UNIVERSIDADE DE SANTIAGO DE COMPOSTELA
DEPARTAMENTO DE ESTATÍSTICA E INVESTIGACIÓN OPERATIVA

Comparing spatial dependence structures

R. M. Crujeiras, R. Fernández-Casal, W. González-Manteiga

Report 06-04

Reports in Statistics and Operations Research

Comparing spatial dependence structures

Rosa M. Crujeiras
Dept. of Statistics and OR
University of Santiago de Compostela

Rubén Fernández-Casal
Dept. of Mathematics
University of A Coruña

Wenceslao González-Manteiga†*
Dept. of Statistics and OR
University of Santiago de Compostela

ABSTRACT

The aim of this work is to establish statistical methodology in order to analyze changes in the dependence structure for different spatial processes or for a process observed on a regular grid at different time moments. We propose a test statistic for testing the hypothesis $H_0 : f_1 = \ldots = f_L$, where each f_l denotes the spectral density of each process, for $l = 1, \ldots, L$. The test is based on a Cramer-von-Mises functional type test introduced in (Vilar-Fernández and González-Manteiga (2004)) for the regression context. A simulation study and real data application are also provided.

Key words: goodness-of-fit; spatial spectral density; biomonitoring studies.

1 Introduction

Many authors have studied the problem of modelling the dependence structure of spatial data, from both parametric and non-parametric approaches. This problem can be focused from the spatial domain, taking the variogram or the covariogram as the target function. On goodness-of-fit testing for dependence structures, Diblasi and Bowman (2001) propose a test for independence and Maglione and Diblasi (2004) extend the former technique for choosing a valid model for a variogram, based on smoothed versions of the observed variables.

*Corresponding author. Postal address: Departamento de Estatística e Investigación Operativa. Facultade de Matemáticas. Campus Sur. C.P. 15782. Santiago de Compostela (Spain). Phone number: +34981563100-13216. Fax number: +34981597054. †E-mail address: wenceslao@usc.es.

The techniques mentioned above consider the dependence modelled through the covariogram or the variogram. An alternative to these techniques is to descrbie the dependence structure using the spectral density (the Fourier Transform of the covariance function). Spectral techniques are a broadly used tool in time series analysis. Despite its extension to higher dimension problems is not straightforward, this approach is gaining acceptance in spatial data analysis. The spatial periodogram is a non-parametric estimator of the spatial spectral density. It inherits all the properties of the time periodogram (asymptotically unbiasedness and independence) in a natural way. In the spatial setting, Fuentes (2002) considered this approach for modelling non-stationary spatial dependence structure.

Considering this spectral scheme, Crujeiras *et al.* (2006) provide two test statistics, using distances on the spectral and on the log-spectral domain. These testing goodness-of-fit testing techniques take advantage of the representation of the spatial periodogram as the response variable in a multiplicative regression model. By a logarithmic transform, the spatial log-periodogram can be written as the exogenous variable in a regression model, where the regression function is the log-spectral density.

In a quite related goodness-of-fit context, King *et al.* (1991) study the problem of comparing two regression curves under independence and Gaussian errors. The general case of comparing $L \geq 2$ regression curves is studied in Dette and Neumeyer (2001). In Vilar-Fernández and González-Manteiga (2004), the authors provide a goodness-of-fit technique for testing the equality of regression curves, under fixed design and dependent errors. Based on the ideas in Vilar-Fernández and González-Manteiga (2004), the goal of our work is to provide a test statistic for testing the hypothesis that the spectral densities of L observations of a spatial random process are equal. In spatial statistics, the design points for different realizations of a process are, in many cases, the same. For instance, when these realizations represent the evolution of a biomonitoring process along time. In this setting, it is not unreasonable to assume that observations are taken on the same set of locations, along time.

On this scope, Zhu *et al.* (2002) establish a statistical methodology to analyze changes in the spatial cumulative distribution function (SCDF), over time. Under shrinking asymptotics (a mixture between increasing and infilling domain asymptotics), following (Lahiri (1999)), the authors prove asymptotic normal distribution of two test statistics, for comparing two time moments. The first statistic is based on the difference between the empirical versions of the SCDF. The second statistic is a weighted integrated squared difference between the empirical counterparts of the SCDF. Both testing techniques are devoted to the detection of differences over time, but not specifically focused on the detection of changes in the dependence structure, as it is our purpose.

Besides, when studying spatio-temporal processes, it may be interest to check

whether spatial dependence remains invariant along time, that is, the process is temporally stationary. With the testing technique we propose, we can check if simple spatio-temporal models can be used to model the dependence structure in the data.

The application of our technique is related to biomonitoring studies. Biomonitoring studies have been hold over the last years in order to determine levels of heavy metal concentration all over Europe. The accumulation of heavy metals over large areas and long time periods may cause chronic damage to living organisms and it must be thoroughly controlled. In the particular case of Galicia (NW Spain), mosses have been used as biomonotors (see Fernandez *et al.* (2000)). Our main concern is the study of Mercury concentrations.

This paper is organized as follows. In Section 2 we provide some background on spatial spectral methods and nonparametric regression. In Section 3, we introduce the test statistic and discuss its application in practice. In Section 4, we provide some simulation results in order to check the performance of the test and Section 5 is devoted to real data application. A brief summary and some discussion are given in Section 6.

2 Some background.

Let Z_l be a zero mean second-order stationary spatial process, observed on a regular grid D_l, for $l = 1, \ldots, L$. That is, $\{Z_l(\mathbf{s}), \mathbf{s} \in D_l = \mathbf{a}_l + D\}$, with $D = \{1, \ldots, d_1\} \times \{1, \ldots, d_2\}$. The case $\mathbf{a}_1 = \ldots = \mathbf{a}_L$ implies that the processes are observed on the same grid of locations. Denote by $N_d = d_1 d_2$ the number of points in any of the grids D_l, with $l = 1, \ldots, L$. The covariance function of the processes are defined by:

$$C_l(\mathbf{u}) = E(Z_l(\mathbf{s}), Z_l(\mathbf{s}+\mathbf{u})), \quad \mathbf{s}, \mathbf{u} \in \mathbb{Z}^2. \tag{1}$$

Assuming that $\sum_{\mathbf{u}} |C_l(\mathbf{u})| < \infty$, by Khinchin's theorem (e.g. Yaglom (1987)), the covariance function of a stationary random process can be written, for $l = 1, \ldots, L$ as:

$$C_l(\mathbf{u}) = \int_{\Pi^2} e^{-i\mathbf{u}^T\boldsymbol{\lambda}} f_l(\boldsymbol{\lambda}) d\boldsymbol{\lambda}, \quad \Pi^2 = [-\pi, \pi] \times [-\pi, \pi] \tag{2}$$

where f_l, the spectral density, is bounded and continuous for all l and T denotes the transpose operator.

