Maxima of discretely sampled random fields, with an application to 'bubbles'

By J.E. Taylor
Department of Statistics, Stanford University, Stanford, CA 94305-4065, U.S.A.
jonathan.taylor@stanford.edu

K.J. Worsley
Department of Mathematics & Statistics, McGill University, Montréal, Québec, Canada H3A 2K6
keith.worsley@mcgill.ca

and F. Gosselin
Département de psychologie, Université de Montréal, Montréal, Québec, Canada H3C 3J7
frederic.gosselin@umontreal.ca

Abstract

A smooth Gaussian random field with zero mean and unit variance is sampled on a discrete lattice, and we are interested in the exceedence probability (P-value) of the maximum in a finite region. If the random field is smooth relative to the mesh size, then the P-value can be well approximated by results for the continuously sampled smooth random field (Adler, 1981; Worsley, 1995a; Taylor and Adler, 2003). If the random field is not smooth, so that adjacent lattice values are nearly independent, then the usual Bonferroni bound is very accurate. The purpose of this paper is to bridge the gap between the two, and derive a simple, accurate upper bound for mesh sizes in between. The result uses a new improved Bonferroni-type bound based on discrete local maxima (DLM). We give an application to the 'bubbles' technique for detecting areas of the face used to discriminate fear from happiness.

Some key words: Bonferroni; Random fields; Bubbles; Euler characteristic.

1 Introduction

Let $Z(x)$ be a D-dimensional stationary Gaussian zero mean, unit variance, random field sampled on a uniform rectilinear lattice (see Figure 1). We are interested in good approximations to

$$P = \mathbb{P}\left(\max_{x \in S} Z(x) > t\right) \tag{1}$$

where $S \subset \mathbb{R}^D$. Such a situation arises frequently in brain mapping, where $Z(x)$ is a 3D image of brain activity in response to a stimulus, sampled on a uniform lattice of voxels. The search

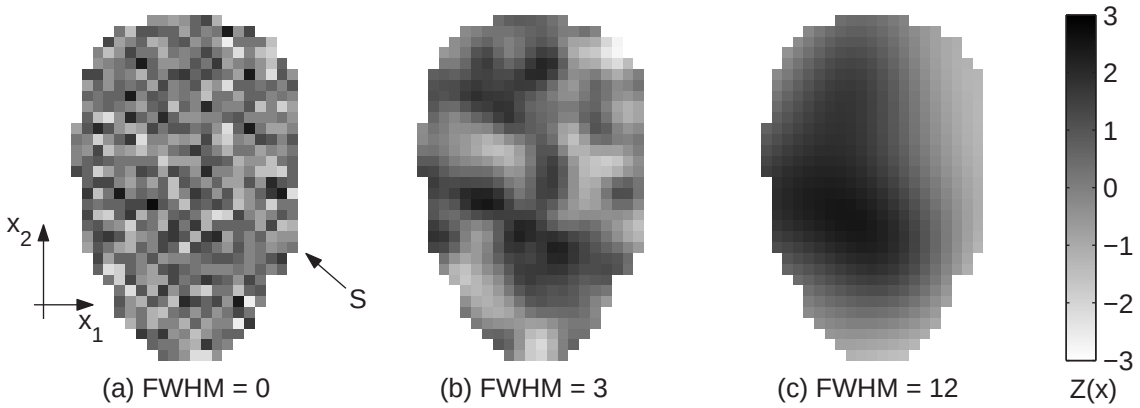

Figure 1: The problem. A smooth Gaussian random field $Z(x)$ is sampled on a rectilinear lattice, and we want to find a good approximation to the P-value of the maximum inside a search regions S. For unsmoothed data (a), Bonferroni is accurate; for very smooth data (c), the expected Euler characteristic method is very accurate; we seek a good approximation in between (b). The data shown is actually a blow-up of the central portion of Figure 5.

region S is usually the whole brain. We are interested in detecting those few isolated regions in S where $\mathbb{E}(Z(x)) > 0$, that is, those regions where brain activity has occurred. This is done by thresholding the image at a high value t chosen so that if the entire image were null, the exceedence probability is controlled to some small value such as 0.05. The threshold t is then determined by equating (1) to 0.05 and solving for t (see Worsley, 2003, and references therein). A 2D example is the 'bubbles' technique for revealing those areas of an image used in a visual discrimination task (Gosselin & Schyns, 2001; Schyns *et al.*, 2002; Chauvin *et al.*, 2004). This will be analysed in detail in Section 8.

The simplest bound on the P-value is the Bonferroni bound (BON):

$$P \leq P_{\text{BON}} = N\overline{\Phi}(t) \tag{2}$$

where N is the number of lattice points inside S and $\overline{\Phi}(z) = \mathbb{P}(Z > z)$, $Z \sim \text{N}(0,1)$. If the random field is not smooth, so that its sampled values are roughly independent, then (2) is very accurate, but if the random field is very smooth, then it becomes too conservative.

