ESTMATING AND TESTING ZONES OF ABRUPT CHANGES FOR SPATIAL DATA

BY

Denis Allard
Edith Gabriel
Jean-Noël Bacro

Research Report No. 2
January 2005

Estimating and Testing Zones of Abrupt Changes for Spatial Data

Denis Allard[1], Edith Gabriel[1] and Jean-Noël Bacro[2]

[1]Institut National de la Recherche Agronomique, Unité de Biométrie
Domaine Saint-Paul, Site Agroparc 84914 Avignon, Cedex 9, France
[2] UMR CNRS 5149, Equipe de probabilités et statistiques, Université Montpellier II
Place Eugène Bataillon, 34095, Montpellier, Cedex 5, France

Summary. We propose a method for detecting the zones where a variable, irregularly sampled in the plane, changes abruptly. Our general model is that under the null hypothesis the variable of interest is the realization of a stationary Gaussian process with constant expectation. The alternative is modeled as a discontinuity of the expectation along a set of curves. A local test is first built using distribution properties of the estimated gradient under the null hypothesis. The points where the local test is rejected define the potential Zones of Abrupt Change (ZACs). These local tests are then aggregated into a single global test. The theory that links the local tests and the global test is based on asymptotic distributions of excursion sets of nonstationary χ^2 fields. We first establish new results about the curvature at local maxima of these fields and give the asymptotic distribution of the area of the connected components of the excursion sets for high thresholds used to test the global significance of the potential ZACs. Issues arising when implementing the method in practice are discussed and the method is evaluated by a simulation study. Its application to a soil data set from an agricultural field shows how it can be used in an exploratory analysis for proposing a subdivision of the domain. Finally, it is compared to other existing methods which estimate local rates of change.

KEY WORDS: Gradient estimation; Changecurve; Excursion set; Gaussian random fields; Non-stationary χ^2 random fields; Geostatistics.

1 Introduction

In this paper we propose a method for detecting the zones where a variable, irregularly sampled in the plane, changes abruptly. Our general model will be that the variable of interest is a realization of a process $Z(x)$ on a domain $\mathcal{D}$, whose mean function may present abrupt change. Under the null hypothesis H_0 the process is stationary and has a constant expectation on $\mathcal{D}$. The alternative H_1 is a discontinuity of $E[Z(x)]$ along a set of curves Γ. Detecting the discontinuities of $E[Z(x)]$ amounts to reject the null hypothesis. We are not interested in

mapping the zones where the gradient is high under the null, but in detecting the zones where the gradient corresponds to departure from the null. Note that there is no assumption about the shape of the curves. They can be regular curves in the plane, intersecting or not. Because we consider the case of irregular sparse sampling, it will not be possible to estimate precisely these curves, but rather the zones of sharp variations. Hence we prefer to refer to the estimation of Zones of Abrupt Change (ZACs).

Since ZACs can have a complex geometry in the plane, methods for estimating ZACs must be assumption free about their shape. A discontinuity of the expectation field along a straight line is illustrated in Figure 1a, along a sine wave in Figure 1b. The discontinuity jump is not constant: it has high values in the center of the image and tends to zero near the edges.
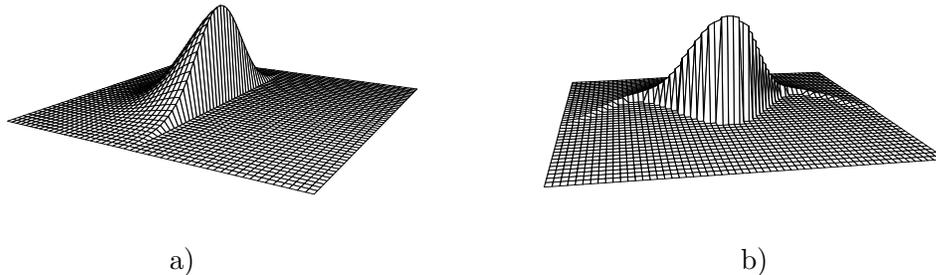

a)   b)

Figure 1: Typical examples of abrupt changes to be estimated

Mapping the zones where a variable under study changes abruptly is a problem encountered in many biological, ecological or agricultural applications. A first example is the study of populations in biology [Womble (1951), Barbujani et al. (1989)]: abrupt changes in allele frequencies of populations can be linked to the boundaries between different populations. A second example, used in this paper to illustrate the methodology, comes from agricultural applications. Precision agriculture aims at defining a location-dependent management within an agricultural field, for nitrogen fertilization for example. The agricultural field is thus divided into homogeneous zones, on the basis of soil samples (Ciba Foundation Symposium on Precision Agriculture, 1997).

Estimating ZACs is a problem related to boundary estimation, edge detection, or change

curve estimation. Edge estimation plays an important role in signal and image processing and there is an extensive body of literature about this subject. Many methods have been investigated, including Markov random fields [e.g., Geman and Geman (1984), Besag (1986) and Besag et al. (1995)], wavelets (Wang, 1998) and non parametric methods (Qiu, 1998).

The problem of estimating discontinuity curves in the plane with irregularly located samples has been much less addressed. Womble (1951) and Barbujani et al. (1989) considered an empirical method based on the upper quantile of the norm of the gradient computed on a grid. By construction, this method always detects the same proportion of grid points with high gradient, without assessing their significance.

Hall et al. (2001) proposed a non parametric tracking method based on spatial approximations of the local likelihood that the discontinuity curve passes through a given point in the plane, as a function of that point. Since it is based on tracking, it needs a starting point in the vicinity of the discontinuity. Also, local likelihood tracking implicitly assumes the existence of a discontinuity curve and therefore does not provide a way of testing its existence.

Godtliebsen et al. (2002) proposed a scale-space approach for finding significance in spatial data, which is a spatial generalization of the kernel method called SiZer developed in Chaudhuri and Marron (1999, 2000) in one dimension. It computes the kernel smoothing for a range of bandwidth h and assesses the significance of the slope for all h. Visualization is done by movies (time corresponding to bandwidth) and with pictures using streamlines along significant gradient vectors and contours. This method requires heavy computations and visualization is not easy to interpret.

Shen et al. (2002) considered the related problem of detecting the presence of a signal f on a spatial domain $\mathcal{D}$, i.e. testing the null hypothesis of whether f is 0 versus the alternative that it is nonzero somewhere. Their method assumes that the data are modeled as a Gaussian process $Z(x) = f(x) + \epsilon(x)$ where ϵ is a zero centered Gaussian stationary process. It uses a wavelet decomposition of the data observed on a grid and tests simultaneously the null hypothesis of whether the wavelet coefficients are 0 versus the alternative that they are not. They proposed a method for controlling the probability of type I errors when performing the simultaneous tests

3

on the wavelet coefficients.

Banerjee et al. (2003) proposed a Bayesian approach for estimating the gradient surface of a spatial process that is irregularly sampled. This approach provides estimates of the posterior distribution of the gradient. However, it does not provide a way to test globally the presence of ZACs. We will return to this point in Section 6.2.

Here we propose an approach based on the distribution of the estimated gradient that links the local tests of "high gradient" to a global test for testing H_0 vs. H_1. We will assume that under the null hypothesis $Z(x)$ is second order stationary, i.e. has a constant expectation and a covariance function $\text{Cov}(x, y)$ that only depends on $x - y$. The sketch of our method is the following. We first perform an optimal interpolation of the gradient using the geostatistical paradigm. The variable Z being irregularly and presumably sparsely sampled, the estimated bivariate gradient is nonstationary. We then define a local test statistic related to this estimated gradient, which enables us to detect all points where the null hypothesis is locally rejected. This set of points defines the potential ZACs. It does not necessarily define curves and it is usually disconnected. We then aggregate these local tests in a global test for each connected component of the potential ZACs, using the asymptotic distribution of their area under the null hypothesis. This method involves two different levels of significance. Let us denote $1 - \eta$ the global level of significance and $1 - \alpha$ the local one. The global probability of type I errors, η, results from the simultaneous consideration of multiple local tests at the level $1 - \alpha$. It will be shown that on the plane the value α that achieves the global level η is a function of η and the covariance function only. However, because in practice we work on a grid, α will also depend on other parameters such as the mesh of the grid on which computations are performed and the sampling pattern. We will show that it is possible to find the appropriate α that controls a specified level η, say $\eta = 0.05$, and that our method is quite robust with respect to the above mentioned parameters.

The theory that links the level α to η on the plane is based on the excursion sets of χ^2 fields, a theory developed in Adler (1981, 2000), Aronowich and Adler (1988), Worsley (1994) and Cao (1999). Standard χ^2 fields are defined as the sum of squared, independent and identically distributed Gaussian stationary random fields. In our problem, these basic requirements do not

4

hold because we deal with the gradient of an interpolated field. In consequence, the χ^2 field considered here is the sum of squared, nonstationary, not independent Gaussian random fields. We will then need to develop theoretical results for this nonstationary χ^2 field.

This paper is organized as follows. The method is presented in Section 2 along with all the theoretical results that are needed. In Section 3 we show how the method must be implemented in practice. Performances and robustness are evaluated through a simulation study in Section 4. In Section 5, the method is illustrated on the exploratory analysis of a soil data set in the context of precision agriculture. In Section 6 we compare this method with the approach in Banerjee et al. (2003). We conclude with possible extensions and a discussion.

2   The method

2.1   General framework

Let $Z = (Z(x_1), \ldots, Z(x_n))'$ be a sample of some real valued, Gaussian random field $Z(x)$, at the locations $x_1, \ldots, x_n$ of some domain $\mathcal{D}$ of $\mathbf{R}^2$. Here and in the rest of the paper, V' (resp. $\mathbf{A}'$) denotes the transpose of a vector V (resp. of a matrix $\mathbf{A}$). We will assume the following:

$\mathcal{A}_1$:   the covariance function $c(h)$ of $Z(x)$ exists and is stationary.

$\mathcal{A}_2$: $c(h)$ is infinitely differentiable when $||h|| > 0$.

Assumption $\mathcal{A}_1$ is usual in spatial statistics. In the discussion, Section 6, we will show that at a very low computation cost this assumption can be weakened to a less demanding assumption. For the sake of clarity we assume here the more restrictive assumption $\mathcal{A}_1$ and develop our method in this case. Assumption $\mathcal{A}_2$ is less usual, because it is about the regularity of the covariance function away from 0. It is a mild condition and is verified for most used covariance functions, e.g. for the Matérn class of covariance function $c(h; \kappa, \phi) \propto (\phi||h||)^\kappa K_\kappa(\phi||h||)$, where $K_\kappa(\cdot)$ is the modified Bessel function of order κ and ϕ is a scale parameter. When $\kappa = 0.5$ we obtain the exponential covariance function. See Stein (1999) for further details on the Matérn class in the context of spatial interpolation. $\mathcal{A}_2$ is however not verified for the popular spherical covariance function $c(h) = \sigma^2\{1 - 1.5||h||/b + 0.5(||h||/b)^3\}$, for $||h|| \leq b$ and $c(h) = 0$ if $||h|| > b$,

because it is not twice differentiable for $||h|| = b$. Note also that no assumption is required for $h = 0$. In particular continuity or differentiability at the origin is not required.

The null hypothesis of absence of ZAC is formally defined as $H_0 : E[Z(x)] = \mu$ for all $x \in \mathcal{D}$, and the alternative hypothesis is modeled as $H_1 :$ the set of the discontinuities of $E[Z(x)]$ define some non void set of curves $\Gamma \subset \mathcal{D}$. No assumption is made about the number or the shape of Γ. In particular Γ can be a single curve or a set of curves. They can be closed curves, or self intersecting curves. We will consider hereafter that $\mu = 0$ under the null hypothesis and the more general case where $\mu \neq 0$ is unknown will be discussed later in Section 6.