The classical nonparametric estimator of the spectral density is the periodogram, which is given by:

$$I_l(\boldsymbol{\lambda}_\mathbf{k}) = \frac{1}{(2\pi)^2 N_d} \left| \sum_{\mathbf{s} \in D_l} Z_l(\mathbf{s}) e^{-i\mathbf{s}^T\boldsymbol{\lambda}_\mathbf{k}} \right|^2, \tag{3}$$

where $\mathbf{s}^T\boldsymbol{\lambda}_\mathbf{k}$ denotes the scalar product in $\mathbb{R}^2$. The periodogram is usually computed at the set of bidimensional Fourier frequencies, $\boldsymbol{\lambda}_\mathbf{k}^T = (\lambda_{k_1}, \lambda_{k_2})$:

$$\lambda_{k_i} = \frac{2\pi k_i}{d_i}, \quad k_i = 0, \pm 1, \ldots, \pm n_i = \lfloor \frac{d_i - 1}{2} \rfloor, \quad i = 1, 2 \tag{4}$$

and denote by $N = (2n_1 + 1)(2n_2 + 1)$ the number of Fourier frequencies. The periodogram (3) can be also written in terms of the sample covariances as:

$$I_l(\boldsymbol{\lambda}_\mathbf{k}) = \frac{1}{(2\pi)^2} \sum_{\mathbf{u} \in \mathcal{U}} \hat{C}_l(\mathbf{u}) e^{-i\mathbf{u}^T\boldsymbol{\lambda}_\mathbf{k}}, \quad l = 1, \ldots, L \tag{5}$$

where $\mathcal{U} = \{\mathbf{u} = (u_1, u_2); u_i = 1 - d_i, \ldots, d_i - 1, i = 1, 2\}$ and the sample covariances, for Z_l with $l = 1, \ldots, L$, are given by:

$$\hat{C}_l(\mathbf{v}) = \frac{1}{N_d} \sum_{\mathbf{s} \in D_l(\mathbf{v})} Z_l(\mathbf{s}) Z_l(\mathbf{s} + \mathbf{v}), \quad D_l(\mathbf{v}) = \{\mathbf{s} \in D_l; \mathbf{s} + \mathbf{v} \in D_l\}. \tag{6}$$

We will suppose that the spatial process Z_l can be represented as:

$$Z_l(\mathbf{s}) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_{jk}^l \varepsilon_l(s_1 - j, s_2 - k), \tag{7}$$

where the error variables ε_l are independent and identically distributed as $N(0, \sigma_{\varepsilon_l}^2)$, for $l = 1, \ldots, L$. Note that any Gaussian stationary process can be represented as in (7). Then, the corresponding spectral density f_l can be written as:

$$f_l(\boldsymbol{\lambda}) = |A_l(\boldsymbol{\lambda})|^2 f_{\varepsilon_l}(\boldsymbol{\lambda}), \quad \boldsymbol{\lambda} \in \Pi^2 \tag{8}$$

where $f_{\varepsilon_l}(\boldsymbol{\lambda}) = \dfrac{\sigma_{\varepsilon_l}^2}{(2\pi)^2}$ and

$$A_l(\boldsymbol{\lambda}) = \sum_{j=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} \psi_{jk}^l e^{-i(j,k)\boldsymbol{\lambda}}, \quad (j,k)\boldsymbol{\lambda} = j\lambda_1 + k\lambda_2.$$

In this case, the periodogram for each process Z_l, with $l = 1, \ldots, L$, admits the following representation:

$$I_l(\boldsymbol{\lambda}_\mathbf{k}) = f_l(\boldsymbol{\lambda}_\mathbf{k}) V_\mathbf{k}^l + R_N^l(\boldsymbol{\lambda}_\mathbf{k}), \tag{9}$$

where the variables $V_\mathbf{k}^l$ are i.i.d. standard exponential distributed, and $V_\mathbf{k}^l$ and $V_\mathbf{k}^{l'}$, with $l \neq l'$ are also independent. The residual term $R_N(\boldsymbol{\lambda}_\mathbf{k})$ is uniformly bounded (see Crujeiras *et al.* (2006)). Applying logarithms in (9) we have:

$$Y_\mathbf{k}^l = m_l(\boldsymbol{\lambda}_\mathbf{k}) + z_\mathbf{k}^l + r_\mathbf{k}^l, \quad l = 1, \ldots, L \tag{10}$$

where $m_l = \log f_l$ is the log-spectral density, the variables $z_\mathbf{k}^l = \log V_\mathbf{k}^l$ are i.i.d. with density function $h(x) = e^{x - e^x}$, and the residual term $r_\mathbf{k}^l$ is given by:

$$r_\mathbf{k}^l = \log\left(1 + \frac{R_N^l(\boldsymbol{\lambda}_\mathbf{k})}{f_l(\boldsymbol{\lambda}_\mathbf{k}) V_\mathbf{k}^l}\right).$$

Our main purpose is to test whether the spatial spectral densities are the same, or equivalently, in terms of the spatial log-spectral densities:

$$\begin{array}{ll} H_0: & m_1 = \ldots = m_L, \\ H_a: & m_l \neq m_j, \text{ for some } l \neq j. \end{array} \tag{11}$$

In this context, the comparisson can be made by considering nonparametric estimators of the spatial log-spectral densities. Several nonparametric estimators of the spatial log-spectral density could be obtained considering a smoothed combination of log-periodogram values, that is:

$$\hat{m}_l(\boldsymbol{\lambda}_\mathbf{k}) = \sum_{\mathbf{i}} W_\mathbf{i}^l(\boldsymbol{\lambda}_\mathbf{k}) Y_\mathbf{i}^l. \tag{12}$$

The weights $W_\mathbf{i}^l$ can be defined as Gasser-Muller weights, for instance:

$$W_\mathbf{i}^l(\boldsymbol{\lambda}) = |H|^{-1/2} \int_{A_\mathbf{i}} K(H^{-1/2}(\boldsymbol{\lambda} - \boldsymbol{\mu}))d\boldsymbol{\mu}, \tag{13}$$

where K is a bidimensional kernel function, H is a bidimensional bandwidth matrix and the integration region is given by:

$$A_\mathbf{i} = [a_{i_1-1}, a_{i_1}] \times [a_{i_2-1}, a_{i_2}], \quad \boldsymbol{\lambda}_\mathbf{i} \in A_\mathbf{i}, \quad \cup_i A_i = A, \quad A_i \cap A_j = \oslash, i \neq j.$$

