Assessing the significance of the correlation between the components of a bivariate Gaussian random field

[1]M. Bevilacqua, [2]R. Vallejos, [1,3]D. Velandia

[1]*Instituto de Estadística, Universidad de Valparaíso, Chile.*

[2]*Departamento de Matemática, Universidad Técnica Federico Santa María, Chile.*

[3]*Facultad de Ciencias Básicas, Universidad Tecnológica de Bolívar, Colombia.*

Corresponding author: R. Vallejos, `ronny.vallejos@usm.cl`

Research Article

Abstract

Assessing the significance of the correlation between the components of a bivariate random field is of great interest in the analysis of spatial data. This problem has been addressed in the literature using suitable hypothesis testing procedures or using coefficients of spatial association between two sequences. In this paper, testing the association between autocorrelated variables is addressed for the components of a bivariate Gaussian random field using the asymptotic distribution of the maximum likelihood estimator of a specific parametric class of bivariate covariance models. Explicit expressions for the Fisher information matrix are given for a separable and a nonseparable version of the parametric class, leading to an asymptotic test. A simulation study compares the type I error and the power of the proposed test with the modified t test (Clifford et al., 1989). The empirical evidence supports our proposal, and as a result, in most of the cases, the new test performs better than the modified t test even when the bivariate covariance model is misspecified or the distribution of the bivariate randon field is not Gaussian. Finally, to illustrate how the proposed test works in practice, we study a real dataset concerning the relationship between arsenic and lead from a contaminated area in Utah, USA.

Keywords: Cross-covariance estimation; Geostatistics; Increasing domain; Hypothesis testing; Power function.

1 Introduction

In the analysis of spatial data, the quantification of spatial associations between two variables has been addressed in several appliad areas where the association between two random fields can give answers to specific problems. For example, Blanco-Moreno et al. (2006) analized spatial and temporal patterns of lolium rigidum-avena sterilis mixed populations in a cereal field while Ojeda et al. (2012) proposed to use the codispersion coefficient to define a measure of similarity between images. The assessment of the correlation between two spatial processes has been tackled using at least two different perspectives. With the first perspective, the problem is assessed using a hypothesis testing approach, mainly by transforming the *t* test in a suitable way to include the spatial information available for each process (Clifford et al., 1989; Haining, 1991; Dutilleul, 1993). Recently, a computational method based on permutations and smoothing of the original variables has been suggested in the context of biodiversity (Viladomat et al., 2014). With the second perspective, the association between two spatial processes is assessed by considering coefficients of spatial association (Matheron, 1965; Tjøstheim, 1978; Lee, 2001), which have been increasingly used in several applied areas, such as hydrology and soil sciences (Goovaerts, 1997; Pringle and Lark, 2006; Córdoba et al., 2013). In particular, the codispersion coefficient, first introduced by Matheron (1965), has received attention in recent years because it allows for the quantification of the existing spatial association between two processes in a particular direction (Ojeda et al., 2012; Cuevas et al., 2013). These and other procedures have been implemented computationally (Osorio et al., 2014), in practice facilitating the analysis of correlation between two real spatial sequences. Applications and extensions of the codispersion coefficient in time series, image processing, and multivariate geostatistics can be found in Vallejos (2008, 2012) and Vallejos et al. (2015).

A third approach assumes that the bivariate spatial process is a bivariate Gaussian random field (BGRF) that has been observed in a spatial domain. Because, under Gaussianity, the mean and the covariance structure completely characterize the distribution of the bivariate process, the dependence within and between the two processes is completely described using the bivariate cross covariance function. The linear model of coregionalization has been used for many years as a parametric model for such functions (Wackernagel (2003), Zhang (2007)). It is obtained by accounting for every component of the bivariate random field as a linear combination of mutually uncorrelated random fields. Nevertheless, as outlined by Gneiting et al. (2010) and Porcu et al. (2013), this model possess some drawbacks. For instance, it is not possible to recover the smoothness of the latent processes, as the smoothness of the components is dominated by the roughest of the latent components representing them. To overcome these drawbacks, new and more flexible models have been proposed in recent years (see Genton and Kleiber, 2014, for an excellent review). For example, Gneiting et al. (2010) and Apanasovich et al. (2012) introduced the bivariate Matérn model, and Daley et al. (2014) proposed a bivariate model with a compactly supported correlation function of the Wendland type (Wendland, 1995). The models proposed by the previous authors share a common general construction for a bivariate parametric model, which consists of modeling the marginal and the cross covariance functions using a specific univariate covariance model with possibly different parameters, while the marginal correlation between the components is described using the so-called colocated correlation parameter (Gneiting et al., 2010). In a special case of this general construction (separable model), the two components of the BGRF share the same correlation structure.

Using this kind of general construction, we propose a parametric test for assessing the significance of the correlation between the components of a BGRF. Specifically, our test is based on the asymptotic distribution of the maximum likelihood (ML) estimator of the colocated correlation parameter. Because a zero correlation is a necessary and sufficient condition for independence in the Gaussian case, this method can be used to test the hypothesis of independence or, in general, to test the significance of the correlation between the components of a BGRF. Following a result stated by Mardia and Marshall (1984) in the context of an increasing domain asymptotic framework, we find explicit expression for the Fisher information matrix associated to the aforementioned general construction. An interesting result is that, when the model is of a separable form, the ML asymptotic distribution of the colocated correlation parameter is free from any spatial dependence, whereas in the general case, the asymptotic distribution is affected by the spatial dependence.

We conducted several simulation studies to explore the power of the tests and the type I error for the Clifford et al. test and our proposal considering bivariate Matérn and Wendland models. In the simulation study, we also investigate the robustness of the proposed test against different types of misspecification of the bivariate covariance model and against possible deviations from the Gaussian distribution. Specifically, we consider bivariate random fields of the chi-square and skew Gaussian type, which have recently been proposed in the literature (Zhang and El-Shaarawi , 2010; Ma , 2011).

We use an example with real data to illustrate the practical scope of our proposal. The dataset consists of georeferenced samples from a contaminated area in Utah, USA, in which the variables of interest are arsenic (As) and lead (Pb). Based on preliminaries exploratory data analysis, we apply the proposed test to determine if the correlation between As and Pb is greater than a certain threshold, a particular case of the general hypothesis testing formulation, given by Clifford et al. (1989) for spatial depdendent sequences. The computation of the suggested test is implemented using existing R packages, which facilitates the use of the test in practice.

The paper is organized as follows. In Section 2, we describe the general parametric class for the bivariate cross-covariance function that is used in the rest of the paper. Section 3 describes the the test based on the asymptotic distribution of the colocated correlation parameter. In Section 4, we evaluate the performance of the test via Monte Carlo simulations, using the Clifford et al. test as benchmark, where the power function and the type I error are used to evaluate the behavior of the tests. In the same section, we evaluate the performance of the proposed test against different types of covariance model misspecifications and against deviations from the Gaussian distribution. The real data example is described in Section 5. Section 6 presents a discussion, which includes problems to be studied in future research. Finally in appendix A we give an example of the application of the results of Mardia and Marshall (1984) extended to the bivariate case and in appendix B we give explicit expressions for the Fisher information matrix in the separable and nonseparable case.

2 Parametric bivariate covariance models

For the remainder of the paper, we denote using $\mathbf{Z}(\boldsymbol{s}) = \{(Z_1(\boldsymbol{s}), Z_2(\boldsymbol{s}))^T\}$, a BGRF with continuous spatial index $\boldsymbol{s} \in \mathbb{R}^d$. The assumption of Gaussianity implies that the first and second moments determine uniquely the finite dimensional distributions. In particular, we shall suppose weak stationarity throughout, so that the mean vector $\boldsymbol{\mu} = \mathbb{E}(\mathbf{Z})$ is constant, and because in the Gaussian distribution the covariance estimation is not affected by the mean, we assume $\boldsymbol{\mu} = (0, 0)^T$ without loss of generality. The covariance function between $\mathbf{Z}(\boldsymbol{s}_1)$ and $\mathbf{Z}(\boldsymbol{s}_2)$, for any pair $\boldsymbol{s}_1, \boldsymbol{s}_2$ in the spatial domain, is represented by a mapping

$\boldsymbol{C}:\mathbb{R}^d \to M_{2\times 2}$ defined through

$$\boldsymbol{C}(\boldsymbol{h}) = [C_{ij}(\boldsymbol{h})]_{i,j=1}^2 = [\text{cov}\,(Z_i(\boldsymbol{s}_1), Z_j(\boldsymbol{s}_2))]_{i,j=1}^2\,, \quad \boldsymbol{h} = \boldsymbol{s}_1 - \boldsymbol{s}_2 \in \mathbb{R}^d. \tag{1}$$

