Journal of Nonparametric Statistics
Vol. 00, No. 00, Month 2010, 1–25

RESEARCH ARTICLE

Study of spatial relationships between two sets of variables: A nonparametric approach

Francisco Cuevas[a] Emilio Porcu[b] and Ronny Vallejos[a*]

[a] *Departamento de Matemática, Universidad Técnica Federico Santa María Avenida Espana 1680, Valparaíso, Chile*
[b] *Departamento de Estadística, Universidad de Valparaíso, Valparaíso, Chile*;

(*v3.7 released September 2009*)

We propose a new method for estimating a codispersion coefficient to quantify the association between two spatial variables. Our proposal is based on a Nadaraya-Watson version of the codispersion coefficient through a suitable kernel. Under regularity conditions, we derive expressions for the bias and mean square error for a kernel version of the cross-variogram and establish the consistency of a Nadaraya-Watson estimator of the codispersion coefficient. In addition, we propose a bandwidth selection method for both, the variogram and the cross-variogram. Monte Carlo simulations support the theoretical findings, and as a result, the new proposal performs better than the classic Matheron's estimator. The proposed method is useful for quantifying spatial associations between two variables measured at the same location. Finally, we study forest data concerning the relationship among the tree height, basal area, elevation and slope of Pinus radiata plantations. A two-dimensional codispersion map is constructed to provide insight into the spatial association between these variables.

Keywords: Codispersion coefficient; Kernel; Nadaraya-Watson estimator; Spatial association

AMS Subject Classification: 62H11; 62H20; 62G05

*Corresponding author. Email: ronny.vallejos@usm.cl

ISSN: 1048-5252 print/ISSN 1029-0311 online

1. Introduction

In the analysis of spatial data, the quantification of spatial associations between two variables is an important issue, and considerable effort has been devoted to the construction of appropriate coefficients and tests for the association between two correlated variables. One can approach this issue by removing the spatial association among the observations and then applying existing techniques that have been developed for independent variables (Cliff and Ord 1981, § 7.4). An alternative way to manage the problem is to consider approaches that allow one to take into account the autocorrelation structure of the data. Tjostheim (1978) developed a nonparametric technique to measure the association between spatial variables from the ranks of the observations and the location coordinates of the measurement points. He also provided the asymptotic normality and asymptotic formulae for bias and variance. Later, Tjostheim's measure was generalized by Hubert and Golledge (1982). Clifford et al. (1989) studied a modified test of association based on the correlation coefficient by evaluating an effective sample size that takes into account the spatial structure. Subsequently, Richardson and Clifford (1991) suggested several adjustments to the original test in order to improve its power. Furthermore, Richardson et al. (1992) discussed how applications of this procedure could be used to study the relationships between lung cancer in men and employment in different industries. Extensions of this work can be found in Haining (1991), Dutilleul (1993), Dutilleulet al. (2008) and Fotheringham and Rogerson (2009). Richardson and Guihenneuc-Jouyaux (2009) reviewed several techniques to assess the spatial association between two geographical variables in the context of spatial epidemiology.

The codispersion coefficient (Matheron 1965) is a measure of association between two spatial variables and has been used in several applications (Chiles and Delfiner 1999; Goovaerts 1994, 1997, 1998). Such a measure is a normalized version of the cross-variogram, being a crucial instrument for multivariate spatial prediction (Cressie 1993; Ver Hoef and Barry 1998). Rukhin and Vallejos (2008) studied the codispersion coefficient from both theoretical and applied viewpoints, and established, for arbitrary lags, the consistency and limiting distribution of the sample coefficient. As a consequence of the asymptotic normality, these authors also addressed the hypothesis testing problem of independence between two spatial sequences. The codispersion coefficient has also been studied in time series to address how two time sequences change concurrently; it is a geometrically natural comovement coefficient since it compares proportional slopes at matched pairs of points across sequences (Croux et al. 2001). Expressions for the asymptotic variance of the coefficient were derived and the performance of the coefficient under additive contamination was examined by Vallejos (2008). Recently established theoretical results for the codispersion coefficient are valid for processes defined on a rectangular grid. However, there are many applications in which the spatial variables are defined on non-regular grids, and where the asymptotic properties of this coefficient remain unknown. Moreover, for two simultaneous autoregressive (SAR) processes, it is not difficult to show that the codispersion coefficient is constant and that it does not provide any information about the spatial association between the processes. Thus, determining more flexible coefficients of association between two spatial processes is important.

The goal of the current paper is to develop a nonparametric version of the codispersion coefficient. Extensions of this nature have previously been considered in the spatial statistics literature. For example, nonparametric estimations of the semi-variogram were examined by Garcia-Soidan et al. (2004), Yu, Mateu and Porcu (2007), and Garcia-Soidan (2007), among others. First, an example that motivated the present work is introduced in Section 2. In Section 3 we propose a nonparametric (Nadaraya-Watson) estimator for the cross-variogram between two spatial processes. Asymptotic expressions for the expected value, variance and mean square error of the proposed estimator will be provided. Next, the codispersion coefficient is computed using the nonparametric version of the cross-variogram and the existing estimators of the semi-variogram of each variable. Under regularity conditions, the consistency of the kernel estimation of the codispersion coefficient is established. In Section 4 a bandwidth selection method associated with the later estimation is proposed. Simulation studies are performed to support the theoretical results. We report the results of the Monte Carlo studies in Section 5. The proposed estimator for the codispersion coefficient is useful for quantifying the spatial association between forest variables based on a study of Pinus radiata plantations in the south of Chile. Through the use of codispersion maps, we explore in Section 6 the spatial association of these variables. Finally, in Section 7 we discuss the implications of the current results and briefly review future research directions.

2. A Motivating example

Here, we present an example of an issue that motivated the present work. Pinus radiata is one of the most widely planted species in Chile; it is planted on a wide array of soil types and in a variety of regional climates. Two important measures of plantation development are the dominant tree height and the basal area. Snowdon argues convincingly that both measures are correlated with regional climate and local growing conditions (Snowdon 2001). The variogram was used to characterize the spatial dependence of each variable. However, the assessment of the spatial association between tree height, tree basal area and other regional climate variables is of great interest for the quantification of spatial dependence and the detection of those directions in which there is either high or low degree of spatial association.

In the present article, we consider the relationship among the tree height, basal area, elevation and slope of Pinus radiata plantations. The study site is located in the sector *Escuadrón*, south of Concepción in the southern portion of Chile (36°54' S, 73°54' O) and has an area of 1244.43 hectare. In addition to more mature stands, we were also interested in the area containing young (i.e., four year old) stands of Pinus radiate, with an average density of 1600 trees per hectare. The basal area and dominant tree height at the year of plantation establishment (1993, 1994, 1995, and 1996) were used to represent the stand attributes. The three variables were obtained from 200 m^2 circular sample plots and point-plant sample plots. For the latter type of sample, four quadrants are established around the sample point; the four closest trees in each quadrant (16 trees in total) are then selected and measured in a clockwise direction. The samples were located systematically using a mean distance of 150 meters between samples. The total number of plots available for this study was 468 (Figure 1). In addition to the tree height and basal area, the coordinates, elevation and slope were recorded for each site. In this study, the year of plantation was not consider relevant in terms of assessing the spatial association between pairs of variables that are of interest. However, the year of plantation establishement could have a significant effect that might be consider in further studies.

Commonly, a data set as the forest data set described above is considered as a spatial point pattern having marks attached to each location. In this paper the locations are considered fixed. This is supported by the way in which the locations were settled down when the trees were planted on a rectangular grid. The sampling schemme in this case preserved this structure as is shown in Figure 1.

3. Definitions and main results

Throughout the paper we shall consider intrinsically stationary random fields $\{X(\boldsymbol{s}),\ \boldsymbol{s} \in D \subset \mathbb{R}^d\}$ with semi-variogram defined as

$$\gamma_X(\boldsymbol{k}) = \frac{1}{2}\mathrm{var}\{X(\boldsymbol{s}+\boldsymbol{k}) - X(\boldsymbol{s})\},$$

where $\boldsymbol{k} \in \mathbb{R}^d$ denotes the spatial lag. For n sampling sites $\boldsymbol{s}_1, \boldsymbol{s}_2, \dots, \boldsymbol{s}_n$, a natural and unbiased estimator based on the method of moments is the empirical semi-variogram (Matheron 1963) given by

$$\widehat{\gamma}_X(\boldsymbol{k}) = \frac{1}{2|N(\boldsymbol{k})|}\sum_{N(\boldsymbol{k})}(X(\boldsymbol{s}_i) - X(\boldsymbol{s}_j))^2, \tag{1}$$

where $N(\boldsymbol{k}) = \{(\boldsymbol{s}_i, \boldsymbol{s}_j) : ||\boldsymbol{s}_i - \boldsymbol{s}_j|| \in T(\boldsymbol{k}), 1 \le i, j \le n\}$, $T(\boldsymbol{k})$ is a tolerance region around $\boldsymbol{k}$, and where $|\cdot|$ denotes cardinality of a set. Alternatively to the empirical semi-variogram, robust estimators have been proposed (Cressie and Hawkins 1980; Genton 1998). Garcia-Soidan (2007) proposed a Nadaraya-Watson type estimator for the semi-variogram defined as

$$\breve{\gamma}_{X_h}(\boldsymbol{k}) = \frac{\sum_{i=1}^n \sum_{j=1}^n K\left(\frac{\boldsymbol{k}-(\boldsymbol{s}_i-\boldsymbol{s}_j)}{h}\right)(X(\boldsymbol{s}_i) - X(\boldsymbol{s}_j))^2}{2\sum_{i=1}^n \sum_{j=1}^n K\left(\frac{\boldsymbol{k}-(\boldsymbol{s}_i-\boldsymbol{s}_j)}{h}\right)}, \tag{2}$$

where h represents a bandwidth parameter and $K : \mathbb{R}^d \longrightarrow \mathbb{R}$ is a symmetric and strictly positive density function. For such an estimator, Garcia-Soidan (2007) establishes, under regularity conditions, consistency and asymptotic normality, and addresses the inadequate behavior of estimator (2) near the endpoints.

Let $\{(X(\boldsymbol{s}), Y(\boldsymbol{s})) : \boldsymbol{s} \in D \subset \mathbb{R}^d\}$ be a bivariate intrinsically stationary random field on D with cross-variogram $2\gamma_{XY}(\cdot) : \mathbb{R}^d \to \mathbb{R}$ defined through (Cressie 1993, pp. 67)

$$2\gamma_{XY}(\boldsymbol{k}) = \mathbb{E}[(X(\boldsymbol{s}+\boldsymbol{k}) - X(\boldsymbol{s}))(Y(\boldsymbol{s}+\boldsymbol{k}) - Y(\boldsymbol{s}))], \tag{3}$$

for all $\boldsymbol{s}, \boldsymbol{s}+\boldsymbol{k} \in D$, and with marginal variograms $2\gamma_X$, $2\gamma_Y$ as defined through Equation (3). The codispersion coefficient (Matheron 1965) is a normalized version of (3) and defined through

$$\rho_{XY}(\boldsymbol{k}) = \frac{\gamma_{XY(\boldsymbol{k})}}{\sqrt{\gamma_X(\boldsymbol{k})\gamma_Y(\boldsymbol{k})}}.$$

Rukhin and Vallejos (2008) found a closed form for the coefficient $\rho_{XY}(\cdot)$ for spatial autoregressive processes under particular assumptions on the correlation structure of the errors and when considering a rectangular lattice. These expressions are also valid for time series models; however, the selection of a parametric model for the cross-variogram function is not trivial. Indeed, Chiles and Delfiner (1999, pp. 330) show that a multivariate model with exponential variograms $\gamma_X(k) = \gamma_Y(k) = 1 - e^{-b|k|}, b > 0$, and a cross-variogram of the form $\gamma_{XY}(k) = \rho(1 - e^{c|k|}), c > 0$, is not valid. Moreover, the only case in which the model can be valid is when $b = c$, with the consequence that $\rho_{XY}(k) = \rho$. An extensive discussion about efficient approaches to construct valid variograms matrix functions can be found in Gneiting et al. (2010) as well as in Stein (2007). One way to avoid such a drawback is to consider a nonparametric approach to assess the estimation of both the cross-variogram and the codispersion coefficient.

