Pub. Inst. Stat. Univ. Paris
LII, fasc. 3, 2008, 27 à 42

MODEL TESTING FOR SPATIALLY CORRELATED DATA

Rosaria Ignaccolo[1] and Nunziata Ribecco[2]

[1]*Dipartimento di Statistica e Matematica Applicata*
Università degli Studi di Torino, Italy
e-mail: `ignaccolo@econ.unito.it`

[2]*Dipartimento di Statistica*
Università degli Studi di Bari, Italy
e-mail: `ribecco@dss.uniba.it`

Abstract

The distribution of a variable observed over a domain depends on the underlying process and also on the geographical locations at which the variable has been measured. In this paper, we fit a model to the distribution supposing that the observations are generated by a stationary strong-mixing random field. Indeed, after estimating the density of the considered variable, we construct a test statistic in order to verify the goodness of fit of the observed spatial data. The proposed class of tests is a generalization of the classical chi-square-test and of the Neyman smooth test. In the framework of increasing domain asymptotics, we analyse the large sample behaviour of the test. The limiting distribution is a linear combination of χ_1^2 r.v.s where the coefficients are the eigenvalues of a matrix Σ essentially related to the spectral density of the random field. Finally some indications about the implementation are provided.

Keyword: *goodness of fit, correlated data, spatial process, mixing random field*

1 Introduction

Usually the fit of the model to observations is tested by the classical χ^2 test (Rogers (1974), Ripley (1981), Cressie (1993)) and often independence cannot be assumed. In many fields — geology, ecology, or forestry, for example — observations are taken at different locations, i.e. are *georeferenced*, and present spatial autocorrelation (Cliff and Ord (1973)). In these cases to check distributional assumptions it is more appropriate to apply goodness of fit tests that take into account spatial dependence. Those goodness of fit tests could also be used to check if a new simulation procedure for random field has produced the desired distribution.

In this paper, we propose to generalize the classical χ^2 test to the case of correlated spatial data. Our framework is based on the nonparametric density estimator by projection, analogously to the cases of time series data in Ignaccolo (2004) and of independent data in Bosq (2002) (see also Bosq's papers referenced there). This class of goodness of fit tests also contains the smooth test of Neyman (1937). To take explicitly into account the spatial autocorrelation among the observed data, we suppose that the observations are generated by a strong-mixing random field and observed on a rectangular region $I_{\boldsymbol{n}}$.

The paper is organized as follows. In Section 2 we set the notations used in Section 3 to define the class of tests. In Section 4 we analyse its asymptotic behaviour (proofs are postponed to the appendix). We conclude providing some indications about the implementation of the test.

2 Definitions and basic framework

Let $\boldsymbol{X} = (X_{\boldsymbol{t}}, \boldsymbol{t} \in \mathbb{Z}^d)$, with $d \geq 1$, a random field defined on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with values in a measurable space $(E, \mathcal{B})$, i.e. a collection of random variables indexed by the discrete multidimensional variable $\boldsymbol{t} \in \mathbb{Z}^d$. In the special case where $d = 1$, the random field $X_{\boldsymbol{t}}$ is just a discrete stochastic process.

The random field is strictly stationary (or homogeneous) if for any set $A \subset \mathbb{Z}^d$, and for any point $\boldsymbol{v} \in \mathbb{Z}^d$ the joint distribution of the random variables $(X_{\boldsymbol{t}}, \boldsymbol{t} \in A)$ is identical to the joint distribution of $(X_{\boldsymbol{t}}, \boldsymbol{t} \in A + \boldsymbol{v})$, where $A + \boldsymbol{v}$ is the set A "translated" by $\boldsymbol{v}$, that is $A + \boldsymbol{v} = \left\{\boldsymbol{s} \in \mathbb{Z}^d : \boldsymbol{s} = \boldsymbol{t} + \boldsymbol{v},\ \text{with}\ \boldsymbol{v} \in A\right\}$.

Let us consider a strictly stationary random field $\boldsymbol{X} = (X_{\boldsymbol{t}}, \boldsymbol{t} \in \mathbb{Z}^d)$, with $d \geq 1$, defined on $(\Omega, \mathcal{A}, \mathbb{P})$ with values in $(E, \mathcal{B})$ and assume that we observe $\boldsymbol{X}$ on a rectangular region $I_{\boldsymbol{n}}$ defined by $I_{\boldsymbol{n}} = \{\boldsymbol{i} : \boldsymbol{i} \in \mathbb{Z}^d, 1 \leq i_l \leq n_l, l = 1, \ldots, d\}$. We write $\boldsymbol{n} \to \infty$ if $\min_l n_l \to \infty$ and we set $n^* = \prod_{l=1}^{d} n_l = |I_{\boldsymbol{n}}|$ the cardinality of $I_{\boldsymbol{n}}$, that is the total number of points where the process is observed.
To take into account the dependence between the observations we suppose that they are generated by a *strong-mixing* (or weak dependent) spatial process.
If $\Lambda \subset \mathbb{Z}^d$, let $\mathcal{A}_\Lambda$ be the σ-algebra generated by the X_ρ, $\rho \in \Lambda$. If $\Lambda_1, \Lambda_2 \subset \mathbb{Z}^d$, let $dist(\Lambda_1, \Lambda_2) = \inf\{dist(\rho_1, \rho_2) : \rho_1 \in \Lambda_1, \rho_2 \in \Lambda_2\}$ and

$$\alpha(\mathcal{A}_{\Lambda_1}, \mathcal{A}_{\Lambda_2}) = \sup_{A_1 \in \mathcal{A}_{\Lambda_1}, A_2 \in \mathcal{A}_{\Lambda_2}} |\mathbb{P}(A_1 \cap A_2) - \mathbb{P}(A_1)\mathbb{P}(A_2)| .$$

The random field $\boldsymbol{X}$ is said α-mixing (or strong mixing) if

$$\alpha_{uv}(n) = \sup\{\alpha(\mathcal{A}_{\Lambda_1}, \mathcal{A}_{\Lambda_2}) : dist(\Lambda_1, \Lambda_2) \geq n, |\Lambda_1| \leq u, |\Lambda_2| \leq v\} \overset{n\to\infty}{\longrightarrow} 0,$$

for any integers $u, v \geq 0$ (see Doukhan (1994)).

