

STRUCTURAL ANALYSIS AND MODELING CHOICE FOR AIR QUALITY DATA

Cappello C.¹ & De Iaco S.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. Non-separable models are receiving a lot of attention, since they are more flexible to handle empirical covariances showed up in applications. Different forms of non-separability for space-time covariance functions have been recently defined in literature. In particular, wide classes of non-separable spatio-temporal covariance functions are able to capture positive and negative non-separability.

In this paper, a review of various well-known non-separable classes of space-time covariance models and their classification according to the notion of non-separability are presented. The inspection of the box-plots of the corresponding non-separability index, classified for spatial and temporal lags, could be helpful to discuss the type of non-separability as well as to select appropriate covariance model for the data. These aspects together with other geometric features of the covariance surface are related to the capability of a model to reflect the possible different variability in space and time and to the power of describing the interaction between space and time.

In the case study, after removing the spatio-temporal trend, residuals are used for structural analysis. Marginals behavior, the asymptotic behavior of the sample covariance surface and the empirical non-separability index are inspected through the implementation of some functions in the R environment. Box-plots of sample non-separability ratios, classified for spatial lags and temporal lags, are proposed as an appropriate diagnostic tool to detect different forms of non-separability. Moreover, after selecting some suitable space-time covariance models, a fitting procedure based on the least squares of the errors between the empirical and the theoretical non-separability indexes is introduced.

Keywords. space-time covariance models, non-separability index, environmental data.

1 Introduction

Although the separable covariance functions have represented one of the first attempts to model the spatio-temporal variability, they are often suitable to be used in environmental applications. For this reason, various classes of non-separable space-time covariance

functions have been introduced in the literature.

Starting from the separable covariance models, some parametric families of space-time covariance functions have been generated (De Iaco et al., 2001, 2002; Ma, 2002, 2003). Otherwise, various classes of non-separable space-time covariances have been constructed through different approaches by Cressie and Huang (1999), Gneiting (2002), Rodriguez and Diggle (2010), among others.

Many statistical tests for separability have been proposed in literature and are based on parametric models (Shitan and Brockwell, 1995; Guo and Billard, 1998; Brown et al., 2000), likelihood ratio tests and subsampling (Mitchell et al., 2005) or spectral methods (Scaccia and Martin, 2005; Fuentes, 2006). However, the statistical tests for separability mentioned above can help to decide for a separable or a non-separable model, but they are not able to suggest the type of non separability and a specific class of models. Rodriguez and Diggle (2010) introduced a first measure of non-separability for a spatial-temporal covariance function; afterwards De Iaco and Posa (2013) proposed a generalization of this index. Classifying models with respect to the type of non-separability might help in choosing the class of models more suitable for the spatio-temporal data to be analyzed. Then, after selecting some appropriate space-time covariance models, a fitting procedure based on the least squares of the errors between the empirical and the theoretical non-separability indexes can be used. Furthermore, the implementation of the non-separability index and its representation in the R environment represents a useful tool in modeling and describing the evolution of a spatio-temporal processes (Bivand and Gebhardt, 2000).

2 Positive and negative non-separability

Let Z be a second order stationary spatio-temporal random variable. A stationary space-time covariance function C is separable if there exist stationary, purely spatial and purely temporal correlation functions ρ_s and ρ_t , respectively, such that:

$$C(\mathbf{h}_s, h_t) = \sigma^2 \rho_s(\mathbf{h}_s) \rho_t(h_t) \quad \forall (\mathbf{h}_s, h_t) \in D \times T \subseteq \mathbb{R}^{d+1},$$

where $\sigma^2 = C(\mathbf{0}, 0)$ and $\rho_s(\mathbf{0}) = \rho_t(0) = 1$.