The analogue of Matheron's estimator for the cross-variogram is obtained through

$$\widehat{\gamma}_{XY}(\boldsymbol{k}) = \frac{1}{2|N(\boldsymbol{k})|} \sum_{N(\boldsymbol{k})} (X(\boldsymbol{s}_i) - X(\boldsymbol{s}_j))(Y(\boldsymbol{s}_i) - Y(\boldsymbol{s}_j)), \tag{4}$$

where $N(\boldsymbol{k})$ is defined as in Equation (1). The corresponding empirical estimator of the codispersion based on (4) is given by

$$\widehat{\rho}_{XY}(\boldsymbol{k}) = \frac{\widehat{\gamma}_{XY}(\boldsymbol{k})}{\sqrt{\widehat{\gamma}_X(\boldsymbol{k})\widehat{\gamma}_Y(\boldsymbol{k})}}. \tag{5}$$

The analogue of the Nadaraya-Watson type estimator for the cross-variogram is instead given by

$$\breve{\gamma}_{XY_h}(\boldsymbol{k}) = \frac{\sum_{i=1}^n \sum_{j=1}^n K\left(\frac{\boldsymbol{k}-(\boldsymbol{s}_i-\boldsymbol{s}_j)}{h}\right)(X(\boldsymbol{s}_i) - X(\boldsymbol{s}_j))\,(Y(\boldsymbol{s}_i) - Y(\boldsymbol{s}_j))}{2\sum_{i=1}^n \sum_{j=1}^n K\left(\frac{\boldsymbol{k}-(\boldsymbol{s}_i-\boldsymbol{s}_j)}{h}\right)}, \tag{6}$$

where $K(\cdot)$ is a kernel function as in Equation (2). Here, we extend the results of Garcia-Soidan (2007) for the semi-variogram to the case of the estimator at (6). In the current case, we find asymptotic expressions for the bias and mean square error of (6). The required assumptions for the results that we present below are the following:

- (A1) $D = D_n = \lambda D_0$ for some increasing sequence $\lambda = \{\lambda_n\}_n^\infty$, and for some fixed and bounded region $D_0 \subset \mathbb{R}^d$ containing a sphere with positive d-dimensional volume.
- (A2) f_0 is a density function in D_0. We assume that $\alpha_1 < f_0(\boldsymbol{x}) < \alpha_2$ for all $\boldsymbol{x} \in D_0$ and for some $\alpha_1, \alpha_2 > 0$.
- (A3) The spatial locations will be set as $\mathbf{s}_i = \lambda \mathbf{u}_i$, for $1 \leq i \leq n$, where $\mathbf{u}_1, \ldots, \mathbf{u}_n$ are independent realizations of a random vector with density function f_0. The subjacent vectors variables will be denoted by $\boldsymbol{U}_1, \ldots, \boldsymbol{U}_n$.
- (A4) f_i is the density function of the random vector $(\boldsymbol{U}_1 - \boldsymbol{U}_2, \ldots, \boldsymbol{U}_1 - \boldsymbol{U}_{i+1})$, for $i \in \mathbb{N}, i \geq 1$. $f_1(\mathbf{0}) > 0$ and f_i is a continuously differentiable function in a neighborhood of $\mathbf{0}^+$ for all $i \in \mathbb{N}, 1 \leq i \leq 4l - 1$, and $l \leq 1$.
- (A5) K is a $d-$variate symmetric, bounded kernel with compact support, with covariance matrix given by $c_K I_d$, for some $c_K > 0$, where I_d the identity matrix of order d.
- (A6) $\{h + \lambda^{-1} + (nh^d)^{-1} + \lambda^d n^{-1})\} \longrightarrow 0$, as $n \longrightarrow \infty$.
- (A7) $\mathbb{E}[X(\boldsymbol{s})^{4m}] < +\infty$ for all $\boldsymbol{s} \in D$, and $m \geq 1$. Moreover, for all $i \in \mathbb{N}, 1 \leq i \leq m$,
  - There exists a function $g_i : \mathbb{R}^{(4i-1)d} \to \mathbb{R}$ continuous and differentiable such that

for all $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_{4i} \in D$,

$$\mathbb{E}\left[\prod_{j=1}^{2i}((X(\mathbf{s}_{2j-1}) - X(\mathbf{s}_{2j}))^2 - \mathbb{E}((X(\mathbf{s}_{2j-1}) - X(\mathbf{s}_{2j}))^2))\right] = g_i(\mathbf{s}_1 - \mathbf{s}_2, \ldots, \mathbf{s}_1 - \mathbf{s}_{4i}).$$

- One has for $x > 0$,

$$I_i(x) = \int_{A_i(x)} |g_i(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots, \boldsymbol{\xi}_{4i})| d\boldsymbol{\xi}_1 d\boldsymbol{\xi}_2 \ldots d\boldsymbol{\xi}_{4i-1} < \infty,$$

where $A_i(x)$ is the set of all vectors $(\mathbf{s}_1, \ldots, \mathbf{s}_{4i-1}) \in \mathbb{R}^{(4i-1)d}$ satisfying $||\mathbf{s}_1|| \leq x$ when $||\mathbf{s}_{2j} - \mathbf{s}_{2j+1}|| \leq x$ for all $j \in \mathbb{N}, 1 \leq j \leq 2i - 1$.

Condition (A7) also holds for $Y(\boldsymbol{s})$ not necessarily with the same function g_i. Under assumptions $(A1) - (A7)$ Garcia-Soidan (2007) found asymptotic expressions for the bias and mean square error of estimator (2). In order to establish the same results for the estimator defined through Equation (6) it is necessary to add the following assumptions:

(H1) There exists a bounded continuous differentiable function $\psi : \mathbb{R}^{3d} \to \mathbb{R}$ such that for all $\mathbf{s}_1, \ldots, \mathbf{s}_n \in D$,

$$\begin{aligned}\mathrm{Cov}[(\mathrm{X}(\boldsymbol{s}_\mathrm{i}) - \mathrm{X}(\boldsymbol{s}_\mathrm{j}))(\mathrm{Y}(\boldsymbol{s}_\mathrm{i}) - \mathrm{Y}(\boldsymbol{s}_\mathrm{j})), (\mathrm{X}(\boldsymbol{s}_\mathrm{k}) - \mathrm{X}(\boldsymbol{s}_\mathrm{l}))(\mathrm{Y}(\boldsymbol{s}_\mathrm{k}) - \mathrm{Y}(\boldsymbol{s}_\mathrm{l}))]\\ = \psi(\mathbf{s}_i - \mathbf{s}_j, \mathbf{s}_i - \mathbf{s}_k, \mathbf{s}_i - \mathbf{s}_l).\end{aligned}$$

(H2) For all $x > 0$, $\eta(x) = \int_{L(x)} |\psi(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \boldsymbol{\xi}_3)| \, d\boldsymbol{\xi}_1 d\boldsymbol{\xi}_2 d\boldsymbol{\xi}_3 \leq \infty$, where $L(x)$ corresponds to the set of all $(\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3, \mathbf{s}_4) \in \mathbb{R}^{4d}$ such that $||\mathbf{s}_1|| < x$ when $||\mathbf{s}_2 - \mathbf{s}_3|| \leq x$ and $||\mathbf{s}_3 - \mathbf{s}_4|| \leq x$.

Theorem 3.1 : *Let $\{(X(\boldsymbol{s}), Y(\boldsymbol{s})) : \boldsymbol{s} \in D \subset \mathbb{R}^d\}$ be a bivariate intrinsically stationary random field. Suppose that Assumptions (A1)-(A7), (H1), (H2) hold with $l = 1$. In addition assume that $\gamma_{XY}(\boldsymbol{k})$ is three times continuously differentiable in a neighborhood of $\boldsymbol{k} \neq 0$. Then*

$$\begin{aligned}\mathbb{E}[\breve{\gamma}_{XY_h}(\boldsymbol{k})] &= \gamma_{XY}(\boldsymbol{k}) + h^2 \frac{c_K}{2} \triangle \gamma_{XY}(\boldsymbol{k}) + o(h^2),\\ Var[\breve{\gamma}_{XY_h}(\boldsymbol{k})] &= \frac{d_K A_{1,d}(\boldsymbol{k})}{2 f_1(\mathbf{0})} n^{-2} \lambda^d h^{-d} + \frac{f_3(\mathbf{0}, \mathbf{0}, \mathbf{0}) A_{2,d}(\boldsymbol{k})}{(2 f_1(\mathbf{0}))^2} \lambda^{-d}\\ &\quad + o(n^{-2} \lambda^d h^{-d} + \lambda^{-d} + h^4),\end{aligned}$$

where c_K is as in (A5), $d_K = \int_{\mathbb{R}^d} K^2(\boldsymbol{\xi}) d\boldsymbol{\xi}$, $A_{1,d}(\boldsymbol{k}) = \psi(\boldsymbol{k}, \mathbf{0}, \boldsymbol{k})$, $A_{2,d}(\boldsymbol{k}) = \int_{\mathbb{R}^d} \psi(\boldsymbol{k}, \boldsymbol{\xi}, \boldsymbol{\xi} + \boldsymbol{k}) d\boldsymbol{\xi}$, $f_3(\mathbf{0}, \mathbf{0}, \mathbf{0})$ is as in (A4), $f_1(\mathbf{0})$ is as in (A3) and $\triangle \gamma_{XY}(\boldsymbol{k}) = \sum_{j=1}^d \left. \frac{\partial^2 \gamma_{XY}(\boldsymbol{l})}{\partial l_j^2} \right|_{\boldsymbol{k}}$.

For a neater exposition, the proofs of all results stated in Sections 3 and 4 are left to the Appendix A.

An expression for the mean square error of estimator (6) can be derived as an immediate consequence of Theorem 3.1. In fact, if $\lambda = O(nh^{d/2})$, then

$$\mathrm{MSE}[\breve{\gamma}_{XY}(\boldsymbol{k})] = O(n^{-8/(8+d)}).$$

Assumptions (A1)-(A4) are related to the domain D, the locations $\boldsymbol{s}_i \in D$ and functions f_i. These assumptions are the usual requirements to perform infill asymptotics. An example in which we chose all quantities involved in these assumptions will be given in Section 5 in a simulation framework. Assumption (A5) is general in the sense that does not require differentiability as in Garcia-Soidan (2007). Assumption (A6) is required for the asymptotic variance of estimator (6) in Theorem 3.1. In order to have a finite covariance between the increments of processes $X(\cdot)$ and $Y(\cdot)$, condition (A7) is assumed as regards the random processes. Assumptions (H1) and (H2) are not too restrictive. An example in which these conditions are satisfied can be found in Appendix A.

A kernel type estimator of the codispersion coefficient is

$$\breve{\rho}_{XY_{\boldsymbol{h}}}(\boldsymbol{k}) = \frac{\breve{\gamma}_{XY_{h_1}}(\boldsymbol{k})}{\sqrt{\breve{\gamma}_{X_{h_2}}(\boldsymbol{k})\breve{\gamma}_{Y_{h_3}}(\boldsymbol{k})}}, \tag{7}$$

where $\boldsymbol{h} = (h_1, h_2, h_3)$, $\breve{\gamma}_{XY_{h_1}}(\boldsymbol{k})$ is as in (6) and $\breve{\gamma}_{X_{h_2}}(\boldsymbol{k})$ is as in (2). For $\breve{\rho}_{XY_{\boldsymbol{h}}}(\boldsymbol{k})$ we can establish the following result.