3 Construction of the class of functional tests

We want to test the simple hypothesis H_0 : "$X_{\boldsymbol{i}}$ has distribution μ" ($X_{\boldsymbol{i}} \sim \mu$), where μ is a probability measure on $(E, \mathcal{B})$ completely specified. Let $\mathcal{P}$ be a family of probability measures on $(E, \mathcal{B})$ dominated by μ and let ν be the generic element of $\mathcal{P}$. We denote by f the probability density of ν with respect to μ, $f = \frac{d\nu}{d\mu}$, and we assume that $f \in L^2(\mu)$, that is a separable Hilbert space with its scalar product $\langle f, g\rangle = \int fg \, d\mu$. Then the density f admits a Fourier expansion and it can be estimated by truncating this expansion, that is taking its orthogonal projection on a subspace of $L^2(\mu)$. This permits estimation of f estimating only a finite number k of Fourier coefficients. Optimality properties of the projection density estimator are obtained when $k = k(n)$, that is the dimension of the subspace where f is projected depends on n (see Bosq and Lecoutre (1987) for further details). Here we consider k fixed and we do not focus on the density estimator's properties, but on the test statistic defined in (2). Before applying the test proposed here, as a preliminary step, the parameter k could be chosen looking at the number of estimated coefficients that are "large enough", as suggested in Bosq (2005, p.61). In the future, this data-driven truncation index could be incorporated in our test statistic.

For a fixed positive integer k let $\{e_0, e_1, \ldots, e_k\}$ be an orthonormal system with $e_0 \equiv 1$, in $L^2(\mu)$, which generates a subspace $E_k = \text{span}\{e_0, e_1, \ldots, e_k\}$, with $\dim(E_k) = k + 1$.
The real valued function $K(x,t) = \sum_{j=0}^{k} e_j(x)e_j(t)$, with $(x,t) \in E \times E$ is the *reproducing kernel* of E_k (see Grenander (1981), Fortet (1995), Berlinet and Thomas-Agnan (2003)).
The density estimator of f by projection on E_k is defined as

$$f_n(t) = \frac{1}{n^*} \sum_{i \in I_n} K(X_i, t) = \sum_{j=0}^{k} \hat{a}_{jn}\, e_j(t) \tag{1}$$

where $\hat{a}_{jn} = \frac{1}{n^*} \sum_{i \in I_n} e_j(X_i)$ is the unbiased estimator of the Fourier coefficient $a_j = \langle f, e_j \rangle$.
Let us consider the $L^2(\mu)$–distance $dist(f_n, 1) = \|f_n - 1\|$ between the estimated density and the hypothesized density $f_0 = \frac{\mathrm{d}\mu}{\mathrm{d}\mu} = 1$ under H_0, where $\|\cdot\|$ denotes the $L^2(\mu)$–norm.
Now we consider the statistic

$$T_n = \sqrt{n^*}(f_n - 1)$$

and its $L^2(\mu)$–norm $\|T_n\| = \sqrt{n^*}\, \|f_n - 1\|$ obtaining

$$\|T_n\|^2 = n^* \left\| 1 + \sum_{j=1}^{k} \hat{a}_{jn} e_j(t) - 1 \right\|^2 = n^* \sum_{j=1}^{k} \hat{a}_{jn}^2. \tag{2}$$

We want to test $H_0 : X_i \sim \mu$ versus $H_1 : X_i \sim \nu \neq \mu$, that is $H_0 : f = 1$ versus $H_1 : f \neq 1$ considering the densities and also $H_0 : \quad a_j = 0\ \forall j \geq 1$ versus $H_1 : \exists j \geq 1 : a_j \neq 0$ with respect to the Fourier coefficients a_j. Indeed we shall limit ourselves to consider the alternative hypothesis $H_1(k) : X_i \sim \nu \neq \mu$ with ν such that there exists a $j \in 1, \ldots, k$ for which e_j is ν-integrable and $a_j = \int e_j\, \mathrm{d}\nu \neq 0$.
The hypothesis H_0 states that, after $a_0 = 1$, all Fourier coefficients are null. Instead according to H_1 there exists a Fourier coefficient (other than a_0) different from zero, and $H_1(k)$ states that it is among the first k coefficients after a_0.
Since the test is based on the deviation of the estimated density from the hypothesized density, it rejects H_0 for large values of $\|T_n\|^2$. We shall prove in Section

4 that, under H_0, $\|T_n\|^2$ converges in distribution to a linear combination of r.v.'s $U_j^2 \sim \chi_1^2$ where the coefficients λ_j^2 are the eigenvalues of the matrix Σ defined in (4).
So it is possible to carry out a test with rejection region $\{\|T_n\|^2 > w\}$ and asymptotic size $\alpha \in]0,1[$ with w given by $\mathbb{P}\left(\sum_{j=1}^k \lambda_j^2 U_j^2 > w\right) = \alpha$, but the method requires the estimation of eigenvalues and the determination of quantiles of the random variable $\sum_{j=1}^k \lambda_j^2 U_j^2$.