De Iaco and Posa (2013) proposed a generalization of the two definitions of non-separability suggested by Rodriguez and Diggle (2010) in order to distinguish between pointwise and uniformly positive and negative non-separability. In particular, given a spatio-temporal covariance, let $\rho(\mathbf{h}_s, h_t; \Theta)$ be the the corresponding spatio-temporal correlation function, De Iaco and Posa (2013) proposed the following index of non-separability:

$$r(\mathbf{h}_s, h_t; \Theta) = \frac{\rho(\mathbf{h}_s, h_t; \Theta)}{\rho(\mathbf{h}_s, 0) \rho(\mathbf{0}, h_t; \Theta)}. \quad (1)$$

Then, a covariance is uniformly positive non-separable if

$$r(\mathbf{h}_s, h_t; \Theta) > 1, \quad \forall (\mathbf{h}_s, h_t) \in D \times T \subseteq \mathbb{R}^{d+1}, \quad (\mathbf{h}_s, h_t) \neq (\mathbf{0}, 0), \quad \forall \Theta, \quad (2)$$

while it is uniformly negative non-separable if

$$r(\mathbf{h}_s, h_t; \Theta) < 1, \quad \forall (\mathbf{h}_s, h_t) \in D \times T \subseteq \mathbb{R}^{d+1}, \quad (\mathbf{h}_s, h_t) \neq (\mathbf{0}, 0), \quad \forall \Theta. \quad (3)$$

On the other hand, if $r(\mathbf{h}_s, h_t; \Theta) > 1$, for some $(\mathbf{h}_s, h_t; \Theta)$, the covariance function is pointwise positive non-separable at the same $(\mathbf{h}_s, h_t; \Theta)$. Alternatively, the covariance C is pointwise negative non-separable at $(\mathbf{h}_s, h_t; \Theta)$, if $r(\mathbf{h}_s, h_t; \Theta) < 1$ for the same $(\mathbf{h}_s, h_t; \Theta)$. As pointed out by De Iaco and Posa (2013), a covariance function which is uniformly positive (negative) non-separable is also pointwise positive (negative) non-separable, but the converse is not true. According to the above definition (1), some classes of covariance models could be classified with reference to the type of non-separability in

- uniformly non separable models, that is the class of product-sum models (De Cesare et al., 2001; De Iaco et al., 2001), the Gneiting class of space-time covariance models (Gneiting, 2002) and Cressie-Huang class of models (1999);
- models with different non-separability indexes, that is the spherical metric model, the class of space-time stationary covariance functions generated by positive mixtures (Ma, 2002), the class of integrated models (De Iaco et al., 2002) and the class of models proposed by Rodriguez and Diggle (2010).

In the following some examples of space-time covariance models have been introduced and the corresponding non-separability indexes have been computed and illustrated through box-plots, classified for different spatial and temporal lags. For this aim, some functions have been implemented in the R environment.

2.1 Uniformly non separable models

The product-sum model, (De Cesare et al., 2001; De Iaco et al., 2001) is characterized by uniform negative non-separability, hence the space-time covariance function is not greater than the product of the corresponding marginals.

In particular, given the following product-sum model

$$\rho(\mathbf{h}_s, h_t; \Theta) = k_1 \exp(-a \|\mathbf{h}_t\|) (1/(b \|\mathbf{h}_s\|^2 + 1)) + k_2 (1/(b \|\mathbf{h}_s\|^2 + 1)) + k_3 \exp(-a \|\mathbf{h}_t\|), \quad (4)$$

where $\Theta = (a, b, k_1, k_2, k_3) = (1, 1, 0.2, 0.4, 0.4)$, the corresponding non-separability ratios have been computed. As shown in Fig. 1, lower values of the non-separability ratio are obtained as spatial and temporal lags increase.

The Gneiting class of spatio-temporal covariance models (Gneiting, 2002) is characterized by uniform positive non-separability, namely the space-time covariance is not lower than the product of the corresponding marginals.