Theorem 3.2 : *Let $\{X(\boldsymbol{s}) : \boldsymbol{s} \in D \subset \mathbb{R}^d\}$ and $\{Y(\boldsymbol{s}) : s \in D \subset \mathbb{R}^d\}$ be two intrinsically stationary processes. Suppose that for both processes assumptions (A1)-(A7) and (H1)-(H2) hold, with different kernels functions in (A5) and different g_i functions in (A7). Then*

$$\breve{\rho}_{XY_{\boldsymbol{h}}}(\boldsymbol{k}) \xrightarrow{\mathcal{P}} \rho_{XY}(\boldsymbol{k}), \ \textit{as } n \to \infty.$$

The problem of unsatisfactory behavior of the estimators near the endpoints has been treated in the literature. The procedures proposed are based on appropriate modifications of the estimator considered, in order to achieve consistency as well as to retain rates of convergence. For a detailed discussion the reader is deferred to Garcia-Soidan et al. (2004).

Bandwidth selection plays an important role in kernel estimation. In order to study the computational implementation of estimator (7) and later use it to analyze real data sets, the bandwidth selection associated with estimators (2) and (6) shall be addressed in the next section.

4. Bandwidth selection

There are several criteria to consider when selecting a bandwidth, depending on the area of study and type of data. Ruppert et al. (1995) studied an effective band selector for regression analysis. In econometrics, some bandwidth selection criteria are discussed in Li and Racine (2006). Recently, Kohler et al. (2011) compared bandwidth selection methods for kernel regression. For an example of bandwidth selection in spatial statistics, see Hallin et al. (2004). It is apparent from (7) that the bandwidth of the nonparametric estimator of the codispersion coefficient depends on three different bandwidth parameters: two associated with the semi-variograms of $X(\cdot)$ and $Y(\cdot)$ and one associated with the cross-variogram between the processes.

4.1. *Bandwidth Selection for the Semi-Variogram*

Garcia-Soidan et al. (2004) suggested a method to compute the bandwidth associated with the semi-variogram of an intrinsically stationary process. This criterion is based on the minimization of an asymptotic mean square error (AMSE). We suggest to use asymptotic versions of the integrated square error (AMISE). By Theorem 3.1, we obtain

$$\begin{aligned}\text{AMISE}[\breve{\gamma}_{X_h}(\boldsymbol{k})] = h^4\frac{c_K^2}{4}\int_{\mathbb{R}^d}(\triangle\gamma_X(\boldsymbol{\xi}))^2\,d\boldsymbol{\xi} + \frac{d_K}{2f_1(\mathbf{0})}n^{-2}\lambda^d h^{-d}\int_{\mathbb{R}^d}B_{1,d}(\boldsymbol{\xi})d\boldsymbol{\xi} \\ + \frac{f_3(\mathbf{0},\mathbf{0},\mathbf{0})}{(2f_1(\mathbf{0}))^2}\lambda^{-d}\int_{\mathbb{R}^d}B_{2,d}(\boldsymbol{\xi})d\boldsymbol{\xi},\end{aligned} \tag{8}$$

where $B_{1,d}(\boldsymbol{k}) = g_1(\boldsymbol{k},\mathbf{0},\boldsymbol{k})$, $B_{2,d}(\boldsymbol{k}) = \int_{\mathbb{R}^d} g_1(\boldsymbol{k},\boldsymbol{\xi},\boldsymbol{k}+\boldsymbol{\xi})d\boldsymbol{\xi}$ and $\triangle\gamma_X(\boldsymbol{k}) = \sum_{j=1}^d \frac{\partial^2\gamma_X(\boldsymbol{l})}{\partial l_j^2}\Big|_{\boldsymbol{k}}$.

One of the major drawbacks concerning the right hand side of (8) is the fact that the integrals are defined on $\mathbb{R}^d$, but in most of the cases, $(\triangle\gamma(\boldsymbol{k}))^2$ is not bounded in a neighborhood of $\mathbf{0}$, because the variograms may not be differentiable at the origin. To avoid this concern, we suggest to use a twice differentiable variogram at the origin and an appropriate selection of the support.

Theorem 4.1: *The bandwidth h_{AMISE} that minimize $\text{AMISE}[\breve{\gamma}_{X_h}(\boldsymbol{k})]$ is given by*

$$h_{\text{AMISE}} = \left(\frac{d}{4}\frac{C_2}{C_1}\right)^{\frac{1}{d+4}},$$

where

$$\begin{aligned}C_1 &= \frac{c_K^2}{4}\int_{\mathbb{R}^d}(\triangle\gamma_X(\boldsymbol{\xi}))^2\,d\boldsymbol{\xi},\\ C_2 &= \frac{d_K}{2f_1(\mathbf{0})}n^{-2}\lambda^d\int_{\mathbb{R}^d}B_{1,d}(\boldsymbol{\xi})d\boldsymbol{\xi},\end{aligned}$$

c_K, d_K, and $f_1(\mathbf{0})$ are as in Theorem 3.1 and $B_{1,d}$ is as in Equation (8).

In order to illustrate the computation of the bandwidth given in Theorem 4.1, we assume $D = B(\mathbf{0},\phi)$, the ball of $\mathbb{R}^d$ centered at $\mathbf{0}$ with radius ϕ, i.e. $D = \{\boldsymbol{s}\in\mathbb{R}^d : ||\boldsymbol{s}|| \le \phi\}$ and consider an isotropic semi-variogram $\gamma_X(||\boldsymbol{k}||)$. Then

$$\begin{aligned}\int_D \gamma^2(\boldsymbol{\xi})d\boldsymbol{\xi} &= \int_D \gamma_X^2(||\boldsymbol{\xi}||)d\boldsymbol{\xi},\\ &= S_{n-1}(1)\int_0^{\phi}\gamma_X^2(\xi)\xi d\xi,\end{aligned} \tag{9}$$

where $S_{n-1}(1)$ is the volume of the $n-1$ dimensional unit sphere. Using

$$\begin{aligned}
\frac{\partial||\boldsymbol{k}||}{\partial k_i} &= \frac{k_i}{||\boldsymbol{k}||}, \\
\frac{\partial \gamma_X(||\boldsymbol{k}||)}{\partial k_i} &= \gamma_X^{'}(||\boldsymbol{k}||)\frac{k_i}{||\boldsymbol{k}||}, \\
\frac{\partial^2 \gamma_X(||\boldsymbol{k}||)}{\partial k_i^2} &= \gamma_X^{''}(||\boldsymbol{k}||)\frac{k_i^2}{||\boldsymbol{k}||^2} + \frac{\gamma_X^{'}(||\boldsymbol{k}||)}{||\boldsymbol{k}||}\left(1 - \frac{k_i^2}{||\boldsymbol{k}||^2}\right)
\end{aligned}$$

where $\gamma'(||\boldsymbol{k}||)$ and $\gamma''(||\boldsymbol{k}||)$ respectively denote the first and second derivative of the variogram with respect to $\boldsymbol{k}$, one obtains

$$\triangle \gamma_X(||\boldsymbol{k}||) = \sum_{i=1}^{d} \frac{\partial^2 \gamma_X(||\boldsymbol{k}||)}{\partial k_i^2} = \gamma_X^{''}(||\boldsymbol{k}||) + (n-1)\frac{\gamma_X^{'}(||\boldsymbol{k}||)}{||\boldsymbol{k}||}.$$

Moreover,

$$\begin{aligned}
\int_D (\triangle\gamma(\boldsymbol{\xi}))^2 d\boldsymbol{\xi} &= \int_D (\triangle\gamma_X(||\boldsymbol{\xi}||))^2 d\boldsymbol{\xi} \\
&= \int_D \left(\gamma_X^{''}(||\boldsymbol{\xi}||) + (n-1)\frac{\gamma_X^{'}(||\boldsymbol{\xi}||)}{||\boldsymbol{\xi}||}\right)^2 d\boldsymbol{\xi} \\
&= S_{n-1}(1)\int_0^{\phi}\left(g^{''}(\xi) + (n-1)\frac{g^{'}(\xi)}{\xi}\right)^2 \xi d\xi. \qquad (10)
\end{aligned}$$

Replacing expressions (9) and (10) in (8) one gets

$$h_{\text{AMISE}} = \left(\frac{dn^{-2}\lambda^d d_k}{c_k^2 2 f_1(0)} \frac{\int_0^{\phi}\gamma_X^2(\xi)\xi d\xi}{\int_0^{\phi}\left(\gamma_X^{''}(\xi) + (n-1)\frac{\gamma_X^{'}(\xi)}{\xi}\right)^2 \xi d\xi}\right)^{\frac{1}{4+d}}. \qquad (11)$$

In Table 1, the values of $\triangle\gamma_X(||\boldsymbol{s}||)$ for different isotropic parametric semi-variograms models are given. In particular, when an Epanechnikov kernel is used, $f_1(0)$ is uniformly distributed, $\lambda^d = n^{\frac{8}{8+d}}$ (Garcia-Soidan 2007) and $d = 2$, h_{AMISE} in (11) becomes

$$h_{\text{AMISE}} = n^{-1/5}2^{2/3}3^{1/3}\left(\frac{\int_0^{\phi}\gamma_X^2(\xi)\xi d\xi}{\int_0^{\phi}\left(\gamma_X^{''}(\xi) + (n-1)\frac{\gamma_X^{'}(\xi)}{\xi}\right)^2 \xi d\xi}\right)^{\frac{1}{6}}.$$

If there exist geometric anisotropy characterized by a positive definite matrix A, the extension of rule (8) is straightforward considering that in this case $D = \{\boldsymbol{s} \in \mathbb{R}^d : \sqrt{\boldsymbol{s}^{'}A\boldsymbol{s}} \leq \phi\}$,

Table 1. Values of $0 \leq t \mapsto \triangle\gamma_X(t)$ for isotropic models ($||\boldsymbol{k}|| =: t$). In the Matérn model $\Gamma(\kappa)$ is the gamma function and K_κ is a modified Bessel function of the second kind of order κ.

Model	$\gamma_X(t)$	$\triangle\gamma_X(t)$
Exponential	$\sigma^2 - \sigma^2 \exp\left(-\frac{t}{\phi}\right)$	$-\frac{e^{-\frac{t}{\phi}}\sigma^2(t+\phi-n\phi)}{t\phi^2}$
Gaussian	$\sigma^2 - \sigma^2 \exp\left(-\left(\frac{t}{\phi}\right)^2\right)$	$\frac{2e^{-\frac{t^2}{\phi^2}}\sigma^2\left(-2t^2+n\phi^2\right)}{\phi^4}$
Exp - Power	$\sigma^2 - \sigma^2 \exp\left(-\left(\frac{t}{\phi}\right)^\kappa\right)$	$-\frac{e^{-\left(\frac{t}{\phi}\right)^\kappa}\kappa\sigma^2\left(2-n+\kappa\left(-1+\left(\frac{t}{\phi}\right)^\kappa\right)\right)\left(\frac{t}{\phi}\right)^\kappa}{t^2}$
Wave	$\sigma^2 - \frac{\sigma^2\phi}{t}\sin\left(\frac{t}{\phi}\right)$	$\frac{\sigma^2\left(-t(-3+n)\phi\cos\left[\frac{t}{\phi}\right]+\left(t^2+(-3+n)\phi^2\right)\sin\left[\frac{t}{\phi}\right]\right)}{t^3\phi}$
Matérn	$\frac{\sigma^2}{2^{\kappa-1}\Gamma(\kappa)}\left(\frac{t}{\phi}\right)^\kappa K_\kappa\left(\frac{t}{\phi}\right)$	$-\frac{2^{1-\kappa}\sigma^2\left(\frac{t}{\phi}\right)^{\kappa-1}\left(tK_{\kappa-2}\left(\frac{t}{\phi}\right)-n\phi K_{\kappa-1}\left(\frac{t}{\phi}\right)\right)}{\phi^3\Gamma(\kappa)}$
Wendland C^2	$\sigma^2\left(1-\left((\upsilon+1)\left(\frac{h}{\phi}\right)+1\right)\left(1-\left(\frac{h}{\phi}\right)\right)^{\upsilon+1}\right)$	$\frac{\sigma^2(1+\upsilon)(2+\upsilon)\left(1-\frac{h}{\phi}\right)^\upsilon(h(d+\upsilon)-d\phi)}{(h-\phi)\phi^2}$

$$\int_D \gamma^2(\boldsymbol{\xi})d\boldsymbol{\xi} = |A|^{-1/2}S_{n-1}(1)\int_0^\phi \gamma_X^2(\xi)\xi d\xi,$$

$$\triangle\gamma(\boldsymbol{s}) = \gamma_X^{''}(||A^{1/2}\boldsymbol{s}||) + (\mathrm{tr}(A)-1)\frac{\gamma_X^{'}(||A^{1/2}\boldsymbol{s}||)}{||A^{1/2}\boldsymbol{s}||},$$

$$\int_D (\triangle\gamma(\boldsymbol{\xi}))^2 d\boldsymbol{\xi} = |A|^{-1/2}S_{n-1}(1)\int_0^\phi \left(\gamma_X^{''}(\xi) + (\mathrm{tr}(A)-1)\frac{\gamma_X^{'}(\xi)}{\xi}\right)^2 \xi d\xi,$$

where $\mathrm{tr}(\cdot)$ represents the trace operator.