Particular cases. Let $\{A_0, A_1, \ldots, A_k\}$ be a finite partition of E with $p_j = \mu(A_j) > 0$, $j = 0, \ldots, k$. For $e_j(\cdot) = p_j^{-1/2} \mathbb{1}_{A_j}(\cdot)$, the system $\{e_0, e_1, \ldots, e_k\}$ is orthonormal and it generates a subspace $E_k \subseteq L^2(\mu)$ that contains every constant function. In this case the density estimator is the histogram and one easily obtains $\|T_n\|^2 = \sum_{j=0}^k \frac{\left[\sum_{i \in I_n} \mathbb{1}_{A_j}(X_i) - n^* p_j\right]^2}{n^* p_j}$ that is the test statistic used in the classical Pearson's χ^2 test.
Moreover, with an orthonormal system related to the Legendre polynomials the statistic $\|T_n\|^2$ coincides with the Neyman statistic (see also Rayner and Best (1989)).

4 Large sample behaviour

The following notations are used throughout the paper. For each j we define the zero-mean real valued r.v.'s

$$Y_{ij} = e_j(X_i) - \mathbb{E}\left(e_j(X_i)\right), \quad i \in I_n$$

and we set

$$(n^*)^{-1/2} S_{nj} = (n^*)^{-1/2} \sum_{i \in I_n} Y_{ij} = \sqrt{n^*}\left(\hat{a}_{jn} - a_j\right)$$

with $n^* = |I_n|$, noting that $a_j = \langle f, e_j \rangle = \mathbb{E}(e_j(X_i)) = \mathbb{E}(\hat{a}_{jn})$.
Moreover we shall use

$$(n^*)^{-1/2} S_n = \begin{pmatrix} (n^*)^{-1/2} S_{n1} \\ \vdots \\ (n^*)^{-1/2} S_{nk} \end{pmatrix} = \sqrt{n^*}\left[\begin{pmatrix} \hat{a}_{1n} \\ \vdots \\ \hat{a}_{kn} \end{pmatrix} - \begin{pmatrix} a_1 \\ \vdots \\ a_k \end{pmatrix}\right] = \sqrt{n^*}\left(A_n - a\right)$$

and, for $\boldsymbol{i} \in I_{\boldsymbol{n}}$,

$$\boldsymbol{Y_i} = (Y_{\boldsymbol{i}1}, \ldots, Y_{\boldsymbol{i}k})^T$$

and the linear combination $V_{\boldsymbol{i}} = \sum_{j=1}^{k} c_j Y_{\boldsymbol{i}j} = \mathbf{c}^T \boldsymbol{Y_i}$ with $\boldsymbol{c} = (c_1, \ldots, c_k)^T \in \mathbb{R}^k$. Under $H_1(k)$ we consider the quadratic form

$$n^*(\boldsymbol{A}_n - \boldsymbol{a})^T(\boldsymbol{A}_n - \boldsymbol{a}) = n^* \sum_{j=1}^{k} (\hat{a}_{j\boldsymbol{n}} - a_j)^2 := \left\| T_{\boldsymbol{n}}' \right\|^2$$

that coincides with $\|T_{\boldsymbol{n}}\|^2$ under H_0 because $a_j = 0$ for all j.
The following theorem provides the limiting distribution of $\left\|T_{\boldsymbol{n}}'\right\|^2$ under some conditions necessary to apply the CLT theorem for random fields (Bolthausen, 1982).

Theorem 1. *Consider a sequence $I_{\boldsymbol{n}}$ of finite subsets of $\mathbb{Z}^d$ that increases to $\mathbb{Z}^d$ and is such that $\lim_{\boldsymbol{n}\to\infty} \frac{|\partial I_{\boldsymbol{n}}|}{|I_{\boldsymbol{n}}|} = 0$, where $\partial I_{\boldsymbol{n}}$ is the boundary of $I_{\boldsymbol{n}}$. If*

1. $\sum_{r=1}^{\infty} r^{d-1}\alpha_{uv}(r) < \infty$ *for* $u + v \leq 4$ *and* $\alpha_{1,\infty}(r) = o(r^{-d})$*;*
2. *for some* $\delta > 0$ $\mathbb{E}(|V_{\boldsymbol{i}}|^{2+\delta}) < \infty$ *and* $\sum_{r=1}^{\infty} r^{d-1}\alpha_{1,1}(r)^{\delta/(2+\delta)} < \infty$*;*
3. $\sigma^2 > 0$ *where* $\sigma^2 = \sum_{\boldsymbol{i}\in\mathbb{Z}^d} \mathbb{E}(V_{\boldsymbol{0}}V_{\boldsymbol{i}})$*;*

then

$$\left\| T_{\boldsymbol{n}}' \right\|^2 \xrightarrow{d} \sum_{j=1}^{k} \lambda_j^2 U_j^2 \tag{3}$$

where the r.v.'s $U_j \sim \mathcal{N}(0,1)$ are independent and λ_j^2 are the eigenvalues of the matrix $\Sigma = (\sigma_{jl})_{j,l=1,\ldots,k}$ with

$$\sigma_{jl} = \sum_{\boldsymbol{v}\in\mathbb{Z}^d} \sigma_{jl}(\boldsymbol{v}) \tag{4}$$

where $\sigma_{jl}(\boldsymbol{v})$ is the crossed covariance function of the stationary bivariate spatial process $(Y_{\boldsymbol{t}j}, Y_{\boldsymbol{t}l})_{\boldsymbol{t}\in\mathbb{Z}^d}$.

Under H_0, as previously seen, $\left\|T_{\boldsymbol{n}}'\right\|^2$ coincides with $\|T_{\boldsymbol{n}}\|^2$. Hence the convergence (3) holds:

$$\|T_{\boldsymbol{n}}\|^2 \xrightarrow{\text{d}} \sum_{j=1}^{k} \lambda_j^{*2} U_j^2$$

and the coefficients in the linear combination of the limiting distribution have particular values because $Y_{\boldsymbol{i}j} = e_j(X_{\boldsymbol{i}})$, being $a_j = \mathbb{E}(e_j(X_{\boldsymbol{i}})) = 0$ under the null hypothesis.