In particular, given the following Gneiting model

$$\rho(\mathbf{h}_s, h_t; \Theta) = \frac{\sigma^2}{(a \|\mathbf{h}_t\|^{2\alpha} + 1)^{\beta d/2}} \exp \left(-\frac{c \|\mathbf{h}_s\|^{2\gamma}}{(a \|\mathbf{h}_t\|^{2\alpha} + 1)^{\beta\gamma}} \right), \quad (5)$$

where $\Theta = (a, c, d, \sigma^2, \alpha, \beta, \gamma) = (2, 0.4, 2, 1, 1, 1, 1)$, the corresponding non-separability ratios have been computed. As shown in Fig. 2, higher values of the non-separability

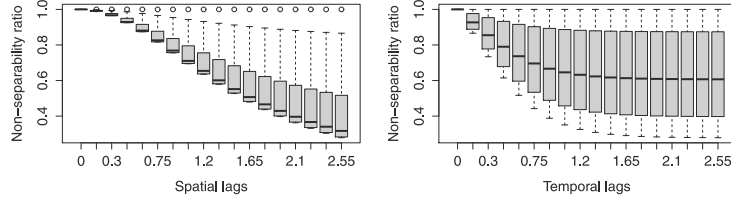


Figure 1: box plots of non-separability ratio, based on the product-sum model (4) and classified for spatial lags and temporal lags.

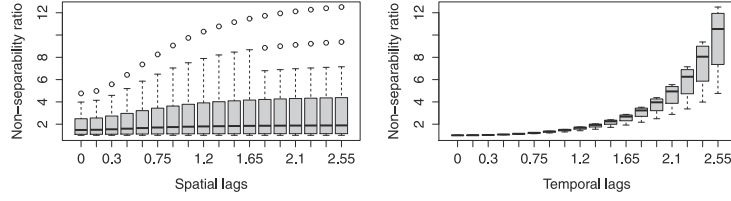


Figure 2: box plots of non-separability ratio, based on the Gneiting covariance model (5) and classified for spatial lags and temporal lags.

ratio are obtained as spatial and temporal lags increase.

Finally, Cressie-Huang class of spatio-temporal covariance models (Cressie and Huang, 1999), is characterized by different types of non-separability.

In particular, given the following Cressie-Huang model

$$\rho(\mathbf{h}_s, h_t; \Theta) = \frac{\sigma^2 c^{d/2}}{(a^2 h_t^2 + 1)^{1/2} (a^2 h_t^2 + c)^{d/2}} \exp \left[-b \left(\frac{a^2 h_t^2 + 1}{a^2 h_t^2 + c} \right)^{1/2} \|\mathbf{h}_s\| \right] \quad (6)$$

the corresponding non-separability ratios have been computed. As shown in Fig. 3, if $c = 0.1$ the model is characterized by a uniformly positive non-separability (Fig. 3-a),

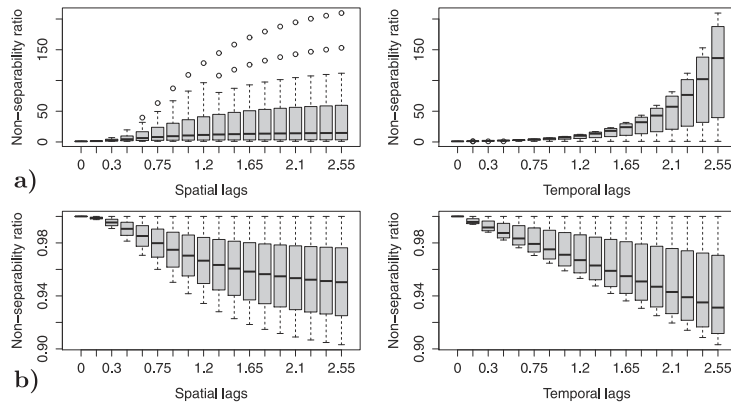


Figure 3: box plots of non-separability ratio, based on the Cressie-Huang covariance model (6) and classified for spatial lags and temporal lags. (a) Parameters: $a = 1, b = 1, c = 0.1, d = 2, \sigma^2 = 1$. (b) Parameters: $a = 1, b = 1, c = 1.1, d = 2, \sigma^2 = 1$

on the other hand, if $c = 1.1$ the model is characterized by a uniformly negative non-separability (Fig. 3-b)).

