A flexible class of covariance functions for multivariate spatial processes

Rafael S. Erbisti[1] and Thaís C. O. Fonseca[2] and Mariane B. Alves[3]

Abstract

In this work we present full Bayesian inference for a new flexible nonseparable class of cross-covariance functions for multivariate spatially referenced data. A Bayesian test is proposed for separability which is much more interpretable than parameters related to separability. Inference is based on an approximation for the likelihood function making inference scalable for several multivariate components. Spatial models have been increasingly applied in several areas, such as environmental science, climate science and agriculture. These data are usually available in space, time and possibly for several processes. In this context the modeling of dependence is crucial for correct uncertainty quantification and reliable predictions. In particular, for multivariate spatial data we need to specify a valid cross-covariance function, which defines the dependence between the components of a response vector for all locations in the spatial domain. However, cross-covariance functions are not easily specified and the computational burden is a limitation for model complexity. In this work, we propose a nonseparable covariance function that is based on the convex combination of separable covariance functions and on latent dimensions representation of the vector components. The covariance structure proposed is valid and flexible. To treat the computational limitation we approximate the full covariance matrices using a decomposition based on the Kronecker product of two separable matrices of minor dimensions. These approximations have been applied to the likelihood function in order to obtain fast estimation of parameters but we still keep the interpretation and flexibility of the multivariate nonseparable model. The effects of using the approximation are evaluated for simulated datasets in terms of prediction error. The usefulness of our proposal is illustrated with simulated datasets.

Keywords: geostatistics, multivariate spatial models, cross-covariance functions, nonseparable multivariate spatial covariance functions, latent dimensions, Bayesian inference.

---

[1]Department of Statistics, Universidade Federal do Rio de Janeiro

[2]Department of Statistics, Universidade Federal do Rio de Janeiro. (To whom correspondence should be addressed; e-mail: thais@im.ufrj.br)

[3]Department of Statistics, Universidade Federal do Rio de Janeiro

1 Introduction

Realistic modeling of multivariate data observed over space and time is of great interest in several areas of application such as environmental science, climate science and agriculture. Often in geostatistical modeling, the data is considered a partial realization of a random function $Y(\mathbf{s})$, $\mathbf{s} \in D \subseteq \Re^d$. Furthermore, in many applications several quantities are measured for each location $\mathbf{s}$ resulting in random vector $\mathbf{Y}(\mathbf{s})$, $\mathbf{Y}(\mathbf{s}) \in \Re^p$. The main goal of this work is to contribute with realistic modeling of multivariate spatial data while keeping computational feasibility and predictive power. In that context, the proposed covariance model is valid, flexible and inference is scalable to large datasets with several components measured across space.

Although complexity in spatial models is a computational problem some features have to be taken care in the realistic analysis of spatial data. Firstly, the spatial multivariate modeling of data depends on the idea that data which are closer in space is more correlated than data further apart. Also, vector components are usually better predicted considering the component dependence of this vector. These general ideas are directly related to the cross-covariance function of spatial multivariate data, that is, $Cov(Y_j(\mathbf{s}), Y_{j'}(\mathbf{s}'))$, $\mathbf{s}, \mathbf{s}' \in D \subseteq \Re^d$ which models the spatial dependence of $Y_j(.)$ and $Y_{j'}(.)$. The covariance functions considered need to be valid. Thus, construction of new realistic covariance functions usually rely on mathematical simplifications which are not necessarily led by good fitting to data.

An usual simplifying assumption in spatial data modeling is that the cross-covariance functions are separable. Separability assumes that the covariance function for different processes and spatial locations can be computed as the product of a purely spatial and a component covariance functions. This is not a realistic assumption for different processes across space. For two fixed locations $\mathbf{s}$ and $\mathbf{s}'$ the respective component covariance should be proportional. That is, when the spatial location varies, the covariance pattern for different components remains the same. Cressie and Huang (1999) discusses some shortcomings of separable models in the context of spatiotemporal processes and point out that separable

1

models are often chosen for convenience rather than for fitting the data well. Stein (2005) present results about the limited kind of behaviours which these classes represent in practice. A consequence of the separability assumption is that the different p processes will have the same spatial range. This is a very restrictive assumption.

Another restrictive assumption that is a consequence of separability is the symmetry of covariance functions. Separability implies full symmetry, thus a covariance function which is not symmetric is also nonseparable. In applied setting, symmetry is not realistic. For instance, processes which are influenced by air flows might have asymmetric covariance functions.

Several authors have proposed models to relax the separability assumption of cross-covariance functions. The linear model of coregionalization (Wackernagel, 1998) defines the spatial process as a linear combination of independent spatial processes, thus, $Y_j(\mathbf{s}) = \sum_{u=1}^{K} W_u^{(j)}$; $W_u^{(j)}(\mathbf{s}), u = 1, \ldots, K$, are uncorrelated spatial processes. In this approach, if each process $W_u^{(j)}$ has spatial correlation $\rho_u(\mathbf{s})$ then the resulting cross-covariance function is $C_{ij}(\mathbf{h}) = \sum_{u=1}^{K} b_{ij}^u \rho_u(\mathbf{h})$, with $\mathbf{h}$ the spatial separation vector.

A different proposal considers multidimensional scaling ideas (Cox and Cox, 2000). Following this idea, Apanasovich and Genton (2010) proposed a multivariate spatiotemporal model based on latent dimensions which represent distances between components. The authors represent the vector of components as coordinates in a $k-$dimensional space. Any valid covariance function can be used considering the latent component distances and spatial distances to define cross-covariances. Moreover, the authors present results of a simulated study where the model compares favorably to the coregionalization set-up which seems to lack flexibility for some scenarios. The approach of Apanasovich and Genton (2010) depends on the specification of nonseparable covariance functions. In the paper they considered the function proposed in Gneiting (2002). The functions presented in Gneiting (2002) are not interpretable or intuitive. In this paper, we follow the multidimensional scaling approach and consider an interpretable class of nonseparable covariance functions.

In a recent paper, Cressie and Zammit-Mangion (2015) proposed the conditional approach to derive multivariate models. The construction is based on partitioning the vector of spatial processes so that the joint distribution is specified through univariate conditional distributions. This is convenient as the modeler just needs to specify univariate covariance functions and an integrable function of p arguments. Obviously, the results will depend on the chosen conditioning and this is not always an easy modeling decision.

In that context, this work extends the class of nonseparable covariance functions proposed in Fonseca and Steel (2011) to the modeling of component and spatial dependence and considers the multidimensional scaling ideas to define latent distances for components as in Apanasovich and Genton (2010). The general class proposed is able to model different ranges in space and asymmetry of covariances. Furthermore, the proposed class allows for different degrees of smoothness across space for different components of the multivariate random vector. Also, the proposed class has subclasses which can possess a covariance function with the same differentiability properties as the Matérn Class. Similar to the conditional approach of Cressie and Zammit-Mangion (2015), the proposed covariance depends on the definition of univariate covariances and a bivariate joint density function. It is advantageous compared to the conditional approach as it depends on a bivariate density function even if p is large.

The remainder of the paper is organized as follows. Section 2 presents definitions and characteristics about multivariate process modeling. A new class of multivariate spatial covariances is presented in Section 3. In Section 4, a new covariance function is defined. Inference on these models will be conducted from a Bayesian perspective through Markov chain Monte Carlo (MCMC) methods, that will be described in Section 5. To allow for fast estimation of parameters even for a large vector of components, a fast algorithm is proposed to compute the likelihood function to allow for scalable modeling of large spatial data with several components in Section 6. In Section 7, a Bayesian test of separability is presented to measure the level of separability between space and components. Finally, Section 8 presents the conclusions.