4.2. *Bandwidth Selection for the Cross-Variogram*

Because of the difficulties in selecting an appropriate parametric model for the cross-variogram, the bandwidth selection method described for the semi-variogram cannot be easily emulated to produce a rule for computing the bandwidth associated with the cross-variogram. However, a simple approach to manage this problem is to consider that the bandwidth of the cross-variogram is in some way related to the variability of processes $X(\cdot)$ and $Y(\cdot)$. Denoting these bandwidths as h_X and h_Y, respectively, an estimation of the bandwidth for the cross-variogram is

$$\widehat{h}_{XY} = \frac{\widehat{\sigma}_X^2\widehat{h}_X + \widehat{\sigma}_Y^2\widehat{h}_Y}{\widehat{\sigma}_X^2 + \widehat{\sigma}_Y^2}, \tag{12}$$

where in practice, $\widehat{\sigma}_X^2$ and $\widehat{\sigma}_Y^2$ can be obtained by fitting a parametric model to the empirical semi-variogram (e.g., one from the models described in Table 1). The bandwidth suggested in Equation (12) can be generalized through quasi arithmetic compositions (Porcu et al. 2009). Part of the material we describe here is largely readapted from the mentioned reference, where such compositions are treated in a more general framework and then devoted to the composition of covariance functions. The result in Proposition

4.3 is instead new and its proof deferred to the Appendix.

Definition 4.2: (Quasi-arithmetic compositions) Let Φ be the class of continuous functions φ with proper inverse. If $\widehat{h}_X : \mathbb{R}_+ \to \mathbb{R}_+$ and $\widehat{h}_Y : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\widehat{h}_X + \widehat{h}_Y \subset D(\varphi^{-1})$ for some $\varphi \in \Phi$, the *quasi-arithmetic composition of* $\widehat{h}_X$ *and* $\widehat{h}_Y$ *with generating function* φ *and weight* $\theta \in [0,1]$ *is defined as the functional*

$$\psi_\theta\left(\widehat{h}_X, \widehat{h}_Y\right) := \varphi\left(\theta\varphi^{-1}\circ\widehat{h}_X + (1-\theta)\varphi^{-1}\circ\widehat{h}_Y\right), \tag{13}$$

where $\circ$ *denotes composition.*

Throughout the paper, we refer to $\varphi \in \Phi$ as the generating functions of ψ_θ with weight θ.

Ordering relations as well as a minimal element can be found among the set of quasi-arithmetic compositions of two $\widehat{h}_X$, $\widehat{h}_Y$ fixed functions indexed by convex generating functions. For a given $0 < \theta < 1$, we have $\mathrm{G}_\theta(\widehat{h}_X, \widehat{h}_Y) := \widehat{h}_X^\theta \widehat{h}_Y^{1-\theta}$, which is the quasi-arithmetic composition associated with $\varphi(t) = \exp(-t)$ that generates a geometric average, with the conventions $\ln(0) = -\infty$ and $\exp(-\infty) = 0$. Also, we have $\mathrm{A}_\theta(\widehat{h}_X, \widehat{h}_Y) = \theta\widehat{h}_X + (1-\theta)\widehat{h}_Y$, that is the quasi-arithmetic composition associated with $\varphi(t) = M(1-t/M)_+$ that generates the arithmetic average, where $M > \max\left(\widehat{h}_X, \widehat{h}_Y\right)$. Finally, we have $\mathrm{H}_\theta(\widehat{h}_X, \widehat{h}_Y) = \frac{\theta}{\widehat{h}_X} + \frac{1-\theta}{\widehat{h}_X}$, the quasi-arithmetic composition associated with $\varphi(t) = 1/t$ and generating the harmonic average, with the conventions $1/\infty = 0$, $1/0 = \infty$ and $0/0 = 0$.

Many pleasant properties of such a framework can be deduced from the results in Hardy et al. (1934) as well as those presented in Porcu et al. (2009). Here, we present a new result that gives a characterization for the association measure proposed in this paper.

Proposition 4.3: *Let* $0 < \theta < 1$. *Let* $\widehat{h}_X$ *and* $\widehat{h}_Y$ *be a pair of mappings from* $\mathbb{R}_+$ *into* $\mathbb{R}_+$. *If* $\widehat{h}_X$ *and* $\widehat{h}_Y$ *are upperly bounded, then for any generating function* $\varphi \in \Phi$, *the matrix*

$$\Sigma := \begin{pmatrix} \psi_\theta\left(\widehat{h}_X, \widehat{h}_X\right) & \psi_\theta\left(\widehat{h}_X, \widehat{h}_Y\right) \\ \psi_\theta\left(\widehat{h}_Y, \widehat{h}_X\right) & \psi_\theta\left(\widehat{h}_Y, \widehat{h}_Y\right) \end{pmatrix}_{i,j=1}^2,$$

is positive definite.

The bandwidth estimation provided by Theorem 4.1 and (12) allows us to implement computationally the Nadaraya-Watson estimator of the codispersion given in (7). Other results can be deduced from the arguments in Porcu et al. (2009) and the references therein. We report them, adapted to our case, for the sake of a self expository manuscript.

For any pair of mappings $g_1, g_2 : \mathbb{R}^2 \to \mathbb{R}$, Let us write $g_1 \le g_2$ whenever $g_1(x_1,x_2) \le g_2(x_1,x_2)$ for any $(x_1,x_2) \in \mathbb{R}^2$. The following facts are true and represent a special case of Proposition 1 in Porcu et al. (2009):

Let $\varphi, \varphi_i \in \Phi$, $i = 1,2$ and let us use the abuse of notation $\psi_\theta^{(i)}$ for the quasi arithmetic composition associated to φ_i. Then,

- if $\varphi^{-1}\circ\varphi$ is convex, then $\psi_\theta^{(1)} \le \psi_\theta^{(2)}$;

- if $\varphi^{-1} \circ \varphi$ is concave, then $\psi_\theta^{(1)} \geq \psi_\theta^{(2)}$;
- If φ is convex, then $\psi_\theta \leq G_\theta \leq A_\theta$;
- As a special case, we get the classic inequality between mean values $H_\theta \leq G_\theta \leq A_\theta$.

As a last remark, the special case $\theta = 1/2$ is called Archimedean composition and has been well known under the framework of copula modelling, for which many references can be found again in Porcu et al. (2009).

5. Simulations

We report finite-sample simulations results for the nonparametric codispersion coefficient described in (7) . The simulations were performed in R (R Development Core Team 2012, version 2.11.1). We consider $d = 2$ and the locations of both processes to belong to the region $D_n = \lambda_n D_0$, where $\lambda_n = n^{8/(8+d)d}$, $D_0 = [0,1] \times [0,1]$, and $\lambda_n = 150^{2/5}, 300^{2/5}, 500^{2/5}$. The random observations considered in this study were generated from a zero mean bivariate Gaussian random field with a covariance matrix associated with a parsimonious bivariate Matérn covariance function (Gneiting et al. 2010) defined through

$$
\begin{aligned}
C_{11}(\boldsymbol{k}) &= \sigma_1^2 M(\boldsymbol{k}|\nu_1, a), \\
C_{22}(\boldsymbol{k}) &= \sigma_2^2 M(\boldsymbol{k}|\nu_2, a), \\
C_{12}(\boldsymbol{k}) = C_{21}(\boldsymbol{k}) &= \rho_{12}\sigma_1\sigma_2 M(\boldsymbol{k}, \frac{1}{2}(\nu_1 + \nu_2), a),
\end{aligned}
$$

where $M(\boldsymbol{k}|\nu, a) = \frac{2^{1-\nu}}{\Gamma(\nu)}(a||\boldsymbol{k}||)^\nu K_\nu(a||\boldsymbol{k}||)$, K_ν is a modified bessel function of the second kind, $a > 0$, and the correlation coefficient is bounded by the ratio of the harmonic mean and the arithmetic mean of the marginal smoothness parameters, i.e.

$$
|\rho_{12}| < \frac{(\nu_1\nu_2)^{1/2}}{\frac{1}{2}(\nu_1 + \nu_2)}.
$$

Several scenarios for inspecting the performance of the kernel estimator of the codispersion coefficient were explored. To quantify the quality of the estimations, the simulation mean, standard deviation and bias were used. The estimator (7) was implemented for $K_2(x, y) = K(x)K(y)$, where $K(\cdot)$ corresponds to the Epanechnikov kernel. The bandwidth associated with estimator (6) was computed according to the formula (12). The experimental results were obtained from five hundred runs, where in each run, samples of size 150, 300, and 500 were generated for both processes. The samples were drawn from a uniform distribution on the square D_n. In each run a bivariate Gaussian random field with a parsimonius Matérn covariance structure was generated for $\nu_1 = 0.5, \nu_2 = 1.5$, $a = 1$, $\sigma_1 = \sigma_2 = 1$, and $\rho_{12} = 0.3$. The kernel cross-variogram and codispersion coefficient were computed for $\boldsymbol{k} = (\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})$ and $\boldsymbol{k} = (\sqrt{2}, \sqrt{2})$. The results of the simulation study are summarized in Table 2.

There are clear patterns for the performance of the Nadaraya-Watson versions of the cross-variogram and codispersion coefficient (Table 2). As is expected, in both cases, the simulation standard deviation is decreasing when the sample size increases. The same behavior is observed for the bias of the estimations when $\boldsymbol{k} = (\sqrt{2}, \sqrt{2})$. Otherwise, there

Table 2. Mean, standard deviation, and bias associated with estimators (4)-(6) and (7) for different sample sizes.