The sets A_i in the partition of A must be Jordan measurable and $\max_i \mu(A_i) = \mathcal{O}(N^{-1})$ (see Müller (1988)). Another options are Priestley-Chao weights:

$$W_\mathbf{i}^l(\boldsymbol{\lambda}) = \frac{\pi^2}{N} K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_\mathbf{i}) = \frac{\pi^2}{N|H|^{1/2}} K(H^{-1/2}(\boldsymbol{\lambda} - \boldsymbol{\lambda}_\mathbf{i})), \tag{14}$$

and Nadaraya-Watson weights:

$$W_\mathbf{i}^l(\boldsymbol{\lambda}) = \frac{K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_\mathbf{i})}{\sum_\mathbf{i} K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_\mathbf{i})} = \frac{K(H^{-1/2}(\boldsymbol{\lambda} - \boldsymbol{\lambda}_\mathbf{i}))}{\sum_\mathbf{i} K(H^{-1/2}(\boldsymbol{\lambda} - \boldsymbol{\lambda}_\mathbf{i}))}. \tag{15}$$

Another alternative consists of considering a local-linear estimator for the spatial log-spectral density. A pilot local-linear estimator for $m_l(\boldsymbol{\lambda})$, with $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)^T$, is obtained by multivariate local linear least squares regression by minimizing:

$$\sum_\mathbf{k} \left(Y_\mathbf{k}^l - (\beta_0, \beta_{10}, \beta_{01}) \begin{pmatrix} 1 \\ \lambda_1 - \lambda_{k1} \\ \lambda_2 - \lambda_{k2} \end{pmatrix} \right)^2 K_H(\boldsymbol{\lambda} - \boldsymbol{\lambda}_\mathbf{k}), \tag{16}$$

and the nonparametric estimator is given by $\hat{m}^l(\boldsymbol{\lambda}) = \hat{\beta}_0$, where $(\hat{\beta}_0, \hat{\beta}_{10}, \hat{\beta}_{0,1})$ is the argument that minimizes expression (16).

The previous expressions for a nonparametric estimator of the log-spectral density come from the nonparametric regression context. Back to model (10),

and taking into account the distribution of the error variables $z_{\mathbf{k}}^l$, the log-likelihood associated with this model, ignoring the term $r_{\mathbf{k}}^l$, is given by:

$$\sum_{\mathbf{k}} \left(Y_{\mathbf{k}}^l - m_l(\boldsymbol{\lambda}_{\mathbf{k}}) - e^{Y_{\mathbf{k}}^l - m_l(\boldsymbol{\lambda}_{\mathbf{k}})} \right), \quad l = 1, \ldots, L. \tag{17}$$

From a nonparametric approach, consider the estimator obtained for the spatial log-spectral density m_l by a multidimensional local linear kernel estimator. For $\mathbf{x} \in \mathbb{R}^2$, we approximate $m_l(\boldsymbol{\lambda}_{\mathbf{k}})$ by the plane $a_l + \mathbf{b}_l^T(\boldsymbol{\lambda}_{\mathbf{k}} - \mathbf{x})$. Therefore, a local log-likelihood function based on (17) is given by:

$$\sum_{\mathbf{k}} \left(Y_{\mathbf{k}}^l - a_l - \mathbf{b}_l^T(\boldsymbol{\lambda}_{\mathbf{k}} - \mathbf{x}) - e^{Y_{\mathbf{k}}^l - a_l - \mathbf{b}_l^T(\boldsymbol{\lambda}_{\mathbf{k}} - \mathbf{x})} \right) K_H(\boldsymbol{\lambda}_{\mathbf{k}} - \mathbf{x}), \quad l = 1, \ldots, L \tag{18}$$

and take the maximum local log-likelihood estimator $\hat{m}_l(\mathbf{x})$ of $m_l(\mathbf{x})$ as $\hat{a}_l$ in the maximizer $(\hat{a}_l, \hat{\mathbf{b}}_l)$ of (18). These nonparametric estimators are used to illustrate the simulation study and the real data application.

In some cases, we may know some features of the dependence structures. For instance, by applying a goodness-of-fit test as those proposed in Crujeiras *et al.* (2006), before testing the equality of the spatial spectral densities, we could assess whether the spectral densities belong to the same parametric family. In that case, the testing problem would be stated as:

$$\begin{array}{ll} H_0: & \theta_1 = \ldots = \theta_L, \\ H_a: & \theta_l \neq \theta_j, \quad \text{for some } l \neq j, \end{array} \tag{19}$$

with $m_{\theta_l} = \log f_{\theta_l}$, $l = 1, \ldots, L$ and $f_{\theta_l} \in \mathcal{F}_\Theta$, where $\mathcal{F}_\Theta$ denotes a parametric family of spectral densities.

3 A new test for comparing spatial log-spectral densities.

As we have already commented in the introduction, Zhu *et al.* (2002) develop hypothesis testing methods to detect a difference in a spatial random process, at two different time points. The testing techniques are based on the SCDF. This random function provides a spatial statistical summary of the random field and it is defined as:

$$F_{\infty,Z}(z; R) = \frac{1}{|R|} \int_R \mathbf{1}(Z(\mathbf{s}) \leq z) d\mathbf{s}, \tag{20}$$

where $\{Z(\mathbf{s}), \mathbf{s} \in R\}$, with $\mathbf{s}$ a continuous spatial index, $R \subset \mathbb{R}^d$, $|R|$ denotes the volume of R and $\mathbf{1}$ is the indicator function. For a finite sample of the process

$\{Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_N)\}$, the empirical counterpart of the SCDF in (20), namely the empirical cumulative distribution function (ECDF) is given by:

$$F_{N,Z}(z; R) = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}(Z(\mathbf{s}_i) \leq z). \tag{21}$$

In Lahiri (1999), theoretical results on the SCDF are given, considering a shrinking asymptotic framework (Cressie (1993), pp.100-101). With the motivation of detecting changes or trends in ecological resources over time, and for the particular case of two time points, the authors derive the large-sample distribution of a normalized test statistic based on the difference of the two ECDFs. A second procedure quantifies the change using a weighted integrated squared difference between the SCDFs.

In the spatial context, we do not know more references on this topic of spatial processes comparison. In our case, we are interested in detecting changes on the dependence structure and, for that purpose, we will consider a spectral approach.

3.1 The test.

Consider the following test statistic, based on a L^2-distance:

$$Q = \sum_{l=2}^{L} \left(\sum_{j=1}^{l-1} \left(\int_{\Pi^2} \left(\widehat{m_l}(\boldsymbol{\lambda}) - \widehat{m_j}(\boldsymbol{\lambda}) \right)^2 \omega(\boldsymbol{\lambda}) d\boldsymbol{\lambda} \right) \right), \tag{22}$$

where ω is a weighting function in Π^2. This weighting function ω is chosen in such a way the edge effect on the estimation is avoided. In our context, we consider a weighting function that filters frequencies around the origin and those one with components near to $\frac{2\pi n_i}{d_i}$. Besides, it is in these cases where the log-periodogram values may present a higher variability. This edge-effect bias is also corrected by the local linear estimator.