The function $\boldsymbol{C}(\boldsymbol{h})$ is called bivariate covariance function. Here, $M_{2\times 2}$ is the set of squared, symmetric and positive definite matrices. The functions $C_{ii}(\boldsymbol{h})$ $i = 1, 2$ are the marginal covariance functions of the Gaussian random fields $Z_i(\boldsymbol{s})$, $i = 1, 2$, while $C_{12}(\boldsymbol{h})$ is called cross covariance function between $Z_1(\boldsymbol{s})$ and $Z_2(\boldsymbol{s})$ at the spatial lag $\boldsymbol{h} \in \mathbb{R}^d$. The mapping $\boldsymbol{R}:\mathbb{R}^d \to M_{2\times 2}$ defined through $\boldsymbol{R}(\boldsymbol{h}) = [R_{ij}(\boldsymbol{h})]_{i,j=1}^2$ with

$$R_{ij}(\boldsymbol{h}) = \frac{C_{ij}(\boldsymbol{h})}{\sqrt{C_{ii}(\boldsymbol{0})C_{jj}(\boldsymbol{0})}}$$

is called bivariate correlation function, $R_{ii}(\boldsymbol{h})$ being the marginal correlation functions of the Gaussian random fields $Z_i(\boldsymbol{s})$, $i = 1, 2$, $R_{12}(\boldsymbol{h})$ being the cross correlation function between the fields $Z_1(\boldsymbol{s})$ and $Z_2(\boldsymbol{s})$ and $R_{12}(\boldsymbol{0})$ expressing the marginal correlation between the two components. The mapping $\boldsymbol{C}$ (and, consequently, $\boldsymbol{R}$) must positive definite, which means that, for a given realization $\boldsymbol{Z} = (\boldsymbol{Z}(\boldsymbol{s}_1)^T, \ldots, \boldsymbol{Z}(\boldsymbol{s}_n)^T)^T$, the $(2n) \times (2n)$ covariance matrix $\Sigma := [\boldsymbol{C}(\boldsymbol{s}_i - \boldsymbol{s}_j)]_{i,j=1}^n$ is positive definite.

We shall assume throughout that the mapping $\boldsymbol{C}$ comes from a parametric family of bivariate covariances $\{\boldsymbol{C}(\cdot;\boldsymbol{\theta}), \boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq R^p\}$, with $\boldsymbol{\Theta}$ an arbitrary parametric space. Recent literature has been engaged on offering new models for bivariate covariances and for a through review the reader is referred to Genton and Kleiber (2014) with their exhaustive list of references. One of them is the linear model of coregionalization, that has been popular for over thirty years (Wackernagel, 2003). It consists of representing the bivariate Gaussian field as a linear combination of q independent univariate fields, with $q = 1, 2$. The resulting bivariate covariance function takes the form:

$$\boldsymbol{C}(\boldsymbol{h};\boldsymbol{\theta}) = \left[\sum_{k=1}^{q} \psi_{ik}\psi_{jk} R_k(\boldsymbol{h}, \boldsymbol{\psi}_k)\right]_{i,j=1}^2, \tag{2}$$

with $A := [\psi_{lm}]_{l,m=1}^{2,q}$ being a $2 \times q$ dimensional matrix with full rank, and with $R_k(\boldsymbol{h}; \cdot)$ being a univariate parametric correlation model. Clearly, we have $\boldsymbol{\theta} = (\mathrm{vec}(A)^T, \boldsymbol{\psi_1}^T, \dots, \boldsymbol{\psi_q}^T)^T$. Note that when $q = 2$, the marginal correlation is given by a nonlinear function of the parameters, i.e. $R_{12}(\mathbf{0}; \boldsymbol{\theta}) = \frac{\psi_{11}\psi_{21} + \psi_{12}\psi_{22}}{\sqrt{(\psi_{11}^2 + \psi_{12}^2)(\psi_{21}^2 + \psi_{22}^2)}}$. A criticism expressed about this model by Gneiting et al. (2010) is that, when $\psi_{ik} \neq 0$ for each i, k, the smoothness of any component defaults to that of the roughest latent process.

Another general parametric class, called *separable*, is obtained through the following bivariate covariance function:

$$\boldsymbol{C}(\boldsymbol{h}, \boldsymbol{\theta}) = [\rho_{ij}\sigma_i\sigma_j R(\boldsymbol{h}, \boldsymbol{\psi})]_{i,j=1}^2, \quad \rho_{ii} = 1, |\rho_{12}| < 1, \tag{3}$$

where $R(\boldsymbol{h}; \boldsymbol{\psi})$ is a univariate parametric correlation model, $\boldsymbol{\theta} = (\sigma_1^2, \sigma_2^2, \boldsymbol{\psi}^T, \rho_{12})^T$, $\sigma_i^2 > 0$, $i = 1, 2$ are the marginal variances and ρ_{12}, the colocated correlation parameter, expresses the marginal correlation between $Z_1(\boldsymbol{s})$ and $Z_2(\boldsymbol{s})$.

This type of construction assumes that the two components of the BGRF share the same correlation structure. Therefore, the model is not able to capture different spatial dependences and/or the smoothness of each of the component fields. A generalization of (3), that here we call *nonseparable*, which allows to overcome this drawback is:

$$\boldsymbol{C}(\boldsymbol{h}, \boldsymbol{\theta}) = \left[\rho_{ij}\sigma_i\sigma_j R(\boldsymbol{h}; \boldsymbol{\psi}_{ij})\right]_{i,j=1}^2, \quad \rho_{ii} = 1. \tag{4}$$

where $\boldsymbol{\theta} = (\sigma_1^2, \sigma_2^2, \boldsymbol{\psi}_{11}^T, \boldsymbol{\psi}_{12}^T, \boldsymbol{\psi}_{22}^T, \rho_{12})^T$. In this general approach, the difficulty lies in deriving conditions on the model parameters that result in a valid multivariate covariance model. For instance Gneiting et al. (2010) proposed the model (4) with $R(\boldsymbol{h}, .)$ equal to the Matérn correlation model :

$$R(\boldsymbol{h}; \boldsymbol{\psi}) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\boldsymbol{h}}{\beta}\right)^{\nu} \mathcal{K}_{\nu}\left(\frac{\boldsymbol{h}}{\beta}\right), \tag{5}$$

where $\boldsymbol{\psi} = (\beta, \nu)^T$, $\beta > 0$ is the scale parameter and $\nu > 0$ indexes differentiability at the origin. The authors find necessary and sufficient conditions on the colocated correlation parameter ρ_{12} in order the model (4) to be valid. Another example of the model (4) can be found in Daley et al. (2014), where $R(\boldsymbol{h};.)$ is a compactly supported correlation model of the Wendland type (Wendland, 1995):

$$R(\boldsymbol{h};\boldsymbol{\psi}) = \left(1 + (\nu+1)\frac{\boldsymbol{h}}{\beta}\right)\left(1 - \frac{\boldsymbol{h}}{\beta}\right)_{+}^{\nu+1} \tag{6}$$

with $\beta > 0$ and $\nu > \frac{(d+1)}{2} + 2$. In this case, the authors provide sufficient conditions for the parameter ρ_{12} for the validity of the model. A comparison in terms of flexibility between (2) and (3) or (4) can be found in Bevilacqua et al. (2015).

A benefit of the general class (3) or (4) with respect to (2) is that the colocated correlation parameter express the marginal correlation between the components, that is, $R_{12}(\mathbf{0};\boldsymbol{\theta}) = \rho_{12}$ so when $\rho_{12} = 0$, the components of the bivariate random field are independent; hence, the colocated correlation parameter can be used to build a test of independence or, in general, to assess the significance of the correlation between the components of the random field. This problem is addressed in the next section.

3 A parametric test

Since we are assuming that the state of truth is represented by some parametric family of bivariate covariances $\{\boldsymbol{C}(\cdot;\boldsymbol{\theta}), \boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq R^p\}$, we may use the abuse of notation $\Sigma(\boldsymbol{\theta})$ for the covariance matrix Σ, in order to emphasize the dependence on the unknown parameters vector. Specifically we assume that the parametric bivariate covariance model is of the type (3), or (4) so $\boldsymbol{\theta} = (\sigma_1^2, \sigma_2^2, \boldsymbol{\psi}^T, \rho_{12})^T$ or $\boldsymbol{\theta} = (\sigma_1^2, \sigma_2^2, \boldsymbol{\psi}_{22}^T, \boldsymbol{\psi}_{11}^T, \boldsymbol{\psi}_{12}^T, \rho_{12})^T$ depending if a separable or a nonseparable bivariate covariance model is considered. For a realization from a BGRF, the log-likelihood, up to an additive constant, can be written as

$$l_n(\boldsymbol{\theta}) = -\frac{1}{2}\log|\Sigma(\boldsymbol{\theta})| - \frac{1}{2}\boldsymbol{Z}^{\intercal}[\Sigma(\boldsymbol{\theta})]^{-1}\boldsymbol{Z}.$$

Consequently, $\widehat{\boldsymbol{\theta}}_n := \text{argmax}_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} l_n(\boldsymbol{\theta})$ is the maximum likelihood estimator of $\boldsymbol{\theta}$. Mardia and Marshall (1984) provide conditions for the univariate case for the consistency and the asymptotic normality of the ML estimator. Under these conditions, $\widehat{\boldsymbol{\theta}}_n$ is consistent and asymptotically normal, with covariance matrix equal to the inverse of the Fisher Information matrix, with the following expression:

$$F_n(\boldsymbol{\theta}) = \left[\frac{1}{2}\text{tr}\left(\Sigma(\boldsymbol{\theta})^{-1}\frac{d\Sigma(\boldsymbol{\theta})}{d\boldsymbol{\theta}_i}\Sigma(\boldsymbol{\theta})^{-1}\frac{d\Sigma(\boldsymbol{\theta})}{d\boldsymbol{\theta}_j}\right)\right]_{i,j=1}^{p}. \tag{7}$$

That is, $\widehat{\boldsymbol{\theta}}_n \xrightarrow{p} \boldsymbol{\theta}$ and $\widehat{\boldsymbol{\theta}}_n \approx \mathcal{N}(\boldsymbol{\theta}, F_n(\boldsymbol{\theta})^{-1})$ as $n \longrightarrow \infty$.