2.2 Models with different non-separability indexes

In the following some classes of space-time covariance functions, characterized by different non-separability indexes have been discussed.

Given the following spherical metric model, whose analytic expression is given below

$$\rho(h; a_1, a_2, b) = \begin{cases} \left[1 - 1.5 \frac{h}{b} + 0.5 \left(\frac{h}{b} \right)^3 \right], & 0 \leq h \leq b \\ 0, & h > b, \end{cases} \quad (7)$$

where $h = (a_1 \|\mathbf{h}_s\|^2 + a_2 |h_t|^2)^{0.5}$ and $b \in \mathbb{R}_+$ is the spatio-temporal range, it is easy to show that it is non-uniformly non-separable, as shown in Fig. 4. In fact, given different combinations of the parameters a_1, a_2 and b there always exist lags $(\mathbf{h}_s, h_t) \neq (\mathbf{0}, 0)$ such that $r(\mathbf{h}_s, h_t; a_1, a_2, b)$ associated to the model (7) is not always greater or less than 1. In particular, the non-separability index r is greater than 1 for small spatio-temporal lags, it becomes less than 1 when the distance h tends to the range b .

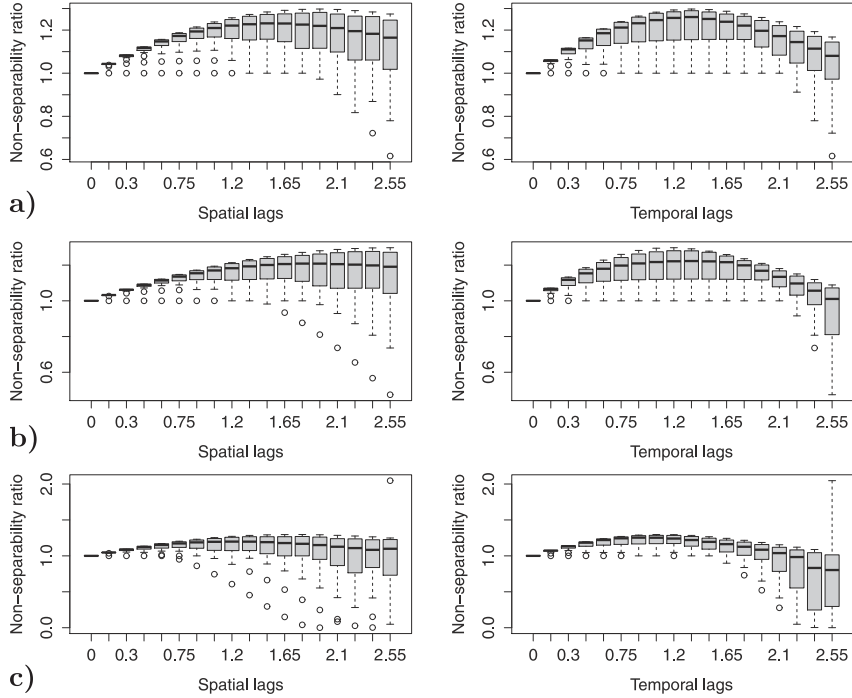


Figure 4: box plots of non-separability ratio, defined for the spheric model (7) and classified for spatial lags and temporal lags. (a) Parameters: $a_1 = 1$, $a_2 = 2$ and $b = 5$. (b) Parameters: $a_1 = 0.2$, $a_2 = 1$ and $b = 3$. (c) Parameters: $a_1 = 0.2$, $a_2 = 0.5$ and $b = 2$.