$\nu_1 = 0.5, \nu_2 = 1.5,\ a = 1,\ \rho_{12} = 0.3,\ \gamma_{XY}(\boldsymbol{k}) = 0.1194278,\ \rho_{XY}(\boldsymbol{k}) = 0.2922169,\ \boldsymbol{k} = (\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})$			
		$n = 150$	
	mean	s. d.	bias
$\breve{\gamma}_{XY_h}(\mathbf{k})$	0.1145731	0.0652170	-0.0048547
$\breve{\rho}_{XY_h}(\mathbf{k})$	0.2809323	0.1353071	-0.0112850
		$n = 300$	
	mean	s. d.	bias
$\breve{\gamma}_{XY_h}(\mathbf{k})$	0.1284656	0.0533544	0.0090378
$\breve{\rho}_{XY_h}(\mathbf{k})$	0.2957279	0.1049523	0.0035110
		$n = 500$	
	mean	s. d.	bias
$\breve{\gamma}_{XY_h}(\mathbf{k})$	0.11846810	0.04569728	-0.0009597
$\breve{\rho}_{XY_h}(\mathbf{k})$	0.28139460	0.09107407	-0.0108220
$\nu_1 = 0.5, \nu_2 = 1.5,\ a = 1,\ \rho_{12} = 0.3,\ \gamma_{XY}(\boldsymbol{k}) = 0.2160805,\ \rho_{XY}(\boldsymbol{k}) = 0.3015092,\ \boldsymbol{k} = (\sqrt{2}, \sqrt{2})$			
		$n = 150$	
	mean	s. d.	bias
$\breve{\gamma}_{XY_h}(\mathbf{k})$	0.1830274	0.1548548	-0.0330530
$\breve{\rho}_{XY_h}(\mathbf{k})$	0.2679031	0.1973018	-0.0336060
		$n = 300$	
	mean	s. d.	bias
$\breve{\gamma}_{XY_h}(\mathbf{k})$	0.2176805	0.1247773	0.0016000
$\breve{\rho}_{XY_h}(\mathbf{k})$	0.2940617	0.1416656	-0.0074480
		$n = 500$	
	mean	s. d.	bias
$\breve{\gamma}_{XY_h}(\mathbf{k})$	0.2163228	0.0997228	0.0002420
$\breve{\rho}_{XY_h}(\mathbf{k})$	0.3006988	0.1121421	-0.0008100

is not a clear pattern for the bias. the kernel estimations of the semi-variograms (not reported here) were not as good as the kernel estimations of the cross-variogram and codispersion coefficient.

Rukhin and Vallejos (2008) studied the performance of the empirical estimator of the codispersion coefficient with respect to the correlation ρ_{12} between the processes. Here, we conducted a Monte Carlo simulation experiment to evaluate the performance of estimators (4), (5), (6), and (7) as a function of ρ_{12}. One hundred and fifty points were randomly generated from a bivariate uniform distribution on the region D_n. Afterward, a single realization from a Gaussian process with a full bivariate Matérn covariance structure defined through (Gneiting et al. 2010)

$$\begin{aligned} C_{11}(\boldsymbol{k}) &= \sigma_1^2 M(\boldsymbol{k}|\nu_1, a_1), \\ C_{22}(\boldsymbol{k}) &= \sigma_2^2 M(\boldsymbol{k}|\nu_2, a_2), \\ C_{12}(\boldsymbol{k}) &= C_{21}(\boldsymbol{k}) = \rho_{12}\sigma_1\sigma_2 M(\boldsymbol{k}, \nu_{12}, a_{12}), \end{aligned}$$

was assigned to each of the selected locations with $\sigma_1 = \sigma_2 = 1$, $a_1 = a_2 = a_{12} = 1$, $\nu_1 = 0.5$, $\nu_2 = 1.5$, and $\nu_{12} = 1$. In each run, the AMISE associated with the estimations (4), (5), (6), and (7) was computed for $\rho = 0; 0.2; 0.4; 0.6; 0.8$. Table 3 shows a decreasing trend of the AMISE associated with the kernel codispersion coefficient as a function of ρ. Moreover, the kernel codispersion coefficient shows superior performance relative

Table 3. AMISE (mean and standard deviation) associated with estimators (4)-(6) and (7) versus ρ_{12}.

ρ_{12}	0	0.2	0.4	0.6	0.8
AMISE$[\widehat{\gamma}_{XY}(\mathbf{k})]$	0.8831 (0.9930)	1.0310 (1.3807)	0.8831 (0.9930)	2.5129 (2.0581)	0.8831 (0.9930)
AMISE$[\breve{\gamma}_{XY_h}(\mathbf{k})]$	1.1498 (1.0453)	1.0614 (1.4014)	0.9287 (1.0453)	2.4981 (1.8343)	3.1784 (2.5377)
AMISE$[\widehat{\rho}_{XY}(\mathbf{k})]$	0.9287 (0.6722)	0.9601 (0.5523)	1.1498 (0.6722)	0.5286 (0.4274)	1.1498 (0.6722)
AMISE$[\breve{\rho}_{XY_h}(\mathbf{k})]$	1.0712 (0.7048)	0.8795 (0.5501)	1.0712 (0.7048)	0.4812 (0.4203)	0.1766 (0.0824)

to the empirical estimator for $\rho \geq 0.2$. The standard deviations of both estimators of the codispersion are comparable except for $\rho = 0.8$. There are no clear patterns for the performance of the kernel cross-variogram relative to the empirical estimator. These results are consistent with the results reported by Rukhin and Vallejos (2008) and highlights the importance of the semi-variogram estimation on the codispersion.

We carried out a third Monte Carlo simulation study to explore the performance of estimators (4), (5), (6), and (7) when the smoothness parameters of the correltion function vary. Under the same conditions as in the previous simulation studies five hundred simulation runs were computed and associated to the locations that are on the region D_n, $n = 150$. The random observtions were drawn from a bivariate Gaussian distribution with a parsimonious Matérn correlation structure. Three sets of smoothing parameters were considered: $\nu_1 = \nu_2 = 0.5$; $\nu_1 = 0.5, \nu_2 = 2.5$; and $\nu_1 = \nu_2 = 2.5$. For each run the mean value and the stardard deviation of the AMISE were recorded. The average values and the standard deviations of the AMISE over these replicates can be found in Table 4.

The performance of the kernel codispersion coefficient is better than that of the empirical estimator except for one case ($\nu_1 = 0.5, \nu_2 = 0.5$) in which the standard deviation of the empirical estimator is smaller than the standard deviation of the kernel codispersion coefficient. In general, we observe that the mean values and standard deviations associated with all estimations do not change drastically when $\nu_1 = \nu_2$.

Putting together all the information collected from the three simulations studies reported in this section, we think that the kernel estimation of the codispersion is a good alternative to the empirical estimator with overall better performance under the setup described above.

6. An Application to forest variables

In this section, we illustrate how the proposed nonparametric approach can be used to estimate the codispersion coefficient for the four forest variables previously described in Section 2. The sample locations where the observations were collected are shown in Figure 1. The goal of this approach is to estimate the spatial relationship between all possible pairs of the variables of basal area, height, elevation, and slope. The first two variables characterize the growing process of the Pinus radiata and are specifically associated to the trees. The last two (environmental) variables are related to the terrain. Figure 2 shows a simple bilinear interpolation and the corresponding contours for the four variables. To use the proposed rule (12) to compute the bandwidth for the semi-variograms and cross-variogram, the parameters of the density $f_1(\mathbf{0})$ need to be estimated. To acomplish this, consider the distribution

$$f_0(x, y) = \frac{1}{(b-a)(d-c)} \mathbb{1}_B(x, y), \tag{14}$$

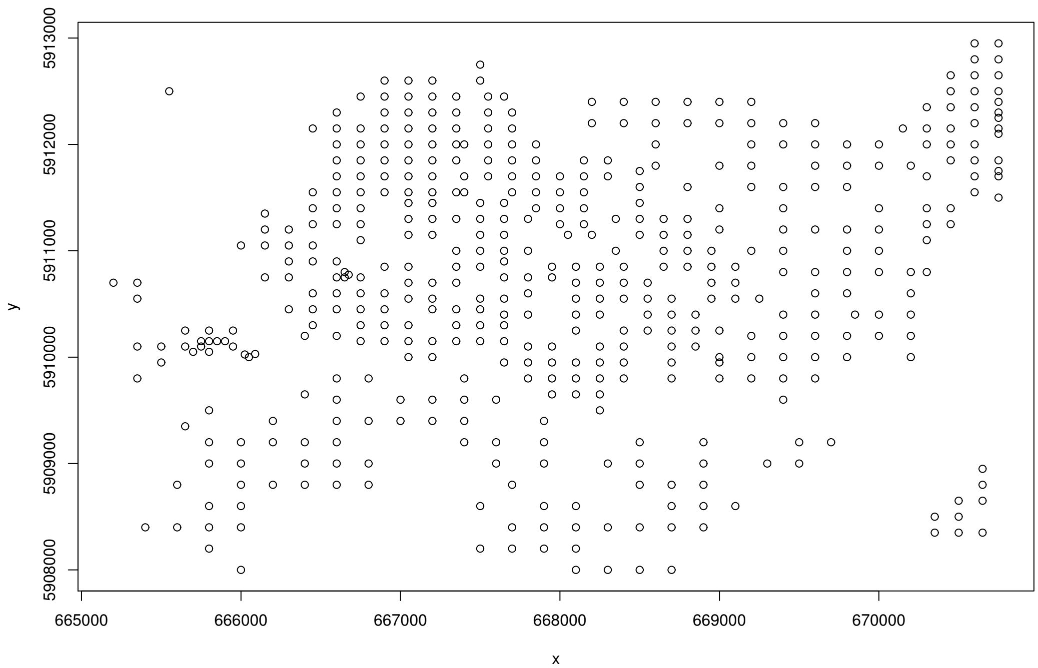


Figure 1. 468 sample locations where the variables were collected in the south of Chile (36°54' S, 73°54' O).

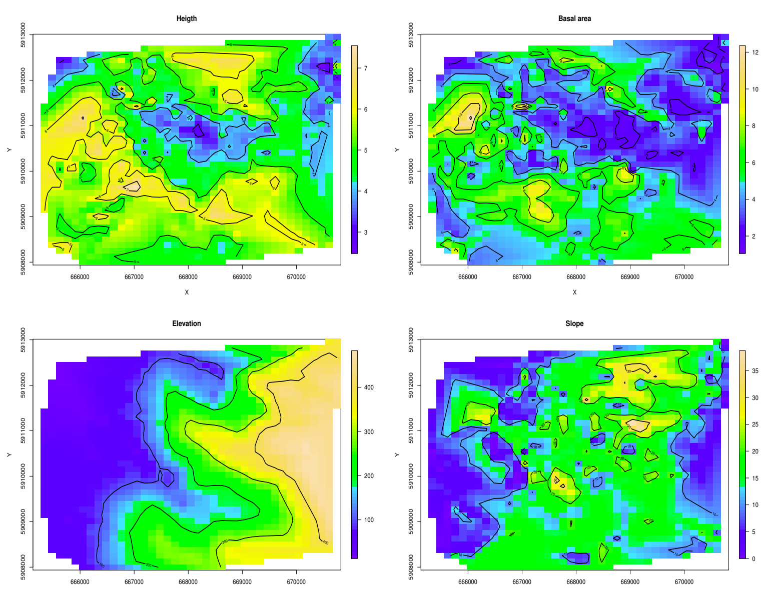


Figure 2. Bilinear interpolation of the variables under study.

Table 4. AMISE (mean and standard deviation) associated with estimators (4)-(6) and (7) for different values of the smoothness parameters.