In Appendix A, we extend the conditions of Mardia and Marshall (1984) to the bivariate case. In general, it is not easy to check such conditions because they are based on the eigenvalues of the covariance matrix and its derivatives. Despite this, we show in Appendix A that these conditions are verified for a bivariate separable exponential model. The model is later used in the numerical example presented in Section 4.

Then testing the independence or assessing the strength of correlation between the two components of a BGRF with covariance (3) or (4), leads to the following hypothesis testing problems:

$$H_0 : \rho_{12} = 0 \quad \text{versus} \quad H_1 : \rho_{12} \neq 0, \tag{8}$$

$$H_0 : \rho_{12} \leq k \quad \text{versus} \quad H_1 : \rho_{12} > k, \tag{9}$$

where k belongs to the feasible parameter space of the bivariate correlation model. Given the maximum likelihood estimate $\widehat{\boldsymbol{\theta}}_n = (\widehat{\sigma}_1^2, \widehat{\sigma}_2^2, \widehat{\boldsymbol{\psi}}^T, \widehat{\rho}_{12})^T$ or $\widehat{\boldsymbol{\theta}}_n = (\widehat{\sigma}_1^2, \widehat{\sigma}_2^2, \widehat{\boldsymbol{\psi}}_{11}^T, \widehat{\boldsymbol{\psi}}_{12}^T, \widehat{\boldsymbol{\psi}}_{22}^T, \widehat{\rho}_{12})^T$ these tests are based on the asymptotic null distribution:

$$\frac{\widehat{\rho}_{12} - \rho_{12}}{se(\widehat{\rho}_{12})} \approx \mathcal{N}(0,1), \tag{10}$$

where $se(\widehat{\rho}_{12})$ denotes the standard error of $\widehat{\rho}_{12}$ given by $se(\widehat{\rho}_{12}) = \sqrt{F_n(\widehat{\boldsymbol{\theta}}_n)^{-1}_{\rho_{12}}}$, and $\rho_{12} = 0$ or $\rho_{12} = k$, depending on whether test (8) or (9) is considered. Here, $F_n(\widehat{\boldsymbol{\theta}}_n)^{-1}_{\rho_{12}}$ is the element on the diagonal of $F_n(\widehat{\boldsymbol{\theta}}_n)^{-1}$ associated to the colocated correlation parameter. We consider bivariate covariance model (4) in an increasing order of complexity:

(A) A separable model that is a bivariate model with covariance model (3). In this case, the Fisher information has a simple form (see Appendix B, separable case), and it is possible to calculate explicitly the inverse of $F_n^{-1}(\boldsymbol{\theta})$. Then, the asymptotic distribution of $\widehat{\rho}_{12}$ is as in (10) with $se(\widehat{\rho}_{12}) = \sqrt{\frac{(\widehat{\rho}_{12}^2-1)^2}{n}}$. It is interesting to note that in this case, the asymptotic distribution does not depend on the spatial correlation structure, i.e., on the choice of $R(\boldsymbol{h}, \cdot)$ but only on the estimated colocated correlation parameter and n.

(B) A nonseparable model that is a bivariate model with covariance model (4) with the constraints $\boldsymbol{\psi}_{12} = f(\boldsymbol{\psi}_{11}, \boldsymbol{\psi}_{22})$ where f is a differentiable function. In this case, the cross parameters are assumed to be functions of the marginal parameters. This constraint is assumed to reduce the number of parameters to be estimated and for simplifying the model. For instance, Gneiting et al. (2010) assumes $\nu_{12} = \nu_{11} + \nu_{22}$ for the smoothness parameter of a Bivariate Matérn covariance, which they call parsimonious bivariate model. Apanasovich and Genton (2010) in their multivariate Gneiting model consider different types of such constraints as, for example, the following one on the scale parameter: $\beta_{12} = \frac{\beta_{11}+\beta_{22}}{2}$. We consider the latter in our simulation study. It turns out from the expression of the Fisher information matrix obtained in Appendix B that the asymptotic variance of the correlation parameter depends on the spatial dependence (see Appendix B, nonseparable case: constrained version).

(C) A nonseparable model that is a bivariate model with covariance model (4). Note that when $\rho_{12} = 0$, $\boldsymbol{\psi}_{12}$ cannot be estimated. From a Fisher information perspective, it means that the Fisher information matrix is singular. For this reason, under this setting, the test (8) is not feasible and the test (9) can be considered only for $k \neq 0$.

In this case, if the covariance parameters are assumed known, $se(\widehat{\rho}_{12})$ is equal to case B (see Appendix B, nonseparable case: full version).

4 Simulation study

In this section, we investigate the performance of the proposed test using the Clifford et al. test as a benchmark. Specifically we report finite-sample simulation results to compare the actual probability of type I error with the nominal and power functions for both tests. Overall, we consider three different levels for the nominal value of the Type I Error and for the power function ($\alpha = 0.01, 0.05, 0.1$). In our simulations, we consider a spatial regular grid on the unit square equally spaced by $1/17$ such that the total number of locations sites is 324, with a total of 648 observations.

We first consider the performance of the proposed test under the standard setting where, by a standard setting, we mean that for a BGRF with a given covariance model of the type (3) or (4), we perform the test using the correct model. Then, we explore the performance of the test under different types of misspecification of the covariance model and under deviations from the Gaussian distribution. Under the standard setting and covariance model misspecification, we consider both hypothesis (8) and (9), while in the case of deviations from the Gaussian distribution, we only considered the test of independence (8). The Clifford et al. test has been used as a benchmark only when testing hypothesis (8).

4.1 Standard setting

Here, we investigate the performance of the proposed test for a BGRF with bivariate covariance model of type (3) or (4). The first bivariate covariance model considered is the following:

$$C_{ij}(\boldsymbol{h}, \boldsymbol{\theta}) = \rho_{ij}\sigma_i\sigma_j e^{-3\frac{||\boldsymbol{h}||}{\beta_{ij}}}, \quad \rho_{ii} = 1, \quad i, j = 1, 2. \tag{11}$$

This model is obtained by choosing in (4) an $R(\boldsymbol{h}, \cdot)$ equal to the Matérn model (5) with $\nu_{ij} = 0.5$ for $i, j = 1, 2$. In this case, $\boldsymbol{\theta} = (\sigma_1^2, \sigma_2^2, \beta_{11}, \beta_{12}, \beta_{22}, \rho_{12})^\top$. Note that the bivariate covariance model is parametrized in terms of marginal practical ranges, that is, $C_{ii}(\boldsymbol{h}, \boldsymbol{\theta}) < 0.05$ when $\|\boldsymbol{h}\| > \beta_{ii}$. The hypothesis (8) was tested under the separable (case A) and nonseparable (case B) cases, while the hypothesis (9) was tested under the nonseparable case (case C). To accomplish this, we consider the following scenarios:

1. For specification A, we considered the hypothesis (8) and simulated 2000 instances of zero mean BGRF setting $\sigma_i^2 = 1$, $\beta = \beta_{ij} = 0.2$ for $i, j = 1, 2$, and we performed ML estimations, obtaining $\widehat{\boldsymbol{\theta}} = (\widehat{\sigma}_1^2, \widehat{\sigma}_2^2, \widehat{\beta}, \widehat{\rho}_{12})^\top$. Then, we rejected H_0 if $\frac{|\widehat{\rho}_{12}|}{se(\widehat{\rho}_{12})} > q_{1-\frac{\alpha}{2}}$ where q_γ is the upper quantile of order γ of the standard normal distribution. The number of times that H_0 is rejected divided by the number of simulations gives the empirical probability of the type I error. In addition, to evaluate the power of the test, we simulated 2000 zero mean BGRF instances under H_1, i.e., under dependence between the two components of the BGRF. Specifically, we simulated and estimated under the same previous setting but with increasing correlation values between the random fields, $\rho_{12} = 0.05, 0.15, 0.25$. The results are shown in Table 1.