2.2 Estimation of the gradient

Since under the null hypothesis $Z(x)$ is a zero mean Gaussian random function, the optimal predictor of $Z(x)$ at an unsampled location x is the linear combination of the n samples, called simple kriging in geostatistics (see e.g. Cressie, 1993),

$$Z^*(x) = C(x)' \mathbf{C}^{-1} Z, \tag{1}$$

where $C(x) = (c(x - x_1), \ldots, c(x - x_n))'$ and $\mathbf{C} = E[ZZ']$ is the covariance matrix with elements $\mathbf{C}_{[kl]} = c(x_k - x_l)$, $k, l = 1, \ldots, n$. The covariance function of the predictor is

$$\mathrm{Cov}(Z^*(x), Z^*(y)) = C(x)' \mathbf{C}^{-1} E[Z\,Z'] \mathbf{C}^{-1} C(y) = C(x)' \mathbf{C}^{-1} C(y),$$

which shows that the field $Z^*(x)$ is not stationary. We have the following result regarding the field $Z^*(x)$.

Proposition 1 *For a covariance function verifying $\mathcal{A}_1$ and $\mathcal{A}_2$, the random field $Z^*(x)$ is infinitely differentiable for all x, except possibly at the sample locations, both almost surely and in the mean squared sense.*

This proposition is a direct consequence of standard results on random fields. Rapidly sketched, the argument is the following: for a given realization Z, the field $Z^*(x) = C(x)' \mathbf{C}^{-1} Z$ inherits the differentiability property of the functions $c(x - x_k)$, $k = 1, \ldots, n$. Hence it is infinitely

6

differentiable for all x, except maybe for points where $x - x_k = 0$, i.e. sample locations. This implies that $Z^*(x)$ is almost surely infinitely differentiable, except possibly at sample locations. Mean squared infinite differentiability of $Z^*(x)$ is then implied by the fact that Z is a Gaussian vector with finite variance. More details can be found in Stein (1999) and Yaglom (1986).

At sample locations, $Z^*(x)$ is continuous if $c(h)$ is continuous at 0, and it is r times differentiable if $c(h)$ is $2r$ times differentiable at 0. If $c(h)$ has a discontinuity at the origin (called a nugget effect in the geostatistical literature), $Z^*(x)$ is discontinuous at sample locations. Since the set of sample locations is a set of null measure in $\mathcal{D}$, it is not a serious limitation. From now on, we will consider that x is not one of the sample points: $x \in \mathcal{D} \backslash \{x_1, \ldots, x_n\}$. According to Proposition 1, the gradient $W(x)$ of the estimated field $Z^*(x)$ exists and

$$W(x) = (W_1(x), W_2(x))^{'} = (\partial_1 Z^*(x), \partial_2 Z^*(x))^{'}, \tag{2}$$

with

$$\partial_i Z^*(x) = \partial_i C(x)^{'} \mathbf{C}^{-1} Z,$$

where ∂_i denotes the partial derivative along the coordinate $i \in \{1, 2\}$. If $c(h)$ is twice differentiable for $h = 0$, $W(x)$ is the best linear unbiased predictor of the gradient of $Z(x)$. If $c(h)$ is not twice differentiable for $h = 0$, which is for example the case for an exponential covariance function, $Z(x)$ has no derivative, but following Chilès and Delfiner (1999, p.314) in this case we accept by extension $W(x)$ as a predictor of the local gradient. Banerjee et al. (2003) provides a complete distribution theory of gradient processes under the assumption of a stationary Gaussian process.

2.3   Local detection of a discontinuity

In a first step we test locally the presence of a discontinuity at a point $x \in \mathcal{D} \backslash \{x_1, \ldots, x_n\}$. We formally define the local test as follows $H_0(x) : E[Z(y)] = 0$ for all y in a small neighborhood around x, versus the alternative $H_1(x) : x \in \Gamma$, where Γ is the discontinuity curve. The local test will be rejected if the local gradient of the estimated surface is above some critical value, as

7

presented below. Let us denote $\mathbf{\Sigma}(x)$ the covariance matrix of $W(x)$:

$$\mathbf{\Sigma}(x) = E[W(x)W'(x)] = \partial C(x)'\mathbf{C}^{-1}\partial C(x),$$

where $\partial C(x)$ is the $n \times 2$ matrix $(\partial_1 C(x), \partial_2 C(x))$. Since $Z(x)$ is a Gaussian field we have that $W(x) \sim \mathcal{N}_2(0, \mathbf{\Sigma}(x))$. At a point x where $\mathbf{\Sigma}(x)^{-1}$ exists, we define

$$T(x) = W(x)'\mathbf{\Sigma}(x)^{-1}W(x). \tag{3}$$

Note that the statistics $T(x)$ is invariant to rotations. The matrix $\mathbf{\Sigma}(x)$ is invertible if none of the vectors $(\partial_i C(x))$, $i = 1, 2$ are vectors of 0s, i.e. if the point x is not such that $c(x - x_k) = 0$ for all $k = 1, \ldots, n$. In such a case, $Z^*(x)$ and $W(x)$ are identically null, and no statement at all can be made about the gradient. From now on, we will restrict ourselves to the set of points x, denoted $\mathcal{D}_0$, such that $\mathbf{\Sigma}(x)^{-1}$ exists. For $x \in \mathcal{D}_0$, $T(x)$ has a $\chi^2(2)$ distribution. The local null hypothesis is rejected if $T(x) > t_{1-\alpha} = \chi^2_{1-\alpha}(2)$ at the confidence level $1 - \alpha$. This procedure can be applied on the nodes of a grid superimposed on the domain $\mathcal{D}$ under study, which enables us to map the regions where the local test is rejected. The set of points were the local test is rejected defines the potential ZACs.

We illustrate this first step on the following example. On the unit square, a sample of a standard Gaussian random field with an exponential covariance function $c(h) = \exp(-||h||/b)$ with $b = 0.1$ is simulated on 100 randomly located samples $x_1, \ldots, x_n$, thus defining a vector Z. To avoid potential numerical instabilities due to very close samples, these locations are generated according to a Strauss process with a hard-core distance equal to 0.02. A discontinuity is introduced by adding a constant a to all samples whose abscissa is less than 0.4. The case $a = 0$ corresponds to H_0 (absence of discontinuity). A realization of this process is depicted Figure 2a for $a = 2.5$. The circles indicate the sample locations and their size is proportional to the value. Figure 2b shows the local tests at the confidence level $1 - \alpha = 0.9994$. This level is not arbitrary: it corresponds to a global level $\eta = 0.05$, as it will be explained in Section 3. Each pixel of a 60×60 grid is colored if the test statistic T in this pixel is above $t_{0.9994} = 14.8$. Black pixels correspond to significant ZACs according to the theory developed in the next section, while grey pixels correspond to non significant ones. A large vertical stripe of colored pixels is

8

visible along the discontinuity. When the same procedure was applied with $a = 0$, there were no pixel with a test statistic above $t_{1-\alpha}$.
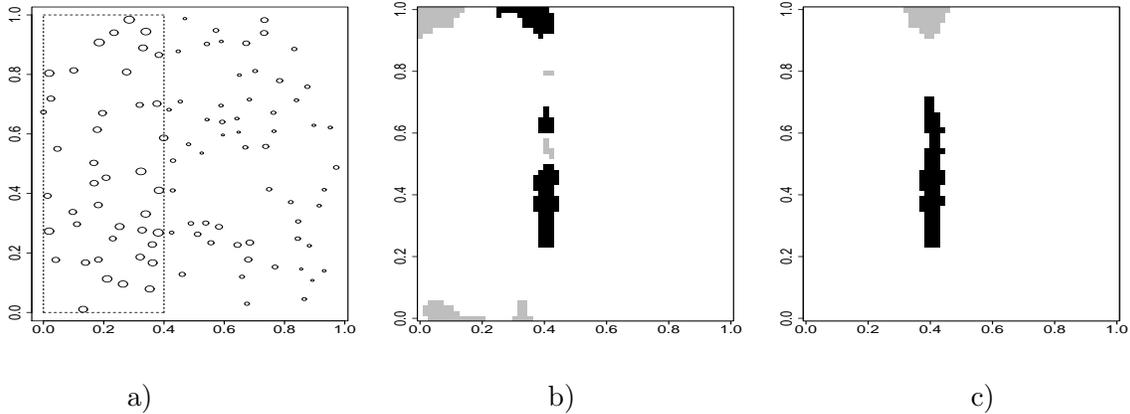

Figure 2: a) 100 randomly located samples of a standard stationary Gaussian random field with the addition of a constant within the rectangle in dashed lines; b) Detection of the ZACs using the true covariance; significant ZACs are in black; non significant ones are in grey; c) Detection of the ZACs with an estimated covariance (last iteration).

If the random field is stationary, we should find a proportion of about α pixels in which $H_0(x)$ is rejected (i.e. 2.16 pixels in the example above). The tests in neighboring grid nodes are not independent, so we also expect these pixels to be close to each other. On the contrary, if there is a discontinuity (or a sharp variation) of the expectation field, we must expect the grid nodes where $H_0(x)$ is rejected to be much more numerous and located on, or near the discontinuity, as illustrated in Figure 2b.

The final step is now to propose a global test for rejecting H_0 versus H_1. This necessitates to simultaneously test the dependent simple tests $H_{0,s}$ versus $H_{1,s}$, where s is an index for the grid nodes in the potential ZACs. We will follow an approach similar to the one used in functional RMI (Worsley, 2001) and obtain an asymptotic distribution for the area of the connected components of the potential ZACs for testing its significance.

2.4 Global test for potential ZACs: theoretical results

Potential ZACs correspond to the excursion set of the field $T(x)$ above a level $t_{1-\alpha}$. In this section it will be shown that under H_0 the area of a connected component, say $S_{t_{1-\alpha}}$, of the excursion set above $t_{1-\alpha}$ is related to the curvature at the local maximum in the connected component, according to:

$$\frac{t_{1-\alpha}S_{t_{1-\alpha}}|\mathbf{\Lambda}|^{1/2}}{\pi} \xrightarrow{\mathcal{L}} E(2) \tag{4}$$

as $t_{1-\alpha} \to \infty$, where $E(2)$ is an exponential random variable with expectation 2 and $\mathbf{\Lambda}$ is a 2×2 matrix defined below. We first need to establish some theoretical results about the nonstationary field $T(x)$ under the null hypothesis, in particular about its curvature at a local maximum. Then, we will prove equation (4).

2.4.1 The curvature of $T(x)$

Let us first introduce some new notations: for a twice differentiable field $T(x)$, $\dot{T}(x)$ will denote the vector $(\partial_1 T(x), \partial_2 T(x))'$ and $\ddot{\mathbf{T}}(x)$ is the 2×2 matrix whose elements are $\partial^2_{kl}T(x)$, with $1 \le k, l \le 2$.