$a_1 = a_2 = a_{12} = 1$, $\rho_{12} = 0.3$, $n = 150$, $\nu_{12} = \frac{1}{2}(\nu_1 + \nu_2)$			
		ν_1	ν_2
AMISE$[\widehat{\gamma}_{XY}(\mathbf{k})]$	0.853115 (0.5698374)	0.5	0.5
AMISE$[\breve{\gamma}_{XY_h}(\mathbf{k})]$	0.904262 (0.5286199)	0.5	0.5
AMISE$[\widehat{\rho}_{XY}(\mathbf{k})]$	0.721578 (0.4374963)	0.5	0.5
AMISE$[\breve{\rho}_{XY_h}(\mathbf{k})]$	0.689808 (0.4110937)	0.5	0.5
		ν_1	ν_2
AMISE$[\widehat{\gamma}_{XY}(\mathbf{k})]$	1.0735390 (0.8620709)	0.5	2.5
AMISE$[\breve{\gamma}_{XY_h}(\mathbf{k})]$	1.0720410 (0.8190623)	0.5	2.5
AMISE$[\widehat{\rho}_{XY}(\mathbf{k})]$	0.9287875 (0.4988746)	0.5	2.5
AMISE$[\breve{\rho}_{XY_h}(\mathbf{k})]$	0.8404175 (0.5141708)	0.5	2.5
		ν_1	ν_2
AMISE$[\widehat{\gamma}_{XY}(\mathbf{k})]$	0.8531148 (0.5698374)	2.5	2.5
AMISE$[\breve{\gamma}_{XY_h}(\mathbf{k})]$	0.9042624 (0.5286201)	2.5	2.5
AMISE$[\widehat{\rho}_{XY}(\mathbf{k})]$	0.7215775 (0.4374963)	2.5	2.5
AMISE$[\breve{\rho}_{XY_h}(\mathbf{k})]$	0.6898077 (0.4110935)	2.5	2.5

where $B = \{(x, y) \in \mathbb{R}^2 : a \leq x \leq b \text{ y } c \leq y \leq d\}$. A straightforward calculcation shows that that if (X_1, Y_1) and (X_2, Y_2) are two independent and identically distributed random vectors with density $f_0(\cdot, \cdot)$, then the density of the vector $(X_1 - X_2, Y_1 - Y_2)$ is given by

$$f(w_1, w_2) = \frac{T_{a,b}(w_1)T_{a,b}(w_2)}{(a-b)^2(c-d)^2},$$

where

$$T_{a,b}(t) = \begin{cases} 0, & t \leq a - b \\ t + (b - a), & a - b \leq t \leq \\ (b - a) - t, & 0 < t \leq b - a \\ 0, & t \geq b - a. \end{cases}$$

Then,

$$f_1(0, 0) = \frac{1}{(b-a)(d-c)}.$$

Denoting the coordinates by $(\lambda^{-1}x, \lambda^{-1}y)$, the maximum likelihood estimations of the parameters of (14) are $a = \min_i\{\lambda^{-1}x_i\}$, $b = \max_i\{\lambda^{-1}x_i\}$, $c = \min_i\{\lambda^{-1}y_i\}$, and $d = \max_i\{\lambda^{-1}y_i\}$. Replacing the available data, we obtained $a = 59381.69, b = 59828.03, c = 527400.8$, and $d = 527824.8$. To fit semi-variograms to each variable, estimator (2) was used with a separable kernel of the form $K_2(x, y) = K(x)K(y)$, where $K(\cdot)$ is the Epanechnikov kernel. The bandwidth values obtained by using (11) for the variograms and using (12) for the cross-variogram are $\widehat{h}_X = 173.92$ (basal area), $\widehat{h}_X = 247.30$ (height), $\widehat{h}_X = 88.43$ (slope), $\widehat{h}_X = 242.62$ (elevation), $\widehat{h}_{XY} = 187.34$ (basal area-height), $\widehat{h}_{XY} = 94.75$ (basal area-slope), $\widehat{h}_{XY} = 242.52$ (basal area-elevation), $\widehat{h}_{XY} = 231.91$

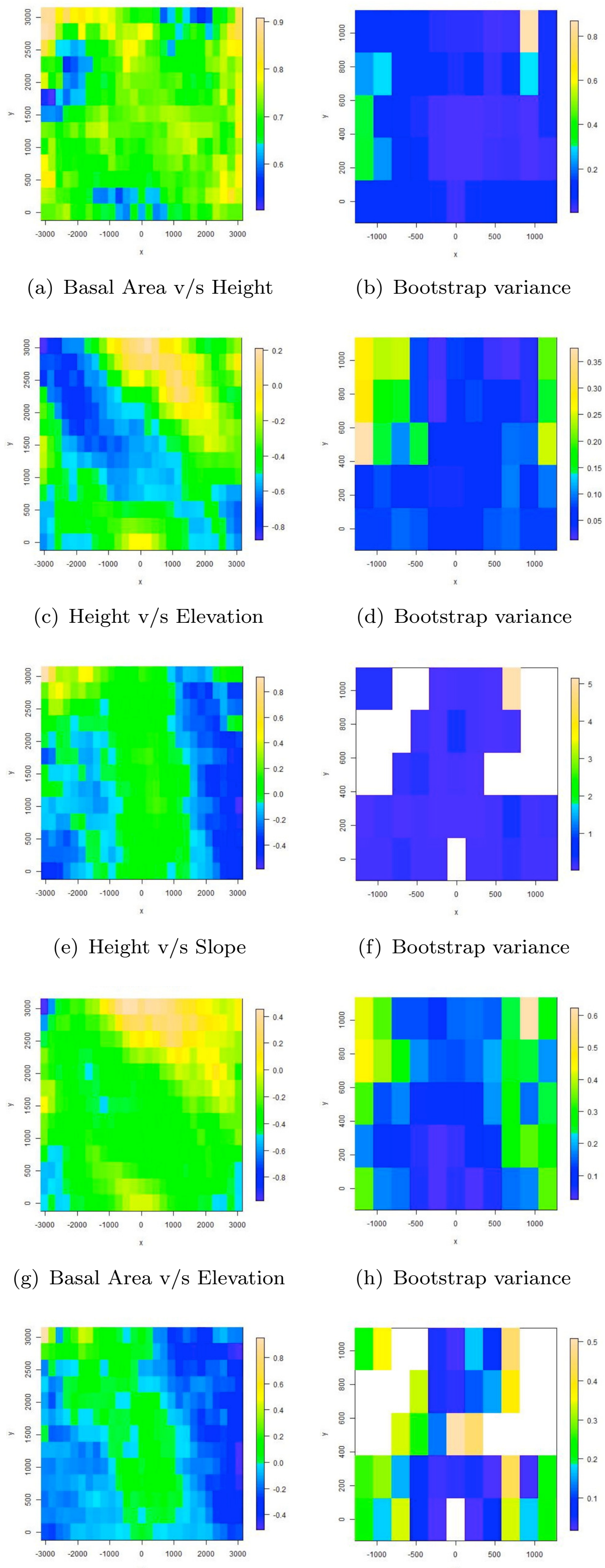
(height-slope), $\widehat{h}_{XY} = 246.98$ (height-elevation). As with the spatial analysis of one variable, here, we constructed a codispersion map to provide better insight into the spatial associations between all pairs of variables in several different directions on a two-dimensional space. Figure 3 shows the codispersion maps that were created from the variables of interest. Note from Figure 3(a) that the codispersion coefficient between basal area and height is high (greater than 0.6) in all directions. The smallest values values are obtained in all directions for a small lag distance. The largest values occur for lag distances bigger than 3000 meters for the angles between 45°and 135°. Figure 3(c) shows that the codispersion coefficients between height and elevation are negatively correlated (-0.6) for angles between 90°and 155°. The same pattern can be observed for the codispersion coefficient for angles between 0°and 90°and lag distances greater than 2000 meters. This is consistent with the terrain features, specifically with the existence of a river that passes through the lower areas of the study region. Hence, associations between large tree heights and small elevations are expected. The values of the codispersion coefficient between the variables height and slope show that the spatial association is very small for an angle of 90°at any distance lag. There is clear vertical banding in the middle of the codispersion map (Figure 3(e)) highlighting the small values of the association between these variables in the center of the image. We also find a weak inverse association (-0.3) at the edges of the codispersion map, for lag distances greater than 2000 meters forming another vertical pattern for an angle of 90°. Figure 3(g) shows an inverse correlation (-0.4) along the 135°line. The spatial association for lag distances greater than 3000 meters and for directiond between 0 °ans 90 °is is negligible. Figure 3(i) shows small spatial association along the 90°line at aby distance lag. Negative spatial association is observed for lag distances greater than 1000 meters for an angle of 90°. Again, certain patterns are present at the edges of the image in the same way as we pointed out for the variables height and slope. The codispersion map between elevation and slope is not shown here because the main goal is to study the spatial association between environmental and terrain variables.

In order to provide a measure of uncertainty for the codispersion coefficient for the forestry data analyzed in this section, the block bootstrap introduced by Sherman (1996) for statistics computed from a spatial lattice was implemented. The spatial locations shown in Figure 1 were divided into six non overlapping blocks to ensure that in each block will be enough points to estimate the variance of the codispersion coefficient for distance lags less than 1187.69 meters, which corresponds to the mimimum distance over all the maximum distances between points computed from the six blocks considered in this study. The bootstrap variance associated with the codispersion coefficients between the pairs of variables of interest was computed following the guidelines given in Sherman (1996), the results are displayed in Figure 3. Values of the variance for those directions in which the number of points were not enough to get an estimation were labeled with white color as is shown in Figure 3(f) and 3(j).

Using the point estimations and variances of the codispersion coefficient provided above, further statistical inference can be conducted in addition to the analysis carried out in this paper.

7. Discussion and conclusions

We have introduced a nonparametric approach for estimating the codispersion between two spatial processes. The proposal is a Nadaraya-Watson type estimator, which is a



(a) Basal Area v/s Height (b) Bootstrap variance

(c) Height v/s Elevation (d) Bootstrap variance

(e) Height v/s Slope (f) Bootstrap variance

(g) Basal Area v/s Elevation (h) Bootstrap variance

(i) Basal Area v/s Slope (j) Bootstrap variance

Figure 3. Codispersion map between all pairs of variables of interest, and the corresponding block bootstrap variance computed using Sherman's approach.

ratio between a kernel estimation of the cross-variogram and the square root of the semivariograms. The introduction of this type of estimator is relevant for attempts to quantify spatial associations between two spatial sequences, such as the forest variables included in this paper.

The present study was supported by asymptotic results that led to the consistency of the proposed estimator. Under regularity conditions, explicit expressions for the bias and mean square error were developed for the cross-variogram and the consistency of the kernel estimator of the codispersion was established. We also derive a rule to compute the bandwidth for the kernel variogram and suggest a simple way to compute the bandwidth associated with the kernel cross-variogram, both proposals were supported by theoretical results.

The finite sample simulation results described in Section 5 showed that the kernel estimator of the codispersion performs better than the widely used empirical estimator of the codispersion first introduced by Matheron. After considering several different scenarios (not shown here), the kernel estimator of the codispersion was not sensitive to the choice of the kernel function. The computation of the kernel estimator is very intensive and demanding. The study of efficient computational routines specially adapted for large data sets is necessary and is a matter for further research.

The study of the asymptotic distribution of the kernel estimator of the codispersion was not covered in this paper, but is required to establish confidence limits of the codispersion for a given lag distance. Although the extension of this result is not straightforward for the case of the codispersion, the additional assumptions required to obtain asymptotic normality of the kernel estimator of the semivariogram are restrictive. Instead, for the practical example discussed in Section 6, the computation of the variance of the coefficient was approached by existing block bootstrap techniques, developed for statistics computed from processes that are measured over rectangular grids.

The codispersion map developed for the forest variables studied in Section 6 highlights the wide range of possibilities for researchers working with two or more spatial variables measured at the same locations. The spatial association between two processes can be computed and visualized in the same plot for different directions, highlighting the smallest and largest values on two-dimensional space. Specifically, the spatial relationships found for the forest variables are of interest for people who perform forest inventories. Several related problems arise from a correlation analysis of these variables. For example, it is not clear how to decrease the effective sample size as a function of correlation in later studies, or how to use the information of the process $X(\cdot)$ to study the process $Y(\cdot)$. In this context, the nonparametric codispersion coefficient may help to answer these questions by quantifying the spatial association as a function of the lag distance, in two-dimensional space. When working with more than two spatial variables, a codispersion matrix may be defined to study problems analogous to dimension reduction techniques, such as principal components analysis. These topics are a matter for future research.

Finally, R-code to compute the kernel estimation of the codispersion described in the article has been developed. A brief description of the routine and examples can be found in Web appendix B.

Acknowledgements

Francisco Cuevas was supported by DGIP, UTFSM and Fondecyt grant N°11075095. Ronny O. Vallejos was partially supported by UTFSM grant 12.10.03, and Fondecyt grant N°1120048. The authors are grateful to Marcelo Miranda, from PUC, Chile, for providing the dataset that was used in the application discussed in Section 6. We are grateful to two anonymous referees and one member of the editorial board for their comments and suggestions which led to an improved presentation.

References

Chilés, J.P., Delfiner, P. (1999), Geostatistics: Modeling Spatial Uncertainty, New York: Wiley.

