0.0195 & 0.0147 \end{bmatrix}. \quad (6)$$

Finally, the *ST-LCM* for PM_{10} , rainfall and wind speed daily measurements is given below

$$\Gamma(\mathbf{h}_s, h_t) = \mathbf{B}_1 g_1(\mathbf{h}_s, h_t) + \mathbf{B}_2 g_2(\mathbf{h}_s, h_t), \quad (7)$$

where the matrices $\mathbf{B}_l, l = 1, 2$, are as in (6) and the space-time variograms $g_l(\mathbf{h}_s, h_t), l = 1, 2$, are modelled as in (4), for $l = 1$, and as in (5) for $l = 2$.

3.5 Particle pollution predictions

First of all, the reliability of the fitted ST-LCM has been assessed by jackknife technique, then the same model has been used to predict PM_{10} for three days after the last available

date. In particular, the 29th, 30th and 31st of March have been considered for jackknife analysis, hence PM_{10} daily concentrations recorded during these days over the study area have been compared with the predicted ones. The correlation coefficients between true and predicted values have been equal to 0.73, 0.78 and 0.77 for the 29th, 30th and 31st of March, respectively. The jackknife results evidently encourage the analyst to use the above ST-LCM in order to predict PM_{10} daily concentrations by spatio-temporal cokriging. It is important to point out that a modified version of the COK2ST, proposed by De Iaco et al. (2010), which allows using the ST-LCM based on different basic models, has been implemented either to perform jackknife technique and to obtain predictions over a grid. Finally, for the 1st, the 2nd and the 3rd of April 2012, predictions of PM_{10} daily averages have been computed over a grid which covers the domain under study (48×34 nodes with a separation distance equal to 3 km). Fig. 5 shows the contour maps of the predicted PM_{10} daily concentrations. It is evident that the south-east area is characterized by the

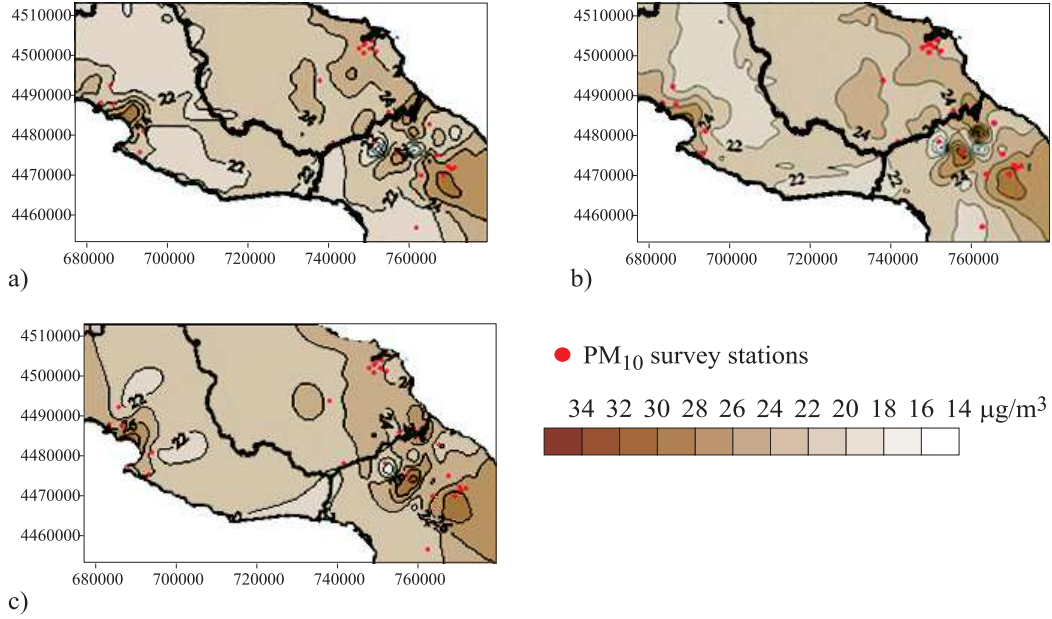


Figure 5: Contour maps for predicted PM_{10} daily concentrations for a) the 1st, b) the 2nd and c) the 3rd of April 2012

greater PM_{10} predicted values with respect to the values predicted for the other parts of the region. However, for the considered period (1-3 April 2012), particle pollution does not exceed the threshold value fixed by the national law.