2 Multivariate process modeling

In the context of multivariate spatial processes, the main goal is usually to model the dependence among several variables measured across a spatial domain of interest, in order to obtain realistic predictions. Denote $\mathbf{Y}(\mathbf{s})$ the $p-$dimensional vector of variables at location $\mathbf{s} \in D$. Thus, the direct covariance function measures the spatial dependence for each component individually, while the cross-covariance function between two random functions measures the component dependence at the same location and the component dependence within two different locations.

Assuming that $\mathbf{Y}(\mathbf{s})$ is a spatially stationary process, that is

$$E[Y_i(\mathbf{s})] = m_i, \quad Cov[Y_i(\mathbf{s}), Y_j(\mathbf{s}+\mathbf{h})] = C_{ij}(\mathbf{h}), \quad \forall \mathbf{s}, \mathbf{s}+\mathbf{h} \in D; i,j = 1,2,...,p,$$

the cross-covariance function of $\mathbf{Y}(\mathbf{s})$ is defined as

$$E[(Y_i(\mathbf{s}) - m_i)(Y_j(\mathbf{s}+\mathbf{h}) - m_j)] = C_{ij}(\mathbf{h}), \quad \mathbf{s}, \mathbf{s}+\mathbf{h} \in D; i,j = 1,2,...,p. \tag{1}$$

The requirement of positive definiteness of $C_{ij}(\cdot)$ is a limitation in the definition of realistic covariance functions for multivariate spatial processes. As a result, several simplifications are called for in practice such as stationarity and separability. Separability states that

$$C_{ij}(\mathbf{s}, \mathbf{s}') = a_{ij}\rho(\mathbf{s}, \mathbf{s}'), \tag{2}$$

with $\boldsymbol{A} = \{a_{ij}\}$ a positive definite $p \times p$ matrix and $\rho(\cdot, \cdot)$ a valid correlation function. Let $\mathbf{Y}$ be a vectorized version of $Y_{ik} = Y_i(\mathbf{s}_k)$, $k = 1, \cdots, n; i = 1, \cdots, p$. Then the covariance matrix is $\boldsymbol{\Sigma} = \mathbf{R} \otimes \mathbf{A}$, with $R_{kl} = \rho(\mathbf{s}_k, \mathbf{s}_l)$, $k, l = 1, \cdots, n$. The condition of positive definiteness is respected if $\mathbf{R}$ and $\mathbf{A}$ are positive definite. This specification is computationally advantageous as inverses and determinants are obtained from smaller matrices, that is, $\boldsymbol{\Sigma}^{-1} = \mathbf{R}^{-1} \otimes \mathbf{A}^{-1}$ and $|\boldsymbol{\Sigma}| = |\mathbf{R}|^p|\mathbf{A}|^n$.

However, this model has theoretical limitations (Banerjee et al., 2004). Firstly, it is an intrinsic model implying that the correlation between two components $Y_i(\mathbf{s}_k)$ and $Y_j(\mathbf{s}_l)$ is a_{ij},

that is, it does not depend on the locations $\mathbf{s}_k$ and $\mathbf{s}_l$. Secondly, note that as the covariance is defined by one spatial correlation function $\rho(\cdot, \cdot)$, the spatial range will be the same for all components. This last feature can be perceived through the following argument: consider the univariate spatial processes $\{Y(\mathbf{s}) : \mathbf{s} \in D\}$ and $\{X(\mathbf{s}) : \mathbf{s} \in D\}$, $D \subset \Re^2$, therefore $\mathbf{Y} = [Y(\mathbf{s}_1), Y(\mathbf{s}_2), ..., Y(\mathbf{s}_n)]^T$ and $\mathbf{X} = [X(\mathbf{s}_1), X(\mathbf{s}_2), ..., X(\mathbf{s}_n)]^T$.

It is possible to express the following linear relationship for any point in D:

$$E[\mathbf{Y}|\mathbf{X}] = \beta_0 + \beta_1 \mathbf{X}. \tag{3}$$

Consider the stacked $2n \times 1$ vector $(\mathbf{X}, \mathbf{Y})^T$, following a multivariate Normal distribution and a separable covariance structure as in (2), that is,

$$\begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} \sim N_{2n}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad \boldsymbol{\Sigma} = \mathbf{A} \otimes \mathbf{R},$$

implying that $\mathbf{X} \sim N_n(\boldsymbol{\mu}_x, a_{11}\mathbf{R})$ and $\mathbf{Y} \sim N_n(\boldsymbol{\mu}_y, a_{22}\mathbf{R})$. It follows directly that $\mathbf{Y}|\mathbf{X} \sim N_n(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$, with $\boldsymbol{\mu}^* = \boldsymbol{\mu}_y - \frac{a_{12}}{a_{11}}\boldsymbol{\mu}_x + \frac{a_{12}}{a_{11}}\mathbf{X}$ and $\boldsymbol{\Sigma}^* = \left(a_{22} - \frac{a_{12}^2}{a_{11}}\right)\mathbf{R}$, which is equivalent to $\mathbf{Y}|\mathbf{X} \sim N_n(\beta_0 + \beta_1\mathbf{X}, \sigma^2\mathbf{R})$, with $\beta_0 = \boldsymbol{\mu}_y - \frac{a_{12}}{a_{11}}\boldsymbol{\mu}_x$, $\beta_1 = \frac{a_{12}}{a_{11}}$ and $\sigma^2 = a_{22} - \frac{a_{12}^2}{a_{11}}$.

If we assume, reversely, $\mathbf{X} \sim N_n(\boldsymbol{\mu}_x, a_{11}\mathbf{R})$ and $\mathbf{Y}|\mathbf{X} \sim N_n(\beta_0 + \beta_1\mathbf{X}, \sigma^2\mathbf{S})$, with $\mathbf{S}$ any spatial correlation matrix, the covariance structure for $\mathbf{Y}$ is

$$\begin{aligned} Cov[Y_i, Y_j] &= \sigma^2 \mathbf{S}_{ij} + \beta_1^2 a_{11} \mathbf{R}_{ij} \\ &= a_{22}\mathbf{S}_{ij} - \frac{a_{12}^2}{a_{11}}\mathbf{S}_{ij} + \frac{a_{12}^2}{a_{11}}\mathbf{R}_{ij}. \end{aligned} \tag{4}$$

Then (4) equals $a_{22}\mathbf{R}$, reducing to the separable specification if and only if $\mathbf{S} = \mathbf{R}$, that is $\mathbf{Y}|\mathbf{X}$ has the same spatial structure as $\mathbf{X}$.

More flexible structures are obtained via the coregionalization approach, which in its simplest form is $\mathbf{Y}(\mathbf{s}) = \mathbf{A}\mathbf{w}(\mathbf{s})$, with $\mathbf{A}$ a $p \times p$ matrix and the components of $\mathbf{w}(\mathbf{s})$, $w_j(\mathbf{s})$, $j = 1, 2, ..., p$, independent and identically distributed spatial processes. If the processes $w_j(\mathbf{s})$ are stationary with zero mean and unit variances and $Cov(w_j(\mathbf{s}), w_j(\mathbf{s}')) = \rho(\mathbf{s} - \mathbf{s}')$, then $E(\mathbf{Y}(\mathbf{s})) = 0$ and the cross-covariance function of $\mathbf{Y}(\mathbf{s})$ is $\boldsymbol{\Sigma}_{\mathbf{Y}(\mathbf{s}), \mathbf{Y}(\mathbf{s}')} \equiv C(\mathbf{s} - \mathbf{s}') =$

$\rho(\mathbf{s} - \mathbf{s}')\mathbf{A}\mathbf{A}^T$ which is separable. A more general form for the coregionalization model considers independent processes $w_j(\mathbf{s})$ however they are not identically distributed. The covariance matrix is given by

$$\boldsymbol{\Sigma}_{\mathbf{Y}(\mathbf{s}),\mathbf{Y}(\mathbf{s}')} \equiv C(\mathbf{s} - \mathbf{s}') = \sum_{j=1}^{p} \rho_j(\mathbf{s} - \mathbf{s}')\mathbf{T}_j$$

with $\mathbf{T}_j = \mathbf{a}_j\mathbf{a}_j^T$, $\mathbf{a}_j$ the $j-th$ column of $\mathbf{A}$. The resulting covariance is nonseparable but is stationary.