Cliff, A., Ord, J. (1981), Spatial Processes: Models and Applications, London: Pion Ltd.

Clifford, P., Richardson, S., and Hémon, D. (1989), "Assessing the significance of the correlation between two spatial processes", *Biometrics*, 45, 123–134.

Cressie, N., Hawkins, D. (1980) "Robust estimation of the variogram: I", *Mathematical Geology*, 2, 115–125.

Cressie, N. (1993), Statistics for spatial data, New York: Wiley.

Croux, C., Forni, M., and Reichlin, L. (2001), "A measure of comovement for economic variables: theory and empirics", *Review of Economics and Statistics*, 83, 232–241.

Dutilleul, P. (1993), "Modifying the t test for assessing the correlation between two spatial processes", *Biometrics*, 49, 305–314.

Dutilleul, P., Pelletier, B., and Alpargu, G. (2008), "Modified F tests for assessing the multiple correlation between one spatial process and several others", *Journal of Statistical Planning and Inference*, 138, 1402–1415.

Fotheringham, S., Rogerson, P. (2009), The SAGE Handbook of Spatial analysis, London: SAGE.

García-Soidán P., Febrero, M., and Gonzáles, W. (2004), "Nonparametric kernel estimation of an isotropic variogram", *Journal of Statistical Planning and Inference*, 121, 65–92.

García-Soidán P. (2007), "Asymptotic normality of the Nadaraya-Watson semivariogram estimators", *Test*, 16, 479–503.

Genton, M.G. (1998), "Highly robust variogram estimation", *Mathematical Geology*, 30, 213–221.

Gneiting, T., Kleiber, W. and Schlather, M. (2010), "Matérn cross-covariance functions for multivariate random fields", *Journal of the American Statistical Association*, 105, 1167–1177.

Goovaerts, P. (1994), "Study of spatial relationships between two sets of variables using multivariate geostatistics, *Geoderma*, 62, 93–107.

Goovaerts, P. (1997), Geostatistics for Natural Resources Evaluation, Oxford: Oxford University Press.

Goovaerts, P. (1998), "Ordinary kriging revisited", *Mathematical Geology*, 30, 21–42.

Hallin, M., Lu, Z., and Tran, L.T. (2004), "Kernel density estimation for spatial processes: the L_1 theory", *Journal of Multivariate Analysis*, 88, 61–75.

Haining, R. (1991), "Bivariate correlation with spatial data", *Geographical Analysis*, 23, 210–227.

Hardy, G. H, Littlewood, J.E., Pólya, G. (1934), Inequalities, Cambridge: Cambridge Univ. Press.

Hubert, L. J., Golledge, R.G. (1982), "Measuring association between spatially defined variables: Tjostheim index and some extensions", *Geographical Analysis*, 4, 273–278.

Kohler, M., Schindler, A., Sperlich, S. (2011), "A Review and Comparison of Bandwidth

Selection Methods for Kernel Regression" *Courant Research Centre: Poverty, Equity and Growth- Discussion Papers*, 95, 1–37.

Li, Q., Racine, J.S. (2006), Non Parametric Econometrics, New York: Princeton University Press.

Matheron, G. (1963), Principles of geostatistics, *Economic geology*, 58, 1246–1266.

Matheron, G. (1965), Les Variables Régionalisées et leur Estimation, Paris: Masson.

Porcu, E., Mateu, J., and Cristakos, G. (2009). "Quasi-arithmetic means of covariance functions with potential applications to space-time data", *Journal of Multivariate Analysis*, 100, 1830–1844.

R Development Core Team (2012), R: A language and environment for statistical computing, *R Foundation for Statistical Computing*, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Richardson, S., Clifford, P. (1991), "Testing association between spatial processes", In *spatial statistics and imaging*, ed. A. Possolo, IMS Lecture Notes 20, 295–308.

Richardson, S., Guihenneuc, C., and Lasserre, V. (1992), "Spatial linear models with autocorrelated error structure", *The Statistician*, 41, 539–539.

Richardson, S., Guihenneuc-Jouyaux, C. (2009), "Impact of Cliff and Ord (1969-1989) on spatial epidemiology", *Geographical Analysis*, 41, 444–451.

Ruppert, D., Sheather, S.J., and Wand, M.P. (1995), "An Effective Bandwidth Selector for Local Least Squares Regression", *Journal of the American Statistical Association*, 90, 1257–1270.

Rukhin, A., Vallejos, R. (2008), "Codispersion coefficient for spatial and temporal series", *Statistics and Probability Letters*, 78, 1290–1300.

Sherman, M. (1996), "Variance estimation for statistics computed from spatial lattice data", *Journal of the Royal Statistical Society*, 58, 509-523.

Snowdon, P. (2001), "Short-term predictions of growth of Pinus radiata with models incorporating indices of annual climatic variation", *Forest Ecology and Management*, 152, 1–11.

Stein, M. L. (2007), " Space-time covariance functions", *Journal of the American Statistical Association*, 100, 310-321.

Tjostheim, D. (1978), "A measure of association for spatial variables", *Biometrika*, 65, 109–114.

Vallejos, R. (2008), "Assessing the association between two spatial or temporal sequences", *Journal of Applied Statistics*, 35, 1323–1343.

Ver Hoef, J.M., Barry, R.P. (1998), "Constructing and fitting models for cokriging and multivariate spatial prediction", *Journal of Statistical Planning and Inference*, 69, 275–294.

Yu, K., Mateu, J., and Porcu, E. (2007), "A kernel-based method for nonparametric estimation of variograms",*Statistica Neerlandica*, 61, 173-197.

Appendix A. Proof of the Results in Section 3 and 4

Proof of Theorem 3.1

According to Assumption (A3) the spatial locations are given by $\boldsymbol{s}_i = \lambda \boldsymbol{u}_i$. Following Garcia-Soidan (2007), let us write $\boldsymbol{U}_{i,j} = \boldsymbol{U}_i - \boldsymbol{U}_j$ and define the following quantities

$$
\begin{aligned}
a_1(\mathbf{k}) &= \sum_{i \neq j} K\left(\frac{\mathbf{k} - \lambda \boldsymbol{U}_{i,j}}{h}\right),\\
a_2(\mathbf{k}) &= \sum_{i \neq j} K\left(\frac{\mathbf{k} - \lambda \boldsymbol{U}_{i,j}}{h}\right)(\gamma_{XY}(\lambda \boldsymbol{U}_{i,j}) - \gamma_{XY}(\mathbf{s})),\\
a_3(\mathbf{k}) &= \sum_{i \neq j} \left(K\left(\frac{\mathbf{k} - \lambda \boldsymbol{U}_{i,j}}{h}\right)\right)^2 \psi(\lambda \boldsymbol{U}_{i,j}, \mathbf{0}, \lambda U_{i,j}),\\
a_4(\mathbf{k}) &= \sum_{(i,j,k) \in B_1} K\left(\frac{\mathbf{k} - \lambda \boldsymbol{U}_{i,j}}{h}\right) K\left(\frac{\mathbf{k} - \lambda \boldsymbol{U}_{i,k}}{h}\right) \psi(\lambda \boldsymbol{U}_{i,j}, \mathbf{0}, \lambda \boldsymbol{U}_{i,k}),\\
a_5(\mathbf{k}) &= \sum_{(i,j,k,l) \in B_2} K\left(\frac{\mathbf{k} - \lambda \boldsymbol{U}_{i,j}}{h}\right) K\left(\frac{\mathbf{k} - \lambda \boldsymbol{U}_{k,l}}{h}\right) \psi(\lambda \boldsymbol{U}_{i,j}, \lambda \boldsymbol{U}_{i,k}, \lambda \boldsymbol{U}_{i,l}),
\end{aligned}
$$

where $B_1 = \{(i,j,k) : i \neq j, k \text{ y } j \neq k\}$, and $B_2 = \{(i,j,k,l) : i \neq j,k,l \text{ y } j \neq k,l \text{ y } k \neq l\}$. Then almost surely for $\mathbf{s} \neq 0$ we obtain

$$
\begin{aligned}
a_1(\mathbf{k}) &= f_1(\mathbf{0}) n^2 \lambda^{-d} h^d + o(n^2 \lambda^{-d} h^d),\\
a_2(\mathbf{k}) &= \frac{1}{2} f_1(\mathbf{0}) c_K \ \triangle \gamma_{XY}(\boldsymbol{k}) \ n^2 \lambda^{-d} h^{d+2} + o(n^2 \lambda^{-d} h^{d+2}),\\
a_3(\mathbf{k}) &= f_1(\mathbf{0}) d_K \psi(\mathbf{k}, \mathbf{0}, \mathbf{k}) n^2 \lambda^{-d} h^d + o(n^2 \lambda^{-d} h^d),\\
a_4(\mathbf{k}) &= f_2(\mathbf{0}, \mathbf{0}) \psi(\mathbf{k}, \mathbf{0}, \mathbf{k}) n^3 \lambda^{-2d} h^{2d} + o(n^3 \lambda^{-2d} h^{2d}),\\
a_5(\mathbf{k}) &= f_3(\mathbf{0}, \mathbf{0}, \mathbf{0}) n^4 \lambda^{-3d} h^{2d} \int_{\mathbb{R}^d} \psi(\mathbf{k}, \boldsymbol{\xi}, \boldsymbol{\xi} + \mathbf{k}) d\boldsymbol{\xi} + o(n^4 \lambda^{-3d} h^{2d}).
\end{aligned}
$$

Then, the bias is

$$
\begin{aligned}
\mathbb{E}[\breve{\gamma}_{XY_h}(\boldsymbol{k})] - \gamma_{XY}(\boldsymbol{k}) &= \mathbb{E}[\mathbb{E}[\breve{\gamma}_{XY_h}(\boldsymbol{k}) / \boldsymbol{U}_1, \ldots, \boldsymbol{U}_n]] - \gamma_{XY}(\boldsymbol{k}),\\
&= \mathbb{E}\left[\frac{a_2(\boldsymbol{k})}{a_1(\boldsymbol{k})}\right],\\
&= \frac{c_{K_d}}{2} \triangle \gamma_X(\boldsymbol{k}) \ h^2 + o(h^2). \qquad \text{(A1)}
\end{aligned}
$$

Similarly, using the convergence orders given for $a_1(\boldsymbol{k}) - a_5(\boldsymbol{k})$ we have

$$
Var[\breve{\gamma}_{XY_h}(\boldsymbol{k})] = \mathbb{E}[Var[\breve{\gamma}_{XY_h}(\boldsymbol{k}) / \boldsymbol{U}_1, \ldots, \boldsymbol{U}_n]] + Var[\mathbb{E}[\breve{\gamma}_{XY_h}(\boldsymbol{k}) / \boldsymbol{U}_1, \ldots, \boldsymbol{U}_n]].
$$

Now, from equation(A1) one obtains $Var[\mathbb{E}[\hat{\gamma}_{XY_h}(\boldsymbol{k})/\boldsymbol{U}_1,\ldots,\boldsymbol{U}_n]] = o(h^4)$. Finally,

$$\begin{aligned}
\mathbb{E}[Var[\breve{\gamma}_{XY_h}(\boldsymbol{k})/\boldsymbol{U}_1,\ldots,\boldsymbol{U}_n]] &= \frac{\mathbb{E}[2a_3(\boldsymbol{k}) + 4a_4(\boldsymbol{k}) + a_5(\boldsymbol{k})]}{4a_1^2(\boldsymbol{k})} \\
&= \frac{d_K A_{1,d}(\mathbf{k})}{2f_1(\mathbf{0})} n^{-2}\lambda^d h^{-d} + \frac{f_3(\mathbf{0},\mathbf{0},\mathbf{0})A_{2,d}(\mathbf{k})}{(2f_1(\mathbf{0}))^2}\lambda^{-d} \\
&+ o(n^{-2}\lambda^d h^{-d} + \lambda^{-d}).
\end{aligned}$$