For simplicity, consider the testing problem $H_0 : m_1 = m_2$ vs. $H_a : m_1 \neq m_2$. In the general case of L processes, we proceed similarly. Assume that both Z_1 and Z_2 have been observed on grids with the same design. This implies that the corresponding Fourier frequencies are the same in both cases. Using Riemann approximation, Q can be approximated by $\hat{Q}$, which is given by:

$$\hat{Q} = \frac{(2\pi)^2}{N} \sum_{\mathbf{k}} \left(\widehat{m_1}(\boldsymbol{\lambda}_\mathbf{k}) - \widehat{m_2}(\boldsymbol{\lambda}_\mathbf{k}) \right)^2 \omega(\boldsymbol{\lambda}_\mathbf{k}). \tag{23}$$

In order to perform the test in practice, the distribution of the test statistic under the null hypothesis H_0 is needed. The asymptotic behaviour of $\hat{Q}$, under H_0, could be establish but usually, the convergence of this type of test statistic to its limit distribution is slow (see, for instance, some works in the regression

setting as Härdle and Mammen (1993)). Therefore, this distribution must be approximated by simulation. For that purpose, it is necessary to obtain an estimation of the spatial log-spectral density under $H_0 : m_1 = m_2 = m$. In the case we consider, the design points (that is, the grid of Fourier frequencies) are the same, and we could build a pilot estimator of the log-spectral density under H_0, namely $\hat{m}$, as an average of the log-periodograms:

$$\tilde{Y}_{\mathbf{k}} = \frac{\left(Y_{\mathbf{k}}^1 + Y_{\mathbf{k}}^2\right)}{2}. \tag{24}$$

If our null hypothesis can be relaxed by the fact that $f_{\theta_1}, f_{\theta_2} \in \mathcal{F}_\Theta$, as in problem (19), then, an estimation of the log-spectral density under $H_0 : m_{\theta_1} = m_{\theta_2} = m_\theta$ can be given by a parametric estimator $m_{\hat{\theta}}$. The parameter vector θ can be estimated, under H_0, by a Whittle log-likelihood approach:

$$\hat{\theta} = \arg\max_{\theta} \sum_{\mathbf{k}} \left(\tilde{Y}_{\mathbf{k}} - m_\theta(\boldsymbol{\lambda}_{\mathbf{k}}) - e^{\tilde{Y}_{\mathbf{k}} - m_\theta(\boldsymbol{\lambda}_{\mathbf{k}})} \right). \tag{25}$$

Whittle estimators are not consistent for dimensions higher or equal to two, and in the case of dimension two, these estimators show a non negligible bias. Different alternatives to achieve consistency in this estimation procedure can be found in Guyon (1982), Dahlhaus and Künsch (1987) or Crujeiras *et al.* (2006).

In order to compute the test statistic (23), we must fix a bandwidth matrix H. The selection of the bandwidth matrix parameter is a crucial step in nonparametric estimation and testing. Nevertheless, the choice of optimal bandwidth matrices in multidimensional testing problems remains unsolved and usually, the standard approach consists of examining the behaviour of the test over a range of bandwidths.

Instead of trying a range of bandwidths, an automatic bandwidth selection criteria could be also used. For instance, we could take an optimal bandwidth matrix for the estimation problem, under H_0.

Since log-periodogram values are asymptotically independent, for a large enough sample, we may expect good approximations of $\hat{H}$ by using a cross-validation criteria. For the testing problem (19), the bandwidth matrix $\hat{H}$ may be selected such that:

$$\hat{H} = \arg\min_{H} \sum_{\mathbf{k}} \left(\hat{m}^{-\mathbf{k}}(H, \boldsymbol{\lambda}_{\mathbf{k}}) - m_{\hat{\theta}}(H, \boldsymbol{\lambda}_{\mathbf{k}}) \right)^2, \tag{26}$$

where $\hat{m}^{-\mathbf{k}}(H, \cdot)$ is the nonparametric estimator of the spatial log-spectral density obtained when ignoring the frequency $\boldsymbol{\lambda}_{\mathbf{k}}$ for obtaining the nonparametric estimator of m at this frequency.

In the nonparametric testing problem (11), the bandwidth matrix could be obtained from:

$$\hat{H} = \arg\min_{H} \sum_{\mathbf{k}} \left(\hat{m}^{-\mathbf{k}}(H, \boldsymbol{\lambda}_{\mathbf{k}}) - \tilde{Y}_{\mathbf{k}} \right)^2 . \tag{27}$$

3.2 Bootstrap procedure for calibrating p-values.

Consider the testing problem (19). This *a priori* information will simplify the algorithm for callibrating the p-value of the test in practice. An estimation of the test statistic, under H_0, can be given by a Monte Carlo approach. In order to calibrate the p-value of the test statistic $\hat{Q}$, the following algorithm can be employed in practice.

Algorithm 1.

Step 1. Compute the observed test statistic $\hat{Q}^{obs}$.

Step 2. Draw two random samples of size $d_1 \times d_2$, with the log-spectral density under H_0, that is, $m_{\hat{\theta}}$.

Step 3. Compute the test statistic for these generated random samples $\hat{Q}^{(b)}$.

Step 4. Repeat *Step 1* and *Step 2* B times and obtain the tests statistic $\hat{Q}^{(1)}, \ldots, \hat{Q}^{(B)}$.

Step 5. Compute the p-value of the test statistic as the percentage of bootstrap replicates $\{\hat{Q}^{(1)}, \ldots, \hat{Q}^{(B)}\}$ that exceed the observed value $Q^{(obs)}$.

In this algorithm, a parametric estimation of the spatial log-spectral density is needed in *Step 2*. This parametric estimator is usually obtained by a Whittle-log likelihood approach given in (25).

Also in *Step 2*, in order to generate a random sample from a spatial process with a certain spatial spectral density, one could use an specific algorithm, for instance, when we consider linear-by-linear process (see Section 4). When an specific algorithm is not available, then we must use a standard technique for the simulation of spatial processes. In this case, Cholesky factorization based method (see Cressie (1993), for example) could be used. Another alternative is spectral simulation procedures, as the Modified Fourier Integral Method (Crujeiras and Fernández-Casal (2006)).

3.3 A completely nonparametric Bootstrap procedure for calibrating p-values.

In the nonparametric testing problem, when we ignore whether the spectral densities belong to the same parametric familiy, a totally nonparametric algorithm for approximating the p-value of the test must be considered.

Algorithm 2.

Step 1. Compute the observed test statistic $\hat{Q}^{obs}$.

Step 2. Draw two random samples of size N_d, with the log-spectral density under $H_0 : m_1 = m_2$ as follows:

2.a Obtain a non parametric estimation $\hat{m}$ of the log-spectral density. For example compute the log-periodograms, taking the average at each frequency (as in (24)) and smooth this average to get $\tilde{Y}^s_{\mathbf{k}}$ (see Robinson (2006)).