To study the rate of convergence, we consider the Kolmogorov distance between distribution functions

$$\Delta_n = \sup_{u \in \mathbb{R}} \left| \mathbb{P}\left(\left\| T_{\boldsymbol{n}}^{'} \right\|^2 \leq u \right) - \mathbb{P}\left(\|U\|^2 \leq u \right) \right| \qquad (5)$$

with $\|U\|^2 = \sum_{j=1}^{k} \lambda_j^2 U_j^2$ (since $U = \sum_{j=1}^{k} \lambda_j U_j e_j$).

Theorem 2. *If the mixing coefficients are exponentially decreasing, that is $\alpha_{uv}(n) = O\left(\mathrm{e}^{-an}\right)$ for $a > 0$, denoting $\sigma_n^2 = Var\left(\sum_{\boldsymbol{i} \in I_{\boldsymbol{n}}} V_{\boldsymbol{i}}\right)$, the distance Δ_n satisfies*

$$\Delta_n = O\left([\log \sigma_n]^{d[(1+\delta)\wedge 2]} \sigma_n^{-(\delta \wedge 1)} \right)$$

where $x \wedge y = \min(x, y)$, while if the mixing coefficients are arithmetically decreasing, that is $\alpha_{uv}(n) = O\left(n^{-a}\right)$ for $a > 0$, Δ_n satisfies

$$\Delta_n = O\left(\sigma_n^{-\xi}\right)$$

with $\xi = (\delta \wedge 1)\frac{2(b-1)}{2b+(\delta \wedge 1)}$ and $b = \frac{a\delta(\delta \wedge 1)}{2d(2+\delta)[(1+\delta)\wedge 2]}$.

Independent data. In the case of iid data, the crossed covariances in the sum (4) are null, except when $\boldsymbol{v} = \boldsymbol{0}$; hence $\sigma_{jl} = \mathbb{E}\left(Y_{\boldsymbol{0}j} Y_{\boldsymbol{0}l}\right)$. Then under the null hypothesis the element of Σ^* (that is Σ under H_0) becomes

$$\sigma_{jl}^* = \begin{cases} 0 & \text{for } j \neq l \\ 1 & \text{for } j = l \end{cases}$$

because $\sigma_{jj}^* = \mathbb{E}\left([e_j(X_{\boldsymbol{0}})]^2\right) = \int e_j^2(x)\,\mathrm{d}\mu(x) = \|e_j\|^2 = 1$ and so $\Sigma^* = \mathbb{I}_k$ where $\mathbb{I}_k$ is the identity matrix of order k. Then for all j one has $\lambda_j^{*2} = 1$ and the limiting distribution is χ_k^2, as we have in the general case (see Bosq (2002)) where $\Delta_n = O\left(n^{-1/2}\right)$.

5 Implementation

The limiting distribution for the test depends on the eigenvalues λ_j^2 of the unknown matrix Σ. So to estimate these eigenvalues we have to estimate the elements σ_{jl} and we propose to use the estimate of the cross-spectral density function (Section 5.1). With the estimated eigenvalues we have a linear combination of χ^2 r.v.'s with 1 degree of freedom. Exact significance points for selected values of k and of the coefficients were published by several authors. Moreover, different evaluations of these quantiles have been proposed, by approximation of series expansions or by numerical methods. For further details see Johnson et al. (1994) and Mathai and Provost (1992).

On the other hand, we propose here a class of tests and to run one of them we have to choose an orthonormal system; for that a suggestion follows in Section 5.2.

5.1 Estimating the eigenvalues by spectral density

For the stationary bivariate spatial process $(Y_{tj}, Y_{tl})_{t\in\mathbb{Z}^d}$ we can define the spectral density functions matrix $f(\omega) = (f_{jl}(\omega))_{j,l=1,\dots,k}$ whose elements are the cross spectral density functions defined by

$$f_{jl}(\omega) = \frac{1}{(2\pi)^d} \sum_{v\in\mathbb{Z}^d} \sigma_{jl}(v) e^{-i\langle\omega, v\rangle}$$

where $\omega = (\omega_1, \dots, \omega_d) \in [-\pi, \pi]^d$ and $\langle\omega, v\rangle = \sum_{i=1}^d \omega_i v_i$ represents the standard scalar product in d-dimensional euclidean space.

Observing that $\sigma_{jl} = (2\pi)^d f_{jl}(\mathbf{0})$, an estimate of σ_{jl} can be obtained evaluating the estimate of the spectral density at the origin, that is

$$\hat{\sigma}_{jl} = (2\pi)^d \hat{f}_{jl}(\mathbf{0}). \tag{6}$$

We use a Bartlett type or Lag-window estimator constructed as follows.

Let $M_1, \dots, M_d$ be some positive integers and consider all rectangular regions $\Lambda_M = \{t \in I_n : 1 \le u_l \le t_l \le U_l \le n_l\}$, where u_l, U_l satisfy $U_l - u_l + 1 = M_l$, for $l = 1, \dots, d$.

By calculating the periodogram over these smaller rectangular regions and averaging all these results, we obtain a consistent (under some regularity conditions) estimator for the spectral density function.