■

An Example in Which Conditions (H1) and (H2) Are Satisfied

Consider a $2n-$variate Gaussian random vector $(X(\mathbf{s}_1),\ldots,X(\mathbf{s}_n),Y(\mathbf{s}_1),\ldots,Y(\mathbf{s}_n))^T$ with mean $\mathbf{0}$ and covariance matrix

$$\Sigma = \begin{pmatrix} \Sigma_1 & \Sigma_{12} \\ \Sigma_{21} & \Sigma_2 \end{pmatrix},$$

where $(\Sigma_1)_{ij} = \sigma_X^2 - \gamma_X(\mathbf{s}_i - \mathbf{s}_j)$, $(\Sigma_2)_{ij} = \sigma_Y^2 - \gamma_Y(\mathbf{s}_i - \mathbf{s}_j)$ for all $i,j = 1,\ldots,n$, $\gamma_X(\infty) < \infty$, and $\gamma_Y(\infty) < \infty$. Let us denote the cross-covariance function between $X(\cdot)$ and $Y(\cdot)$ as $C_{XY}(\boldsymbol{s}_i - \boldsymbol{s}_j)$ such that the cross-covariance matrix $(\Sigma_{12})_{ij} = \sigma_{XY} - C_{XY}(\mathbf{s}_i - \mathbf{s}_j)$ for all $i,j = 1,\ldots,n$, and assume that $\Sigma_{21} = \Sigma_{12}^T$. By properties of the covariance function we can write

$$\begin{aligned}
&Cov[(X(\boldsymbol{s}_i) - X(\boldsymbol{s}_j))(Y(\boldsymbol{s}_i) - Y(\boldsymbol{s}_j)), (X(\boldsymbol{s}_k) - X(\boldsymbol{s}_l))(Y(\boldsymbol{s}_k) - Y(\boldsymbol{s}_l))] = \\
&\mathbb{E}(X(\mathbf{s}_i)X(\mathbf{s}_k)Y(\mathbf{s}_i)Y(\mathbf{s}_k) - X(\mathbf{s}_j)X(\mathbf{s}_k)Y(\mathbf{s}_i)Y(\mathbf{s}_k) - X(\mathbf{s}_i)X(\mathbf{s}_l)Y(\mathbf{s}_i)Y(\mathbf{s}_k) \\
&+ X(\mathbf{s}_j)X(\mathbf{s}_l)Y(\mathbf{s}_i)Y(\mathbf{s}_k) - X(\mathbf{s}_i)X(\mathbf{s}_k)Y(\mathbf{s}_j)Y(\mathbf{s}_k) + X(\mathbf{s}_j)X(\mathbf{s}_k)Y(\mathbf{s}_j)Y(\mathbf{s}_k) \\
&+ X(\mathbf{s}_i)X(\mathbf{s}_l)Y(\mathbf{s}_j)Y(\mathbf{s}_k) - X(\mathbf{s}_j)X(\mathbf{s}_l)Y(\mathbf{s}_j)Y(\mathbf{s}_k) - X(\mathbf{s}_i)X(\mathbf{s}_k)Y(\mathbf{s}_i)Y(\mathbf{s}_l) \\
&+ X(\mathbf{s}_j)X(\mathbf{s}_k)Y(\mathbf{s}_i)Y(\mathbf{s}_l) + X(\mathbf{s}_i)X(\mathbf{s}_l)Y(\mathbf{s}_i)Y(\mathbf{s}_l) - X(\mathbf{s}_j)X(\mathbf{s}_l)Y(\mathbf{s}_i)Y(\mathbf{s}_l) \\
&+ X(\mathbf{s}_i)X(\mathbf{s}_k)Y(\mathbf{s}_j)Y(\mathbf{s}_l) - X(\mathbf{s}_j)X(\mathbf{s}_k)Y(\mathbf{s}_j)Y(\mathbf{s}_l) - X(\mathbf{s}_i)X(\mathbf{s}_l)Y(\mathbf{s}_j)Y(\mathbf{s}_l) \\
&+ X(\mathbf{s}_j)X(\mathbf{s}_l)Y(\mathbf{s}_j)Y(\mathbf{s}_l)) \\
&- \mathbb{E}((X(\boldsymbol{s}_i) - X(\boldsymbol{s}_j))(Y(\boldsymbol{s}_i) - Y(\boldsymbol{s}_j)))\mathbb{E}((X(\boldsymbol{s}_k) - X(\boldsymbol{s}_l))(Y(\boldsymbol{s}_k) - Y(\boldsymbol{s}_l))).
\end{aligned}$$

In the Gaussian case the following identity holds

$$\mathbb{E}(X(\mathbf{s}_i)X(\mathbf{s}_j)X(\mathbf{s}_k)X(\mathbf{s}_l)) = \sigma_{ij}\sigma_{kl} + \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk},$$

where σ_{ij} corresponds to the covariance between $X(\mathbf{s}_i)$ and $X(\mathbf{s}_j)$. Then a straightforward calculation shows that

$$\begin{aligned}
&\psi(\mathbf{s}_i-\mathbf{s}_j,\mathbf{s}_i-\mathbf{s}_k,\mathbf{s}_i-\mathbf{s}_l) = \gamma^2_{XY}(\mathbf{s}_i-\mathbf{s}_k)+\gamma^2_{XY}(\mathbf{s}_j-\mathbf{s}_k)+\gamma^2_{XY}(\mathbf{s}_i-\mathbf{s}_l)+\gamma^2_{XY}(\mathbf{s}_j-\mathbf{s}_l)\\
&-4\sigma_{XY}+\frac{1}{2}\sqrt{g_X(\mathbf{s}_i-\mathbf{s}_j,\mathbf{s}_i-\mathbf{s}_k,\mathbf{s}_i-\mathbf{s}_l)g_Y(\mathbf{s}_i-\mathbf{s}_j,\mathbf{s}_i-\mathbf{s}_k,\mathbf{s}_i-\mathbf{s}_l)}\\
&+C_{XY}(\mathbf{s}_j-\mathbf{s}_k)C_{XY}(\mathbf{s}_l-\mathbf{s}_i)+C_{XY}(\mathbf{s}_i-\mathbf{s}_k)C_{XY}(\mathbf{s}_l-\mathbf{s}_j)\\
&+C_{XY}(\mathbf{s}_j-\mathbf{s}_l)C_{XY}(\mathbf{s}_i-\mathbf{s}_k)+C_{XY}(\mathbf{s}_i-\mathbf{s}_l)C_{XY}(\mathbf{s}_j-\mathbf{s}_k),
\end{aligned}$$

where g_X and g_Y correspond to the g function given in (A7) for the processes $X(\cdot)$ and $Y(\cdot)$ respectively. Thus Assumption (H1) is satisfied.

To inspect Assumption (H2) we consider again the Gaussian case. By using the inequalities $|\gamma_{XY}(\mathbf{k})|^2 \le \gamma_X(\mathbf{k})\gamma_Y(\mathbf{k})$ and $|C_{XY}(\mathbf{k})|^2 \le C_X(\mathbf{0})C_Y(\mathbf{0})$, it can be shown that

$$\begin{aligned}
&\psi(\mathbf{s}_i-\mathbf{s}_j,\mathbf{s}_i-\mathbf{s}_k,\mathbf{s}_i-\mathbf{s}_l) \le 4C_X(\mathbf{0})C_Y(\mathbf{0})+\gamma_X(\mathbf{s}_i-\mathbf{s}_k)\gamma_Y(\mathbf{s}_i-\mathbf{s}_k)\\
&+\gamma_X(\mathbf{s}_j-\mathbf{s}_k)\gamma_Y(\mathbf{s}_j-\mathbf{s}_k)+\gamma_X(\mathbf{s}_i-\mathbf{s}_l)\gamma_Y(\mathbf{s}_i-\mathbf{s}_l)+\gamma_X(\mathbf{s}_j-\mathbf{s}_l)\gamma_Y(\mathbf{s}_j-\mathbf{s}_l)-4\sigma_{XY}\\
&+\frac{1}{2}\sqrt{g_X(\mathbf{s}_i-\mathbf{s}_j,\mathbf{s}_i-\mathbf{s}_k,\mathbf{s}_i-\mathbf{s}_l)g_Y(\mathbf{s}_i-\mathbf{s}_j,\mathbf{s}_i-\mathbf{s}_k,\mathbf{s}_i-\mathbf{s}_l)}. \qquad \text{(A2)}
\end{aligned}$$

Then, integrating both sides of (A2) over the region $L(x)$, the condition (H2) is obtained.

Proof of Theorem 3.2

Under the Assumptions of the Theorem the following is true: $\breve{\gamma}_{X_{h_2}}(\mathbf{k})$ converges in probability to $\gamma_X(\boldsymbol{k})$, $\breve{\gamma}_{Y_{h_3}}(\mathbf{k})$ converges in probability to $\gamma_Y(\boldsymbol{k})$, and $\breve{\gamma}_{XY_{h_1}}(\boldsymbol{k})$ converges in probability to $\gamma_{XY}(\boldsymbol{k})$. Since $\sqrt{xy}$ is a continuous function, it follows that $\sqrt{\breve{\gamma}_{X_{h_2}}(\mathbf{k})\breve{\gamma}_{Y_{h_3}}(\mathbf{k})}$ converges in probability to $\gamma_X(\boldsymbol{k})\gamma_Y(\boldsymbol{k})$. Because $\breve{\gamma}_{XY_{h_1}}(\boldsymbol{k})$ converges in probability to $\gamma_{XY}(\boldsymbol{k})$ (Theorem 1), we obtain $\breve{\rho}_{XY_h}(\mathbf{k}) = \frac{\breve{\gamma}_{XY_{h_1}}(\mathbf{k})}{\sqrt{\breve{\gamma}_{X_{h_2}}(\boldsymbol{k})\breve{\gamma}_{Y_{h_3}}(\mathbf{k})}} \xrightarrow{\mathcal{P}} \rho_{XY}(\boldsymbol{k})$. ∎

Proof of Theorem 4.1

A straightforward calculation shows that

$$\frac{\partial \mathrm{AMISE}[\breve{\gamma}_{X_h}(\boldsymbol{k})]}{\partial h} = 4C_1h^3 - dC_2h^{-(d+1)}.$$

Since AMISE$[\breve{\gamma}_{X_h}(\boldsymbol{k})]$ is a convex function with respect to h we can find the minimum by solving the equation $\frac{\partial AMISE[\breve{\gamma}_{X_h}(\boldsymbol{k})]}{\partial h} = 0$. This implies that $4C_1h^3 = dC_2h^{-(d+1)}$. Thus $h_{AMISE} = \left(\frac{d}{4}\frac{C_2}{C_1}\right)^{\frac{1}{d+4}}$. ∎

Proof of Proposition 4.4

We give a proof of the constructive type. The assertion is verified whenever the determinantal inequality

$$\psi_\theta\left(\widehat{h}_X, \widehat{h}_X\right) \times \psi_\theta\left(\widehat{h}_Y, \widehat{h}_Y\right) \geq \psi_\theta\left(\widehat{h}_X, \widehat{h}_Y\right) \times \psi_\theta\left(\widehat{h}_Y, \widehat{h}_X\right) \tag{A3}$$

holds for any upperly bounded pair $(\widehat{h}_X, \widehat{h}_Y)$ and for any generating function $\varphi \in \Phi$. Direct inspection shows that $\psi_\theta(\widehat{h}_X, \widehat{h}_X) = \widehat{h}_X$ for any function $\widehat{h}_X$ and generating function $\varphi \in \Phi$ (and similarly for $\widehat{h}_Y$), and for any $0 \leq \theta \leq 1$. Also, from the arguments in Hardy et al. (1934) we have $\psi_\theta(\widehat{h}_X, \widehat{h}_Y) \leq \mathrm{G}_\theta(\widehat{h}_X, \widehat{h}_Y)$, so that the right hand side of Equation (A3) shall be bounded by $\widehat{h}_X \widehat{h}_Y$. The proof is completed. ■