2.b Apply the Inverse Fourier Transform on $\tilde{I}^s(\boldsymbol{\lambda}_{\mathbf{k}}) = e^{\tilde{Y}^s_{\mathbf{k}}}$ and get an estimation of the covariance function $\hat{C}(\mathbf{u})$, with $\mathbf{u} \in \{\mathbf{u} = (u_1, u_2) : u_1 = 1 - d_1, \ldots, d_1 - 1, u_2 = 1 - d_2, \ldots, d_2 - 1\}$.

2.c Obtain two realizations of the process, on a grid of size N_d, from the estimated covariances, $\hat{C}$.

Step 3. Compute the test statistic for these generated random samples $\hat{Q}^{(b)}$.

Step 4. Repeat *Step 1* and *Step 2* B times and obtain the tests statistic $\hat{Q}^{(1)}, \ldots, \hat{Q}^{(B)}$.

Step 5. Compute the p-value of the test statistic as the percentage of bootstrap replicates $\{\hat{Q}^{(1)}, \ldots, \hat{Q}^{(B)}\}$ that exceed the observed value $\hat{Q}^{(obs)}$.

In *Step 2.a* we must take into account that, in order to generate a sample on a grid $\{1, \ldots, d_1\} \times \{1, \ldots, d_2\}$ the covariances $\hat{C}(\mathbf{u})$ must cover a wider grid of size $\{1, \ldots, k_1\} \times \{1, \ldots, k_2\}$, with $k_i = 2d_i - 1$, $i = 1, 2$ (see Priestley (1981)). Therefore, the spectrum must be computed in a finner grid of frequencies. In *Step 2.c*, under the assumption of Gaussian data, Cholesky factorization method could be applied.

These algorithms can be easily generalized to the general case of checking for differences within a collection of $L > 2$ processes, or $L > 2$ observations of the same process.

4 Simulation results.

We illustrate the performance of the test statistic considering a particular class of spatial processes, the bidimensional autoregressive process (from now on $BAR(1)$):

$$Z_l(i,j) = \beta_1^l Z(i-1,j) + \beta_2^l Z(i,j-1) - \beta_1^l \beta_2^l Z(i-1,j-1) + \varepsilon_l(i,j), \quad l = 1, 2, \tag{28}$$

where $\varepsilon_l(i,j)$ are independent identically distributed Gaussian random variables, with zero mean and variance $\sigma^2_{\varepsilon_l}$. This is a particular case in the class of linear-by-linear processes, introduced by Martin (1979), also known as the doubly-geometric process. Parameters β_1^l and β_2^l, for $l = 1, 2$, belong to $[0, 1)$ in order

	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
$\boldsymbol{\beta}^1 = \boldsymbol{\beta}^2 = (0.0, 0.0)$	0.010	0.050	0.093
$\boldsymbol{\beta}^1 = \boldsymbol{\beta}^2 = (0.3, 0.3)$	0.007	0.033	0.076
$\boldsymbol{\beta}^1 = \boldsymbol{\beta}^2 = (0.6, 0.6)$	0.010	0.041	0.081
$\boldsymbol{\beta}^1 = \boldsymbol{\beta}^2 = (0.9, 0.9)$	0.048	0.107	0.193

Table 1: Size of the test, with **Algorithm 1**. 20×20 grid. $\boldsymbol{\beta}^j$ parameter vector, in model (28), for the sample from Z_j, $j = 1, 2$. Significance level α.

to guarantee stationarity. The spectral densities corresponding to Z_1 and Z_2 are given by:

$$f_l(\boldsymbol{\lambda}) = \frac{\sigma_\varepsilon^2}{(2\pi)^2} \cdot \frac{1}{1 + (\beta_1^l)^2 - 2\beta_1^l \cos(\lambda_1)} \cdot \frac{1}{1 + (\beta_2^l)^2 - 2\beta_2^l \cos(\lambda_2)}, \quad l = 1, 2. \tag{29}$$

In order to study the performance of the test, in terms of size and power, we consider different values for β_1^l and β_2^l, from 0.0 (corresponding to the case of independence) to 0.9. One thousand replicates of the process are generated on a 20×20 regular grid. Random sample generations of this process are obtained as in Alonso *et al.* (1996).

We set the null hypothesis that Z_1 and Z_2 are $BAR(1)$ processes with the same dependence structure, that is, testing problem (19). Therefore, *Algorithm 1* is implemented in this case. A multiplicative Epanechnikov bidimensional kernel is considered. The weighting filters the frequencies near the origin and those with the largest components, in order to avoid the edge effect. The bandwidth parameter is chosen using the cross-validation criteria (26). We consider diagonal bandwidth matrices, whose elements are proportional to the spacing between frequencies, that is:

$$H = r \cdot diag\left(\frac{2\pi}{n_1}, \frac{2\pi}{n_2}\right). \tag{30}$$

The nonparametric estimator for the spatial log-spectral density is obtained from the local-linear method, specified in equation (16).

The size of the test is shown in Table 1, at three different significance levels: 0.01, 0.05 and 0.10. $\boldsymbol{\beta}^j = (\beta_1^j, \beta_2^j)$, for $j = 1, 2$, denote the parameters in Z_1 and Z_2, respectively. The percentage of rejections of the test statistic is computed from 1000 simulations. Some results on the power of the test are shown in Table 2. The test shows a good behaviour in all cases.

When no a priori knwoledge on the form of the spectral densities is available, then *Algorithm 2* must be implemented. Under the same conditions on the size of the grid, kernel function, bandwidth selection and number of Bootstrap

	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
$\boldsymbol{\beta}^1 = (0.0, 0.0), \boldsymbol{\beta}^2 = (0.05, 0.05)$	0.015	0.054	0.112
$\boldsymbol{\beta}^1 = (0.0, 0.0), \boldsymbol{\beta}^2 = (0.1, 0.1)$	0.017	0.065	0.133
$\boldsymbol{\beta}^1 = (0.0, 0.0), \boldsymbol{\beta}^2 = (0.2, 0.2)$	0.067	0.196	0.293
$\boldsymbol{\beta}^1 = (0.0, 0.0), \boldsymbol{\beta}^2 = (0.3, 0.3)$	0.290	0.550	0.670
$\boldsymbol{\beta}^1 = (0.0, 0.0), \boldsymbol{\beta}^2 = (0.6, 0.6)$	0.990	0.990	1.000
$\boldsymbol{\beta}^1 = (0.0, 0.0), \boldsymbol{\beta}^2 = (0.9, 0.9)$	1.000	1.000	1.000
$\boldsymbol{\beta}^1 = (0.3, 0.3), \boldsymbol{\beta}^2 = (0.6, 0.6)$	0.220	0.500	0.590
$\boldsymbol{\beta}^1 = (0.3, 0.3), \boldsymbol{\beta}^2 = (0.9, 0.9)$	0.910	0.980	0.990
$\boldsymbol{\beta}^1 = (0.6, 0.6), \boldsymbol{\beta}^2 = (0.9, 0.9)$	0.200	0.380	0.520

Table 2: Power of the test, with **Algorithm 1**. 20×20 grid. $\boldsymbol{\beta}^j$ parameter vector, in model (28), for the sample from Z_j, $j = 1, 2$. Significance level α.