Then we can take as Lag-window estimator

$$\tilde{f}_{jl}(\omega) = \frac{1}{(2\pi)^d} \sum_{v\in\mathbb{Z}^d} W(v) \tilde{\sigma}_{jl}(v) e^{-i\langle\omega, v\rangle} \tag{7}$$

where $\tilde{\sigma}_{jl}(\boldsymbol{v}) = \frac{1}{n^*}\sum_{\boldsymbol{v}\in\mathbb{Z}^d} \tilde{Y}_{\boldsymbol{t},j}\tilde{Y}_{\boldsymbol{t}+\boldsymbol{v},l}$ with

$$\tilde{Y}_{\boldsymbol{t},j} = \begin{cases} Y_{\boldsymbol{t},j} & \text{if } \boldsymbol{t} \in I_{\boldsymbol{n}} \\ 0 & \text{else} \end{cases}$$

and the window is defined by

$$W(\boldsymbol{v}) = \left(\prod_{l=1}^{d}\left(1 - \frac{|v_l|}{M_l}\right)\right)^+$$

where $(x)^+ = \max(x, 0)$ is the positive part function.

We follow the work of Politis and Romano (1996) for a (not crossed) spectral density, and we suppose the validity of the following condition $\mathbf{A_0}$, with $M = \prod_{l=1}^{d} M_l$:

($\mathbf{A_0}$) *There are positive constants* c_* *and* c^* *such that* $c_* < M_j/M_l < c^*$, $c_* < n_j/n_l < c^*$ $\forall j, l = 1, \ldots, d$ *and* $M = o(n^*)$ *for* $n^* \to \infty$.

Under some moment and weak dependence conditions and under the assumption $\mathbf{A_0}$ it can be shown that

$$\mathsf{Bias}(\tilde{f}_{jl}(\boldsymbol{\omega})) = O(M^{-1/d}) \tag{8}$$

and

$$\mathsf{Var}(\tilde{f}_{jl}(\boldsymbol{\omega})) = O(M/n^*) \tag{9}$$

This is a typical behaviour for a Lag-window estimator, but however Politis and Romano (1996) propose a "bias-corrected" flat-top estimator that has the same order of magnitude for the variance, but a smaller bias. Then, following Politis-Romano (1996) and adapting their reasoning to crossed covariance functions and cross spectral densities, we define the flat-top estimator for $f_{jl}(\boldsymbol{\omega})$ in (10).

Let us consider the unbiased estimator for the crossed covariance $\sigma_{jl}(\boldsymbol{v})$ as

$$\hat{\sigma}_{jl}(\boldsymbol{v}) = \frac{1}{\prod_{l=1}^{d}(n_l - |v_l| + 1)}\sum_{\boldsymbol{v}\in\mathbb{Z}^d} \tilde{Y}_{\boldsymbol{t},j}\tilde{Y}_{\boldsymbol{t}+\boldsymbol{v},l}.$$

For a constant $c \in (0,1)$ and denoting $m_l = cM_l$, $\forall l = 1, \ldots, d$, we define the flat-top cross spectral density estimator as

$$\hat{f}_{jl}(\boldsymbol{\omega}; c) = \frac{1}{1-c}\tilde{f}_{jl,M}(\boldsymbol{\omega}) - \frac{c}{1-c}\tilde{f}_{jl,m}(\boldsymbol{\omega}) \tag{10}$$

where

$$\tilde{f}_{jl,M}(\boldsymbol{\omega}) = \frac{1}{(2\pi)^n} \sum_{\boldsymbol{v}\in\mathbb{Z}^d} W_M(\boldsymbol{v})\hat{\sigma}_{jl}(\boldsymbol{v})\mathrm{e}^{-\mathrm{i}\langle\boldsymbol{\omega},\boldsymbol{v}\rangle}$$

$$\tilde{f}_{jl,m}(\boldsymbol{\omega}) = \frac{1}{(2\pi)^n} \sum_{\boldsymbol{v}\in\mathbb{Z}^d} W_m(\boldsymbol{v})\hat{\sigma}_{jl}(\boldsymbol{v})\mathrm{e}^{-\mathrm{i}\langle\boldsymbol{\omega},\boldsymbol{v}\rangle}$$

and the windows are defined by

$$W_M(\boldsymbol{v}) = \left(1 - \max_{1\leq l\leq d} \frac{|v_l|}{M_l}\right)^+ \qquad \text{and} \qquad W_m(\boldsymbol{v}) = \left(1 - \max_{1\leq l\leq d} \frac{|v_l|}{m_l}\right)^+ .$$

So we have constructed an estimator that is the linear combination of two Lag-window estimators with differing bandwidths; the resultant pyramidal window has its top 'chopped off' (see Figure 3, p. 45, in Politis and Romano (2006)), that is $W_{M,m}(\boldsymbol{v}) = \frac{1}{1-c}W_M(\boldsymbol{v}) - \frac{c}{1-c}W_m(\boldsymbol{v})$.

We pose the following weak dependence conditions based on the second order moment in increasing order of strength (i.e. $A_3 \Rightarrow A_2 \Rightarrow A_1$), that can be viewed as different smoothness conditions of the spectral density.

($\mathbf{A_1}$) *There are (finite) positive constants $B, c_1, \ldots, c_d$ and a positive constant $K > d$ such that $|\sigma_{jl}(\boldsymbol{v})| \leq B\left(\max_l \frac{|v_l|}{c_l}\right)^{-K}$.*

($\mathbf{A_2}$) *There are (finite) positive constants $K, B, c_1, \ldots, c_d$ such that $|\sigma_{jl}(\boldsymbol{v})| \leq B\exp\left\{-K\max_l \frac{|v_l|}{c_l}\right\}$.*

($\mathbf{A_3}$) *There are (finite) positive constants $c_1, \ldots, c_d$ such that $\sigma_{jl}(\boldsymbol{v}) = 0$, if $\max_l \frac{|v_l|}{c_l} \geq 1$.*

The introduction of the constants $c_1, \ldots, c_d$ allows for the possibility that the cross covariances vanish along different directions in $\mathbb{Z}^d$. Another reasonable condition for the choice of M_l is the following:

($\mathbf{A_M}$) *There are (finite) positive constants $c_1, \ldots, c_d$ such that $\frac{M_i}{M_j} = \frac{c_i}{c_j}$ for $i, j = 1, \ldots, d$.*

Various combinations of these conditions yield different, variable, orders of the bias of the estimator, that is, we have different performances of the family of lag window estimator $\left\{\hat{f}_{jl}(\boldsymbol{\omega};c),\ c\in(0,1)\right\}$ as summarized in the theorem below (obtained following Politis and Romano (1996) results).