In this work we follow the multidimensional scaling framework and the latent dimensions proposed in Apanasovich and Genton (2010). The vector of components are represented as coordinates in a $k-$dimensional space, for an integer $1 \leq k \leq p$, that is, the $i-th$ component is represented as $\xi_i = \{\xi_{i1}, ..., \xi_{ik}\}^T$.

This approach can be used for any valid covariance function $\Sigma_{ij} = C\{(\mathbf{s}, \xi_i), (\mathbf{s}', \xi_j)\}$. For any $\mathbf{s}$, $\mathbf{s}'$ there is $C_{\mathbf{s},\mathbf{s}'}(.)$ such that $C_{ij}(\mathbf{s}, \mathbf{s}') = C_{\mathbf{s},\mathbf{s}'}(\xi_i, \xi_j)$ for some $\xi_i, \xi_j \in \Re^k$.

The latent coordinates may be treated as parameters and estimated from data. Moreover, it is possible to consider the simplification $\delta_{ij} = \|\xi_i - \xi_j\|$. This approach is similar to the multidimensional scaling (Cox and Cox, 2000) with latent distances δ_{ij}'s, where for fixed locations $\mathbf{s}$ and $\mathbf{s}'$, small δ_{ij}'s are converted into strong positive cross-correlation. Notice that large values of δ_{ij}'s might mean small correlation or negative correlation. A further discussion about this issue is presented in the conclusions. A review of the main approaches to building a valid multivariate cross-covariance function is presented in Genton and Kleiber (2015).

In this work we consider an intuitive proposal which is based on mixing separable functions as in Fonseca and Steel (2011). Furthermore we consider an approximation of the likelihood which allows estimation for large datasets. In Section 6 we present a sensitivity study to investigate the errors obtained with the likelihood approximation.

3 Multivariate spatial modeling based on mixtures

In this section we present a new class of multivariate spatial covariances which are flexible and intuitive depending only on the specification of univariate functions in space. We consider the latent dimension approach of Apanasovich and Genton (2010) to model cross-dependencies between components of a spatial vector. Furthermore we define the nonseparable function based on the spatiotemporal mixture approach of Fonseca and Steel (2011). Thus, only univariate valid spatial functions need to be specified.

Fonseca and Steel (2011) consider $(\mathbf{s}, t) \in D \times T$, $D \subseteq \Re^d$, $T \subseteq \Re$, as space-time coordinates varying continuously on $D \times T$ and $Z_1(\mathbf{s})$, $Z_2(t)$ uncorrelated processes, $\{Z_1(\mathbf{s}) : \mathbf{s} \in D\}$ denoting a purely spatial process with covariance $C_1(\mathbf{s})$ and $\{Z_2(t) : t \in T\}$ a purely temporal process with covariance $C_2(t)$. The mixture representation of the covariance structure of $Z(\mathbf{s}, t)$ is defined as follows: assume that (U, V) is a nonnegative bivariate random vector following a joint distribution $G(u, v)$, independent of $\{Z_1(\mathbf{s}) : \mathbf{s} \in D\}$ and $\{Z_2(t) : t \in T\}$. Define the process $Z(\mathbf{s}, t) = Z_1(\mathbf{s}, U)Z_2(t, V)$, where $\{Z_1(\mathbf{s}, u)\}$ remains a purely spatial process for every $u \in \Re_+$ with a stationary covariance function $C_1(\mathbf{s}, u)$ for $\mathbf{s} \in D$ and every $u \in \Re_+$, which is a measurable function of $u \in \Re_+$ for every $\mathbf{s} \in D$. Analogously, let $\{Z_2(t, v)\}$ be a purely temporal process with covariance $C_2(t, v)$, which is a stationary covariance function for $t \in T$ and every $v \in \Re_+$ and a measurable function of $v \in \Re_+$ for every $t \in T$. Thus the corresponding covariance of $Z(\mathbf{s}, t)$ is a convex combination of separable covariance functions. This is a valid and generally nonseparable function

$$C(\mathbf{s}, t) = \int \int C(\mathbf{s}; u)C(t; v)g(u, v)dudv. \tag{5}$$

The proposed idea in the present work is to modify (5) to deal with the spatial multivariate specification. Thus, consider (U, V) independent of the process $\mathbf{Y}(\mathbf{s})$. Similar to Fonseca and Steel (2011), the covariance of $\mathbf{Y}(\mathbf{s})$ is a convex combination of separable covariance functions, given by

$$C_{ij}(\mathbf{s}, \xi) = \int \int C(\mathbf{s}; u)C_{ij}(\xi; v)g(u, v)dudv \tag{6}$$

7

with ξ representing a latent dimension as in Apanasovich and Genton (2010) and $\mathbf{s}$ an arbitrary spatial location.

According to Fonseca and Steel (2011), the fundamental step in the definition of this class of functions lies on the representation of the dependence between U and V. Define variograms $\gamma_1(\mathbf{s}) \equiv \gamma_1$ and $\gamma_2(\xi) \equiv \gamma_2$ as continuous functions on $\mathbf{s} \in \Re^d$ and $\xi \in \Re^k$, respectively. Then, it is possible to analytically solve (6), still assuring that the generated covariance is positive definite, defining $C(\mathbf{s}; u) = exp\{-\gamma_1 u\}$ and $C(\xi; v) = exp\{-\gamma_2 v\}$.

Proposition 3.1 *Consider a bivariate nonnegative vector (U, V) with joint moment generator function $M(.,.)$. If the variograms $\gamma_1(\boldsymbol{s}) \equiv \gamma_1$ and $\gamma_2(\xi) \equiv \gamma_2$ are continuous functions of $\boldsymbol{s} \in \Re^d$ and $\xi \in \Re^k$, respectively, and $C(\boldsymbol{s}; u) = exp\{-\gamma_1 u\}$, $C(\xi; v) = exp\{-\gamma_2 v\}$, then (6) implies that*

$$C_{ij}(\boldsymbol{s}, \xi) = M(-\gamma_1, -\gamma_2), \tag{7}$$

which is a valid covariance function.

Majumdar and Gelfand (2007) use Monte Carlo integration to solve an integral similar to (6). Apanasovich et al. (2012) consider a multivariate version of Matérn, presenting a flexible model, allowing for different behaviour for each component. The proposed approach (6) also presents that flexibility.

Following Proposition 3.1, it is possible to build nonseparable structures, based only on the joint distribution of (U, V). Thus, consider the following proposition.

Proposition 3.2 *Consider the independent nonnegative random variables X_0, X_1 and X_2, with moment generator functions M_0, M_1 and M_2. Define U and V as: $U = X_0 + X_1$ and $V = X_0 + X_2$. If $C(\boldsymbol{s}; u) = exp\{-\gamma_1 u\}$ and $C(\xi; v) = exp\{-\gamma_2 v\}$, as in proposition 3.1, then the covariance function resulting from (6) is*

$$C_{ij}(\boldsymbol{s}, \xi) = M_0(-\gamma_1 - \gamma_2)M_1(-\gamma_1)M_2(-\gamma_2). \tag{8}$$

Observe that if U and V are uncorrelated, that is, $U = X_1$ and $V = X_2$, the separable specification is obtained, since $C_{ij}(\mathbf{s}, \xi) = M_1(-\gamma_1)M_2(-\gamma_2)$.