	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
$\boldsymbol{\beta}^1 = \boldsymbol{\beta}^2 = (0.0, 0.0)$	0.014	0.049	0.095
$\boldsymbol{\beta}^1 = \boldsymbol{\beta}^2 = (0.3, 0.3)$	0.011	0.046	0.102
$\boldsymbol{\beta}^1 = \boldsymbol{\beta}^2 = (0.6, 0.6)$	0.015	0.056	0.099
$\boldsymbol{\beta}^1 = \boldsymbol{\beta}^2 = (0.9, 0.9)$	0.057	0.131	0.210

Table 3: Size of the test, with **Algorithm 2**. 20×20 grid. $\boldsymbol{\beta}^j$ parameter vector, in model (28), for the sample from Z_j, $j = 1, 2$. Significance level α.

replicates, we run new simulations. In this case, Z_1 and Z_2 are simulated from model (28) but we do not use the fact that both spectral densities belong to the same family.

Results of the test, using the completely nonparametric algorithm, are given in Table 3 and 4. Apparently, the nonparametric algorithm provides as good results as the parametric one, both in terms of size and power. The behaviour in terms of power for the nonparametric algorithm, in a 40×40 regular grid, is shown in Table 5.

5 Real data application.

The use of mosses as biomonitors has been proved to be a useful way of determining levels of atmospheric deposition since the uptake of heavy metals in mosses occurs mainly from this source. This biomonitoring technique was developed in the 1960s and since then it has been used to determine heavy metal deposition both in large scale studies in different countries or in areas close to industrial zones. This technique was first used in Galicia (NW Spain) in 1995

	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
$\boldsymbol{\beta}^1 = (0.0, 0.0), \boldsymbol{\beta}^2 = (0.05, 0.05)$	0.015	0.054	0.109
$\boldsymbol{\beta}^1 = (0.0, 0.0), \boldsymbol{\beta}^2 = (0.1, 0.1)$	0.015	0.069	0.130
$\boldsymbol{\beta}^1 = (0.0, 0.0), \boldsymbol{\beta}^2 = (0.2, 0.2)$	0.074	0.207	0.317
$\boldsymbol{\beta}^1 = (0.0, 0.0), \boldsymbol{\beta}^2 = (0.3, 0.3)$	0.326	0.579	0.700
$\boldsymbol{\beta}^1 = (0.0, 0.0), \boldsymbol{\beta}^2 = (0.6, 0.6)$	0.994	0.999	1.000
$\boldsymbol{\beta}^1 = (0.0, 0.0), \boldsymbol{\beta}^2 = (0.9, 0.9)$	1.000	1.000	1.000
$\boldsymbol{\beta}^1 = (0.3, 0.3), \boldsymbol{\beta}^2 = (0.6, 0.6)$	0.252	0.489	0.638
$\boldsymbol{\beta}^1 = (0.3, 0.3), \boldsymbol{\beta}^2 = (0.9, 0.9)$	0.927	0.983	0.995
$\boldsymbol{\beta}^1 = (0.6, 0.6), \boldsymbol{\beta}^2 = (0.9, 0.9)$	0.265	0.469	0.590

Table 4: Power of the test, with **Algorithm 2**. 20×20 grid. $\boldsymbol{\beta}^j$ parameter vector, in model (28), for the sample from Z_j, $j = 1, 2$. Significance level α.

	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
$\boldsymbol{\beta}^1 = (0.0, 0.0), \boldsymbol{\beta}^2 = (0.05, 0.05)$	0.034	0.112	0.192
$\boldsymbol{\beta}^1 = (0.0, 0.0), \boldsymbol{\beta}^2 = (0.1, 0.1)$	0.202	0.390	0.518
$\boldsymbol{\beta}^1 = (0.0, 0.0), \boldsymbol{\beta}^2 = (0.2, 0.2)$	0.910	0.974	0.988

Table 5: Power of the test, with **Algorithm 2**. 40×40 grid. $\boldsymbol{\beta}^j$ parameter vector, in model (28), for the sample from Z_j, $j = 1, 2$. Significance level α.

(Fernandez *et al.* (2000)), in order to establish a sampling network which allows the whole area to be monitored, following the Scandinavian programme of biomonitoring using moss. In 2004, samples of two types of mosses, *Scleropodium purum* and *Hypnum cupressiforme* were collected on a grid with 148 points. Concentrations of different heavy metals were analyzed, but we will focus our attention on Hg (Mercury).

Mercury is not a common element in earth's crust. However, since mercury does not blend geochemically with elements in the crustal mass, Hg ores can be highly concentrated. Besides, Hg is a bioaccumulative toxin and it is easily absorbed through the skin, respiratory and gastrointestinal tissues, so the exposure to high Hg concentrations produces toxic effects on human beings. Hg depositions are typically associated with chlor-alkali plants and electricity stations. In the particular case of Galicia, there exist two power plants in the N and a chlor-alkali plant in the SW.

In 2004, measurements of Hg (in parts per billion) were taken in March and September, on a regular grid over Galicia. Our maing goal is to check whether the dependence structure in the data observed in March and September is the same. In this case, we consider two nonparametric estimators for the spatial log-spectral density. First, we consider a local linear estimator, given by (16). Secondly, the Whittle estimator from (18) is used. The kernel function is a multiplicative Epanechnikov kernel and the weighting function is chosen to avoid the edge-effect. The algorithm for approximating the p-value of the test statistic is the nonparametric Bootstrap method in Section 3.2.