We perform the change of variable $U(x) = \mathbf{\Sigma}(x)^{-1/2}W(x)$ such that $T(x)$ is the sum of two independent, $\mathcal{N}(0,1)$, squared random variables, $T(x) = U_1^2(x) + U_2^2(x)$. Explicitly,

$$U_1(x) = \frac{W_1(x)}{\sigma_1(x)}, \quad U_2(x) = \frac{1}{\sqrt{1-\rho^2(x)}}\left\{\frac{W_2(x)}{\sigma_2(x)} - \rho(x)\frac{W_1(x)}{\sigma_1(x)}\right\}, \tag{5}$$

where $\sigma_i(x) = \sqrt{\mathbf{\Sigma}_{[ii]}(x)}$, $i = 1, 2$ and $\rho(x) = \mathbf{\Sigma}_{[12]}(x)/\{\sigma_1(x)\sigma_2(x)\}$. Let

$$\rho_{ij}(x,y) = \frac{E[W_i(x)W_j(y)]}{\sigma_i(x)\sigma_j(y)} = \frac{\partial_i C(x)'\mathbf{C}^{-1}\partial_j C(y)}{\sigma_i(x)\sigma_j(y)}$$

be the correlation function between $W_i(x)$ and $W_j(y)$, $i, j = 1, 2$. Then, $\rho_{12}(x,x) = \rho(x)$ and $\rho_{11}(x,x) = \rho_{22}(x,x) = 1$. We have the following results regarding the covariance functions for the fields U_1 and U_2 :

$$E[U_1(x)U_1(y)] = \rho_{11}(x,y),$$

$$E[U_1(x)U_2(y)] = \left\{\rho_{12}(x,y) - \rho(y)\rho_{11}(x,y)\right\}/\sqrt{1-\rho^2(y)},$$

10

$$E[U_2(x)U_2(y)] = \{\rho_{22}(x,y) - \rho(x)\rho_{12}(x,y) - \rho(y)\rho_{21}(x,y) + \rho(x)\rho(y)\rho_{11}(x,y)\}/\sqrt{(1-\rho^2(x))(1-\rho^2(y))}.$$

These equations show that the marginal distributions of $U_1(x)$ and $U_2(x)$ are stationary, but that their covariance functions depend on both x and y. Hence, the fields are not stationary. Neither are they identically distributed ($E[U_1(x)U_1(y)] \neq E[U_2(x)U_2(y)]$) or independent ($E[U_1(x)U_2(y)] \neq 0$).

Proposition 2 *Conditional on $U_i(0)$, $\dot{U}_i(0)$ and $\ddot{U}_i(0)$ ($0 \in \mathcal{D}_0$), the fields $U_i(x)$ defined above have the same distribution as:*

$$U_i(0) + x'\dot{U}_i(0) + \frac{1}{2}x'\ddot{U}_i(0)x + \eta_i(x) \tag{6}$$

in a neighborhood of 0, where $\eta_i(x)$ is a Gaussian field with mean $E[\eta_i(x)] = o(||x||^2)$ and covariance function $E[\eta_i(x)\eta_i(y)] = o(||x||^2||y||^2)$.

Proof : We first re-write $U_i(x) = A_i(x)'\mathbf{C}^{-1}Z$, $i = 1,2$, where $A_i(x)$ are the vectors of length n derived from (5). For lighter notations, we now drop the subscript i. The n (non random) surfaces of $A(x)$ are infinitely differentiable and hence so are the random fields $U(x)$, both almost surely and in the mean squared sense. For all realizations of Z, one can write

$$U(x) = \sum_{k=1}^{n} A_k(x)[\mathbf{C}^{-1}Z]_k = \sum_{k=1}^{n} \left(A_k(0) + x'\dot{A}_k(0) + \frac{1}{2}x'\ddot{A}_k(0)x + R_k(x) \right)[\mathbf{C}^{-1}Z]_k,$$

where $A_k(0)$ is the k-th element of $A(0)$ and similarly for $\dot{A}_k(0)$ and $\ddot{A}_k(0)$, $[\mathbf{C}^{-1}Z]_k$ is the k-th element of the vector $\mathbf{C}^{-1}Z$, $R_k(x) \leq M_k||x||^3$ and M_k is a constant. Hence,

$$U(x) = U(0) + x'\dot{U}(0) + \frac{1}{2}x'\ddot{U}(0)x + R(x)'\mathbf{C}^{-1}Z,$$

with $R(x) = (R_1(x), \ldots, R_n(x))'$. Then, conditional on $V = (U, \partial_1 U, \partial_2 U, \partial_{11}^2 U, \partial_{12}^2 U, \partial_{22}^2 U)(0)'$, we have $U(x) = U(0) + x'\dot{U}(0) + \frac{1}{2}x'\ddot{U}(0)x + \eta(x)$, where $\eta(x)$ is a Gaussian random field with

$$E[\eta(x)] = R(x)'\mathbf{C}^{-1}E[Z \mid V] \text{ and } E[\eta(x)\eta(y)] = R(x)'\mathbf{C}^{-1}E[ZZ' \mid V]\mathbf{C}^{-1}R(y).$$

Hence $E[\eta(x)] = o(||x||^2)$ and $E[\eta(x)\eta(y)] = o(||x||^2||y||^2)$. $\qquad\square$

11

We now move to the structure of the field $T(x)$ near a local maximum. Let $T_{T_0}(x)$ be the field $T(x)$ conditional on the event that $T(x)$ has a local extremum with height T_0 at $x = 0$, i.e. $T(0) = T_0$, $\dot{T}(0) = 0$, (see for example Adler (1981) for a precise definition of such a conditional field). Let us also $\mathbf{\Lambda}_i(x) = E[\dot{U}_i(x)\dot{U}_i(x)'] = \mathrm{Var}(\dot{U}_i(x))$ denote the covariance matrix of the gradient of the field $U_i(x)$. Their expression, given in Section 3.1 and Appendix A, depends only on the covariance function $c(h)$ and its first and second order derivatives.

Theorem 1 *Consider the fields $U_1(x)$, $U_2(x)$ and $T(x)$ defined as above and let $v(x) = U_1^2(x)/T(x)$. Conditional on $T(x)$ having a local extremum at $0 \in \mathcal{D}_0$ of height T_0,*

$$T_0^{-1}\ddot{\mathbf{T}}_{T_0}(0) + 2\{v(0)\mathbf{\Lambda}_1(0) + (1 - v(0))\mathbf{\Lambda}_2(0)\} = o_p(1),$$

as $T_0 \to \infty$, where $\ddot{\mathbf{T}}_{T_0}(x)$ denotes the second order derivative of $T_{T_0}(x)$.

Theorem 2 *Conditional on $T(x)$ having a local maximum at $0 \in \mathcal{D}_0$ of height T_0, for $x \in \mathcal{D}_0$ in a neighborhood of 0,*

$$T_{T_0}(x/\sqrt{T_0}) - T_0 + x'\{v(0)\mathbf{\Lambda}_1(0) + (1 - v(0))\mathbf{\Lambda}_2(0)\}x = o_p(1), \quad as \ \ T_0 \to \infty.$$

Theorem 1 states that for high local extrema of the field $T(x)$, the matrix of second derivatives converges in law to a negative definite deterministic matrix. Hence, the local extremum is a local maximum. The proof of this theorem necessitates several lemmas and is given in appendix. The proof of theorem 2 relies on a straightforward Taylor expansion of $U_1(x)$ and $U_2(x)$ and the use of (6) and Theorem 1. These results are an extension of the results for standard χ^2 fields shown in Cao (1999).

2.4.2 Asymptotic distribution of the area of a potential ZAC

It is now possible to establish an asymptotic law for the area of the connected components of the excursion set of $T(x)$ above a high threshold t. Let us denote A_t the excursion set above a threshold t, $A_t = \{x : T(x) > t\}$ and $\mathcal{C}_t$ one of its connected components. According to Theorem

12

2, for high levels t the field $T(x)$ is approximately an elliptic paraboloïd in the neighborhood of a local maximum, say of height T_m. The area of $\mathcal{C}_t$, denoted S_t, is then the area of the horizontal section of this elliptic paraboloïd at a distance $T_m - t$ from the maximum. The next two theorems provide an asymptotic distribution for the area S_t for high thresholds t. This will allow us to use S_t as a test statistic for the global test.

Theorem 3 *Conditional on $T(x)$ having a local maximum $T_0 > t$ at $0 \in \mathcal{D}_0$,*

$$S_t = \frac{T_0 - t}{T_0}\pi|\mathbf{\Lambda}(0)|^{-1/2} + o_p(1/T_0), \tag{7}$$

as $t \to \infty$, where $\mathbf{\Lambda}(0) = v(0)\mathbf{\Lambda}_1(0) + (1 - v(0))\mathbf{\Lambda}_2(0)$ and $|\cdot|$ denotes the determinant of a matrix and v is defined in Theorem 2.

Proof : According to Theorem 2, conditional on $T(0)$ being a local maximum of height T_0, $T(x)$ has the representation $T_{T_0}(x) = T_0(1 - x'\mathbf{\Lambda}x) + o_p(1)$ on the connected component $\mathcal{C}_t = \{x \, : \, T_0(1 - x'\mathbf{\Lambda}x) + o_p(1) > t\}$ as $t \to \infty$. Its Lebesgue measure is $S_t = \int_{T_0(1-x'\mathbf{\Lambda}x)+o_p(1)>t} dx$ which leads to the result. $\qquad\square$

Theorem 4 *Conditional on $T(x)$ having a local maximum $T_0 > t$ at $0 \in \mathcal{D}_0$,*

$$tS_t \overset{\mathcal{L}}{\to} \pi|\mathbf{\Lambda}(0)|^{-1/2}E(2), \tag{8}$$

as $t \to \infty$, where $E(2)$ is an exponential random variable with expectation 2.

Proof : We first show the following:

$$T_0 - t = E(2) + o_p(1) \tag{9}$$

conditional on $T_0 > t$ as $t \to \infty$. We will use an argument similar to those used by Aldous (1989): the local maxima of $T(x)$ in $\mathcal{D}$ can be considered as a point process. Because $T(x)$ is infinitely differentiable, this point process is locally finite. Let us denote $\lambda(z)dz$ the rate of local maxima of heights in $[z, z + dz[$, i.e. the expectation of the number of such local maxima per area unit. Then, assuming that border effects can be ignored,

$$P(T(x) > t) = e^{-t/2} = \frac{1}{\nu(\mathcal{D})}\int_{\mathcal{D}}\int_t^{+\infty}\lambda(z)\mathbf{1}_{\mathcal{C}_{y,z,t}}(x)dz\,dy,$$

13

where $\nu(\mathcal{D})$ is the area of $\mathcal{D}$ and $\mathcal{C}_{y,z,t}$ denotes a connected component of the excursion set above t having a maximum of height z at y. Now, $x \in \mathcal{C}_{y,z,t} \Leftrightarrow 0 \in \mathcal{C}_{x-y,z,t} \Leftrightarrow y - x \in \check{\mathcal{C}}_{0,z,t}$ with $\check{\mathcal{C}}_{0,z,t}$ being the symmetrical of $\mathcal{C}_{0,z,t}$ around 0. Hence,

$$P(T(x) > t) = \frac{1}{\nu(\mathcal{D})} \int_{\mathcal{D}} \int_t^{+\infty} \lambda(z) \mathbf{1}_{\check{\mathcal{C}}_{0,z,t}}(y - x) dz \, dy.$$

Then,

$$e^{-t/2} = \frac{1}{\nu(\mathcal{D})} \int_t^{+\infty} \lambda(z) S(0, z, t) dz,$$

where $S(0, z, t)$ is the area of $\mathcal{C}_{0,z,t}$. According to Theorem 3, for high t and conditional on a height z, $S(0, z, t) = (z - t)\pi|\mathbf{\Lambda}(0)|^{-1/2}/z + o(1/z)$. Hence, neglecting $o(1/z)$ and after rearrangement,

$$K \int_t^{+\infty} \lambda(z) \frac{z - t}{z} dz = e^{-t/2},$$

where $K = \pi\nu(\mathcal{D})^{-1}|\mathbf{\Lambda}(0)|^{-1/2}$. It is easy to check that $\lambda(z) = (4K)^{-1}z e^{-z/2}$ is the solution of this integral equation. This leads to:

$$P(T_0 > t + u \mid T_0 > t) = \frac{\int_{t+u}^{+\infty} \lambda(z) dz}{\int_t^{+\infty} \lambda(z) dz} = e^{-u/2}(1 + O(u/t)).$$

The rest of the proof is now straightforward. According to (9), as $t \to \infty$, $T_0 = t + O_p(1)$. Plugging this in (7) leads to $S_t = E(2)\pi|\mathbf{\Lambda}(0)|^{-1/2}/t + O_p(1/t)$. $\square$

When $t_{1-\alpha}$ is very large, there is with a high probability at most one connected component on $\mathcal{D}$ under H_0 and testing on each connected component is then equivalent to testing on the entire domain $\mathcal{D}$. This theorem enables us to reject for each connected component of the potential ZAC the null hypothesis H_0 of whether the connected component of the potential ZAC is from a stationary random field versus the alternative H_1. At a global level of confidence $1 - \eta$, H_0 will be rejected if

$$\exp\left(-\frac{t_{1-\alpha} S_{t_{1-\alpha}} |\mathbf{\Lambda}|^{1/2}}{2\pi}\right) < \eta,$$

where $\mathbf{\Lambda}$ is computed at the maximum of $T(x)$ in $\mathcal{C}_{t_{1-\alpha}}$. On the plane, provided that $t_{1-\alpha}$ is large enough this test is exact. Because in practice we work on a grid, some care must be taken regarding the appropriate value α for achieving a desired level η. We address this problem in the next section.