The class generated by proposition 3.2 allows for different parametric representations, as we vary the specifications for X_0, X_1 and X_2. By construction, any non-null correlation between U and V will be positive.

Proposition 3.2 indeed generates a valid correlation structure, since $C_{ij}(\mathbf{0}) = 1$, thus, following Majumdar and Gelfand (2007), in order to obtain a valid covariance structure, define

$$\rho_{ij}(\mathbf{s}, \xi) = \frac{C_{ij}(\mathbf{s}, \xi)}{[C_{ii}(\mathbf{0})C_{jj}(\mathbf{0})]^{1/2}}. \tag{9}$$

Note that $\rho_{ii}(\mathbf{0}) = 1$. Consider a diagonal matrix D_{cov} with elements $[D_{cov}]_{ii} = C_{ii}(\mathbf{0})$. If $R_{ij}(\mathbf{s}, \xi) = D_{cov}^{-1/2}C_{ij}(\mathbf{s}, \xi)D_{cov}^{-1/2}$, then $R_{ij}(\mathbf{s}, \xi)$ is a valid cross-correlation. If we define $D_\sigma^{1/2} = diag(\sigma_1, ..., \sigma_p)$, $\sigma_i > 0$, a valid covariance structure is obtained, given by the matrix $C_\sigma = D_\sigma^{1/2}R_{ij}(\mathbf{s}, \xi)D_\sigma^{1/2}$.

Proposition 3.3 *Consider the nonnegative independent variables X_0, X_1 and X_2, with moment generator functions M_0, M_1 and M_2. Define U and V as: $U = X_0 + X_1$, $V = X_0 + X_2$. If $C(\boldsymbol{s}; u) = \sigma_i exp\{-\gamma_1 u\}$ and $C(\xi; v) = \sigma_j exp\{-\gamma_2 v\}$, then (6) implies that*

$$C_{ij}(\boldsymbol{s}, \xi) = \sigma_i\sigma_j M_0(-\gamma_1 - \gamma_2)M_1(-\gamma_1)M_2(-\gamma_2), \tag{10}$$

which is a valid covariance function.

4   Flexible classes

In this section we present a new covariance function from proposition 3.3. Consider X_0, X_1 and X_2 following gamma distributions as Fonseca and Steel (2011).

Theorem 4.1 *Consider $X_l \sim Gamma(\alpha_l, \lambda_l)$, $l = 0, 1$ and 2, then from proposition 3.3, the cross-covariance function is*

$$C_{ij}(\boldsymbol{s}, \xi) = \sigma_i\sigma_j \left(1 + \frac{\gamma_1 + \gamma_2}{\lambda_0}\right)^{-\alpha_0} \left(1 + \frac{\gamma_1}{\lambda_1}\right)^{-\alpha_1} \left(1 + \frac{\gamma_2}{\lambda_2}\right)^{-\alpha_2} \tag{11}$$

with $\sigma_j > 0$, $j = 1, ..., p$, $\alpha_l > 0$ and $\lambda_l > 0$, $l = 0, 1, 2$.

It is difficult to interpret some parameters in the proposed function (11). We expect to work a function that allows different spatial ranges for each component. The dependence of U and V is governed by the variable X_0, it is important to define a parameter responsible for the behaviour of the correlation between these variables. Remember that if U and V are uncorrelated then the separable case is obtained.

For this, the initial idea was to determine $\lambda_i = 1$, for $i = 0, 1$ e 2, and to work with component variogram $\gamma_2 = \|\xi_i - \xi_j\| = \delta_{ij}$. Furthermore, we introduced an extra parameter in spatial variogram. This parameter varies with the components i and j, i.e, $\gamma_1 = \frac{\|\mathbf{s} - \mathbf{s}'\|}{b_{ij}} = \frac{h}{b_{ij}}$. Therefore, the general model is given by

$$C_{ij}(\mathbf{s}, \xi) = \sigma_i \sigma_j \left(1 + \delta_{ij} + \frac{h}{b_{ij}} \right)^{-\alpha_0} \left(1 + \frac{h}{b_{ij}} \right)^{-\alpha_1} (1 + \delta_{ij})^{-\alpha_2} \tag{12}$$

where δ_{ij} is the latent distance between the components i and j, σ_i is the standard deviation of component i, b_{ij}'s are spatial range parameters, α_l are smoothness parameters, for $l = 1$ and 2, and α_0 is a separability parameter.

Notice that if we work with the same spatial range parameters for all components, that it, $b_{ij} = \phi$, $\forall i, j = 1, 2, ..., p$, we provided a particular case of the general function. Furthermore, if $\alpha_0 = 0$, the separable model is obtained and the resulting covariance function is in the Cauchy Class.

5 Inference

Let $(\mathbf{y}(\mathbf{s}_1), \ldots, \mathbf{y}(\mathbf{s}_n))$ be a matrix of multivariate data observed at spatial locations $\mathbf{s}_1, \ldots, \mathbf{s}_n \in D$, where $\mathbf{y}(\mathbf{s}_i) = (y_1(\mathbf{s}_i), \ldots, y_p(\mathbf{s}_i))'$ is a $p-$dimensional vector. If the Gaussian assumption is made, the likelihood function for the unknown parameters based on n spatial locations is given by

$$l(\mathbf{y}; \boldsymbol{\theta}) = (2\pi)^{\frac{-np}{2}} |\boldsymbol{\Sigma}|^{-1/2} exp \left\{ \frac{-1}{2} (\mathbf{y} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu}) \right\} \tag{13}$$

with $\mathbf{y}$ the vectorized version of $(\mathbf{y}(\mathbf{s}_1), \ldots, \mathbf{y}(\mathbf{s}_n))$ with np observations, $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ the mean vector, $\boldsymbol{\Sigma}$ the covariance matrix with dimension $np \times np$, and $\boldsymbol{\theta}$ the parameter vector. The covariance matrix has components defined by equation (12). In particular for our model specification $\boldsymbol{\theta} = (\boldsymbol{\sigma}, \boldsymbol{\delta}, \boldsymbol{\alpha}, \mathbf{b}, \boldsymbol{\beta})$, with $\boldsymbol{\sigma} = (\sigma_1, ..., \sigma_p)$, $\boldsymbol{\delta}$ the vector of latent variables δ_{ij}, $i \neq j$, $i, j = 1, ..., p$, $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \alpha_2)$, $\mathbf{b}$ the range parameter vector b_{ij}, $i, j = 1, ..., p$, e $\boldsymbol{\beta} = (\beta_{10}, ..., \beta_{p0}, \beta_{11}, ..., \beta_{p1}, ..., \beta_{1q}, ..., \beta_{pq})$, with q the number of covariates.

To complete the Bayesian model specification, the prior distributions must be defined for all parameters in the proposed model (12). Prior independence is assumed for the parameters in the model such that $\sigma_i \sim Ga(c_i, d_i)$, $i = 1, ..., p$, $\delta_{ij} \sim Ga(f_{ij}, g_{ij})$, $i \neq j$, $i, j = 1, \ldots, p$, $\alpha_k \sim Ga(r_k, s_k)$, $k = 0, 1, 2$, $b_{ij} \sim Ga(u_{ij} \times med(d_s), u_{ij})$, $i, j = 1, ..., p$, $med(d_s)$ is the median of the spatial distances, $\boldsymbol{\beta} \sim N_{pq}(\boldsymbol{\lambda}, \boldsymbol{\Lambda})$.

Inference is based on simulations from the complete conditional distributions for sets of parameters. For $\boldsymbol{\beta}$ the complete conditional distribution is Gaussian. For the other parameters in the covariance function the distributions have no closed form and Metropolis-Hastings steps are considered in the Gibbs sampler algorithm. Details on such algorithms are presented in Gamerman and Lopes (2006).