   For this scenario, we also considered the hypothesis (9) with $k = 0.15$. Specifically, we simulated, under H_0 ($\rho_{12} = 0.15$), 2000 zero mean BGRF instances under the same previous simulation setting and using the ML method. We then rejected H_0 if $\frac{\widehat{\rho}_{12} - k}{se(\widehat{\rho}_{12})} > q_{1-\alpha}$. To evaluate the power of the test, a BGRF was simulated, and the corresponding probability was computed under H_1. We simulated, under the same setting, 2000 zero mean BGRF instances but using increasing correlation values between the components of the random field, $\rho_{12} = 0.2, 0.3, 0.4$. Then, we applied our tests in the same way as before. The results are shown in Table 3.

2. For specification B, we considered the hypothesis (8), and we simulated, under H_0

($\rho_{12} = 0$) and under H_1 ($\rho_{12} = 0.05, 0.15, 0.25$), 2000 zero mean BGRF instances setting $\sigma_1^2 = \sigma_2^2 = 1$, $\beta_{11} = 0.2$, $\beta_{22} = 0.1$. We performed ML estimations using the constraint $\beta_{12} = 0.5(\beta_{11} + \beta_{22})$, obtaining $\widehat{\boldsymbol{\theta}} = (\widehat{\sigma}_1^2, \widehat{\sigma}_2^2, \widehat{\beta}_{11}, \widehat{\beta}_{22}, \widehat{\rho}_{12})^\top$. We then applied the proposed test as explained previously to estimate the probability of the type I error and the power. The results are shown in Table 2.

Additionally, the hypothesis (9) was considered with $k = 0.15$. Specifically, under the same simulation setting, we simulated under H_0 ($\rho_{12} = 0.15$) and then under H_1 ($\rho_{12} = 0.2, 0.3, 0.4$), 2000 zero mean BGRF instances, and then, for each iteration, we applied the proposed test as explained before in order to estimate the probability of the type I error and the power. The results are shown in Table 4.

3. For specification C, we considered only the hypothesis (9) with $k = 0.15$. In particular, we simulated, under H_0 ($\rho_{12} = 0.15$) and then under H_1 ($\rho_{12} = 0.2, 0.3, 0.4$), 2000 zero mean BGRF instances setting $\sigma_1^2 = \sigma_2^2 = 1$, $\beta_{11} = 0.2$, $\beta_{22} = 0.1$, $\beta_{12} = 0.15$. We performed ML estimations, obtaining $\widehat{\boldsymbol{\theta}} = (\widehat{\sigma}_1^2, \widehat{\sigma}_2^2, \widehat{\beta}_{11}, \widehat{\beta}_{12}, \widehat{\beta}_{22}, \widehat{\rho}_{12})^\top$. We applied the proposed test as explained previously in order to estimate the probability of the type I error and the power. The results are shown in Table 5.

We replicated the same simulation study used previously (points 1, 2, and 3) with the model:

$$C_{ij}(\boldsymbol{h}, \boldsymbol{\theta}) = \rho_{ij}\sigma_i\sigma_j \left(1 + 5\frac{||\boldsymbol{h}||}{\beta_{ij}}\right)\left(1 - \frac{||\boldsymbol{h}||}{\beta_{ij}}\right)_+^5, \quad \rho_{ii} = 1, \quad i, j = 1, 2, \tag{12}$$

which was obtained from (4), choosing $R(\boldsymbol{h}, \cdot)$ as the Wendland model (6) with $\nu = 4$. We used the same parameter settings for the variances, the colocated parameter and the compact support parameters. The results are displayed in Tables 1, 2, 3, 4 and 5.

Some comments are in order. When considering the test of independence (8) for scenarios 1 and 2 (Tables 1 and 2), a comparison of our test with the Clifford et al. test shows that the estimates of the probability of type I error are very close to the true values for both scenarios and for both the exponential and Wendland bivariate covariance models. Moreover, in both cases, the estimated power function of the new test is greater than that of the Clifford et al. test, highlighting the quality of our proposal in this regard. Note that in the separable case described in Table 1, the probability of type I error and the power across the two covariance models considered are very similar. This is not surprising because, in this case, the asymptotic distribution of the colocated parameter does not depend on the choice of the covariance model. On the other hand, in the nonseparable case, scenario B (Table 2), the test applied using the Wendland model performs slightly better than the exponential model in terms of power. Similar comments can be formulated when considering test (9) (Tables 3, 4 and 5). Recall that in this case, the comparison with the Clifford et al. test is not feasible.

4.2 Bivariate covariance model misspecification

Because the proposed test is based on the assumption that the bivariate parametric covariance model is of type (3) or (4), it is of interest to investigate, through numerical examples, the robustness of the test against possible covariance model misspecifications. In particular, we consider three types of covariance model misspecifications.

- A first source of misspecification is when the true covariance model and the model used in the test belong to the class (4) but with different levels of specification. For instance, the true model can be of the type (4) with specification B, but specification A is used to perform the test. To investigate the performance of the test, under this kind of misspecification and when considering hypothesis (8), we used the 2000 BGRF instances simulated under the previously described scenario 2, that is, a bivariate nonseparable covariance exponential model. We applied the proposed test considering the model used in scenario 1 (a bivariate separable covariance exponential). As before, we considered the cases H_0 ($\rho_{12} = 0$) and H_1 ($\rho_{12} = 0.05, 0.15, 0.25$) to estimate the probability of the type I error and the power of the proposed test. The performance of the test of independence (8) using the wrong covariance model is shown in Table 6 (top) (These results must be compared with the results in Table 2 (top)). It is observed in this case that the estimation of the type I error is quite reasonable, and as expected, there is a small loss in terms of power with respect to the standard setting. Nevertheless, the proposed test performs better than the Clifford et al. test.

We also explored the performance of the test when considering the hypothesis (9) with $k = 0.15$ under this kind of misspecification. We used the 2000 BGRF instances simulated under scenario 2 under H_0 ($\rho_{12} = 0.15$) and then under H_1 ($\rho_{12} = 0.2, 0.3, 0.4$) to estimate the probability of the type I error and the power of the proposed test. The performance of test (9) when using the wrong covariance model is shown in table 7 (top). (These results must be compared with the results in Table 4 (top)). Also in this case, there is a small loss of power, as expected, with respect to the standard setting.

- A second source of misspecification is when the true covariance model is of type (3) or (4) with $R(\boldsymbol{h}, \cdot)$ as parametric correlation model and the test is applied using the covariance model of type (3) or (4) but using a different parametric correlation model. Given this kind of misspecification, hypothesis (8) was considered. We simulated 2000 BGRF instances with a separable Wendland model under scenario 1, and we applied the proposed test using the separable exponential model. Consequently, we considered the cases H_0 ($\rho_{12} = 0$) and H_1 ($\rho_{12} = 0.05, 0.15, 0.25$) to estimate the probability of the type I error and the power of the proposed test. The results regarding the performance of test (8) using the wrong covariance model are shown in Table 6 (middle), and they must be compared with the results in Table 1 (bottom). There is a small loss of power with respect to the standard setting, as expected, when the the type I error is kept constant under the nominal level. Nevertheless the test performs better than the Cliffor f et al. test in terms of power.

We also explored the performance of the test when considering the hypothesis (9), with $k = 0.15$, under this kind of misspecification. We simulated 2000 BGRF instances from a separable Wendland model under the scenario 1, and we applied the proposed test using the separable exponential model. The cases H_0 ($\rho_{12} = 0.15$) and H_1 ($\rho_{12} = 0.2, 0.3, 0.4$) were considered to estimate the probability of the type I error and the power of the proposed test. The results are shown in Table 7 (middle) and they must be compared with the results in Table 3 (bottom). In this case, there is a slight overestimation of the probability of I type error and a small loss in power with respect to the standard setting.

- A third type of misspecification is when the true covariance model and the covariance model used in the test are different. Here, we consider a setting where the true model is the linear model of coregionalization given in equation (2), but a model of the type (3) is used in the test. Precisely, in (2), $R_1(\boldsymbol{h}) = e^{-3\frac{||h||^2}{c_1}}$ and $R_2(\boldsymbol{h}) = \frac{c_2}{20.371||h||}\sin(\frac{20.371||h||}{c_2})$ (Wackernagel, 2003), with $c_1 = 0.2$, $c_2 = 0.15$, $\psi_{ii} = \sqrt{1-x}$, and $\psi_{12} = \psi_{21} = \sqrt{x}$, with $x = 0, 0.00063, 0.00566, 0.01588$. Under this parameter setting, the BGRF has unit variances, with marginal practical ranges equal to 0.2 and 0.15, and $R_{12}(\mathbf{0};\boldsymbol{\theta})$ is approximatively equal to $0, 0.05, 0.15, 0.25$ respectively. To evaluate the empirical probability of the type I error associated with the test (8), we simulated 2000 zero mean BGRF instances under H_0 ($x = 0$). The proposed test was applied assuming that the true model in its separable form is as in equation (11). In addition, to evaluate the power of the test, we simulated 2000

BGRF instances under H_1 ($x = 0.00063, 0.00566, 0.01588$). For each iteration, the proposed test was applied in the same way as described before.