14

3 Practical implementation

We now detail how this procedure is implemented in practice. In particular we address successively the problem of the discretization of the plane by a grid, how the level α for this grid is found and finally the estimation of the covariance function under the alternative.

3.1 Working on a grid

In this section we suppose that the covariance function is known. Its estimation is considered in section 3.3. On each grid node (i, j), the gradient $W(i, j)$, the matrix $\boldsymbol{\Sigma}(i, j)$ and the field $T(i, j)$ are first computed. On the grid nodes where $\boldsymbol{\Sigma}(i, j)$ is invertible, the gradient is multiplied by $\boldsymbol{\Sigma}(i, j)^{-1/2}$, thus leading to $U_1(i, j)$ and $U_2(i, j)$. For these computations, the covariance matrix $\mathbf{C}$ of the data, and for each pixel the covariance vector $C(i, j)$ and its derivatives $D_1(i, j) = \partial_1 C(i, j)$ and $D_2(i, j) = \partial_2 C(i, j)$ are needed. They are computed from the covariance function $c(h)$ of the random field $Z(x)$. We show here how to compute $\boldsymbol{\Lambda}_1$ at the grid node (i, j). Dropping the dependence to (i, j) we have:

$$\boldsymbol{\Lambda}_{1\,[k,l]} = E[\partial_k U_1 \partial_l U_1] = E[\partial_k(W_1/\sigma_1)\partial_l(W_1/\sigma_1)] = \{\partial_k D_1' \mathbf{C}^{-1} \partial_l D_1 - \partial_k \sigma_1 \partial_l \sigma_1\}/\sigma_1^2, \quad k, l = 1, 2,$$

where $\partial_k \sigma_1 = \partial_k D_1' \mathbf{C}^{-1} D_1/\sigma_1$. There is a similar albeit much longer expression for $\boldsymbol{\Lambda}_2$, mainly because of the factor involving $\rho(x)$ that needs also to be differentiated. It is given in appendix. For both $\boldsymbol{\Lambda}_1(i, j)$ and $\boldsymbol{\Lambda}_2(i, j)$ the second order derivatives $\partial_{kl}^2 C(i, j)$ need to be computed.

Then, for the level $1 - \alpha$, the set of the grid nodes whose statistic $T(i, j)$ is above the $1 - \alpha$ quantile of a $\chi^2(2)$ distribution define the potential ZAC. For each of its connected component, its area $S_{t_{1-\alpha}}$ and the values U_1, U_2, v, $\boldsymbol{\Lambda}_1$ and $\boldsymbol{\Lambda}_2$ are computed on the grid node where T is maximum. A p-value is then computed according to Equation (8) :

$$p = \exp\left(-\frac{t_{1-\alpha} S_{t_{1-\alpha}} |\boldsymbol{\Lambda}|^{1/2}}{2\pi}\right), \tag{10}$$

with $\boldsymbol{\Lambda} = v\boldsymbol{\Lambda}_1 + (1 - v)\boldsymbol{\Lambda}_2$ and $v = U_1^2/T$. Equation (10) is thus a numerical approximation of Theorem 4 due to the discretization of the plane. This approximation was explored by means of

a series of 1000 independent simulations under H_0 as described in Section 2. On each of them, the statistic $X = t_{1-\alpha} S_{t_{1-\alpha}} |\mathbf{\Lambda}|^{1/2}/\pi$ was computed on all connected components of potential ZACs. For conciseness considerations, the results are not illustrated here, but a discretization effect was clearly apparent. Since the area of the excursion set decreases as the threshold $t_{1-\alpha}$ increases, the grid must be sufficiently fine to detect small connected components when the level $1 - \alpha$ is high. For a combination of a coarse grid and a high threshold, the discretization effect leads to a lack of small values for the statistic X. When the level $1 - \alpha$ is not high enough, the convergence in law (8) does not hold. In conclusion, the grid must be sufficiently fine to capture all connected components of the excursion set and the level $1 - \alpha$ must be sufficiently high for the convergence in law to hold, but not too high relatively to the mesh of the grid. In all tested situations, we found a 60×60 grid to be sufficiently fine.

3.2 Finding α

On a grid, the appropriate level α achieving a specified level η is thus a trade-off between the necessity of having a high level for the convergence in law to hold on the one hand, and a level that is not to high relatively to the mesh of the grid. Series of simulations not reported here have shown that ZACs are less often detected as $1 - \alpha$ increases due to the discretization effect for high levels and more often detected as n increases, because the gradient is better predicted. We propose to find the appropriate level by Monte-Carlo simulations under H_0: a series of N simulations of a Gaussian field under the null hypothesis, conditional on the same discretization, sample locations and covariance function is performed. The level $\hat{\alpha}$ corresponding to η is then defined as

$$\hat{\alpha} = \sup\{\alpha : M_\alpha \leq \eta N\}, \tag{11}$$

where M_α is the number of simulations (among N) with significant ZACs at the level $1 - \alpha$.

We performed $N = 1000$ simulations with $n = 100$ on a 60×60 grid. An exponential covariance model was simulated with $b = 0.1$ and $b = 0.2$. To each simulation correspond specific sampling locations (randomly drawn according to a Strauss process), hence a specific level $1 - \alpha_i$, $i = 1, \ldots, N$. A series of 100 realizations at the same sample locations are then

generated and each value $\hat{\alpha}_i$ is found according to (11) for $\eta = 0.05$. The necessary number of realizations for a given precision can be found using binomial confidence intervals, but this has not been pursued here. For $b = 0.1$, $1 - \hat{\alpha}_i$ ranges from 0.9973 to 0.9996, the median being 0.9991. It corresponds to thresholds for the χ^2 field from 11.8 to 17.0, the median being 14.0. For $b = 0.2$ these values are $(1 - \hat{\alpha})_{min} = 0.9987$, $(1 - \hat{\alpha})_{max} = 0.9998$ and $(1 - \hat{\alpha})_{med} = 0.9994$. Note that the number of samples, the covariance function and the grid being identical for all simulations, the differences are only due to the sampling pattern. For these very high levels, potential ZACs under H_0 are almost always unique and very small (between 1 and 3 pixels). Hence, at these levels, it is equivalent to consider that the global level of significance $1 - \eta$ is defined for each connected component or globally for the entire domain.

To illustrate this procedure, we go back to the example depicted in Figure 2. Using Monte-Carlo simulations, it was found that $1 - \hat{\alpha} = 0.9994$. The p-value of each connected component of the potential ZACs is now computed using (10). When the p-value is smaller than $\eta = 0.05$, it is considered as significant and colored in black; otherwise it is colored in grey. We define the ZACs as the set of significant connected components of the potential ZACs. On Figure 2b, there are three ZACs along the discontinuity; the p-value of the largest is less than 10^{-10}, the other two are 0.002 and 0.047. The other clumps have a p-value larger than 0.05 (from 0.080 to 0.632).

3.3 Estimation of the covariance function under the alternative

Equation (10) assumes that the covariance function is known. In practice when analyzing data it is unknown and needs to be estimated. The covariance function is estimated using the variogram function, which is widely used in the geostatistical literature, see e.g. Chilès and Delfiner (1999). For a stationary random field $Z(x)$ with variance σ^2, the variogram

$$\gamma(h) = E[\{Z(x) - Z(x + h)\}^2]/2, \tag{12}$$

is related to the covariance function: $\gamma(h) = \sigma^2 - c(h)$. There are numerous methods for estimating the covariance function. We have chosen to use a weighted least square fit, as

advocated by Cressie (1993). It has the advantage of being implemented in many packages, and we have used throughout the `variofit` function of the `geoR` package. Also, it is more robust to departure from the Gaussian assumption than likelihood based methods. The estimated parameters of the covariance function minimize the distance between a given parametric family of variogram models and the experimental variogram $\hat{\gamma}(h) = 1/(2n_h) \sum_{x_i-x_j \sim h} \{Z(x_i) - Z(x_j)\}^2$, where n_h is the number of pairs involved in the sum. We refer to the literature for a more detailed presentation of the variogram and the practice of its estimation.

Under the alternative hypothesis, the presence of discontinuities in the expectation field leads to some difficulties in the estimation of the variogram. The experimental variogram $\hat{\gamma}(h)$ involves differences $Z(x_i) - Z(x_j)$. If the points x_i and x_j are on different sides of a discontinuity this will imply a bias in the estimated parameters of the variogram. In particular the estimated variance $\hat{\sigma}^2$ will be overestimated. This will lead to smaller potential ZACs, higher p-values and ultimately to a loss of power. To solve this difficulty, we propose the following iterative procedure. We first estimate a global covariance function (i.e. using all pairs of samples) and a first set of ZACs. We then re-estimate the covariance function, but all pairs of samples $\{Z(x_i), Z(x_j)\}$ for which the segment $[x_i, x_j]$ intersects a significant ZAC are discarded from the variogram estimation procedure. ZACs are then re-estimated using the new covariance function, and the procedure is iterated until convergence occurs. If the range parameter of the variogram changes, the level α must be re-evaluated using a new series of simulations. Convergence is reached when the set of ZACs remains unchanged. We have no formal proofs of the convergence of the procedure, but convergence was always observed in less than five iterations. This procedure usually decreases dramatically the estimated variance and increases the detected ZACs. Under the null hypothesis, false positive ZACs are detected with a probability η; in this case they are very small and only very few pairs of points are discarded for the re-estimation of the covariance function. On all tested situations, ZACs were unchanged at the second iteration. As illustration, we analyze again the synthetic data pictured in Figure 2 pretending that the covariance function is unknown. The covariance function is estimated and leads to the following overestimated parameters: $(\hat{b}, \hat{\sigma}^2) = (0.28, 3.02)$. The corresponding level for the test is $1 - \hat{\alpha} = 0.9985$. The covariance is

then re-estimated as described above and convergence is reached in four iterations with final estimates $(\hat{b}, \hat{\sigma}^2) = (0.21, 1.35)$ and $1 - \hat{\alpha} = 0.9994$. The covariance parameters are closer to the true values; the variance in particular has been decreased by 60%. The final detected ZAC is depicted Figure 2c. This example shows that under H_1 it is possible to estimate simultaneously the covariance function and the ZACs.

4 Simulation Study

In this section, we explore two important points using a simulation study: the power of the method and the influence of an error of estimation of the covariance functions. The first point can, at least locally, be tackled mathematically. It is currently under scrutiny and deserves a paper on its own. The second problem is much more complicated and is usually considered using computer intensive methods in spatial statistics.