5.1 Prediction

One of the main goals in spatial data analysis is to obtain prediction in new locations or for missing data within the observed data. Let $\mathbf{y}_u$ be the observation vector at unmeasured locations $\mathbf{s}_u \in D$. The prediction of $\mathbf{y}_u$ is based on the predictive distribution $p(\mathbf{y}_u|\mathbf{y}_o)$, with $\mathbf{y}_o$ denoting the vector of observed data. Thus,

$$p(\mathbf{y}_u|\mathbf{y}_o) = \int p(\mathbf{y}_u|\mathbf{y}_o, \boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathbf{y}_o) d\boldsymbol{\theta}. \tag{14}$$

From the Gaussian assumption, the distribution $p(\mathbf{y}_u|\mathbf{y}_o, \boldsymbol{\theta})$ is also Gaussian with parameters $\boldsymbol{\mu}^* = \boldsymbol{\mu}_u + \boldsymbol{\Sigma}_{uo}\boldsymbol{\Sigma}_{oo}^{-1}(\mathbf{y}_o - \boldsymbol{\mu}_o)$ and $\boldsymbol{\Sigma}^* = \boldsymbol{\Sigma}_{uu} - \boldsymbol{\Sigma}_{uo}\boldsymbol{\Sigma}_{oo}^{-1}\boldsymbol{\Sigma}_{ou}$. Assume that $\boldsymbol{\theta}^{(1)}, ..., \boldsymbol{\theta}^{(M)}$ are a sample from the posterior distribution $(\boldsymbol{\theta}|\mathbf{y}_o)$ obtained by MCMC sampling. Thus, the

predictive distribution in (14) may be obtained by the approximation:

$$\widehat{p}(\mathbf{y}_u|\mathbf{y}_o) = \frac{1}{M} \sum_{i=1}^{M} p(\mathbf{y}_u|\mathbf{y}_o, \boldsymbol{\theta}^{(i)}). \tag{15}$$

6   Likelihood computation for large data

With the increase of high-resolution geocoded data the big problem became crucial in the spatial and spatiotemporal setup. For instance, if Gaussianity is assumed, large covariance matrices need to be inverted in the inference procedure and computational effort is of cubic order on the number of locations. This limitation becomes even more important in the case of spatiotemporal or multivariate data. Even low dimensional vectors observed over space may lead to huge covariance matrices, making the inference for unknown parameters not feasible. Thus, a compromise between complexity and parsimony is called for in this context.

We have presented a nonseparable covariance model which results in a full matrix $\boldsymbol{\Sigma}$ which might have high dimension and the computation of likelihoods require the inversion of this matrix. We investigate the use of separable approximations for the matrix $\boldsymbol{\Sigma}$ which will lead to a fast computation of likelihood functions. The approximation will be based on the proposal in Genton (2007).

Genton (2007) investigates the use of singular decompositions of a full matrix in the context of nonseparable spatiotemporal covariance matrices. The work considers a decomposition based on separable matrices which allow for fast inversions and determinant computations. Thus, instead of $np \times np$ matrices, the approximation uses only $n \times n$ and $p \times p$ matrices. We consider the same separable approximation in order to compute likelihoods for the nonseparable multivariate spatial models presented in Section 4. The aim is to obtain matrices $\mathbf{R} \in \Re^{n\times n}$ e $\mathbf{A} \in \Re^{p\times p}$ such that the Frobenius norm[4] of $\|\boldsymbol{\Sigma}-\mathbf{R}\otimes\mathbf{A}\|_F$ is minimized, for a given full covariance matrix $\boldsymbol{\Sigma}$. The author shows that the solution to this problem is given by the singular value decomposition of a permuted version of $\boldsymbol{\Sigma} \in \Re^{np\times np}$.

---

[4]The Frobenius norm of a matrix $\mathbf{B}$ $(\|\mathbf{B}\|_F)$ is given by $\|\mathbf{B}\|_F = \left(\sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij}^2\right)^{1/2}$

The idea is to rearrange $\boldsymbol{\Sigma}$ obtaining another matrix $\Im(\boldsymbol{\Sigma}) \in \Re^{n^2 \times p^2}$, such that the squared sum of $\|\boldsymbol{\Sigma} - \mathbf{R} \otimes \mathbf{A}\|_F$ equals $\|\Im(\boldsymbol{\Sigma}) - vec(\mathbf{R}) \otimes vec(\mathbf{A})^T\|_F$. It is showed that $\|\boldsymbol{\Sigma} - \mathbf{R} \otimes \mathbf{A}\|_F = \|\Im(\boldsymbol{\Sigma}) - vec(\mathbf{R}) \otimes vec(\mathbf{A})^T\|_F$ and $\|\boldsymbol{\Sigma}\|_F = \|\Im(\boldsymbol{\Sigma})\|_F$.

The problem then reduces to finding the rank of the rectangular matrix $\Im(\boldsymbol{\Sigma}) \in \Re^{n^2 \times p^2}$, which solution can also be found in Golub and Van Loan (1996) and is given by:

$$vec(\mathbf{R}) = \sqrt{w_1}\boldsymbol{u}_1 \qquad vec(\mathbf{A}) = \sqrt{w_1}\boldsymbol{v}_1 \qquad (16)$$

with $\boldsymbol{u}_1$ denoting the first column of the matrix $\mathbf{U} \in \Re^{n^2 \times n^2}$ and $\boldsymbol{v}_1$, the first column of $\mathbf{V} \in \Re^{p^2 \times p^2}$.

In order to measure the quality of the approximation, Genton defines an approximation error, denoted by $\kappa_{\boldsymbol{\Sigma}}(\mathbf{R}, \mathbf{A})$, as follows:

$$\kappa_{\boldsymbol{\Sigma}}(\mathbf{R}, \mathbf{A}) = \frac{\|\boldsymbol{\Sigma} - \mathbf{R} \otimes \mathbf{A}\|_F}{\|\boldsymbol{\Sigma}\|_F}. \qquad (17)$$

$\kappa_{\boldsymbol{\Sigma}}(\mathbf{R}, \mathbf{A})$ varies between zero (if $\boldsymbol{\Sigma}$ is separable) and $\sqrt{1 - \frac{1}{r}}$, and is minimized by $\mathbf{R}$ and $\mathbf{A}$ given above. A standardized error index, varying between zero and one is given by:

$$\kappa_{\boldsymbol{\Sigma}}^*(\mathbf{R}, \mathbf{A}) = \frac{\kappa_{\boldsymbol{\Sigma}}(\mathbf{R}, \mathbf{A})}{\sqrt{1 - \frac{1}{r}}}. \qquad (18)$$