The example 4 proposed by Ma (2002)

$$\rho(\mathbf{h}_s, h_t, \Theta) = \frac{\log [1 - \theta_1 \rho_s(\mathbf{h}_s) - \theta_2 \rho_t(h_t) - \theta_{12} \rho_s(\mathbf{h}_s) \rho_t(h_t)]}{\log \theta_3} \quad (8)$$

where $\theta_3 = 1 - (\theta_1 + \theta_2 + \theta_{12})$ and θ_{12} are non-negative constants such that $0 < \theta_1 + \theta_2 + \theta_{12} < 1$, is non-uniformly non-separable, that is the ratio r is greater or less than 1, depending on the lags and/or the parameter of the model (Fig. 5-a)). However, the class of covariance

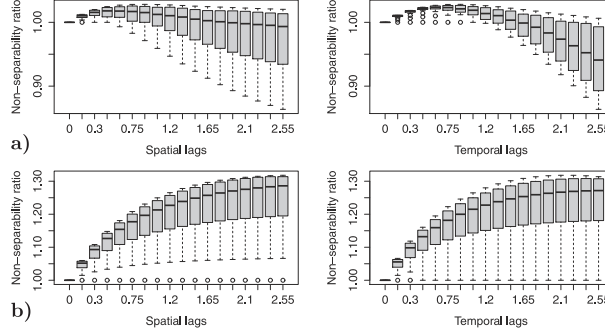


Figure 5: box plots of non-separability ratio, defined for a mixture model proposed by Ma (8) and classified for spatial lags and temporal lags. (a) Parameters: $a_1 = 1$, $a_1 = 2$, $\theta_1 = 0.1$, $\theta_2 = 0.3$ and $\theta_{12} = 0.2$. (b) Parameters: $a_1 = 2$, $a_1 = 2$, $\theta_1 = 0.1$, $\theta_2 = 0.2$ and $\theta_{12} = 0.5$.

functions proposed by Ma (2002) is able to describe uniformly positive and negative non-separability, as shown in Fig. 5-b).

The class of integrated models, proposed by De Iaco et al. (2002) is flexible enough to handle either uniformly positive and negative non-separability or non-uniform non-separability, depending on the lags and/or to the parameters of the chosen models. In particular, consider the following integrated product-sum model:

$$\begin{aligned} \rho(\mathbf{h}_s, h_t, \Theta) &= \int_0^\infty \left[k_1 e^{-x \|\mathbf{h}_s\|^\alpha / b} e^{-x |h_t|^\delta / c} + k_2 e^{-x \|\mathbf{h}_s\|^\alpha / b} + k_3 e^{-x |h_t|^\delta / c} \right] \frac{\beta^{n+1}}{\Gamma(n+1)} x^n e^{-\beta x} dx \\ &= k_1 \frac{\beta^{n+1}}{\left(\frac{\|\mathbf{h}_s\|^\alpha}{b} + \frac{|h_t|^\delta}{c} + \beta \right)^{n+1}} + k_2 \frac{\beta^{n+1}}{\left(\frac{\|\mathbf{h}_s\|^\alpha}{b} + \beta \right)^{n+1}} + k_3 \frac{\beta^{n+1}}{\left(\frac{|h_t|^\delta}{c} + \beta \right)^{n+1}} \end{aligned} \quad (9)$$

where $\Theta = (b, c, \alpha, \delta, k_1, k_2, k_3, \beta, n)$ is the parameters vector.

Let $\Theta = (1, 1, 1, 1, 0.4, 0.3, 0.3, 2, 2)$ the integrated product-sum model describes space-time covariances that are uniformly negative non-separable (Fig. 6-a)); on the other hand, let $\Theta = (1, 1, 2, 2, 0.7, 0, 0.3, 2, 2)$ the integrated product-sum model describes space-time covariances that are uniformly positive non-separable (Fig. 6-b)). Finally, if $\Theta = (1, 1, 1, 1, 0.6, 0.1, 0.3, 1, 1)$ the integrated product-sum model describes space-time covariances that are non-uniformly non-separable (Fig. 6-c)). The class of models proposed by Rodriguez and Diggle (2010) is given below

$$C(\mathbf{h}_s, h_t, \Theta) = \frac{\sigma^2}{2} [\rho_{s,1}(\mathbf{h}_s; \theta_1) \rho_{t,1}(h_t; \phi_1) + \rho_{s,2}(\mathbf{h}_s; \theta_2) \rho_{t,2}(h_t; \phi_2)] \quad (10)$$

where $\rho_{s,1}(\mathbf{h}_s; \theta_1)$, $\rho_{s,2}(\mathbf{h}_s; \theta_2)$, $\rho_{t,1}(h_t; \phi_1)$, $\rho_{t,2}(h_t; \phi_2)$ are, respectively, two non-negative and integrable spatial correlation functions and two non-negative and integrable temporal correlation functions. Note that the above class of covariance functions could be