4.1 Power of the method

On each of the 1000 realizations on which $\hat{\alpha}_i$, $i = 1, \dots, N$, was estimated in section 3.2, a constant a (varying from 0 to 3) is now added in the region where the abscissa is less than 0.4, as described in Section 2.4. Three levels of significance are considered: $1 - \alpha = 0.995, 0.999$ and $1 - \hat{\alpha}_i$. Note that by definition, only $1 - \hat{\alpha}_i$ correctly controls the global level η. Not all significant ZACs correspond to the discontinuity: ZACs detected by the algorithm are considered as "true" if the abscissa of the maximum is in the interval $x_1 \in [.27, .53]$, corresponding to the columns $16, \dots, 32$ of the 60×60 grid. Other ZACs correspond to ZACs that should not be detected. The proportion of simulations with "true" ZACs at the level $1 - \hat{\alpha}_i$ corresponds to the power of the global test. Table 1 reports the number of simulations for which "true" significant ZACs are detected. The following observations can be made: i) the level $1 - \alpha = 0.995$ is too low since it detects 88 simulations with true ZACs under H_0, a number significantly higher than expected; ii) there is not a significant difference in the detection rate using the levels $1 - \hat{\alpha}$ or 0.999. Hence, as a first approximation, it would probably be appropriate to use the latter instead of using the

more computer intensive value $1 - \hat{\alpha}$; iii) the discontinuity is more often detected as a increases, and it is very often detected (more than 80% of the simulations) for $a = 2.5$ which is quite a good result considering that there are only 100 sample values in the domain and that they are not independent; iv) in presence of a discontinuity, the detection rate is higher when the range parameter b increases because a larger range parameter corresponds to a higher regularity of the underlying random field.

Similar simulations not reported here show that a higher density of samples yields to an even better detection rate (e.g., in the same conditions, for $n = 200$, a discontinuity with $a = 2$ is detected for more than 90% of the simulations).

Table 1: Number of simulations (out of 1000) with "true" significant ZACs for different values of $1 - \alpha$ (60×60 grid; $n = 100$, exponential covariance with range parameter b).

	$b = 0.1$			$b = 0.2$		
a	0.995	0.999	$1 - \hat{\alpha}$	0.995	0.999	$1 - \hat{\alpha}$
0	88	28	26	79	22	25
1	259	82	72	292	121	127
1.5	498	258	246	609	367	381
2	778	548	588	887	722	727
2.5	941	835	811	986	945	941
3	993	971	967	999	999	997

4.2   Sensitivity to the covariance estimation

We now explore the sensitivity of the method to misspecification of the covariance function: a Gaussian random field is simulated with a given covariance function, and the method is applied with a different one. Simulations are performed using the same 1000 sampling locations as above for which the levels $1 - \hat{\alpha}_i$ were already estimated for the Exp(0.1) case. On each of these locations, Gaussian fields with different covariances are simulated in addition to the Exp(0.1) realization: two of them are of the same parametric form, with different range parameters ($b = 0.07$ and $b = 0.13$) and one is the so called spherical covariance function $c(h) = 1 - 1.5||h||/b + 0.5(||h||/b)^3$, for $||h|| \leq b$ and $c(h) = 0$ if $||h|| > b$, often used in geostatistics

20

for interpolation purpose. It is continuous at the origin (but not differentiable) and infinitely differentiable when $||h|| > 0$ except for $||h|| = b$ where it is not twice differentiable. Hence, for all points of $\mathcal{D}$ at a distance b of one of the sampling points, the method could not be applied. However, because this set of points is of null measure, no problems were encountered when applying the method blindly. The range parameter of the spherical covariance is chosen equal to 0.3, so that for both models the covariance is almost zero for the same distance $||h|| = 0.3$. The method is applied with an Exp(0.1) covariance function. The detection rate of the ZACs is reported in Table 2 for the levels $1 - \hat{\alpha}_i$ found for the baseline case.

Table 2: Number of simulations (out of 1000) with significant "true" and "other" ZACs detected for different simulated models. The method is applied using the parameters of an exponential covariance with range parameter $b = 0.1$ (60×60 grid; $n = 100$). Simulated models are exponential (resp. spherical) covariance model with parameter b', denoted $E(b')$ (resp. $S(b')$).

	"True" significant ZACs				"Other" significant ZACs			
a	E(0.07)	E(0.1)	E(0.13)	S(0.3)	E(0.07)	E(0.1)	E(0.13)	S(0.3)
0	26	26	22	19	51	43	40	33
1	78	72	81	67	59	49	39	29
1.5	242	246	258	225	75	54	39	33
2	526	508	555	487	108	59	46	40
2.5	824	811	840	804	177	78	56	55
3	967	967	980	964	264	100	63	67

In general there is not an important difference in the detection rate for the four cases considered. When simulating an Exp(0.13), the underlying random field is smoother than for an Exp(0.1), thereby making it easier to detect the discontinuity. Hence the number of simulations with "true" ZACs detected is more important. The average area of the detected ZACs (not reported here) are also larger (20% larger, on average) when the range parameter increases. The number (and the area) of the "other" ZACs detected is usually small, in accordance with the fact that these "other" ZACs are related to type I errors. When the discontinuity is high ($a = 3$), the number of "other significant" ZACs increases, in particular for the shorter range ($b = 0.07$). When looking closely at these simulations, we found that for short range the excursion sets are less smooth, and often break down in several connected components. Among them, some

have their centers outside the limits used for defining "true" ZACs, but very close to it, thereby contributing to increasing the number of simulations with "other" ZACs. This Table shows that the detection of significant ZACs is not too sensitive to a misspecification of the covariance function. Overestimation of the range parameter is a more acceptable error than underestimation in terms of detection: it detects more true ZACs without increasing the false positives away from the true discontinuity. Note also that misspecifying the covariance parametric model (Sph(0.3) instead of Exp(0.1)) is of secondary importance if the range parameter is correctly estimated.

5   Example : A soil Data Set

Precision agriculture aims at defining a location dependent management within an agricultural field, for nitrogen fertilizer for example, in order to better manage production in quantity and quality and minimize nitrogen losses towards ground water. This implies to consider the soil characteristics and their spatial variability and to define homogeneous management zones.

The data considered here were sampled in August 2001 in an agricultural field (10 ha) in Chambry, Northern France, on a pseudo-regular grid with 77 nodes (distance between nodes = 36 m), see Figure 3a. The analysis is performed on soil water content (WC), one of the most important parameters in plant growth models used for managing crops. It was measured on soil cores up to 120 cm. Each soil sample is itself a mixture of three cores taken at 50 cm distance. Similar data, collected in a nearby field are presented and analyzed in Mary *et al.* (2001). WC ranges from 130 mm to 448 mm; the average is 275.6 mm and the standard deviation is 54.6 mm. The median (262 mm) is very close to the average, the first and third quartiles are $Q_1 = 240$ mm and $Q_3 = 318$ mm respectively. The histogram is depicted in Figure 3b. In addition to these data, a map of the four main soil types drawn by soil scientists using independent data was provided (Figure 3c). The North is mainly silt with altered chalk in the subsoil (black). In the South is mainly found chalk and cryoturbed chalk (light and dark grey). In the middle, which is also the summit of the field is sand (white).

An exponential covariance function was estimated with the following parameters: (range, variance) = (30 m, 3176 mm^2). A 62×98 grid with a mesh size of 5 m is used, and simulations

22

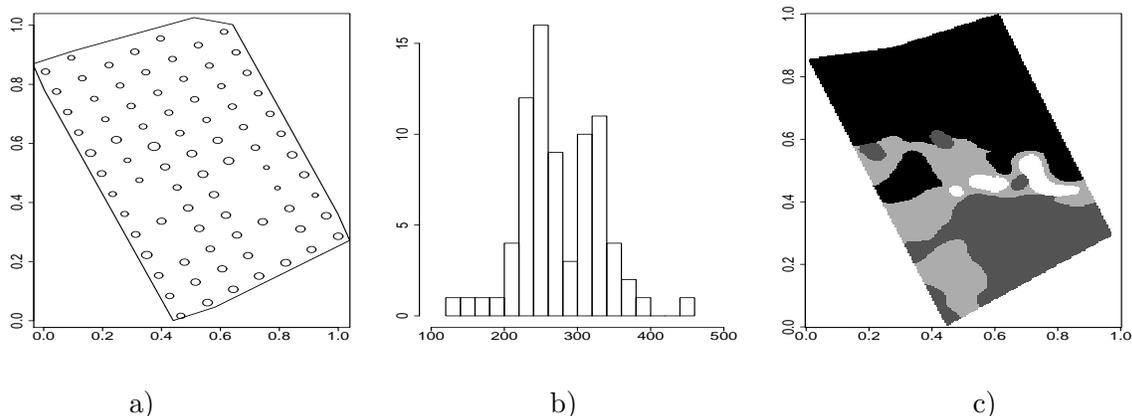a)   b)   c)

Figure 3: a) Proportional view of the location of the samples; b) Histogram of the samples; c) map showing the 4 soil types

conditional on the sample locations with the above parameters have shown that 95% of the realizations have no significant ZACs for $1 - \hat{\alpha} = 0.9984$. Convergence is reached in three iterations only, and the null hypothesis is rejected. The final variogram parameters are (range, variance) = (36 m, 2550 mm^2). Note that the variance parameter has been decreased by about 20%. Figures 4a and 4b show the estimated variograms at the first and at the last iteration. The corresponding ZACs, depicted in Figure 4c and 4d, are easily related to the soil type map (Figure 3c): they overlap the transition area in the middle of the field. The summit zone with the sand in the Eastern part of the field is clearly delineated. The transition between chalk and silt in the Western part is partly detected. In the central part, no ZAC is estimated, as this area corresponds to a smooth transition. The sensitivity to the range parameter was explored by performing the same analysis with smaller and larger range parameters: 25 m and 40 m respectively; convergence is reached to the same set of ZACs.

Seven other dates were available for analysis, from march 2000 to February 2003. On all of them but one, ZACs were detected at the same locations, thereby confirming this first analysis. For October 2000, no ZACs were detected. The reason was that at this date, no sample were taken precisely at the location of the ZAC. Another important variable of interest for agronomists is the mineral nitrogen. Interestingly no ZAC could be detected for any date for this variable,
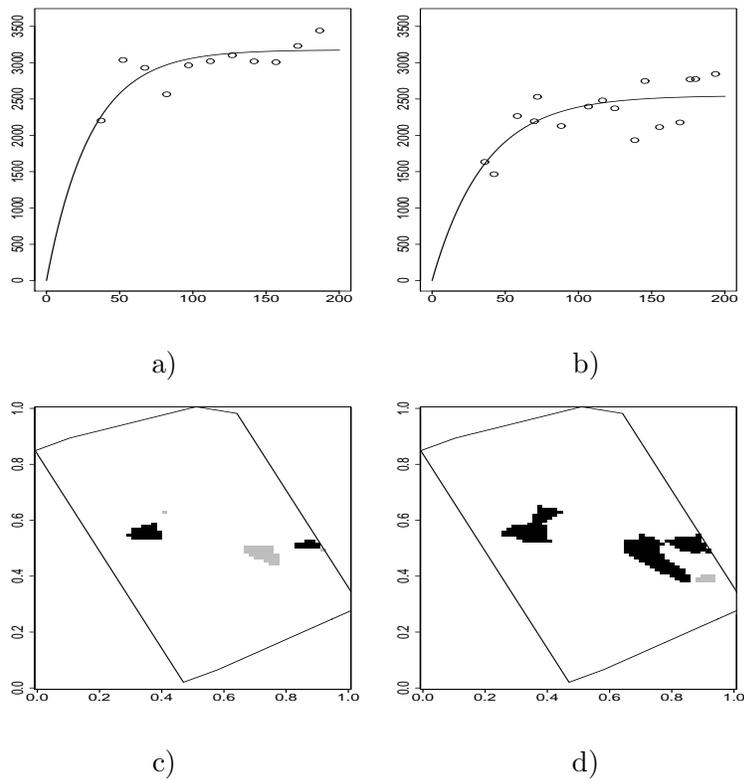
23

Figure 4: Estimating the ZACs on the soil data set. a) Experimental variogram (dots) and fitted model (solid line curve) at the first iteration; b) same at the last iteration; c) ZACs at the first iteration; d) ZACs at the last iteration.

which according to agronomists was to be expected for physical reasons. This analysis indicates that considering the water content only, the main transitions of the soil map could be retrieved. The field can thus be divided into two management zones corresponding to the top and the bottom of the field. A third zone, corresponding to the summit of the field could also be defined, but it would be much smaller and would probably not be interesting under a management point of view.