We also investigated the performance of the test when considering the hypothesis (9) under this kind of misspecification. The previous coregionalization covariance structure with the same parameter setting was used for $x = 0.00566, 0.0101, 0.023, 0.04175$, which corresponds to a BGRF with unit variances and $R_{12}(\mathbf{0}; \boldsymbol{\theta})$ approximatively equal to $0.15, 0.2, 0.3, 0.4$ respectively. To evaluate the empirical type I error probability of test (9) with $k = 0.15$, we simulated 2000 BGRF instances under H_0 ($x = 0.00566$). In each iteration, we applied the proposed test assuming (11) as the true model in its separable form. Moreover, to evaluate the power of the test, we simulated 2000 BGRF instances under H_1 ($x = 0.0101, 0.023, 0.04175$). In each iteration, we applied the proposed test using the wrong model, as described before. The results are shown in Tables 6 and 7 (bottom). For the hypothesis of independence (8), the test using the wrong covariance model performs better than the Clifford et al. test in terms of power and keeps the type I error below the nominal level (Table 6 bottom),and for hypothesis (9), the results are quite reasonable in terms of type I error and power (Table 7 bottom).

4.3 Non-Gaussian distribution

Finally, we investigated the performance of the test under deviations from the Gaussian distribution in two examples. First, we considered a bivariate random field whose marginal distribution is chi-square with p degrees of freedom, where $p \geq 1$ is a positive integer. Following Ma (2011), let $\mathbf{Y}(\boldsymbol{s}) = (Y_1(\boldsymbol{s}), Y_2(\boldsymbol{s}))^T$ be a bivariate random field defined as $\mathbf{Y}(\boldsymbol{s}) = \sum_{k=1}^{p} \mathbf{Z}_k(\boldsymbol{s}) \circ \mathbf{Z}_k(\boldsymbol{s})$, where $\circ$ denotes the Hadamard product between two vectors and for $k = 1, \ldots, p$, $\mathbf{Z}_k(\boldsymbol{s})$ are independent copies of $\mathbf{Z}(\boldsymbol{s}) = (Z_1(\boldsymbol{s}), Z_2(\boldsymbol{s}))^T$, a BGRF with $C_{ij}^{Z}(\boldsymbol{h})$ as a generic element of the bivariate covariance function.

For each $\boldsymbol{s}$ and $i = 1, 2$ fixed, $\mathbf{Y}_i(\boldsymbol{s})/C^Z_{ii}(\mathbf{0})$ is a chi square random variable with p degrees of freedom and the bivariate covariance function of $\mathbf{Y}(\boldsymbol{s})$ depends on that of the underlying BGRF, that is, $\boldsymbol{C}^Y(\boldsymbol{h}) = \left[2p(C^Z_{ij}(\boldsymbol{h}))^2\right]^2_{i,j=1}$. Then, assuming the parametric model in equation (11) in its separable form for $C^Z_{ij}(\boldsymbol{h})$, the bivariate covariance model of $\mathbf{Y}(\boldsymbol{s})$ is given by:

$$\boldsymbol{C}^Y(\boldsymbol{h}, \boldsymbol{\theta}, p) = \left[2p\rho^2_{ij}\sigma^2_i\sigma^2_j e^{-6\frac{||\boldsymbol{h}||}{\beta}}\right]^2_{i,j=1}, \tag{13}$$

and $R^Y_{12}(\mathbf{0}, \boldsymbol{\theta}, p) = \rho^2_{12}$. We simulate 2000 chi-squared bivariate random fields with different degrees of freedom ($p = 1, 15$) for a bivariate covariance model (13), fixing the same parameters as in scenario 1 but with $\rho_{12} = 0$. Our goal is to estimate the empirical probability of the type I error and to evaluate the power of the test for $\rho_{12} = \sqrt{0.05}, \sqrt{0.15}, \sqrt{0.25}$. For each iteration, we assume a BGRF with covariance model (11). The estimation process is carried out using the ML method obtaining $\widehat{\boldsymbol{\theta}} = (\widehat{\sigma}^2_1, \widehat{\sigma}^2_2, \widehat{\beta}, \widehat{\rho}_{12})^\top$. Consequently, we apply the test of independence and the Clifford et al. test, obtaining for $p = 1, 15$ the results that are displayed in Table 8. When $p = 1$, both tests underestimate the probability of type I error, especially when $\alpha = 0.05, 0.1$. This drawback disappeared, as expected, when $p = 10$. In this case, our test performs better than the Clifford et al. test. This result is in agreement with the fact that when increasing the degrees of freedom, the chi-square distribution (rescaled by the degrees of freedom) converges to a Gaussian distribution.

As a second example, we consider a bivariate version of the skew Gaussian random field proposed in Zhang and El-Shaarawi (2010) . Let $\mathbf{Y}(\boldsymbol{s}) = (Y_1(\boldsymbol{s}), Y_2(\boldsymbol{s}))^T$ be a bivariate random field defined as:

$$Y_i(\boldsymbol{s}) = \psi_i|X_i(\boldsymbol{s})| + \sigma_i Z_i(\boldsymbol{s}), \quad i = 1, 2, \tag{14}$$

where $\psi_i, \in \mathbb{R}$ is the asymmetry parameter of the $i - th$ random field and $\sigma_i > 0$. Here, we are assuming that $\mathbf{Z}(\boldsymbol{s}) = (Z_1(\boldsymbol{s}), Z_2(\boldsymbol{s}))^T$ is a BGRF separable parametric covariance model of the type:

$$\boldsymbol{C}^Z(\boldsymbol{h},\boldsymbol{\theta}) = [\rho_{ij}R(\boldsymbol{h},\boldsymbol{\psi})]_{i,j=1}^2\,, \quad \rho_{ii}=1, |\rho_{12}|<1, \tag{15}$$

where $\boldsymbol{\theta}=(\boldsymbol{\psi}^T,\rho_{12})^T$. $X_i(\boldsymbol{s})$, $i=1,2$ is a zero mean unit variance Gaussian random field with $R(\boldsymbol{h},\boldsymbol{\psi})$ as a correlation model. We also assume that $X_i(\boldsymbol{s}) \perp Z_j(\boldsymbol{s})$ for $i,j=1,2$ and $X_1(\boldsymbol{s}) \perp X_2(\boldsymbol{s})$, where $\perp$ denotes independence between random components. In this framework, it is straightforward to show that the covariance function of a bivariate skew Gaussian random field $\mathbf{Y}(\boldsymbol{s})$ is given by:

$$\boldsymbol{C}^Y(\boldsymbol{h},\boldsymbol{\theta},\boldsymbol{\tau}) = \left[C^Y_{ij}(\boldsymbol{h},\boldsymbol{\theta},\boldsymbol{\tau})\right]_{i,j=1}^2 \tag{16}$$

where $C^Y_{ii}(\boldsymbol{h},\boldsymbol{\theta},\boldsymbol{\tau}) = \frac{2\psi_i^2}{\pi}g(R(\boldsymbol{h},\boldsymbol{\psi})) + \sigma_i^2 R(\boldsymbol{h},\boldsymbol{\psi})$ for $i=1,2$ and $C^Y_{12}(\boldsymbol{h},\boldsymbol{\theta},\boldsymbol{\tau}) = \sigma_1\sigma_2\rho_{12}R(\boldsymbol{h},\boldsymbol{\psi})$. Here, $\boldsymbol{\tau}=(\sigma_1,\sigma_2,\psi_1,\psi_2)^T$ and $g(b)=\sqrt{1-b^2}+b\ arcsin(b)-1$, $|b|<1$. Thus,

$$R^Y_{12}(\mathbf{0},\boldsymbol{\theta}) = \frac{\sigma_1\sigma_2\rho_{12}}{\sqrt{\sigma_1^2+\psi_1^2(1-2/\pi)}\sqrt{\sigma_2^2+\psi_2^2(1-2/\pi)}}. \tag{17}$$

Under this construction, if the asymmetry parameters increase, the possible range of marginal correlation between the two random fields decreases.

We simulate 2000 skew Gaussian bivariate random field instances choosing $R(\boldsymbol{h},\boldsymbol{\psi}) = e^{-3\frac{||h||}{\beta}}$ and setting $\sigma_1=\sigma_2=1$, $\psi_1=\psi_2=\psi=1$, $\beta=0.2$, and $\rho_{12}=0, 0.06816901$, $0.204507, 0.3408451$. Under this setting, $R^Y_{12}(\mathbf{0},\boldsymbol{\theta})$ is equal to $0, 0.05, 0.15, 0.25$ respectively. In each iteration, we assume a BGRF with covariance model (11), and our test of independence was applied. We then increased the level of asymmetry by setting $\psi_1 = \psi_2 = \psi = 2.775$ and $\rho_{12} = 0, 0.1899127, 0.5697382, 0.982$, obtaining $R^Y_{12}(\mathbf{0},\boldsymbol{\theta}) = 0, 0.05, 0.15, 0.25$, respectively, and our tests were applied again. The results are shown in Table 9. As expected, when the level of asymmetry increases ($\psi = 2.775$), the test delivers a worse performance. In this case, the Clifford et al. test slightly outperforms our test in terms of power. Decreasing the asymmetry parameter ($\psi = 1$), that is, approaching the Gaussian case, our test performs better than the Cliffordet al. test, as expected. In both cases, the estimation of the probability of type I error is quite reasonable.