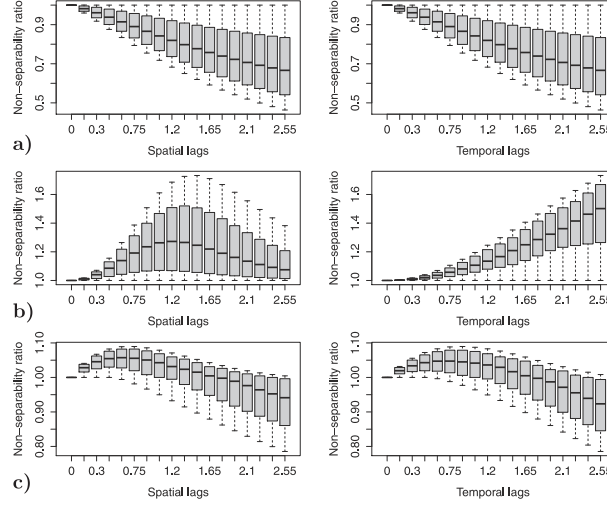


Figure 6: box plots of non-separability ratio, defined for the integrated product-sum model (9) and classified for spatial lags and temporal lags.

considered as a special case of integrated product models.

If $\rho_{s,1}(\mathbf{h}_s, a_{s_1}) = e^{-a_{s_1}\|\mathbf{h}_s\|}$ and $\rho_{s,2}(\mathbf{h}_s, a_{s_2}) = e^{-a_{s_2}\|\mathbf{h}_s\|}$, and similarly $\rho_{t,1}$ and $\rho_{t,2}$ are exponential models with parameters a_{t_1} and a_{t_2} , then the Rodriguez and Diggle model is uniformly positive non-separable when $a_{s_1} > a_{s_2}$ and $a_{t_1} > a_{t_2}$ (Fig. 7-a)), whereas they are

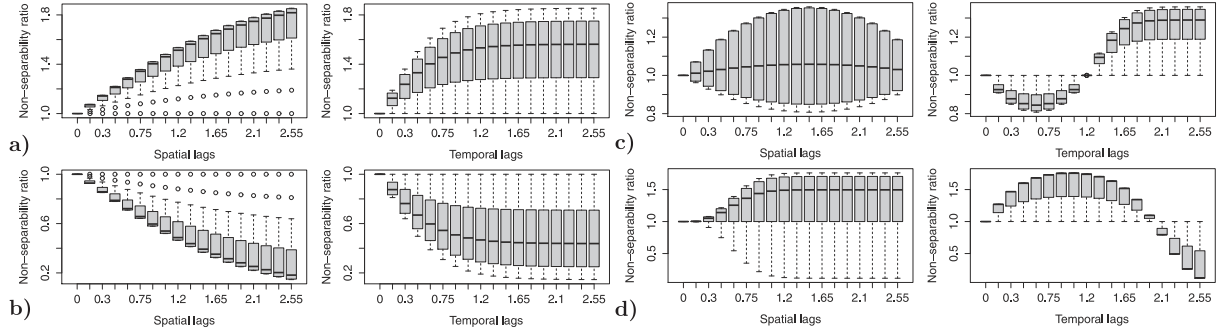


Figure 7: box plots of non-separability ratio, defined for the Rodriguez and Diggle model (10) and classified for spatial lags and temporal lags.

uniformly negative non-separable when $a_{s_1} < a_{s_2}$ and $a_{t_1} > a_{t_2}$ (and viceversa), Fig. 7-b). Otherwise, if $\rho_{s,1}(\mathbf{h}_s, a_{s_1}) = e^{-a_{s_1}\|\mathbf{h}_s\|^2}$ and $\rho_{s,2}(\mathbf{h}_s, a_{s_2}) = e^{-a_{s_2}\|\mathbf{h}_s\|}$, and $\rho_{t,1}$ and $\rho_{t,2}$ are exponential and Gaussian models, respectively, with parameters a_{t_1} and a_{t_2} , then they are non uniformly non-separable (Fig. 7-c),-d)).