6 Extensions, comparison and discussion

6.1 Extensions

Several assumptions were made for establishing our method: in particular, we assumed known expectation, second order stationarity and regularity of the covariance function. We now discuss how some of these assumptions can be weakened.

6.1.1 Unknown expectation

We assumed second order stationarity with known expectation μ and thus simple kriging was used throughout. In general, the expectation is unknown and the spatial optimal predictor, called ordinary kriging (Cressie, 1993; Chilès and Delfiner, 1999), is :

$$\tilde{Z}(x) = C(x)'\mathbf{C}^{-1}Z + \frac{1 - C(x)'\mathbf{C}^{-1}\mathbf{1}}{\mathbf{1}'\mathbf{C}^{-1}\mathbf{1}}\mathbf{1}'\mathbf{C}^{-1}Z, \tag{13}$$

where $\mathbf{1} = (1, \ldots, 1)'$ is a vector of ones of length n. According to Proposition 1, the gradient $\tilde{W}(x)$ of $\tilde{Z}(x)$ exists and $\tilde{W}(x) = \partial C'(x)\mathbf{K}^{-1}Z$, with $\mathbf{K}^{-1} = (\mathbf{C}^{-1} - \mathbf{C}^{-1}\mathbf{1}\mathbf{1}'\mathbf{C}^{-1})/\mathbf{1}'\mathbf{C}^{-1}\mathbf{1}$. The associated covariance matrix is $\tilde{\boldsymbol{\Sigma}}(x) = \partial C'(x)\mathbf{K}^{-1}\mathbf{C}\mathbf{K}^{-1}\partial C(x)$, which yields to $\tilde{\boldsymbol{\Sigma}}(x) = \partial C'(x)\mathbf{K}^{-1}\partial C(x)$, because $\mathbf{K}^{-1}\mathbf{C}\mathbf{K}^{-1} = \mathbf{K}^{-1}$. The field $\tilde{T}(x) = \tilde{W}'(x)\tilde{\boldsymbol{\Sigma}}(x)^{-1}\tilde{W}(x)$ is then a $\chi^2(2)$ field on the set of points x where $\tilde{\boldsymbol{\Sigma}}(x)^{-1}$ exists. The case of an unknown expectation is thus formally identical to the case seen in Section 2, $\mathbf{C}^{-1}$ being replaced by $\mathbf{K}^{-1}$ when computing $\tilde{W}$, $\tilde{U}$, $\tilde{\boldsymbol{\Sigma}}$ and $\tilde{\boldsymbol{\Lambda}}$ and $\tilde{T}$. All propositions and theorems remain valid. In practice, differences between (1) and (13) are small when the data are centered and when the distance to the data is small compared to the range of the covariance function. Some differences were visible when simulating non centered data and predicting outside the sampled domain or near its border.

6.1.2 More general covariances

The two assumptions $\mathcal{A}_1$ and $\mathcal{A}_2$ can be weakened. Instead of assumption $\mathcal{A}_1$, one can assume $\mathcal{A}_1'$: $Z(x)$ has stationary first order differences, i.e. $E[Z(x) - Z(y)] = 0$ and $\text{Var}(Z(x) - Z(y)) = 2\gamma(x - y)$ for all $x, y \in \mathcal{D}$. This weaker assumption is called intrinsic stationarity in the

geostatistical literature and $\gamma(x - y)$ is the variogram of $Z(x)$ introduced in (12). In this case it can be shown that the optimal predictor is again the ordinary kriging (13), in which the vector $C(x)$ whose elements are $c(x - x_i)$ and the matrix $\mathbf{C}$ whose elements are $c(x_i - x_j)$ are replaced by a variogram vector $G(x)$ and a variogram matrix $\mathbf{G}$ whose elements are $\gamma(x - x_i)$ and $\gamma(x_i - x_j)$ respectively. Hence, under assumption $\mathcal{A}'_1$ it is possible to use our method for detecting ZACs using ordinary kriging. Assumption $\mathcal{A}_2$ is about the regularity of the covariance function away from 0. For sake of simplicity we assumed that it was infinitely differentiable for all h with $||h|| > 0$. As already stated, this condition is mild and is verified for most used covariance functions. In fact only third order differentiability of the covariance function is necessary. The spherical covariance function does not verify this less restrictive assumption (derivatives are not continuous for $||h|| = b$), and thus in theory should not be used. However, the set of points where differentiability is not verified is a set of null measure in the plane. Practically speaking, this condition is never a serious limitation.

6.1.3   Sampling design

No spatial regularity of the sampling design is required but the local sampling density has a direct consequence on the power of the method. Clearly, for sparse sampling, we will not be able to detect abrupt changes. The lack of samples makes it simply impossible to test whether the estimated local gradient corresponds to a smooth transition or if it is locally sharp. This leads us to the question of the power of the test. The global power has been explored in the simulation study reported Section 3.2. If the alternative hypothesis is specified, a local power (i.e., the probability $1 - \beta(x)$ to detect on x a discontinuity curve Γ that passes through x) can be approximated. This work is still in progress and is not reported here.

6.2   Comparison and discussion

We now compare our method with the one described in Banerjee et al. (2003) (hereafter denoted BGS) on their very same example: a sample ($n = 100$) of a stationary Gaussian random field with covariance function $c(h) = \sigma^2(1 + \phi||h||) \exp(-\phi||h||)$, with $\sigma^2 = 1.0$ and $\phi = 1.05$, see

26

Section 5 in BGS. BGS compute the posterior distribution of the gradient at two locations and find that the 95% credibility interval of the N-S gradient at the point (3.5, 3.5) does not contain 0. This is used as an indication of a significant N-S slope at this point, while the same analysis on the E-W gradient showed no significant slope. Using the same code (available at www.biostat.umn.edu/~sudiptob/Software/Software.html), we computed the posterior distribution of the gradient at all grid nodes of a 30×30 grid. This computation took 4.5 days on a recent PC running R under a Linux operating system. Figure 5a depicts all pixels on which either the N-S or the E-W gradient 95% credibility interval does not contain 0. This procedure finds the pixels with a gradient in the 5% tail in these directions, but does not assess that these gradients are due to an abrupt change violating the assumption of stationarity. Because the gradients in the two directions are independent (see BGS, Section 3), the real level pictured in Figure 5a is $0.95 \times 0.95 = 0.9025$.

We estimated the parameters of the covariance function as described in Section 3.3 and obtained $(\hat{\sigma}^2, \hat{\phi}) = (0.87, 0.99)$. Figure 5b shows the pixels with a value T above the quantile $t_{0.9025} = 4.66$. The regions highlighted in Figures 5a and 5b are quite similar, at the exception of a large area in the upper part of the domain and a smaller one in the middle of the image. In these areas, the gradient is oriented in the NE-SW direction and is thus not visible on Figure 5a. Our test statistics T being rotation invariant, it is visible on Figure 5b. As discussed in Section 3, one cannot assess the significance of the potential ZACs at the level $1 - \alpha = 0.9025$ because it is too low for the convergence in law (8) to hold. For the estimated parameters above, we found $1 - \hat{\alpha} = 0.9983$. At this level only one pixel had a value T above the quantile $t_{1-\hat{\alpha}} = 12.75$, whose coordinate is (3.5, 3.8), in the same region of higher gradient than the test point used in BGS. The corresponding p-value computed according to (10) is 0.753 and H_0 is not rejected. This analysis required 30 minutes, including the estimation of $\hat{\alpha}$ with 100 realizations. We tested the sensitivity of these results with respect to the estimated parameters and found consistently no significant ZACs. The same analysis was repeated on a 60×60 grid and gave exactly the same results. Our analysis shows clearly that the null hypothesis of no abrupt change must not be rejected.

27

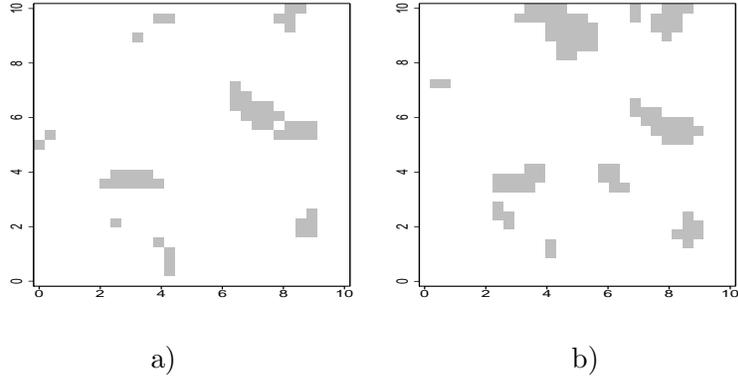a)   b)

Figure 5: Sets of pixels of a 30×30 grid with "high" gradient, i.e. where: a) the 95% credibility interval of the N-S or the E-W gradient does not contain 0, according to BGS; b) the statistic T is above the level $t_{0.9025} = 4.66$, according to our method.

This comparison highlights the major difference between the two methods. BGS provide an attractive approach for estimating locally the posterior distribution of the gradient of the process and thus for reconstructing its surface. However, they do not provide a theory for testing the significance of the gradients at all locations simultaneously. For example, the N-S gradient at the point $(3.5, 3.5)$ is only locally significant, i.e. this point belongs to a zone where the gradient is the highest in the domain. But it should not be globally significant since it originated from a stationary random field: in any realization of a stationary random field, there almost always exist some pixels where the local slope is significantly different from 0 at the level $\alpha = 0.05$ as soon as the domain is larger than the correlation distance. Our method makes a clear distinction between local "high gradient" and a global approach for testing H_0 versus H_1, while the approach in BGS is only local. At the local level, the potential ZACs are the zones of "high gradient". The test statistic used for defining these zones is rotation invariant and is thus easier to use than the directional derivatives used in BGS. The theory exposed in Section 2 makes the link between the local level α and the global level η. It provides a practical way to perform a global test for rejecting H_0 and for detecting ZACs, i.e. zones where the assumption of stationarity of the expectation field is violated.

We now proceed with some discussion about the differences between our method and two

28

other recent methods that can be found in the literature. Compared to the non parametric approach developed in Hall *et al.* (2001), our method finds all the ZACs at once and does not need any starting point. More importantly, our method does not make the implicit assumption that there is a discontinuity in the domain but provides at the same time a test for rejecting the null hypothesis of absence of ZACs and an estimate of these ZACs in case of rejection. We are also able to assign a p-value to each ZAC, thereby discriminating between significant locally high gradient zones and non significant ones. It detects (i.e. estimates *and* tests) a more general class of discontinuity curves, but at the cost of a more particular model of random field. On the one hand, it is more particular in the sense that it relies on a Gaussian assumption for the data. But, if the data cannot be considered as Gaussian, they can be transformed on a Gaussian scale, using a parametric form (e.g. log transform, Cox transform) or a non parametric quantile transform. This, of course, does not ensure the gaussianity of the random field but should be enough for the method to work reasonably well in practice, as illustrated on the agricultural example in Section 5. On the other hand, however, this assumption makes it possible to explicitly take into account the autocorrelation between the samples and to detect discontinuities for a low density of samples in presence of autocorrelation.