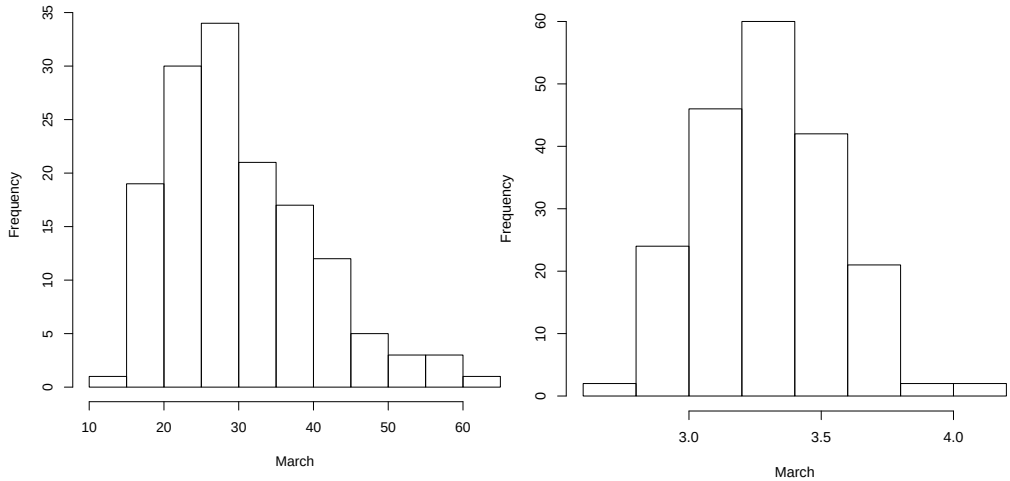

Figure 1: Left panel: histogram of Hg concentrations in March. Right panel: histogram of $log(Hg)$ concentrations in March.

Figure 1 show the histograms of the concentrations of Hg in March (left panel) and the logarithmic transformation of the data (right panel). Similar

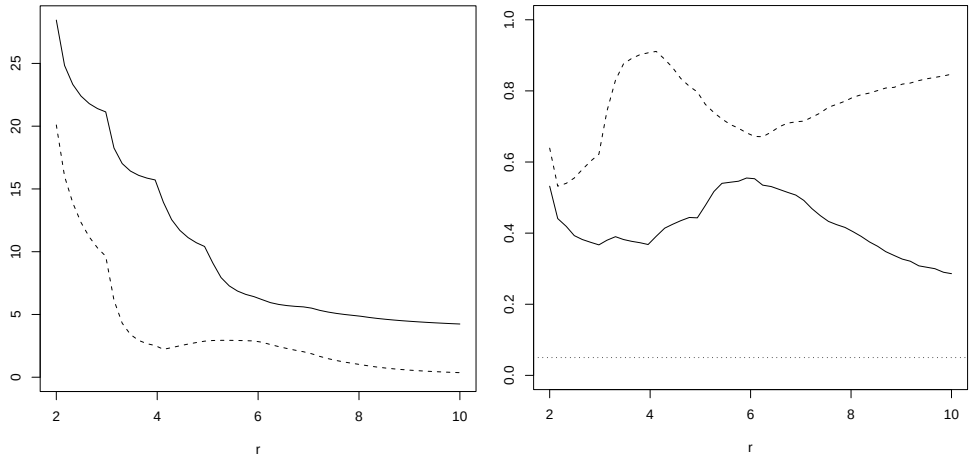

Figure 2: Left panel: test statistics. Right panel: p-values. Solid line: based on the local linear estimator (16). Dashed line: based on local loglikelihood estimator (18). Dotted line: significance level 0.05. r denotes the scaling parameter in (30).

results are obtained for data collected in September. In Figure 2 we show the values of the tests (right panel) and the corresponding p-values (left panel) along a range of bandwidths. There is no evidence that the dependence structure in Hg concentrations has changed from March to September. In Figure (3), we show the results of the tests and the corresponding p-values when applying a logarithmic transform to the data.

6 Summary and discussion.

We propose a testing technique to check whether the spectral densities of L observations of a spatial random process are equal. As a particular case, one could also checked in observations of a process over L time moments exhibit the same dependence structure.

The technique is based on a weighted L_2 distance between nonparametric estimators of the log-spectral densities. These nonparametric estimators can be smoothed versions of the log-periodogram values, using classical weights from nonparametric regression as Gasser-Muller or Nadaraya-Watson. On this context, we consider a local linear least squares estimator of the log-spectral density. Since log-periodogram values are not normally distributed, least squares methods may not be suitable and a better estimator could be obtained by local log-likelihood maximization, as in (18). Nevertheless, from our experience, we can say that tests based on local linear least squares or tests based on local log-likelihood estimators show a similar behaviour.

In order to approximate the p-value of the test in practice, we provide an

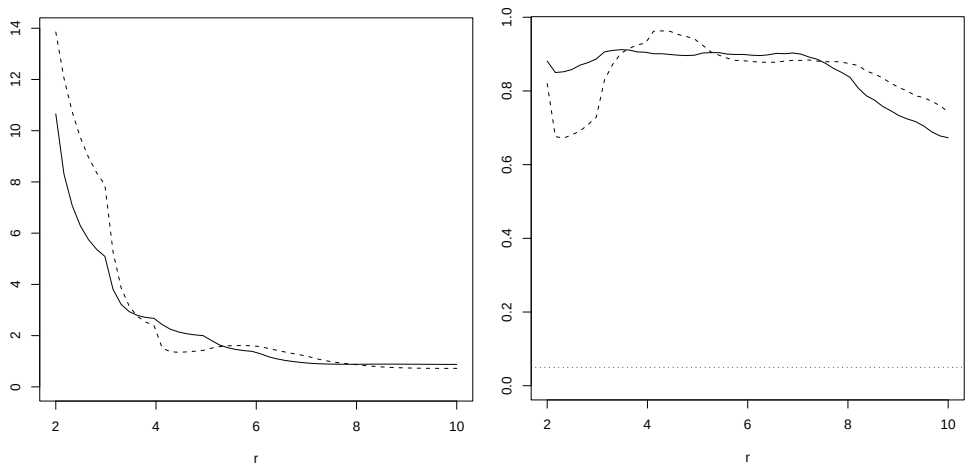

Figure 3: Left panel: test statistics. Right panel: p-values. Solid line: based on the local linear estimator (16). Dashed line: based on local loglikelihood estimator (18). Dotted line: significance level 0.05. r denotes the scaling parameter in (30).

algorithm based on a Bootstrap approach. In the case that the dependence structures of the L observed processes belong to a parametric family, the algorithm for approximating the p-value of the test can be simplified by the use of the Modified Fourier Integral Method for spatial processes simulation. When there is no a priori knowledge on the form of the spectral densities, a completely nonparametric algorithn is also provided. In the simulation study, considering the local-linear nonparametric estimator of the log-spectral density, we can see that the performance of the test is satisfactory, both in terms of size and power.

Finally, the motivation behind the development of this technique is to detect changes in the dependence structure of heavy metal concentrations in mosses. For the particular case of H_g concentrations, measures taken in March 2004 and September 2004 have been compared. There is no significance of a change in the pattern of dependence between these two periods. Therefore, if we were interested in adjusting a spatio-temporal model to this process, we could consider a time-stationary depedence model. See, for instance Fernández-Casal *et al.* (2003).

7 Acknowledgements.

The authors acknowledge the financial support of the Spanish Ministry of Education and Science and FEDER, project MTM2005-00820 and grant BES2003-0581.