Theorem 3 (Politis-Romano).

1. *Under A_0, A_1 and A_M it follows that*

$$\sup_{\boldsymbol{\omega}\in[-\pi,\pi]^d}\left|\mathit{Bias}\left(\hat{f}_{jl}(\boldsymbol{\omega};c)\right)\right| = O\left(M^{1-K/d}\right).$$

2. *Under A_0, A_2 and A_M and for $M_l = \beta c_l \log n_l,\ l=1,\ldots,d$ with $\beta > \frac{d}{2K}$ it follows that*

$$\sup_{\boldsymbol{\omega}\in[-\pi,\pi]^d}\left|\mathit{Bias}\left(\hat{f}_{jl}(\boldsymbol{\omega};c)\right)\right| = O\left(\frac{(\log n^*)^{d-1}}{(n^*)^{\beta K/d}}\right) = o\left(\frac{1}{\sqrt{n^*}}\right).$$

3. *Under A_3, if there are positive constants c_* and c^* such that $c_* < n_j/n_k < c^*\ \forall j,k = 1,\ldots,d$ for $n^*\to\infty$ and there exist constants m_l, M_l for $l = 1,\ldots,d$ such that $M_l \geq m_l \geq c_l$, it follows that*

$$\sup_{\boldsymbol{\omega}\in[-\pi,\pi]^d}\left|\mathit{Bias}\left(\hat{f}_{jl}(\boldsymbol{\omega};c)\right)\right| = 0.$$

For further results and comments see Politis and Romano (1996).

At this point we can come back to (6) and consider an estimator for σ_{jl} as

$$\hat{\sigma}_{jl,c} = \sum_{\boldsymbol{v}\in\mathbb{Z}^d} W_{M,m}(\boldsymbol{v})\hat{\sigma}_{jl}(\boldsymbol{v})$$

that inherits the properties of $\hat{f}_{jl}(\boldsymbol{\omega};c)$.

Nevertheless, the matrix $\hat{\Sigma}_n = (\hat{\sigma}_{jl})_{1\leq j,l\leq k}$ may not be positive definite. Hence we propose (see Ignaccolo (2004)) to correct the elements on the diagonal with

$$\zeta_n = \left|\inf_{\|\boldsymbol{c}\|\leq 1}\boldsymbol{c}^T\hat{\Sigma}_n\boldsymbol{c}\right| + \frac{\tau}{n}$$

where $\boldsymbol{c}\in\mathbb{R}^k$ and τ is a suitable positive constant, defining the new matrix

$$\hat{\Sigma}_n^+ = \hat{\Sigma}_n + \zeta_n \mathbb{I}_k.$$

Considering a vector $\boldsymbol{c} \in \mathbb{R}^k$, with $\|\boldsymbol{c}\| = 1$, one has $\boldsymbol{c}^T\hat{\Sigma}_n^+\boldsymbol{c} = \boldsymbol{c}^T\hat{\Sigma}_n\boldsymbol{c} + \zeta_n > 0$ because when $\boldsymbol{c}^T\hat{\Sigma}_n\boldsymbol{c} < 0$ the definition of ζ_n implies $-\boldsymbol{c}^T\hat{\Sigma}_n\boldsymbol{c} < \zeta_n$.
Moreover, we have proved that $\zeta_n \xrightarrow{\text{a. s.}} 0$ and consequently $\hat{\Sigma}_n^+ \xrightarrow{\text{a. s.}} \Sigma$ (Prop. 4 in Ignaccolo (2004)). So, by continuity, the eigenvalues $\hat{\lambda}_j^2$ of $\hat{\Sigma}_n^+$ converge a.s. to λ_j^2 and we can use them for the limiting distribution.

5.2 The choice of the orthonormal system

To run a test of the proposed class we have to choose an orthonormal system. To use the projection density estimator, Bosq (2005, p.75) underlines the convenience to make an argumented choice, since choosing a system could imply to favour a certain kind of densities.
To link this choice to the distribution considered in H_0, we suggest using the set of orthonormal polynomials linked to it. More precisely, we refer to the Meixner class (see Lancaster (1975) and Rayner and Best (1989) p.140), for which an orthogonal system can be obtained from the recurrence relation $P_{n+1}(x) = (x + n\lambda)P_n(x) + n(-\sigma^2 + (n-1)\gamma)P_{n-1}(x)$ for $n = 0, 1, 2, \ldots$ where $P_{-1}(x) = 0$, $P_1(x) = 1$ and $\sigma^2 = \mathsf{Var}(X)$ with $X = Y - \mathbb{E}(Y)$. The constants λ and γ can be determined from $P_2(0)$ and $P_3(0)$ and to have an orthonormal system $P_n(x)$ must be normalized by dividing by $s_n = \sqrt{\mathbb{E}\left([P_n(x)]^2\right)}$.
The Normal and the Gamma distributions belong to the Meixner class and, in particular, for the gaussian case $\{P_n(x)\}$ is the set of Hermite polynomials. Moreover the Poisson, the Binomial and the Negative Binomial distributions belong to this class too.
Although the probability measure μ is unknown, we could choose a set of orthonormal polynomials linked to the hypothesis H_0 (as we could do to use the projection density estimator), that are the Meixner polynmials.
We observe that the proposed class of tests is defined for continuous variables. But if we are interested in the models describing the positions of units in the space (like the Poisson, the Binomial and the Negative Binomial) and it is not possible to assume the independence between the observations, we suggest to apply the proposed test choosing the polynomials linked to the hypothesized distribution according to Meixner.