In Chaudhuri and Marron (1999, 2000) and Godtliebsen et al. (2002), significance of the features is assessed at all scales, i.e. for all bandwidths. The assumption is that the data $Z(x_i)$ stem from an underlying signal f and independent Gaussian noise : $Z(x_i) = f(x_i) + \epsilon_i$. Under the null hypothesis of a constant signal, the data set is an i.i.d. sample of a random variable. For the applications we have in mind, the assumption of independent noise is unrealistic. In our method, we suppose that it is a stationary random field, whose covariance function is to be estimated. Having estimated the covariance parameters from the data, it is not necessary to detect significant features at all scale. Hence our method requires less computation and is easier to interpret visually. Furthermore, based on a simulation study, we showed that our method is robust with respect to change in the range parameter.

29

Appendix A: $\mathbf{\Lambda}_2$

We give the full expression for the matrix $\mathbf{\Lambda}_2$ used in (8).

$$
\begin{aligned}
\mathbf{\Lambda}_{2\,[kl]} = {} & \{ \partial_k D_2' \mathbf{C}^{-1} \partial_l D_2 / \sigma_2^2 - \partial_k \sigma_2 \partial_l \sigma_2 / \sigma_2^2 + \partial_k \rho \partial_l \rho - \partial_k \rho \partial_l D_2' \mathbf{C}^{-1} D_1 / \sigma_1 \sigma_2 \\
& - \partial_l \rho \partial_k D_2' \mathbf{C}^{-1} D_1 / \sigma_1 \sigma_2 + \rho[(\partial_k \sigma_2 / \sigma_2)(\partial_l D_1' \mathbf{C}^{-1} D_2 / \sigma_1 \sigma_2 + \partial_l \rho) \\
& + (\partial_l \sigma_2 / \sigma_2)(\partial_k D_1' \mathbf{C}^{-1} D_2 / \sigma_1 \sigma_2 + \partial_k \rho) + (\partial_l \sigma_1 / \sigma_1)(\partial_k D_2' \mathbf{C}^{-1} D_1 / \sigma_1 \sigma_2) \\
& + (\partial_k \sigma_1 / \sigma_1)(\partial_l D_2' \mathbf{C}^{-1} D_1 / \sigma_1 \sigma_2) - \partial_k D_1' \mathbf{C}^{-1} \partial_l D_2 / \sigma_1 \sigma_2 - \partial_k D_2' \mathbf{C}^{-1} \partial_l D_1 / \sigma_1 \sigma_2] \\
& + \rho^2 [\partial_k D_1' \mathbf{C}^{-1} \partial_l D_1 / \sigma_1^1 - \partial_k \sigma_1 \partial_l \sigma_1 / \sigma_1^1 - \partial_k \sigma_1 \partial_l \sigma_2 / \sigma_1 \sigma_2 \\
& - \partial_k \sigma_2 \partial_l \sigma_1 / \sigma_1 \sigma_2] \} / (1 - \rho^2) - \rho^2 (\partial_k \rho \partial_l \rho) / (1 - \rho^2)^2.
\end{aligned}
$$

The proof is direct but quite long.

Appendix B: Proof of Theorem 1

In this appendix, we prove Theorem 1. We first present a classical lemma, the proof of which can be found in Miller (1964, p.71).

Lemma 1 *Let $Y = (Y_1, \ldots, Y_4)'$ be a vector of four non centered Gaussian random variables with mean $\mu = (\mu_1, \ldots, \mu_4)'$ and covariance Γ_{ij}, $1 \le i,j, \le 4$. Then, $E[Y_1 Y_2 Y_3] = \mu_1 \mu_2 \mu_3 + \sum_i \mu_i \Gamma_{jk}$ and $E[Y_1 Y_2 Y_3 Y_4] = \mu_1 \mu_2 \mu_3 \mu_4 + \sum_{j \ne i} \mu_i \mu_j \Gamma_{kl} + \sum_{j>i} \Gamma_{ij} \Gamma_{kl}$, where $j \ne i$, $k \ne i,j$ and $l \ne i,j,k$.*

Lemma 2 *Consider $U(x) = (U_1(x), U_2(x))'$ defined in Section 2.4. Then, for $1 \le i \le 2$,*
a) $U_i(x)$ and $\dot{U}_i(x)$ are independent
b) $Cov(\ddot{\mathbf{U}}_{\mathbf{i}}(x), \dot{U}_i(x)\dot{U}_i(x)' \mid U(x)) = \mathbf{0}$,
c) Let us denote $\partial^{(m)} U_i$ the vector of all partial derivatives of order $m \ge 1$ of U_i, $i = 1, 2$. By convention, let $\partial^{(0)} U_i = U_i$. Then, $\partial^{(m)} U_i$ and $\partial^{(n)} U_j$ are independent when $i \ne j$, for $m, n \ge 0$.
d) $\ddot{\mathbf{U}}_i(x)_{|U(x)} \stackrel{\mathcal{L}}{\sim} \mathcal{N}_4 \left(-U_i(x)\mathbf{\Lambda}_i(x), \mathbf{M}_i(\mathbf{\Lambda}_i(x)) \right)$, where $\mathbf{M}_i(\mathbf{\Lambda}_i(x))_{[kl]} = Cov(\partial_{kl}^2 U_i(x), \partial_{k'l'}^2 U_i(x) \mid U_i(x))$ is the conditional covariance matrix of the second order derivatives of $U_i(x)$ and does not depend on $U_i(x)$,

30

e) $\partial^2_{kl}U_i(x)_{|U(x),\dot{U}(x)} \sim \mathcal{N}\left(E_{\partial^2 U_i(x)}, V_{\partial^2 U_i(x)}\right)$, with $E_{\partial^2 U_i(x)} = -U_i\mathbf{\Lambda}_i(x)_{[kl]} + E[\partial^2_{kl}U_i(x)\dot{U}_i(x)]\mathbf{\Lambda}^{-1}_i\dot{U}_i(x)$
and $V_{\partial^2 U_i(x)} = \mathbf{M}_i(\mathbf{\Lambda}_i(x))_{[kl]} + E[\partial^2_{kl}U_i(x)\dot{U}_i(x)]\mathbf{\Lambda}^{-1}_i(x)E[\partial^2_{kl}U_i(x)\dot{U}_i(x)]'$.

Before presenting the proof, it is worth noting that $\dot{U}_i$ and $\ddot{\mathbf{U}}_i$ are not independent. We only have the weaker result b).

Proof : a) This is a direct consequence of the fact that the variance of $U_i(x)$ is constant. Indeed, since $E[U_i(x)^2] = 1$, we have that $\partial_k E[U_i(x)^2] = 0$, hence $2E[U_i(x)\partial_k U_i(x)] = 0$. Since they are Gaussian, absence of correlation is equivalent to independence. For the sake of lighter notations, we now drop the dependence to x.

b) Using Lemma 1 and a),

$$
\begin{aligned}
E[\partial_k U_i\partial_l U_i\partial^2_{k'l'}U_i \mid U] &= E[\partial_k U_i \mid U]\,E[\partial_l U_i\partial^2_{k'l'}U_i \mid U] + E[\partial_l U_i \mid U]\,E[\partial_k U_i\partial^2_{k'l'}U_i \mid U] \\
&\quad + E[\partial^2_{k'l'}U_i \mid U]\,E[\partial_k U_i\partial_l U_i \mid U] \\
&= E[\partial^2_{k'l'}U_i \mid U]\,E[\partial_k U_i\partial_l U_i \mid U],
\end{aligned}
$$

since $E[\partial_k U_i \mid U] = 0$. Hence, $\mathrm{Cov}(\partial^2_{k'l'}U_i, \partial_k U_i\partial_l U_i \mid U_i) = 0$.

c) This result is easily proven using recursively $E[\partial^{(m)}U_i \mid \partial^{(n)}U_j] = \partial E[\partial^{(m-1)}U_i \mid \partial^{(n)}U_j]$ and $E[\partial^{(m-1)}U_i \mid \partial^{(n)}U_j] = E[\partial^{(m-1)}U_i\partial^{(n)}U_j]\mathrm{Var}^{-1}(\partial^{(n)}U_j)\partial^{(n)}U_j$. The first result holds because of the regularity properties of the fields U_i; the second result is the usual regression formula for Gaussian vectors.

d) Using part c) of this lemma, $\ddot{\mathbf{U}}_i$ is independent of U_j when $i \neq j$, $i,j = 1,2$. Hence, conditioning $\ddot{\mathbf{U}}_i$ on U is equal to conditioning $\ddot{\mathbf{U}}_i$ on U_i. We first prove that $E[\ddot{\mathbf{U}}_\mathbf{i} \mid U_i] = -U_i\mathbf{\Lambda}_i$. We start from proposition a). Then, $\partial_l E[U_i\partial_k U_i] = 0$, i.e

$$E[\partial_k U_i\partial_l U_i] = -E[U_i\partial^2_{kl}U_i]. \tag{14}$$

On the other hand, the classical regression of $\partial^2_{kl}U_i$ given U_i is:

$$E[\partial^2_{kl}U_i \mid U_i] = E[\partial^2_{kl}U_i] + \mathrm{Cov}(U_i, \partial^2_{kl}U_i)(U_i - E[U_i])/\mathrm{Var}(U_i) = E[U_i\partial^2_{kl}U_i]U_i \tag{15}$$

since $E[U_i] = 0$, $E[\partial^2_{kl}U_i] = 0$ and $\mathrm{Var}(U_i) = 1$. Putting together (14) and (15) leads to the desired result. We now turn to the conditional covariance

$$\mathrm{Cov}(\partial^2_{kl}U_i, \partial^2_{k'l'}U_i \mid U_i) = E[\partial^2_{kl}U_i\partial^2_{k'l'}U_i \mid U_i] - E[\partial^2_{kl}U_i \mid U_i]E[\partial^2_{k'l'}U_i \mid U_i]$$

31

$$= E[\partial_{kl}^2 U_i \partial_{k'l'}^2 U_i \mid U_i] - U_i^2 \Lambda_{i\,[kl]} \Lambda_{i\,[k'l']}.$$

According to (15),

$$\partial_{kl}^2 U_i \overset{\mathcal{L}}{=} \mathrm{Cov}(\partial_{kl}^2 U_i, U_i) U_i + \sqrt{\mathrm{Var}(\partial_{kl}^2 U_i) - \mathrm{Cov}(\partial_{kl}^2 U_i, U_i)^2}\, \varepsilon_{kl}$$

$$\overset{\mathcal{L}}{=} -\Lambda_{i\,[kl]} U_i + \sqrt{\mathrm{Var}(\partial_{kl}^2 U_i) - \Lambda_{i\,[kl]}^2}\, \varepsilon_{kl},$$

where the equalities are equalities in law and ε_{ij} is a residual $\mathcal{N}(0,1)$ random variable. Hence,

$$\mathrm{Cov}(\partial_{kl}^2 U_i, \partial_{k'l'}^2 U_i \mid U_i) = \sqrt{\mathrm{Var}(\partial_{kl}^2 U_i) - \Lambda_{i\,[kl]}^2}\sqrt{\mathrm{Var}(\partial_{k'l'}^2 U_i) - \Lambda_{i\,[k'l']}^2}\, E[\varepsilon_{kl}\varepsilon_{k'l'}]. \qquad (16)$$

e) According to part c) of this lemma, $\ddot{\mathbf{U}}_i$ is independent of $(U_j, \dot{U}_j)$ for $i \neq j$ and $i, j = 1, 2$. Hence, conditioning $\ddot{\mathbf{U}}_i$ on $(U, \dot{U})$ is equal to conditioning $\ddot{\mathbf{U}}_i$ on $V_i = (U_i, \dot{U}_i)'$. Thus we have

$$E[\partial_{kl}^2 U_i \mid V_i] = E[\partial_{kl}^2 U_i] + \mathrm{Cov}\left(\partial_{kl}^2 U_i, V_i\right) \mathrm{Var}^{-1}(V_i) V_i,$$

with $\mathrm{Cov}\left(\partial_{kl}^2 U_i, V_i\right) = \left(E[\partial_{kl}^2 U_i U_i] \quad E[\partial_{kl}^2 U_i \dot{U}_i]\right)$ and $\mathrm{Var}^{-1}(V_i) = \begin{pmatrix} 1 & 0 \\ 0 & \mathbf{\Lambda}_i^{-1} \end{pmatrix}$. Hence,

$$E[\partial_{kl}^2 U_i \mid V_i] = -\mathbf{\Lambda}_{i[kl]} U_i + E[\partial_{kl}^2 U_i \dot{U}_i] \mathbf{\Lambda}_i^{-1} \dot{U}_i. \qquad (17)$$