From the covariance structure proposed in equation (12) with $\alpha_1 = \alpha_2 = 1$, we investigate the sensitivity of the separability approximation error index as a function of α_0. Note that we use the idea previously applied to the context of nonseparable spatiotemporal covariance matrices in the context of nonseparable multivariate spatial covariance matrices. In Figure 1 we can see that the separability approximation error index is not larger than 5% for equal spatial ranges. From Figure 1(a) note that there is no error when α_0 is zero, which reduces to the separable case. If different spatial ranges are considered, it is possible to see in Figure 1(b) that the error index does not start at zero because if $\alpha_0 = 0$ the separable case is not obtained.
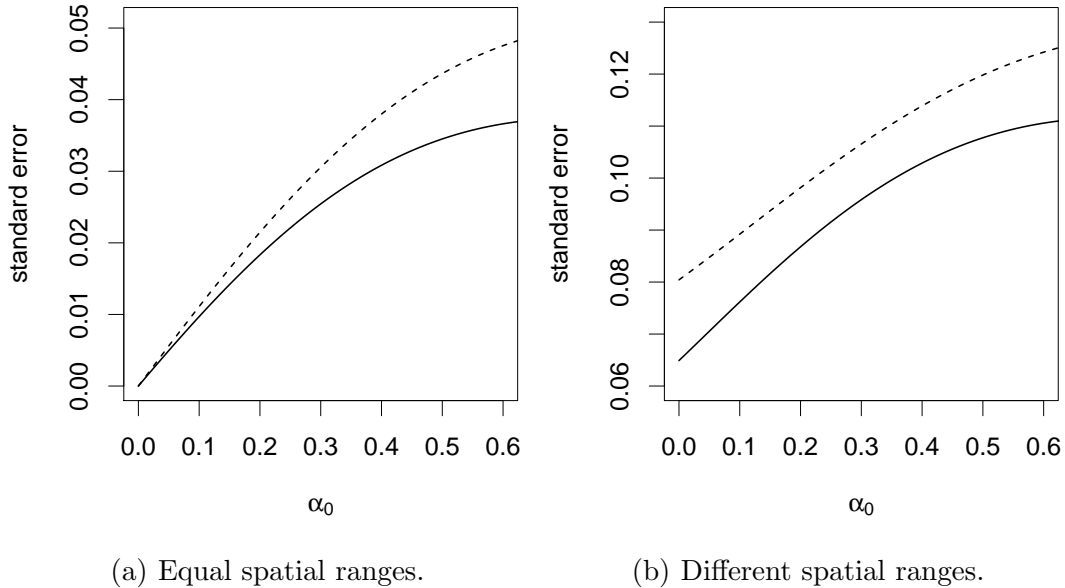
(a) Equal spatial ranges.　　　　(b) Different spatial ranges.

Figure 1: Separability approximation error index as a function of α_0. Full line: $p = 2$; dashed line: $p = 3$.

6.1　Sensitivity study

We present a sensitivity study of the approximation structure investigated by Genton (2007), however, used for the multivariate spatial case. We consider different scenarios and measure the errors obtained in the likelihood approximation. Moreover, we compare the inferential and predictive results obtained as we apply the full proposed nonseparable model with and without separable approximation for the covariance matrix in the likelihood computation, as well as considering a separable model.

Consider a bivariate dataset of 200 spatial locations in the $[0, 1] \times [0, 1]$ square and the following parameter specification $\boldsymbol{\Theta} = (\boldsymbol{\mu}, \delta_{12}, \phi, \alpha_0, \sigma_1, \sigma_2)$ with $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = 0$, $\delta_{12} = 2$, $b_{ij} = b_{ji} = \phi = 0.2$, for all $i, j = 1, ..., p$, $\alpha_0 = 1$, $\sigma_1 = 1$ and $\sigma_2 = 0.75$. In this example, we generate only one dataset in the region of interest. We plot the likelihood contour with both structures, using the separable approximation for the covariance matrix and its full original structure. From Figure 2 it can be seen that the approximate structure is very similar to the full structure. Note that in some cases the approximate likelihood and the exact one are

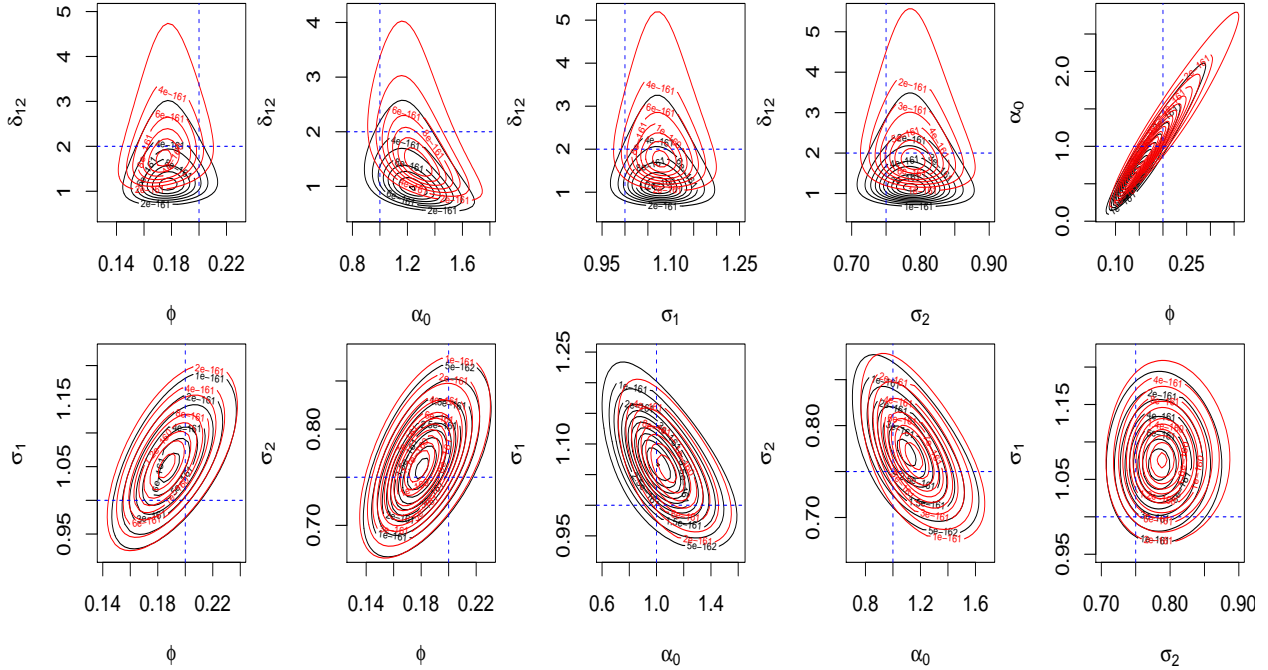almost coincident. It seems that the approximations are satisfactory.

Figure 2: Likelihood contour plots. Black line: full structure. Red line: approximate structure. Dashed blue line: true value of parameters.

We analyzed the necessary time to calculate the likelihood function based on a full covariance matrix and a covariance matrix with approximate structure. We generated $p = 2, 3, 5$ and 8 variables in a dataset with $n = 100, 200, 500, 700$ and 1000 spatial locations in the $[0, 1] \times [0, 1]$ square. In this example, 200 replicates were generated in the region of interest.

Table 1 shows that the separable approach provides important gains in computational efficiency. Note that the time to calculate the likelihood function is substantially lower when we use the approximate structure.

We also analyzed the time reduction using the separable approximations. Figure 3 shows that the time to calculate the likelihood function decreases as the size of the covariance matrix increases. Indeed, if we increase the variable numbers or spatial locations or both, the greater will be the computational gain with the approximate structure.

Finally, we compare the predictive results obtained by the separable model, the separable

n	$p = 2$		$p = 3$		$p = 5$		$p = 8$	
	full	approx.	full	approx.	full	approx.	full	approx.
100	2.3	0.8	3.6	0.6	8.9	0.9	28.0	2.1
200	12.3	3.1	13.9	1.6	44.8	4.2	187.1	11.1
500	74.0	13.4	143.1	12.1	618.6	30.5	2409.5	86.9
700	148.2	20.9	388.4	29.4	1649.7	65.8	6520.7	182.2
1000	374.6	52.6	1020.7	66.2	4673.9	133.5	19180.3	446.2