5 A Real Data Example

Georeferenced data have been selected for illustrative purposes in this example. The dataset consists of soil samples collected in and around the vacant industrially contaminated Murray smelter site (Utah, USA). This area was polluted by airborne emissions and the placement of waste slag from smelting processes. A total of 253 locations were included in the study, and soil samples were taken from each location. Each georeferenced sampling quantity is a pool composite of four closely adjacent soil samples, for which the heavy metals arsenic (As) and lead (Pb) were measured. A complete description of the Murray smelter site dataset can be found in Griffith (2002) and Griffith and Paelinck (2011). The locations and attributes As and Pb for each location are shown in Figure 1.

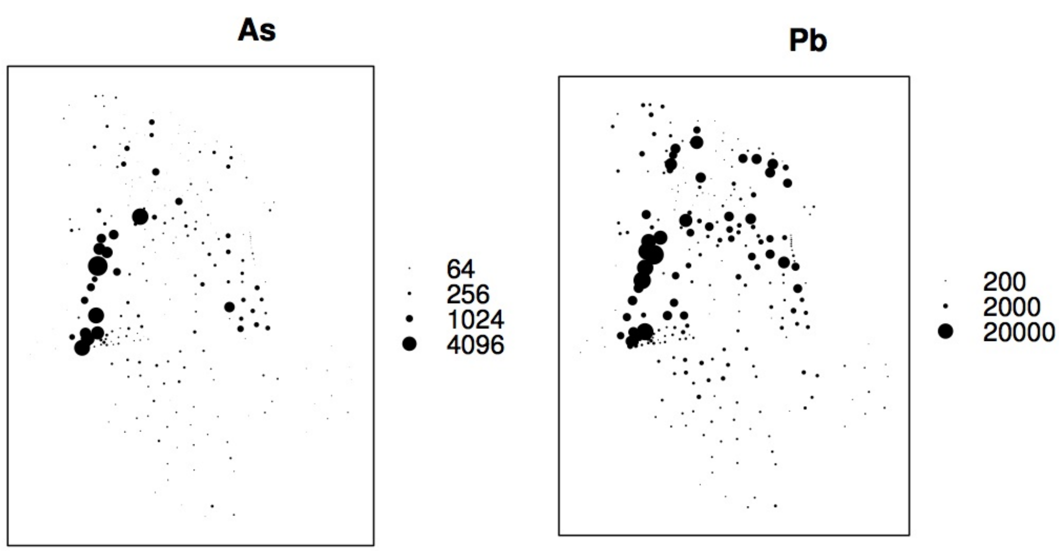

Figure 1: Bubble plots for As (left) and for Pb (right).

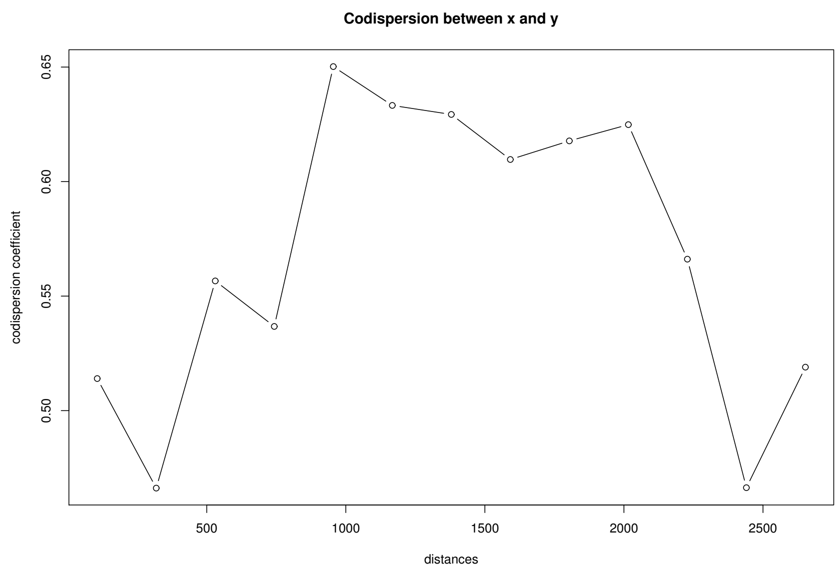

Figure 2: Codispersion coefficient between x=As and y=Pb for the lag distance range from 0 to 2700 m.

The modified t test described in (Clifford et al., 1989) was used to test the absence of spatial association between the variables As and Pb. The R package `SpatialPcak` stipulates $F = 81.9490$, the degrees of freedom 1 and 154.0617 for the numerator and denominator, respectively, of the F distribution, the p-value < 0.0001 and the sample correlation coefficient $r = 0.5893$. Thus, the null hypothesis of no spatial association between the processes is rejected with a 5% level of significance. In addition, the code `summary(murray.ttest)` provides the upper boundaries for each of the thirteen (default) bins used in the computation of the modified t-test, and for each class, the Moran coefficient is also given for both variables (As and Pb).

From Figure 2, it can be emphasized that the values of the codispersion coefficient are in most cases greater than 0.5 (this value has been selected for illustration purposes only). However, it is well known (Rukhin and Vallejos, 2008) that the codispersion coefficient for several spatial models can be written as a constant depending on the spatial parameters of the models times the correlation coefficient between the stochastic errors of each marginal process. Even though the correlation coefficient and the codispersion coefficient are related for a wide class of models, from Figure 2, it is not possible to infer if there is enough evidence in favor of a correlation coefficient greater than a certain value, in this case, 0.5. In practice, it is not simple to test such hypothesis due to the existing spatial dependence in the data. In fact, Clifford et al. test was constructed for this purpose.

An alternative solution to this problem is to assume the data as a realization from a BGRF with a bivariate covariance model of type (4) and perform the following test:

$$\mathrm{H}_0 : \rho_{12} \leq 0.5 \text{ versus } \mathrm{H}_1 : \rho_{12} > 0.5$$

In particular, we consider the model in equation (11) with an increasing level of complexity as the specifications A, B and C described in Section 3. Specifically, for illustration purposes we consider the following:

1. A separable model with a common scale parameter $\beta = \beta_{11} = \beta_{12} = \beta_{22}$.

2. A non-separable model with $\beta_{12} = \frac{\beta_{11}+\beta_{22}}{2}$.

3. A non-separable model without constraints on the parameters.

Table 10 shows the ML estimates of the three models and the corresponding standard errors (in brackets) obtained using the square root of the diagonal entries of the inverse of the Fisher information matrix given in Appendix B for the three cases.

The estimates of the colocated correlation parameter are, respectively, 0.465, 0.488 and 0.517 for specifications A, B, and C. The values of the test statistics $\frac{\hat{\rho}_{12}-0.5}{se(\hat{\rho}_{12})}$ are, respectively, $-0.700, -0.250, 0.374$, and the corresponding respective p-values are $0.77, 0.599, 0.35$. For the three model specifications, the null hypothesis is not rejected considering a significance level of 5%. Even though the codispersion coefficient between As and Pb is greater than 0.5 for most of the values displayed in Figure 2, there is not enough evidence to claim that the correlation between As and Pb is significantly greater than 0.5.

The analysis and the simulation study were carried out using an upcoming release of the R package CompRandFld (Padoan and Bevilacqua, 2015) and the package SpatialPack (Osorio et al., 2014), which are both available on CRAN (http://cran.r-project.org/).

6 Concluding remarks

In this paper, we introduced and discussed a new approach to assess the significance of the correlation between the components of a BGRF. The procedure relies on the maximum likelihood asymptotic distribution of the colocated correlation parameter of a parametric class of bivariate covariance models. One advantage of the test is that the asymptotic variance has a very simple closed form, facilitating the use of the test in practice. The empirical evidence collected from the simulation study using bivariate Matérn and Wendland models in Section 4 highlighted the quality of our proposal with respect to the Clifford et al. test when the true model follows a BGRF. In the simulation study, we also investigated the robustness of the test against possible types of misspecification of the covariance structure. As expected, the results show a small loss in power in this case, but in general, the test is robust against possible misspecification and always performs better than the Clifford et al. test. Moreover, we investigated the behaviour of the test under deviations from the Gaussian distribution. In particular, we considered bivariate random fields of the chi-square and skew Gaussian type. The results of our examples show that strong deviations from Gaussianity can affect the performance of the test, but in general, the test is robust against small deviations from the Gaussian distribution and always performs better than the Clifford et al. test.

The real dataset analyzed in Section 5 indicated that the new test is an interesting alternative to the existing methods and that it provides a new framework to assess the significance between two spatial variables.