Regarding the conditional variance, we have, using the usual conditional variance formula for Gaussian vectors $\mathrm{Var}\left(\partial_{kl}^2 U_i \mid V_i\right) = \mathrm{Var}\left(\partial_{kl}^2 U_i\right) - \mathrm{Cov}\left(\partial_{kl}^2 U_i, V_i\right) \mathrm{Var}^{-1}(V_i) \mathrm{Cov}\left(V_i, \partial_{kl}^2 U_i\right)$, where

$$\mathrm{Var}\left(\partial_{kl}^2 U_i\right) = \mathrm{Var}\left(\partial_{kl}^2 U_i \mid U_i\right) + \mathrm{Cov}\left(\partial_{kl}^2 U_i, U_i\right) \mathrm{Var}^{-1}(U_i) \mathrm{Cov}\left(U_i, \partial_{kl}^2 U_i\right) = \mathbf{M}_i(\mathbf{\Lambda}_i)_{[kl]} + \mathbf{\Lambda}_{i\,[kl]}^2,$$

according to part d) of this lemma. The second term is equal to

$$\left(E[\partial_{kl}^2 U_i U_i] \quad E[\partial_{kl}^2 U_i \dot{U}_i]\right) \begin{pmatrix} 1 & 0 \\ 0 & \mathbf{\Lambda}_i^{-1} \end{pmatrix} \left(\begin{array}{c} E[\partial_{kl}^2 U_i U_i] E[\partial_{kl}^2 U_i \dot{U}_i] \end{array}\right) = \mathbf{\Lambda}_{i\,[kl]}^2 + E[\partial_{kl}^2 U_i \dot{U}_i] \mathbf{\Lambda}_i^{-1} E[\partial_{kl}^2 U_i \dot{U}_i]'.$$

Hence, $\mathrm{Var}\left(\partial_{kl}^2 U_i \mid V_i\right) = \mathbf{M}_i(\mathbf{\Lambda}_i)_{[kl]} + E[\partial_{kl}^2 U_i \dot{U}_i] \mathbf{\Lambda}_i^{-1}{}_{[kl]} E[\partial_{kl}^2 U_i \dot{U}_i]'.$ \hfill $\square$

Lemma 3 For $T(x) = U_1^2(x) + U_2^2(x)$ defined in Section 2.4, conditional on $(U(x), \dot{U}(x))$,

$$\partial_{kl}^2 T(x) \sim \mathcal{N}\left(E_{\ddot{\mathbf{T}}_{kl}}, V_{\ddot{\mathbf{T}}_{kl}}\right),$$

32

with $E_{\ddot{\mathbf{T}}_{kl}} = 2\sum_{i=1}^{2}\left(\partial_k U_i(x)\partial_l U_i(x) - U_i^2(x)\mathbf{\Lambda}_i(x)_{[kl]} + U_i(x)E[\partial_{kl}^2 U_i(x)\dot{U}_i(x)]\mathbf{\Lambda}_i^{-1}(x)\dot{U}_i(x)\right)$, and
$V_{\ddot{\mathbf{T}}_{kl}} = 4\sum_{i=1}^{2} U_i^2(x)\left(\mathbf{M}_i(\mathbf{\Lambda}_i(x))_{[kl]} + E[\partial_{kl}^2 U_i(x)\dot{U}_i(x)]\mathbf{\Lambda}_i^{-1}(x)E[\partial_{kl}^2 U_i(x)\dot{U}_i(x)]'\right)$, where $\mathbf{M}_i(\mathbf{\Lambda}_i(x))$
is defined in lemma 2.

Proof : We start from $\ddot{\mathbf{T}} = 2(\dot{U}_1\dot{U}_1' + \dot{U}_2\dot{U}_2' + U_1\ddot{U}_1 + U_2\ddot{U}_2)$. The conditional expectation
$E[\ddot{\mathbf{T}} \mid U]$ is easily found by a direct application of (17). For the conditional variance, we have

$$
\begin{aligned}
\mathrm{Var}\left(\partial_{kl}^2 T \mid V\right) &= 4\mathrm{Var}\left(\partial_k U_1\partial_l U_1 + \partial_k U_2\partial_l U_2 + U_1\partial_{kl}^2 U_1 + U_2\partial_{kl}^2 U_2 \mid V\right) \\
&= 4\mathrm{Var}\left(U_1\partial_{kl}^2 U_1 + U_2\partial_{kl}^2 U_2 \mid V\right) \\
&= 4U_1^2\mathrm{Var}\left(\partial_{kl}^2 U_1 \mid V\right) + 4U_2^2\mathrm{Var}\left(\partial_{kl}^2 U_2 \mid V\right) + 8U_1 U_2\mathrm{Cov}\left(\partial_{kl}^2 U_1, \partial_{kl}^2 U_2 \mid V\right).
\end{aligned}
$$

The covariance term is equal to 0, according to Lemma 2c. Finally,

$$
\mathrm{Var}\left(\partial_{kl}^2 T \mid U, \dot{U}\right) = 4\sum_{i=1}^{2}\left\{U_i^2\left(\mathbf{M}_i(\mathbf{\Lambda}_{i,[kl]}) + E[\partial_{kl}^2 U_i\dot{U}_i]\mathbf{\Lambda}_i^{-1}{}_{[kl]}E[\partial_{kl}^2 U_i\dot{U}_i]'\right)\right\}. \quad \square
$$

3   Proof of Theorem 1

We want to prove that $T_0^{-1}\ddot{\mathbf{T}}_{T_0}(0) + 2\{v(0)\mathbf{\Lambda}_1(0) + (1-v(0))\mathbf{\Lambda}_2(0)\} = o_p(1)$, as $T_0 \to \infty$, where
T_{T_0} is the field $T(\cdot)$, conditional on the event that there is an extremum of height T_0 at 0. It
is shown by using a multivariate version of Chebychev inequality that we first recall : let X
be a random vector with mean m_X and covariance matrix $\mathbf{C}_X$. Then, $P(\|X - m_X\| \geq a) \leq$
$\mathrm{tr}(\mathbf{C}_X)/a^2$. According to lemma 3,

$$
\begin{aligned}
E[\partial_{kl}^2 T(0) \mid U, \dot{U}] &= 2\sum_{i=1}^{2}\left(-U_i^2(0)\mathbf{\Lambda}_i(0)_{[kl]} + U_i(0)E[\partial_{kl}^2 U_i(0)\dot{U}_i(0)]\mathbf{\Lambda}_i^{-1}(0)\dot{U}_i(0)\right), \\
&= 2\sum_{i=1}^{2}-U_i^2(0)\mathbf{\Lambda}_i(0)_{[kl]} + O\left(\sqrt{T_0}\right),
\end{aligned}
$$

and $\mathrm{Var}\left(\partial_{kl}^2 T(0) \mid U, \dot{U}\right) = O(T_0)$. Hence, $\mathrm{tr}(\mathrm{Var}(\ddot{\mathbf{T}}(\mathbf{0}) \mid U)) = O(T_0)$ as $T_0 \to \infty$, and the
application of the multivariate Chebychev inequality to $T_0^{-1}\ddot{\mathbf{T}}_{|U,\dot{U}}$ leads to the result. $\quad \square$

References

Alder, R. (1981) *The Geometry of Random Fields.* New-York: Wiley.

33

Adler, R. (2000) On excursion sets, tube formulas and maxima of random fields. *The Annals of Applied Probability*, **10**, 1-74.

Aronowich, M. and Adler, R. (1988) Sample path behaviour of χ^2 surfaces at extrema. *Advances in Applied Probabilities*, **18**, 901-920.

Barbujani, G., Oden , N.L. and Sokal, R.R. (1989) Detecting areas of abrupt change in maps of biological variables. *Systematic Zoology*, **38**, 376-389.

Banerjee, S., Gelfand, A.E. and Sirmans, C.F. (2003) Directional rates of change under spatial process models. *Journal of the American Statistical Association*, **98**, 946–954.

Besag, J. (1986) On the statistical analysis of dirty pictures (with discussion). *Journal of the Royal Statistical Society*, Ser. B. **48**, 259–302.

Besag, J., Green, P., Higdon, D. and Mengersen, K. (1995) Bayesian computation and stochastic systems (with discussions). *Statistical Science*, **10**, 3-66.

Cao, J. (1999) The size of the connected components of excursion sets of χ^2, t and F fields. *Advances in Applied Probabilities (SGSA)*, **31**, 579-595.

Ciba Foundation Symposium 210 (1997) *Precision Agriculture: Spatial and temporal variability of environmental quality.* Chichester: Wiley.

Chaudhuri, P. and Marron, J. S. (1999) SiZer for exploration of structure in curves. *Journal of the American Statistical Association*, **94**, 807-823.

Chaudhuri, P. and Marron, J. S. (2000) Scale space view of curve estimation. *Annals of Statistics*, **28**, 408-428.

Chilès, J.-P. and Delfiner, P. (1999) *Geostatistics: modeling spatial uncertainty.* New-York: Wiley.

Cressie, N. (1993) *Statistics for spatial data, Revised Edition.* New-York: Wiley.

Geman, S. and Geman, D. (1984) Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE Trans. Pattern Analysis and Machine Intelligence 6*, 721–741.

Godtliebsen, F. Marron, J.S. and Pizer, S.M. (2002) Significance in Scale-Space for Clustering. In Lawson A.B. and Denison D. (Eds.) *Spatial Cluster Modelling* . Boca Raton: Chapman and Hall/CRC, 24–36.

Hall, P., Peng, L. and Rau, C. (2001) Local likelihood tracking of fault lines and boundaries. *J. R. Statistic. Soc. B*, **63**, 569-582.

Mary, B., Beaudoin, N., Machet, J.M., Bruchou, C. and Ariès F. (2001) Characterization and Analysis of soil variability within two agricultural fields: the case of water and mineral N profiles. In, *Proc. 3rd European Conference on Precision Agriculture* (eds Grenier and Blackmore), pp 431-436.

Miller, K.S. (1964) *Multidimensional Gaussian distribution*, SIAM series in Applied Mathematics.

Qiu, P. (1998) Discontinuous regression surfaces fitting. *The Annals of Statistics*, **26**, 2218-2245.

Shen, X., Hsin-Cheng, H. and Cressie N. (2002) Nonparametric Hypothesis Testing for a Spatial Signal. *Journal of the American Statistical Association*, **97**, 1122-1140.

Stein, M. (1999) *Interpolation of spatial data: some theory for kriging*. Springer-Verlag.

Wang Y. (1998) Change Curve Estimation via Wavelets. *Journal of the American Statistical Association*, **93**, pp 163-172.

Womble, W.H. (1951) Differential systematics. *Science*, **114**, 315-322.

Worsley, K. (1994) Local maxima and the expected Euler characteristic of excursion sets of χ^2, F and t fields. *Adv. Appl. Prob.*, **26**, 13-42.

Worsley, K. (2001) Testing for signals with unknown location and scale in a χ^2 random field, with application to fMRI. *Adv. Appl. Prob. (SGSA)*, **33**, 773-793.

Yaglom, A. M. (1986) *Correlation theory of stationary and related random functions*. Springer-Verlag, Berlin.