Acknowledgements

The authors thank Prof. Denis Bosq, Prof. Fortunato Pesarin and Prof. Michael Stein for helpful comments on a previous version of this paper. The work of Rosaria Ignaccolo was partially supported by MIUR Cofin 2004.

A Proofs of Section 4

Proof of Theorem 1. We consider the linear combination $\boldsymbol{c}^T \boldsymbol{S}_n$ and with the previous notations we get

$$\boldsymbol{c}^T \boldsymbol{S}_n = \sum_{j=1}^{k} c_j S_{nj} = \sum_{j=1}^{k} c_j \sum_{\boldsymbol{i}\in I_n} Y_{\boldsymbol{i}j} = \sum_{\boldsymbol{i}\in I_n} \sum_{j=1}^{k} c_j Y_{\boldsymbol{i}j} = \sum_{\boldsymbol{i}\in I_n} \boldsymbol{c}^T \boldsymbol{Y}_{\boldsymbol{i}} = \sum_{\boldsymbol{i}\in I_n} V_{\boldsymbol{i}}.$$

With the previous assumptions, we can apply the Bolthausen CLT theorem (see Bolthausen (1982) or Rosenblatt (2000) p.56) to the zero-mean real valued random field $V_{\boldsymbol{i}}$ that is stationary and α-mixing (as we can write $V_{\boldsymbol{i}} = g(X_{\boldsymbol{i}})$ with g measurable and $\alpha_V(n) \leq \alpha_X(n)$). Hence we obtain

$$\frac{\sum_{\boldsymbol{i}\in I_n} V_{\boldsymbol{i}}}{\sigma\sqrt{n^*}} \xrightarrow{\mathrm{d}} \mathcal{N}(0,1)$$

that is

$$(n^*)^{-1/2} \boldsymbol{c}^T \boldsymbol{S}_n \xrightarrow{\mathrm{d}} \mathcal{N}(0,\sigma^2).$$

If we can write $\sigma^2 = \boldsymbol{c}^T \Sigma \boldsymbol{c}$, with Σ symmetric and positive definite, we can apply the Crámer–Wold device and we can obtain

$$\sqrt{n^*}(\boldsymbol{A}_n - \boldsymbol{a}) \xrightarrow{\mathrm{d}} \boldsymbol{Z} \sim \mathcal{N}_k\left(\boldsymbol{0}, \Sigma\right).$$

So for the quadratic forms we can have

$$n^*(\boldsymbol{A}_n - \boldsymbol{a})^T \mathbb{I} (\boldsymbol{A}_n - \boldsymbol{a}) \xrightarrow{\mathrm{d}} \boldsymbol{Z}^T \mathbb{I} \boldsymbol{Z}$$

that is $\left\|T_{\boldsymbol{n}}'\right\|^2 \xrightarrow{\mathrm{d}} \sum_{j=1}^k \lambda_j^2 U_j^2$ observing that $\boldsymbol{Z}^T \mathbb{I} \boldsymbol{Z} = \sum_{j=1}^k \lambda_j^2 \left(U_j\right)^2 = \sum_{j=1}^k \lambda_j^2 \chi_1^2$.
So it remains to check that we can write $\sigma^2 = \boldsymbol{c}^T \Sigma \boldsymbol{c}$, identifying Σ.
From the Bolthausen CLT we have $\sigma^2 = \sum_{\boldsymbol{i}\in\mathbb{Z}^d} \mathbb{E}(V_0 V_{\boldsymbol{i}})$ and it holds

$$\mathbb{E}(V_0 V_{\boldsymbol{i}}) = \mathbb{E}\left(\left[\boldsymbol{c}^T \boldsymbol{Y}_0\right]\left[\boldsymbol{c}^T \boldsymbol{Y}_{\boldsymbol{i}}\right]^T\right) = \mathbb{E}\left(\boldsymbol{c}^T \boldsymbol{Y}_0 \boldsymbol{Y}_{\boldsymbol{i}}^T \boldsymbol{c}\right) = \boldsymbol{c}^T \mathbb{E}\left(\boldsymbol{Y}_0 \boldsymbol{Y}_{\boldsymbol{i}}^T\right) \boldsymbol{c} = \boldsymbol{c}^T \Sigma_{0\boldsymbol{i}} \boldsymbol{c}$$

denoting $\Sigma_{0i} = \mathbb{E}\left(\boldsymbol{Y}_0 \boldsymbol{Y}_i^T\right)$. Then the generic element of Σ is $\sigma_{jl} = \sum_{i\in\mathbb{Z}^d} \mathbb{E}\left(Y_{0j} Y_{il}\right)$. By stationarity of $\boldsymbol{Y}_i$ it follows $\mathsf{Cov}(\boldsymbol{Y}_0, \boldsymbol{Y}_v) = \mathsf{Cov}(\boldsymbol{Y}_t, \boldsymbol{Y}_{t+v})$ and for each element $\mathbb{E}\left(Y_{0j} Y_{vl}\right) = \mathbb{E}\left(Y_{tj} Y_{t+v,l}\right)$ so we denote

$$\sigma_{jl}(\boldsymbol{v}) = \mathbb{E}\left(Y_{tj} Y_{t+v,l}\right)$$

the crossed covariance of the stationary bivariate spatial process $(Y_{tj}, Y_{tl})_{t\in\mathbb{Z}^d}$. Moreover the following equality holds

$$\sigma_{jl}(-\boldsymbol{v}) = \mathbb{E}\left(Y_{tj} Y_{t-v,l}\right) = \mathbb{E}\left(Y_{t-v,l} Y_{tj}\right) = \sigma_{lj}(\boldsymbol{v}).$$

So we can write $\sigma_{jl} = \sum_{v\in\mathbb{Z}^d} \sigma_{jl}(\boldsymbol{v})$ and we have

$$\sigma_{lj} = \sum_{v\in\mathbb{Z}^d} \sigma_{lj}(\boldsymbol{v}) = \sum_{v\in\mathbb{Z}^d} \sigma_{jl}(-\boldsymbol{v}) = \sum_{u\in\mathbb{Z}^d} \sigma_{jl}(\boldsymbol{u}) = \sigma_{jl},$$

so that Σ is symmetric.