The proposed test is based on the distribution asymptotic of the colocated correlation parameter under an increasing domain asymptotic framework. Under an infill asymptotic framework, in the univariate spatial case, it is difficult to derive general results on the asymptotic distribution of the maximum likelihood estimator (see Chen et al. , 2000; Zhang , 2004; Loh , 2005), and no results are available for the bivariate case. It would be interesting to study the asymptotic distribution of the colocated correlation parameter, under infill asymptotic, at least for the separable case. As shown in this paper, under an increasing domain asymptotic framework, the asymptotic distribution is not affected by the spatial dependence, and we believe that this result is still valid under an infill asymptotic framework.

A drawback of the proposed test is the computational cost involved in the maximum likelihood estimation when dealing with large datasets. In this case, likelihood approximations such as a composite likelihood approximation (Varin et al., 2011; Bevilacqua and Gaetan , 2014) could be used to build a test with a good balance between computational complexity and performance of the test. This topic is to be investigated in future research.

Acknowledgements

The research work conducted by Moreno Bevilacqua was supported in part by FONDECYT grant 1121408, Chile. Ronny Vallejos was supported in part by FONDECYT grant 1120048, Chile and AC3E grant FB-0008, UTFSM. The authors thank the editor, an associate editor, and two referees for their comments and suggestions which led to an improved presentation.

References

Apanasovich T, Genton MG. 2010. Cross-covariance functions for multivariate random fields based on latent dimensions. *Biometrika* **97**: 15-30.

Apanasovich T, Genton MG, Sun Y. 2012. A valid Matérn class of cross-covariance functions for multivariate random fields with any number of components. *J. Amer. Statist. Assoc*. **107**: 180-193.

Bevilacqua, M., Hering, A. S., and Porcu, E. 2015. On the flexibility of multivariate covariance models: Comment on the Paper by Genton and Kleiber. *Statistical Science*. 30(2) 167-169.

Bevilacqua M, Gaetan C. 2014. Comparing composite likelihood methods based on pairs for spatial Gaussian random fields. *Statistics and Computing*. Springer US. DOI 10.1007/s11222-014-9460-6.

Blanco-Moreno, HM, Chamorro, C, Sanz, FX. 2006. Spatial and temporal patterns of lolium rigidum-avena sterilis mixed populations in a cereal field, *Eur. Weed Res. Soc. Weed Res.* **46**: 207-218.

Chen HS, Simpson DG, Ying Z. 2000. Fixed-domain asymptotics for a stochastic process model with measurement error. *Statist. Sinica* **10**: 141-156.

Clifford P, Richardson S, Hémon D. 1989. Assessing the significance of the correlation bween two spatial processes. *Biometrics* **45**: 123-134.

Córdoba M, Bruno C, Costa J, Balzarini M. 2013. Subfield management class delineation using cluster analysis from spatial principal components of soil variables. *Computers and Electronics Agriculture* **97**: 6-14.

Cuevas F, Porcu E, Vallejos R. 2013. Study of spatial relationships between two sets of variables: A nonparametric approach. *Journal of Nonparametric Statistics* **25**: 695-714.

Daley DJ, Porcu E, Bevilacqua M. 2014. Classes of compactly supported covariance functions for multivariate random fields. (Submitted).

Dutilleul P. 1993. Modifying the t test for assessing the correlation between two spatial processes. *Biometrics* **49**: 305-314.

Genton MG, Kleiber W. 2014. Cross-Covariance Functions for Multivariate Geostatistics. *Statistical Science*, in press.

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

Griffith D. 2002. The geographic distribution of soil-lead concentration: description and concerns. *URISA Journal* **14**: 5-15.

Griffith D, Paelinck JHP. 2011. *Non-Stardard Spatial Statistics*. Springer, New York.

Goovaerts P. 1997. *Geostatistics for Natural Resources Evaluation*. Oxford University Press, Oxford.

Haining R. 1991. Bivariate Correlation with Spatial Data. *Geographical Analysis* **23**: 210-227.

Lee, S., 2001. Developing a bivariate spatial association measure: An integration of Pearson's r and Moran's I. *Journal of Geographical Systems* **3**: 369-385.

Loh WL. 2005. Fixed-domain asymptotics for a subclass of Matérn-type Gaussian random fields. *Ann. Statist.* **33**: 2344-2394.

Ma C. 2011. Covariance matrix functions of vector chi-square random fields in space and time. *IEEE Transactions on communications* **59**: 2554-2561.

Ma, C., 2013. Student's t Vector Random Fields with Power-Law and Log-Law Decaying Direct and Cross Covariances. *Stochastic Analysis and Applications* **31**: 167-182.

Mardia KV, Marshall RJ. 1984. Maximum Likelihood Estimation of Models for Residual Covariance in Spatial Regression. *Biometrika* **71**: 135-146.

Matheron G. 1965. *Les Variables Régionalisées et leur Estimation*. Masson, Paris.

Ojeda S, Vallejos R, Lamberti P. 2012. Measure of similarity between images based on the codispersion coefficient. *Journal of Electronic Imaging* **21**: 023019.

Osorio F, Vallejos R, 2014. SpatialPack: A package to assess the association between two spatial processes. R package version 0.2-3, `URL:http://cran.r-project.org/web/packages/SpatialPack/`

Padoan S, Bevilacqua M. 2015. Analysis of Random Fields Using CompRandFld. *Journal of Statistical Software* **63**: 1-27.

Porcu E, Daley DJ, Buhmann MD, Bevilacqua M. 2013. Radial basis functions with compact support for multivariate geostatistics. *Stochastic Environmental Research and Risk Assessment* **27**: 909-922.

Pringle MJ, Lark RM. 2006. Spatial analysis of model error, illustrated by soil carbon dioxide emissions. *Vadose Zone Journal* **5**: 168-183.

Rukhin A, Vallejos R, 2008. Codispersion coefficient for spatial and temporal series. *Statistics and Probability Letters* **78**: 1290-1300.

Tjøstheim 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.

Vallejos R. 2012. Testing for the absence of correlation between two spatial or temporal sequences. *Pattern Recognition Letters* **33**: 1741-1748.

Vallejos R, Mallea A, Herrera M, Ojeda SM. 2015. A multivariate geostatistical approach for landscape classification from remotely sensed image data. *Stochastic Environmental Research and Risk Assessment* **29**: 369-365.

Varin C, Reid N, Firth D. 2011. An Overview of Composite Likelihood Methods. *Statistica Sinica* **21**: 5-42.

Viladomat J, Mazumder R, McInturff A, McCauley DJ, Hastie T. 2014. Assessing the significance of global and local correlations under spatial autocorrelation: A nonparametric approach. *Biometrics* **70**: 409-418.

Wackernagel H. 2003. *Multivariate Geostatictics. An Intoduction with Applications*. Third Edition. Springer - Verlag Berlin Heidelberg.

Wendland H. 1995. Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree. *Adv. Comput. Math.* **4**: 389-396.

Zhang H. 2004. Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *J. Amer. Statist. Assoc.* **99**: 250-261.

Zhang H. 2007. Maximum-likelihood estimation for multivariate spatial linear coregionalization models. *Environmetrics* **18**: 125-139.

Zhang H. and El-Shaarawi, A. 2010. On spatial skew-Gaussian processes and applications. *Environmetrics* **21**: 33-47.

	P(Type I Error)			Power								
				$R_{12}(\mathbf{0})=0.05$			$R_{12}(\mathbf{0})=0.15$			$R_{12}(\mathbf{0})=0.25$		
α	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1
P	0.011	0.057	0.113	0.057	0.146	0.228	0.577	0.777	0.856	0.977	0.997	0.999
C	0.012	0.050	0.101	0.023	0.096	0.171	0.260	0.480	0.611	0.722	0.891	0.941
	P(Type I Error)			Power								
				$R_{12}(\mathbf{0})=0.05$			$R_{12}(\mathbf{0})=0.15$			$R_{12}(\mathbf{0})=0.25$		
α	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1
P	0.012	0.057	0.112	0.054	0.143	0.224	0.577	0.779	0.855	0.977	0.997	0.998
C	0.008	0.055	0.107	0.032	0.010	0.177	0.298	0.540	0.652	0.796	0.921	0.958

Table 1: Empirical probability of the type I error and power of the parametric test (P) and Clifford's test (C) when testing hypothesis (8) for a separable bivariate exponential model (top) and a separable bivariate Wendland model (bottom) (scenario 1).

	P(Type I Error)			Power								
				$R_{12}(\mathbf{0})=0.05$			$R_{12}(\mathbf{0})=0.15$			$R_{12}(\mathbf{0})=0.25$		
α	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1
P	0.011	0.055	0.107	0.052	0.138	0.219	0.583	0.789	0.864	0.978	0.996	0.999
C	0.010	0.047	0.093	0.038	0.109	0.188	0.412	0.653	0.766	0.928	0.981	0.994
	P(Type I Error)			Power								
				$R_{12}(\mathbf{0})=0.05$			$R_{12}(\mathbf{0})=0.15$			$R_{12}(\mathbf{0})=0.25$		
α	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1
P	0.010	0.045	0.109	0.059	0.156	0.248	0.649	0.846	0.904	0.994	1	1
C	0.012	0.047	0.090	0.037	0.140	0.201	0.512	0.745	0.841	0.975	0.996	0.997

Table 2: Empirical probability of the type I error and power of the parametric test (P) and Clifford's test (C) when testing hypothesis (8) for a nonseparable bivariate exponential model (top) and a nonseparable bivariate Wendland model (bottom) (scenario 2).