3 Case study

In this paper, the new function called `sep_index`, defined for determining non-separability index (1) in the R environment, has been applied in order to analyze the space-time correlation structure of the daily PM_{10} averages, observed at some monitoring stations, from 1998 to 2009.

Structural analysis and non-separability inspection could be performed in the R environment (with packages `spacetime` and `gstat`) through the following steps:

1. create a spatio-temporal object to record the data set;
2. determine sample spatio-temporal variogram surface (Fig. 8-a));
3. inspect marginal variograms, through the new function `vario_marg` (Fig. 9);
4. compute sample spatio-temporal covariance surface (Fig. 8-b));
5. compute the empirical non-separability index, through the new function `sep_index` (Fig. 10).

Note that in step 1 and 2 the required packages of the R environment are `spacetime` and `gstat`, respectively.

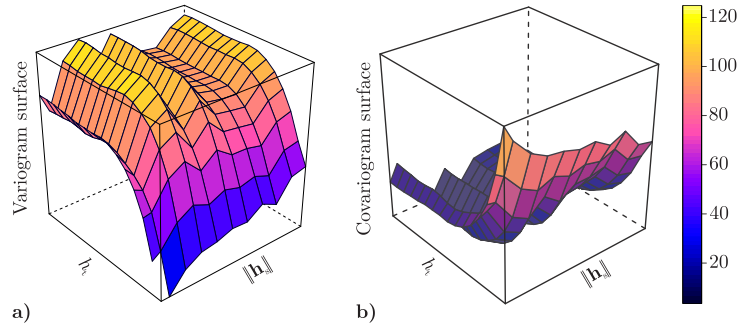


Figure 8: Sample space-time a) variogram surface, b) covariogram surface.

The new function, named `sep_index` requires some parameters to be set: the number

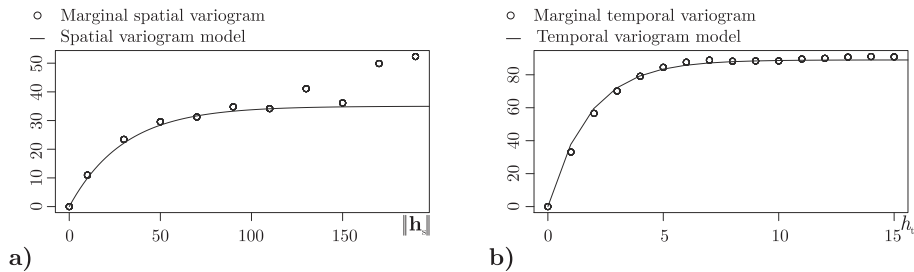


Figure 9: Sample marginal variograms in space (a) and time (b) and their models.

of spatial lags (`ns`), the number of temporal lags (`nt`), the matrix in which the spatio-temporal covariogram values (`cov`) are stored and the value of the global sill (`globalSill`). The function `sep_index` returns the purely spatial and purely temporal sample covariances

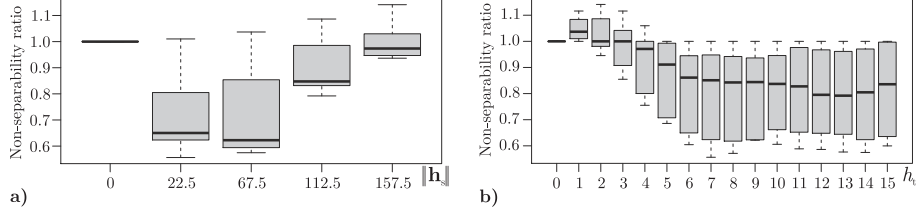


Figure 10: box-plots of sample non-separability ratio, computed for daily averages PM_{10} concentrations and classified for (a) spatial lags and (b) temporal lags.