Table 1: Necessary time (in seconds) to calculate the likelihood function based on a full covariance matrix and an approximate structure. (Intel(R) Core(TM) i7-3630QM, 2.40GHz, 6GB RAM)

approximation for the covariance matrix in the likelihood of the nonseparable model and the results obtained with the nonseparable original covariance structure, without any approximations for the likelihood. For that purpose, we generated five datasets from the nonseparable structure proposed in Section 4. We use a less general function than proposed in equation (12). For each dataset, we generated $p = 2$ variables in $n = 110$ spatial locations in the $[0,1] \times [0,1]$ square considering the following parameter specification $\boldsymbol{\Theta} = (\boldsymbol{\mu}, \delta_{12}, \phi, \alpha_0, \sigma_1, \sigma_2)$ with $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = 0$, $\delta_{12} = 0.8$, $\phi = 0.1$, $\alpha_0 = 1$, $\sigma_1 = 1.5$ and $\sigma_2 = 0.9$. The covariance functions used for the separable and nonseparable models are respectively shown in equations (19) and (20):

$$C_{ij}(\mathbf{s}) = \begin{cases} a_{11} \left(1 + \left(\frac{h}{\phi} \right)^2 \right)^{-1} & (i = j = 1) \\ a_{22} \left(1 + \left(\frac{h}{\phi} \right)^2 \right)^{-1} & (i = j = 2) \\ a_{12} \left(1 + \left(\frac{h}{\phi} \right)^2 \right)^{-1} & (i \neq j), \end{cases} \tag{19}$$
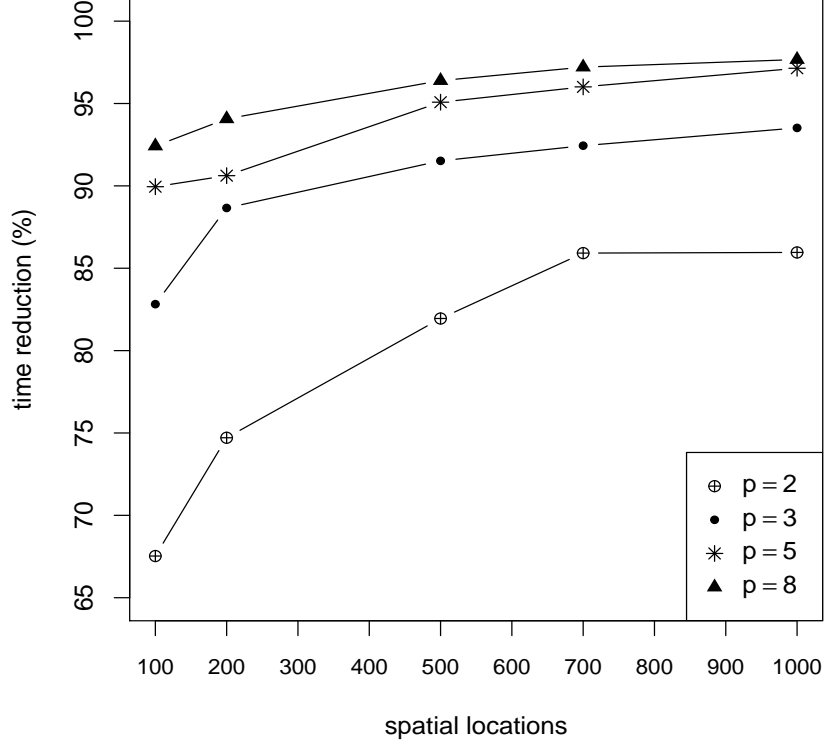
Figure 3: Computational time reduction (in percent) in calculation likelihood function using approximate structure.

$$
C_{ij}(\mathbf{s}, \xi) = \begin{cases}
\sigma_1^2 \left(1 + \frac{h}{\phi}\right)^{-(\alpha_0 + 1)} & (i = j = 1) \\
\sigma_2^2 \left(1 + \frac{h}{\phi}\right)^{-(\alpha_0 + 1)} & (i = j = 2) \\
\sigma_1 \sigma_2 \left(1 + \delta_{12} + \frac{h}{\phi}\right)^{-\alpha_0} \left(1 + \frac{h}{\phi}\right)^{-1} (1 + \delta_{12})^{-1} & (i \neq j).
\end{cases}
\tag{20}
$$

We adopted $T = 30$ independent replicates. The observations were generated from the model $\mathbf{Y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}$. We consider a Gaussian process, so $\mathbf{Y}_t \sim N_{np}(\boldsymbol{X\beta}, \boldsymbol{\Sigma})$, $t = 1, ..., T$, where $\boldsymbol{\Sigma}$ is $np \times np$ covariance matrix and $\boldsymbol{X}$ are independent variables (latitude, longitude and altitude). Furthermore, the information about five spatial locations were removed for prediction. Therefore, we estimate the model using information about $n = 105$ spatial locations. After the estimation of the models for each dataset, we were able to calculate predictive errors. Table 2 shows the results. We can see that in predictive terms the approximation

17

leads to very similar results to the full case. Note that the separable model presents the worst results, it is not able to accommodate the nonseparable structure.

Data	$\|\boldsymbol{y}_{true} - \hat{\boldsymbol{y}}\|$			Average interval score		
	SEP	NSEP APP.	NSEP	SEP	NSEP APP.	NSEP
1	9.557	8.613	8.709	148.334	133.756	133.687
2	11.939	11.580	11.694	145.473	131.614	130.796
3	12.091	10.741	11.091	142.832	131.218	131.098
4	11.602	10.761	10.652	146.319	133.528	132.883
5	11.076	9.830	9.693	142.703	130.902	130.290
Mean	11.253	10.305	10.368	145.132	132.204	131.751

Table 2: Predictive errors. SEP: separable model. NSEP APP: nonseparable approximate model. NSEP: nonseparable model.

7 Bayesian Hypotheses testing for separability

Following Fonseca and Steel (2011), we choose the correlation between U and V as a measure of separability. Indeed, if U and V are uncorrelated, the resulting model is separable, so

$$\rho = \rho(U, V) = \frac{Cov(U, V)}{\sqrt{Var(U)Var(V)}}$$
$$= \frac{\alpha_0}{\sqrt{(\alpha_0 + \alpha_1)(\alpha_0 + \alpha_2)}}.$$

It is easy to see that if $\alpha_0 = 0$ implies $\rho = 0$. Note that $0 \leq \rho \leq 1$, where 0 indicates separability and 1 indicates strong nonseparability. From a frequentist point of view, many authors present a formal method to test separability in the spatiotemporal models (Mitchell et al., 2005, 2006; Fuentes, 2006). The test proposed in this work aims to measure the degree of separability between space and components and we follow the Bayesian paradigm for hypothesis testing.

7.1 Bayesian model choice

The usual continuous prior for positive parameters, as the one considered for α_0 in Section 5, assigns zero probability for the null hypothesis $\alpha_0 = 0$. As an alternative consider the following mixture representation

$$\pi(\alpha_0) = p_0 \mathcal{D}_0 + (1 - p_0)p(\alpha_0),$$

with $\mathcal{D}_0$ the dirac function at $\alpha_0 = 0$ and $p(\alpha_0)$ a continuous distribution for $\alpha_0 > 0$. Thus, p_0 is the prior probability of a separable covariance function. The resulting posterior distribution in this specification is also a mixture

$$\pi(\alpha_0 \mid \mathbf{y}) = \tilde{p}_0 \mathcal{D}_0 + (1 - \tilde{p}_0)p(\alpha_0 \mid \mathbf{y}),$$

with $\tilde{p}_0$ being the posterior probability of separable covariance functions given the data.

The posterior probabilities $\tilde{p}_0$ might be used to select a model (Bayesian model choice) or to predict new observations based on model averaging across both models.