References

Alonso, F. J.; Angulo, J. M.; and Bueso, M. C. (1996). Estimation and smoothing from incomplete data for a class of lattice processes. *Comm. Statist. Theory Methods*, volume 25, no. 3:pp. 667–681. ISSN 0361-0926.

Cressie, N. A. C. (1993). *Statistics for spatial data*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. New York: John Wiley & Sons Inc. ISBN 0-471-84336-9. A Wiley-Interscience Publication.

Crujeiras, R. M. and Fernández-Casal, R. (2006). Simulation of spatial dependence structures. *Technical Report. Department of Statistics and O.R. University of Santiago de Compostela*.

Crujeiras, R. M.; Fernández-Casal, R.; and González-Manteiga, W. (2006). Goodness-of-fit tests for the spatial spectral density. *Technical Report. Department of Statistics and O.R. University of Santiago de Compostela*.

Dahlhaus, R. and Künsch, H. (1987). Edge effects and efficient parameter estimation for stationary random fields. *Biometrika*, volume 74, no. 4:pp. 877–882. ISSN 0006-3444.

Dette, H. and Neumeyer, N. (2001). Nonparametric analysis of covariance. *Annals of Statistic*, volume 29, no. 5:pp. 1361–1400.

Diblasi, A. and Bowman, A. W. (2001). On the use of the variogram in checking for independence in spatial data. *Biometrics*, volume 57, no. 1:pp. 211–218. ISSN 0006-341X.

Fernandez, J.; Rey, A.; and Carballeira, A. (2000). An extended study of heavy metal deposition in galicia based on moss analysis. *Science of Total Environment*, volume 254:pp. 31–44.

Fernández-Casal, R.; González-Manteiga, W.; and Febrero-Bande, M. (2003). Flexible spatio-temporal stationary variogram models. *Statistics and Computing*, volume 13:pp. 127–136.

Fuentes, M. (2002). Spectral methods for nonstationary spatial processes. *Biometrika*, volume 89, no. 1:pp. 197–210. ISSN 0006-3444.

Guyon, X. (1982). Parameter estimation for a stationary process on a d-dimensional lattice. *Biometrika*, volume 69, no. 1:pp. 95–105. ISSN 0006-3444.

Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.*, volume 21, no. 4:pp. 1926–1947. ISSN 0090-5364.

King, E.; Hart, J.; and Wherly, T. (1991). Testing the equality of two regression curves using linear smoothers. *Statistics & Probability Letters*, volume 12:pp. 239–247.

Lahiri, S. (1999). Asymptotic distribution of the empirical spatial cumulative distribution function predictor and prediction bands based on a subsampling method. *Probab. Theory Related Fields*, volume 114:pp. 55–84.

Maglione, D. and Diblasi, A. (2004). Exploring a valid model for the variogram of an isotropic spatial process. *Stoch. Envir. Res. and Risk Ass.*, volume 18:pp. 366–376.

Martin, R. J. (1979). A subclass of lattice processes applied to a problem in planar sampling. *Biometrika*, volume 66, no. 2:pp. 209–217. ISSN 0006-3444.

Müller, H.-G. (1988). *Nonparametric regression analysis of longitudinal data*, volume 46 of *Lecture Notes in Statistics*. Berlin: Springer-Verlag. ISBN 3-540-96844-X.

Priestley, M. B. (1981). *Spectral analysis and time series. Vol. 1.* London: Academic Press Inc. [Harcourt Brace Jovanovich Publishers]. ISBN 0-12-564901-0. Univariate series, Probability and Mathematical Statistics.

Robinson, P. (2006). Nonparametric spectrum estimation for spatial data. *To appear in Journal of Statistical Planning and Inference.*

Vilar-Fernández, J. and González-Manteiga, W. (2004). Nonparametric comparison of curves with dependent errors. *Statistics*, volume 38, no. 2:pp. 81–99.

Yaglom, A. M. (1987). *Correlation theory of stationary and related random functions. Vol. II.* Springer Series in Statistics. New York: Springer-Verlag. ISBN 0-387-96331-6. Supplementary notes and references.

Zhu, J.; Lahiri, S.; and Cressie, N. (2002). Asymptotic inference for spatial cdfs over time. *Statistica Sinica*, volume 12:pp. 843–861.

Reports in Statistics and Operations Research

2004

04-01 Goodness of fit test for linear regression models with missing response data. *González Manteiga, W., Pérez González, A.*
Canadian Journal of Statistics (to appear).

04-02 Boosting for Real and Functional Samples. An Application to an Environmental Problem. *B. M. Fernández de Castro and W. González Manteiga.*

04-03 Nonparametric classification of time series: Application to the bank share prices in Spanish stock market. *Juan M. Vilar, José A. Vilar and Sonia Pértega*.

04-04 Boosting and Neural Networks for Prediction of Heteroskedatic Time Series. *J. M. Matías, M. Febrero, W. González Manteiga and J. C. Reboredo.*

04-05 Partially Linear Regression Models with Farima-Garch Errors. An Application to the Forward Exchange Market. *G. Aneiros Pérez, W. González Manteiga and J. C. Reboredo Nogueira.*

04-06 A Flexible Method to Measure Synchrony in Neuronal Firing. *C. Faes, H. Geys, G. Molenberghs, M. Aerts, C. Cadarso-Suárez, C. Acuña and M. Cano.*

04-07 Testing for factor-by-curve interactions in generalized additive models: an application to neuronal activity in the prefrontal cortex during a discrimination task. *J. Roca-Pardiñas, C. Cadarso-Suárez, V. Nacher and C. Acuña.*

04-08 Bootstrap Estimation of the Mean Squared Error of an EBLUP in Mixed Linear Models for Small Areas. *W. González Manteiga, M. J. Lombardía, I. Molina, D. Morales and L. Santamaría.*

04-09 Set estimation under convexity type assumptions. *A. Rodríguez Casal.*

2005

05-01 SiZer Map for Evaluating a Bootstrap Local Bandwidth Selector in Nonparametric Additive Models. *M. D. Martínez-Miranda, R. Raya-Miranda, W. González-Manteiga and A. González-Carmona.*

05-02 The Role of Commitment in Repeated Games. *I. García Jurado, Julio González Díaz*.

05-03 Project Games. *A. Estévez Fernández, P. Borm, H. Hamers*

05-04 Semiparametric Inference in Generalized Mixed Effects Models. *M. J. Lombardía, S. Sperlich*

2006

06-01 A unifying model for contests: effort-prize games. J. González Díaz

06-02 The Harsanyi paradox and the "right to talk" in bargaining among coalitions. J. J. Vidal Puga

06-03 A functional analysis of NOx levels: location and scale estimation and outlier detection. M. Febrero, P. Galeano, W. González-Manteiga

06-04 Comparing spatial dependence structures. R. M. Crujeiras, R. Fernández-Casal, W. González-Manteiga

Previous issues (2001 – 2003):
http://eio.usc.es/pub/reports.html