	P(Type I Error)			Power								
				$R_{12}(\mathbf{0}) = 0.20$			$R_{12}(\mathbf{0}) = 0.30$			$R_{12}(\mathbf{0}) = 0.40$		
α	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1
P	0.012	0.057	0.104	0.090	0.241	0.359	0.711	0.881	0.930	0.996	1	1
	P(Type I Error)			Power								
				$R_{12}(\mathbf{0}) = 0.20$			$R_{12}(\mathbf{0}) = 0.30$			$R_{12}(\mathbf{0}) = 0.40$		
α	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1
P	0.011	0.056	0.105	0.092	0.239	0.360	0.723	0.886	0.931	0.996	1	1

Table 3: Empirical probability of the type I error and power of the parametric test (P) when testing hypothesis (9) with $k = 0.15$, for a separable bivariate exponential model (top) and a separable bivariate Wendland model (bottom) (scenario 1).

	P(Type I Error)			Power								
				$R_{12}(\mathbf{0}) = 0.20$			$R_{12}(\mathbf{0}) = 0.30$			$R_{12}(\mathbf{0}) = 0.40$		
α	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1
P	0.010	0.051	0.109	0.061	0.160	0.249	0.645	0.824	0.892	0.995	1	1
	P(Type I Error)			Power								
				$R_{12}(\mathbf{0}) = 0.20$			$R_{12}(\mathbf{0}) = 0.30$			$R_{12}(\mathbf{0}) = 0.40$		
α	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1
P	0.009	0.048	0.097	0.100	0.244	0.367	0.782	0.924	0.966	0.999	1	1

Table 4: Empirical probability of the type I error and power of the parametric test (P) when testing hypothesis (9) with $k = 0.15$, for a nonseparable bivariate exponential model (top) and a nonseparable bivariate Wendland model (bottom) (scenario 2).

	P(Type I Error)			Power								
				$R_{12}(\mathbf{0}) = 0.20$			$R_{12}(\mathbf{0}) = 0.30$			$R_{12}(\mathbf{0}) = 0.40$		
α	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1
P	0.011	0.052	0.106	0.082	0.215	0.345	0.564	0.807	0.890	0.987	0.999	1
	P(Type I Error)			Power								
				$R_{12}(\mathbf{0}) = 0.20$			$R_{12}(\mathbf{0}) = 0.30$			$R_{12}(\mathbf{0}) = 0.40$		
α	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1
P	0.010	0.048	0.104	0.093	0.239	0.374	0.735	0.893	0.949	0.998	1	1

Table 5: Empirical probability of the type I error and power of the parametric test (P) when testing hypothesis (9) with $k = 0.15$, for a nonseparable bivariate exponential model (top) and a nonseparable bivariate Wendland model (bottom) (scenario 3)

	P(Type I Error)			Power								
				$R_{12}(\mathbf{0})=0.05$			$R_{12}(\mathbf{0})=0.15$			$R_{12}(\mathbf{0})=0.25$		
α	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1
P	0.010	0.047	0.094	0.046	0.137	0.217	0.561	0.770	0.857	0.976	0.996	0.999
	P(Type I Error)			Power								
				$R_{12}(\mathbf{0})=0.05$			$R_{12}(\mathbf{0})=0.15$			$R_{12}(\mathbf{0})=0.25$		
α	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1
P	0.019	0.067	0.121	0.058	0.170	0.255	0.573	0.767	0.840	0.991	0.967	0.995
	P(Type I Error)			Power								
				$R_{12}(\mathbf{0})=0.05$			$R_{12}(\mathbf{0})=0.15$			$R_{12}(\mathbf{0})=0.25$		
α	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1
P	0.009	0.050	0.098	0.053	0.154	0.231	0.594	0.798	0.870	0.981	0.999	1
C	0.011	0.050	0.093	0.043	0.152	0.228	0.517	0.743	0.840	0.969	0.994	0.998

Table 6: Empirical probability of the type I error and power of the parametric test (P) when testing hypothesis (8) under covariance misspecification of the first type (top), the second type (middle) and the third type (bottom) .

	P(Type I Error)			Power								
				$R_{12}(\mathbf{0})=0.20$			$R_{12}(\mathbf{0})=0.30$			$R_{12}(\mathbf{0})=0.40$		
α	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1
P	0.010	0.048	0.107	0.053	0.151	0.230	0.611	0.810	0.876	0.994	0.999	1
	P(Type I Error)			Power								
				$R_{12}(\mathbf{0})=0.20$			$R_{12}(\mathbf{0})=0.30$			$R_{12}(\mathbf{0})=0.40$		
α	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1
P	0.015	0.061	0.119	0104	0.260	0.364	0.713	0.875	0.92	0.991	0.999	1
	P(Type I Error)			Power								
				$R_{12}(\mathbf{0})=0.20$			$R_{12}(\mathbf{0})=0.30$			$R_{12}(\mathbf{0})=0.40$		
α	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1
P	0.012	0.055	0.106	0.096	0.262	0.370	0.764	0.907	0.955	0.997	1	1

Table 7: Empirical probability of the type I error and power of the parametric test (P) when testing hypothesis (9) with $k = 0.15$, under covariance misspecification of the first type (top), the second type (middle) and the third type (bottom).

	P(Type I Error)			Power								
				$R_{12}(\mathbf{0}) = 0.05$			$R_{12}(\mathbf{0}) = 0.15$			$R_{12}(\mathbf{0}) = 0.25$		
α	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1
P	0.011	0.038	0.074	0.100	0.183	0.262	0.450	0.650	0.715	0.863	0.929	0.953
C	0.009	0.037	0.074	0.093	0.182	0.245	0.456	0.605	0.690	0.823	0.909	0.938
	P(Type I Error)			Power								
				$R_{12}(\mathbf{0}) = 0.05$			$R_{12}(\mathbf{0}) = 0.15$			$R_{12}(\mathbf{0}) = 0.25$		
α	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1
P	0.012	0.049	0.096	0.054	0.151	0.232	0.556	0.774	0.855	0.976	0.997	0.999
C	0.013	0.044	0.096	0.047	0.136	0.219	0.486	0.725	0.818	0.950	0.991	0.997

Table 8: Empirical probability of the type I Error and power of the parametric test (P) and Clifford's test (C) when testing hypothesis (8) for a bivariate chi-square random fields with $p = 1$ (top) and $p = 15$ (bottom) degrees of freedom.

	P(Type I Error)			Power								
				$R_{12}(\mathbf{0}) = 0.05$			$R_{12}(\mathbf{0}) = 0.15$			$R_{12}(\mathbf{0}) = 0.25$		
α	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1
P	0.008	0.047	0.093	0.028	0.108	0.182	0.402	0.624	0.745	0.919	0.975	0.989
C	0.006	0.047	0.087	0.032	0.121	0.205	0.429	0.684	0.780	0.941	0.986	0.993
	P(Type I Error)			Power								
				$R_{12}(\mathbf{0}) = 0.05$			$R_{12}(\mathbf{0}) = 0.15$			$R_{12}(\mathbf{0}) = 0.25$		
α	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1
P	0.009	0.048	0.098	0.042	0.132	0.209	0.464	0.701	0.806	0.954	0.987	0.995
C	0.007	0.054	0.102	0.029	0.104	0.180	0.305	0.549	0.674	0.837	0.947	0.971

Table 9: Empirical probability of the type I Error and power of the parametric test (P) and Clifford's test (C) when testing hypothesis (8) for a a bivariate skew Gaussian random fields with $\psi_1 = \psi_2 = 2.775$ (top) and $\psi_1 = \psi_2 = 1$ (bottom).

Parameters	(A)	(B)	(C)
ρ_{12}	0.465	0.488	0.517
	(0.049)	(0.047)	(0.047)
σ_1^2	1.028	0.981	0.978
	(0.096)	(0.089)	(0.090)
σ_2^2	0.885	0.922	0.885
	(0.082)	(0.095)	(0.088)
β	98.03	–	–
	(12.30)		
β_{11}	–	61.338	78.997
		(12.959)	(13.710)
β_{22}	–	129.20	124.36
		(20.490)	(22.452)
β_{12}	–	–	143.978
			(18.886)
Likelihood	-644.38	-639.09	-630.01

Table 10: Maximum likelihood parameter estimates with associated standard errors (in brackets) for the bivariate covariance model in equation (11) under settings 1, 2, and 3.

Caption of Figure 1

Bubble plots for As (left) and for Pb (right).

Caption of Figure 2

Codispersion coefficient between x=As and y=Pb for the lag distance range from 0 to 2700 m.