□

Proof of Theorem 2. In what follows we use notations and results derived earlier in the proof of Theorem 1. With respect to the rate of convergence we apply Theorems 2 and 3 in Guyon and Richardson (1984) (see also Doukhan (1994) p. 49) to the random field V_i and, posing $N_\sigma \sim \mathcal{N}(0, \sigma^2)$, we observe that

$$\begin{aligned}
&\sup_{t\in\mathbb{R}} \left|\mathbb{P}\left((n^*)^{-1/2}\boldsymbol{c}^T \boldsymbol{S}_n \le t\right) - \mathbb{P}\left(N_\sigma \le t\right)\right| \\
&\quad = \sup_{v\in\mathbb{R}} \left|\mathbb{P}\left(\left\|\boldsymbol{c}^T\left[\sqrt{n^*}(\boldsymbol{A}_n - \boldsymbol{a})\right]\right\|^2 \le v^2\right) - \mathbb{P}\left(\left\|\boldsymbol{c}^T \boldsymbol{Z}\right\|^2 \le v^2\right)\right| \\
&\quad = \sup_{v\in\mathbb{R}} \left|\mathbb{P}\left(n^* \left\|\boldsymbol{A}_n - \boldsymbol{a}\right\|^2 \le \frac{v^2}{\|\boldsymbol{c}\|^2}\right) - \mathbb{P}\left(\|\boldsymbol{Z}\|^2 \le \frac{v^2}{\|\boldsymbol{c}\|^2}\right)\right| \\
&\quad = \Delta_n
\end{aligned}$$

with the position $u = \frac{v^2}{\|\boldsymbol{c}\|^2}$ as defined in (5). □

References

[1] BERLINET, A. and THOMAS-AGNAN, C. (2003). *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Kluwer Academic Publishers, Boston.

[2] BOLTHAUSEN, E. (1982). On the Central Limit Theorem for stationary mixing random fields. *The Annals of Probability*, **10** 1047–1050.

[3] BOSQ, D. (2002). Functional Test of Fit. In: C. Huber-Carol et al.: (Eds.) *Goodness-of-Fit Tests and Model Validity*. Birkhäuser, Boston, 341–356.

[4] BOSQ, D. (2005). *Inférence et prevision en grandes dimensions*. Economica, Paris.

[5] BOSQ, D. and LECOUTRE, J.P. (1987). *Theorie de l'estimation fonctionnelle*. Economica, Paris.

[6] CLIFF, A.D. and ORD, J.K. (1973). *Spatial Autocorrelation*. Pion Ltd., London.

[7] CRESSIE, N.A.C. (1993). *Statistics for Spatial Data*. Wiley, New York.

[8] DOUKHAN, P. (1994). *Mixing. Properties and Examples*. Springer Verlag, New York.

[9] FORTET, R. (1995). *Vecteurs, fonctions et distributions aléatoires dans les hispaces de Hilbert*. Editions Hermès, Paris.

[10] GRENANDER, U. (1981). *Abstract inference*. Wiley, New York.

[11] GUYON, X. and RICHARDSON, S. (1984). Vitesse de convergence du théorème de la limite centrale pour des champs faiblement dépendants. *Z. Wahrscheinlichkeitstheorie verw.*, **66**, 297–314.

[12] IGNACCOLO, R. (2004). Goodness-of-Fit Tests for Dependent Data, *Journal of Nonparametric Statistics*, **16**, 19–38.

[13] IGNACCOLO, R. and RIBECCO, N. (2003). Fitting spatial autocorrelated data, *Atti del Convegno S.Co. 2003*, Universitá Ca' Foscari di Venezia, 241–246.

[14] JOHNSON, N.L., KOTZ, S. and BALAKRISHNAN, N. (1994). *Continuous univariate distributions. Vol. 1*. Wiley, New York.

[15] LANCASTER, H.O. (1975). Joint Probability Distribution in the Meixner Classes. *J. R. Stat. Society, B*, **37**, 434–443.

[16] MATHAI, A.M. and PROVOST, S.B. (1992). *Quadratic forms in random variables*. Marcel Dekker, New York.

[17] NEYMAN, J. (1937). "Smooth" test for goodness of fit. *Skand. Aktuarietidskr.*, **20**, 150–199.

[18] POLITIS, D.N. and ROMANO J.P. (1993). Nonparametric Resampling for Homogeneous Random Fields. *J. of Multivariate Analysis*, **47**, 301–328.

[19] POLITIS, D.N. and ROMANO J.P. (1996). On flat-top kernel spectral density estimators for homogenous random fields. *J. of Statistical Planning and Inference*, **51**, 41–53.

[20] RAYNER, J.C.W. and BEST, D.J. (1989). *Smooth Tests of Goodness of Fit*. Oxford University Press, New York.

[21] RIPLEY, B.D. (1981). *Spatial Statistics*. Wiley, New York.

[22] ROGERS, A. (1974). *Statistical analysis of spatial dispersion*. Pion Ltd., London.

[23] ROSENBLATT, M. (2000). *Gaussian and Non-Gaussian Linear Time Series and Random Fields*. Springer Verlag, New York.