We simulate four different scenarios from separable to very nonseparable structure. In this context, we use the model in equation (20) and generate datasets with $p = 2$ components, $n = 105$ spatial locations and $T = 10$ independent replicates in time. We consider a different degree of separability ρ for each dataset and the same parameter specification $\boldsymbol{\Theta} = (\boldsymbol{\mu}, \delta_{12}, \phi, \sigma_1, \sigma_2)$ with $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = 0$, $\delta_{12} = 1.5$, $\phi = 0.05$, $\sigma_1 = 1.5$ and $\sigma_2 = 0.9$. Figure 4 shows the likelihood function for α_0 based on the degree of separability ρ. Note in Figure 4 that the data gives information regarding the estimation of separability. Furthermore, we expect the probability of separability to be very small when we define a dataset with $\rho = 0.10$. Indeed in the fourth scenario, Figure 4(d), the data indicates probability close to zero for the null hypothesis of separability.
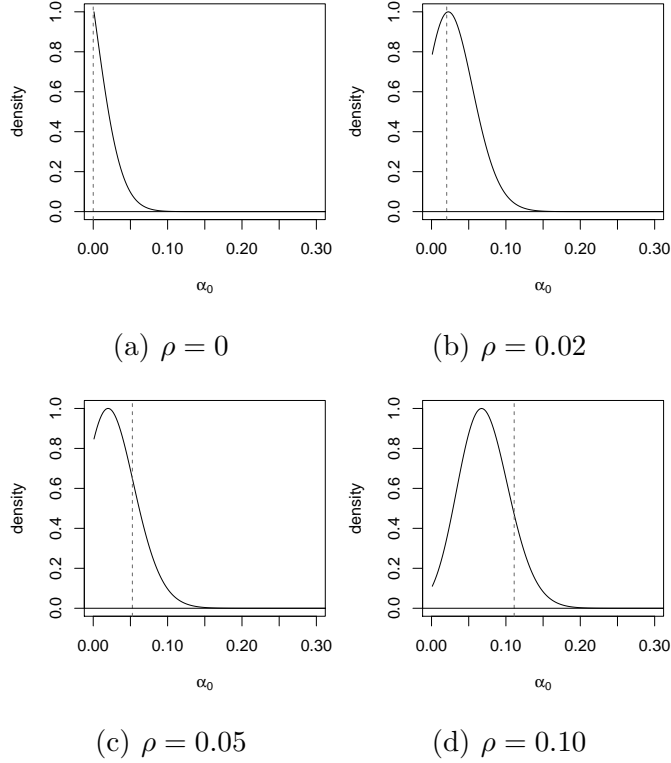
Figure 4: Likelihood functions. Dashed line: true value of α_0.

Table 3 presents the posterior probabilities $\tilde{p}_0$ for each model. In estimation for each scenario we consider the approximation of likelihood presented in Section 6. It is possible to see that the difference between the values of the measure of separability is very subtle but the difference between the posterior probabilities is substantial. Thus, the posterior probability of separability is a much easier to interpret measure for inference regarding separability. Note that ρ values greater than 0.10 indicate a strong nonseparability.

ρ	0	0.02	0.05	0.10
$\tilde{p}_0$	0.738	0.692	0.230	0.046

Table 3: Posterior probabilities $\tilde{p}_0$ for each measure of separability.

8   Conclusion

We have proposed a new flexible class of covariance functions for multivariate spatial processes based on the convex combination of separable covariance functions and on latent dimensions.

The idea of latent distances is the same as the dissimilarity measures of multidimensional scaling. Indeed, the latent distance between two components quantifies the degree of dissimilarity between them. According to Kaufman and Rousseeuw (2005), dissimilarity between two objects i and j, d_{ij}, is a nonnegative measure that is small (close to zero) when they are "near" to each other and large when they are very different[5]. These measures of dissimilarity can be calculated from the correlation between variables, that is, $d_{ij} = 1 - r_{ij}$, with r_{ij} the correlation between variables i and j. From this definition, variables with a strong positive correlation present dissimilarity close to zero, whereas variables with a strong negative correlation are considered very dissimilar (Kaufman and Rousseeuw, 2005). However, variables with strong negative correlation also imply the similar behaviour or patterns, despite an inverted scale of measurement (Khattree and Naik, 2000). The latent distances, δ_{ij}, considered in the model proposed in this work, also assign high values for negatively correlated variables. In this context, in future research we will work with distances that consider a dependency between components, regardless of their direction.

We have described an approximation of the full covariance matrices using the decomposition based on the Kronecker product of two separable matrices of minor dimensions. A sensitivity study was performed showing that the approximate approach provides important gains in computational efficiency. Although taking advantage of approximations to compute the likelihood, our proposal keeps interpretation and flexibility.

In terms of prediction the proposed model presents better results than the separable model, even when the approximation is considered. We conclude that it is better to consider a separable approximation of our nonseparable proposed model than to consider the separable

---

[5]More details on dissimilarity measures are presented in Cox and Cox (2000).

structure.

Finally, we have proposed a Bayesian test to measure the degree of separability between space and components. From the posterior probabilities $\tilde{p}_0$ we can choose the most suitable model. Indeed, the proposed measure is a easier to interpret than the separability parameter itself.

References

Apanasovich, T. V. and Genton, M. G. (2010). "Cross-covariance functions for multivariate random fields based on latent dimensions." *Biometrika*, 97, 1, 15–30.

Apanasovich, T. V., Genton, M. G., and Sun, Y. (2012). "A Valid Matérn Class of Cross-Covariance Functions for Multivariate Random Fields With Any Number of Components." *Journal of the American Statistical Association*, 107, 1, 15–30.

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2004). *Hierarchical Modeling and Analysis for Spatial Data*. Monographs on Statistics and Applied Probability, 1st ed. Chapman & Hall/CRC.

Cox, T. F. and Cox, M. A. A. (2000). *Multidimensional scaling*. Chapman & Hall/CRC.

Cressie, N. and Huang, H.-C. (1999). "Classes of Nonseparable, Spatio-Temporal Stationary Covariance Functions." *Journal of the American Statistical Association*, 94, 448, 1330–1340.

Cressie, N. and Zammit-Mangion, A. (2015). "Multivariate spatial covariance models: a conditional approach." Submited.

Fonseca, T. C. O. and Steel, M. F. J. (2011). "A General Class of Nonseparable Space-time Covariance Models." *Environmetrics*, 22, 2, 224–242.

Fuentes, M. (2006). "Testing for Separability of Spatial-Temporal Covariance Functions." *Journal of Statistical Planning and Inference*, 136, 2, 447–466.

Gamerman, D. and Lopes, H. F. (2006). *Markov Chain Monte Carlo: stochastic simulation for bayesian inference*. 2nd ed. CRC Press.

Genton, M. (2007). "Separable Approximations Of Space-time Covariance Matrices." *Environmetrics*, 18, 681–695.

Genton, M. G. and Kleiber, W. (2015). "Cross-Covariance Functions for Multivariate Geostatistics." *Statistical Science*, 30, 3, 147–163.

Gneiting, T. (2002). "Nonseparable, Stationary Covariance Functions for Space-Time Data." *Journal of the American Statistical Association*, 97, 458, 590–600.

Golub, G. H. and Van Loan, C. F. (1996). *Matrix Computations*. The Johns Hopkins University Press.

Kaufman, L. and Rousseeuw, P. J. (2005). *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley Series in Probability and Statistics, 1st ed. Wiley-Interscience.

Khattree, R. and Naik, D. N. (2000). *Multivariate Data Reduction and Discrimination with SAS Software*. 2nd ed. SAS Publishing.

Majumdar, A. and Gelfand, A. E. (2007). "Multivariate Spatial Modeling for Geostatistical Data Using Convolved Covariance Functions." *Mathematical Geology*, 39, 7, 225–245.

Mitchell, M. W., Genton, M. G., and Gumpertz, M. L. (2005). "Testing for separability of space-time covariances." *Environmetrics*, 16, 819–831.

— (2006). "A likelihood ratio test for separability of covariances." *Journal of Multivariate analysis*, 97, 1025–1043.

Stein, M. L. (2005). "Space-Time Covariance Functions." *Journal of the American Statistical Association*, 100, 469, 310–321.

Wackernagel, H. (1998). *Multivariate geostatistics*. Springer, New York.