matrix, empirical non-separability index ratio and box-plots of sample non-separability ratios, classified for spatial lags and temporal lags. For the environmental data set under study, sample non-separability index is less than 1 for almost all \mathbf{h}_s, h_t (Fig. 10). Hence, it should be suitable to select a covariance model characterized by uniformly negative non-separability, such as the product-sum model or its integrated version (9), or the Rodriguez and Diggle (10) for appropriate choices of the coefficients and the marginals. The preferable modeling choice can be reasonably associated with the minimum value of the mean squared error between the empirical and the theoretical non-separability indexes.

4 Conclusions

Non-separable spatio-temporal covariance models are receiving increased interest from the specialized literature, but some theoretical and practical aspects have to be faced. In this paper some well-known non-separable space-time covariance models have been analyzed and classified according to the notion of non-separability. These results can be helpful to select an appropriate covariance models for describing environmental space-time data. As regards computational aspects the new function `sep_index`, implemented in the R environment, has been described; this function allows to compute the empirical non-separability index and draw the box-plots of sample non-separability ratios, classified for spatial lags and temporal lags.

Acknowledgments

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] Bivand, R.S., Gebhardt, A. (2000), Implementing functions for spatial statistical analysis using the R language. *J. Geogr. Syst.* 2(3), 307–311
- [2] Cressie N., Huang H. (1999), *Classes of nonseparable, spatio-temporal stationary covariance functions*, *Journal of the American Statistical Association*, 94, 1330–1340.
- [3] De Cesare L., Myers D., Posa D. (2001), *Estimating and modeling space-time correlation structures*, *Statistics and Probability letters*, 51(1), 9–14.
- [4] De Iaco S. (2010), *Space time correlation analysis: a comparative study*, *Journal of Applied statistics*, 37(6), 1027–1041.
- [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., Posa D. (2002), *Nonseparable space-time covariance models: some parametric families*, *Mathematical Geology*, 34(1), 23–41.
- [8] De Iaco S., Posa D., Myers D. (2013), *Characteristics of some classes of space-time covariance functions*, *Journal of Statistical Planning and Inference*, 143, 2002–2015.
- [9] Fuentes M. (2006), *Testing for separability of spatial-temporal covariance functions*, *Journal of Statistical Planning and Inference*, 136, 447–466.
- [10] Gneiting T. (2002), *Nonseparable, stationary covariance functions for space-time data*, *Journal of the American Statistical Association*, 97, 590–600.
- [11] Guo J.H., Billard L. (1998), *Some inference results for casual autoregressive processes on a plane*, *Journal of Time Series Analysis* 19, 681–691
- [12] Ma C. (2002), *Spatio-temporal covariance functions generated by mixture*, *Mathematical Geology*, 34(8), 965–975.
- [13] Ma C. (2003), *Families of spatio-temporal stationary covariance models*, *Journal of Statistical Planning and Inference*, 116, 489–501.
- [14] Pebesma E. (2012), *Spacetime:spatio-temporal data in R*, *Journal of statistical software*, 51(7), 1–30.
- [15] Rodriguez A., Diggle, P.J. (2010), *A class of convolution-based models for spatio-temporal processes with non-separable covariance structure*, *Scandinavian Journal of statistics*, 37(4), 553–567.
- [16] Scaccia L., Martin, R. J. (2005), *Testing axial symmetry and separability of lattice processes*, *Journal of Statistical Planning and Inference*, 131, 19–39.
- [17] Shitan, M., Brockwell, P. (1995), *An asymptotic test for separability of a spatial autoregressive model*, *Communication in Statistics: Theory and Methods*, 24(8), 2027–2040.