4 Summary

In this paper, a multivariate geostatistical approach has been discussed and predictions of PM_{10} daily concentrations, during a critical period of the year, over an urban area, have been illustrated. It is worth noting that spatio-temporal cokriging predictions of PM_{10} daily values have been determined by taking into account some meteorological variables, such as rainfall and wind speed, since these meteorological variables play a significant role on particle pollution. The spatio-temporal correlation among the study variables has been described by a ST-LCM based on two independent basic components at two scales of variability in space and time. Moreover, the choice of an appropriate class of models for the basic components, has been supported by the computation of non-separability index. Future researches should aim to an interaction between space-time modeling of air pollution and urban environment representation (traffic network, location of industrial facilities, emission sources and topographic conditions), easily managed in a Geographic Information System (GIS), in order to generate science-based spatio-temporal maps providing useful representations of air quality over the study area.

5 Acknowledgements

This research has been supported by the Interreg project “Development of an Environmental Meteorological Stations Network on the Ionian Islands and Southern Italy-DEMSNIISI” Greece-Italy 2007-2013.

References

- [1] Apasanovich T.V., Genton M.G. (2010), *Cross-covariance functions for multivariate random fields based on latent dimension*, Biometrika, 97,15–30.
- [2] Babak O., Deutsch C.V. (2009), *An intrinsic model of coregionalization that solves variance inflation in collocated cokriging*, Computers & Geosciences, 35(3), 603–614.
- [3] Cappello C., De Iaco S., Posa D. (2013), *Structural Analysis and modeling choice for air quality data*, Proceedings of the International Conference GeoENV 2014, Paris (France).
- [4] Choi J., Fuentes M., Reich B.J., Davis J.M. (2009), *Multivariate spatial-temporal modeling and prediction of speciated fine particles*, Journal of Statistical Theory and Practice, 3(2), 407–418.
- [5] De Iaco S., Posa D. (2013), *Positive and negative non-separability for space-time covariance models*, Journal of Statistical Planning and Inference, 143, 378–391.
- [6] 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

- [7] De Iaco S., Myers D.E., Posa D. (2011), *On strict positive definiteness of product and product-sum covariance models*, Journal of Statistical Planning and Inference, 141, 1132–1140.
- [8] De Iaco S., Palma M., Posa D. (2005), *Modeling and prediction of multivariate space-time random fields*, Computational statistics & data analysis, 48, 525–547.
- [9] De Iaco S., Palma M., Posa D. (2012), *Prediction of particle pollution through space-time multivariate analysis* AStA Adv. Stat. Anal., 97, 133–150.
- [10] De Iaco S., Myers D.E., Palma M., Posa D. (2010), *FORTTRAN programs for spacetime multivariate modeling and prediction*, Computers & Geosciences, 36(5), 636–646.
- [11] De Iaco S., Maggio S., Palma M., Posa D. (2012), *Towards an automatic procedure for modeling multivariate space-time data*, Computers & Geosciences, 41, 1–11.
- [12] De Iaco S., Myers D.E., Palma M., Posa D. (2013), *Using Simultaneous Diagonalization to Identify a SpaceTime Linear Coregionalization Model*, Mathematical Geosciences.
- [13] Deutsch C.V., Journel A.G. (1998) *GSLib: Geostatistical Software Library and User's Guide*, Oxford Univ. Press, New York.
- [14] Emery X. (2010), *Interactive algorithms for fitting a linear model of coregionalization*, Computers & Geosciences, 36(9), 1150–1160.
- [15] Gelfand A.E, Banerjee S., Gamerman D. (2005), *Spatial process modelling for univariate and multivariate dynamic spatial data*, Environmetrics, 16, 465–479.
- [16] Gneiting T. (2002), *Nonseparable, stationary covariance functions for space-time data*, Journal of the American Statistical Association, 97, 590–600.
- [17] Gneiting T., Kleiber W., Schlather M. (2010), *Matérn cross-covariance functions for multivariate random fields*, Journal of American Statistical Association, 105(491), 1167–1177.
- [18] Li B., Genton M.G., Sherman M. (2008), *Testing the covariance structure of multivariate random fields*, Biometrika, 95(4), 813–829.
- [19] Matheron G. (1982), *Pour une analyse krigéante des données régionalisées*, Rapport technique N732, Ecole Nationale Supérieure des Mines de Paris.