If the random field is smooth relative to the mesh size and S is convex or nearly so then very accurate approximations to (1) have recently been found using the expected Euler characteristic (expected EC, or XEC) of the excursion set of a continuously sampled random field (Adler, 1981; Taylor & Adler, 2003). In 2D, the EC counts the number of connected components minus the number of holes in the excursion set, which for high thresholds takes the value 1 if the maximum exceeds the threshold, and 0 otherwise. Hence XEC approximates the P-value of the maximum for high thresholds. For a stationary Gaussian random field the XEC depends on the roughness of the random field measured by

$$\text{Var}\left(\frac{\partial Z(x)}{\partial x}\right) = \Lambda. \tag{3}$$

Then the XEC P_{XEC} is

$$P \approx P_{\text{XEC}} = \sum_{d=0}^{D} \mu_d(S\Lambda^{1/2})\frac{\partial^d \overline{\Phi}(t)}{\partial t^d}, \tag{4}$$

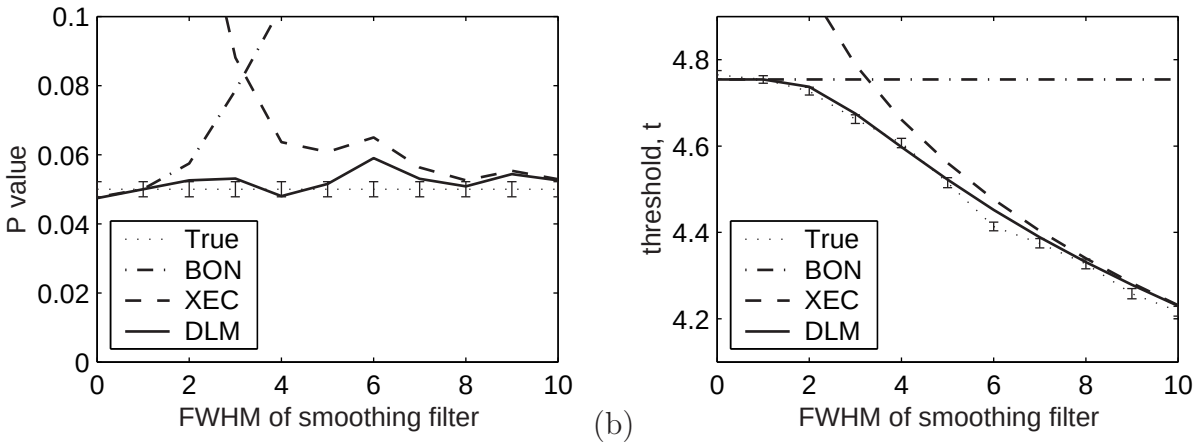

Figure 2: (a) Comparison of Bonferroni (BON), expected Euler characteristic (XEC) and proposed Discrete Local Maxima (DLM) P-values for a 224^2 lattice search region, as a function of filter $FWHM$ relative to mesh size, $FWHM/v$. The thresholds were the P=0.05 thresholds based on 9999 simulations (True). Error bars are ± 1 Sd. (b) Same as in (a), but replacing P-value by threshold.

where $\mu_d(S)$ is the d-dimensional Minkowski functional or intrinsic volume of S (Worsley, 1995a), and $S\Lambda^{1/2} = \{s\Lambda^{1/2} : s \in S\}$. In 2D, $\mu_0(S)$ is the EC of S (=1 if S is simply connected); $\mu_1(S)$ is half the perimeter length of S; $\mu_2(S)$ is the area of S. For $d = D$,

$$\mu_D(S\Lambda^{1/2}) = |S||\Lambda|^{1/2} \tag{5}$$

where $|S|$ is the Lebesgue measure of S. For large thresholds t, the $d = D$ term in (4) is the most important, and

$$P_{\mathrm{XEC}} = |S||\Lambda|^{1/2}(2\pi)^{-(D+1)/2}\exp(-t^2/2)t^{D-1} \cdot (1 + O(1/t)). \tag{6}$$

The error in the approximation (4) is *exponentially* smaller than the $d = 0$ term (Taylor *et al.*, 2005).

In the brain mapping literature, smoothness is conveniently measured by FWHM, defined as follows. If $Z(x)$ is modelled as white noise convolved with an isotropic Gaussian-shaped filter, then FWHM is the Full Width at Half Maximum of this filter. FWHM is $\sqrt{8\log 2}$ times the standard deviation of the Gaussian filter, and if $\Lambda = \lambda I$ where I is the $D \times D$ identity matrix, it can be shown that

$$FWHM = \sqrt{4\log 2/\lambda}. \tag{7}$$

The XEC P-values (6) are then roughly proportional to the volume of S divided by $FWHM^D$, so that as the random field becomes less smooth, then these P-values become larger.

Figure 2 illustrates the point. Gaussian random fields were simulated on a 256^2 lattice by smoothing white noise with an isotropic Gaussian filter. For $FWHM < 1$ lattice step BON is accurate; for $FWHM > 5$ lattice steps XEC is accurate.

A simple solution is to take the best of both, i.e. the minimum of the two P's or thresholds. This will give P-values that are never more than twice the true value in the example in Figure 2. This method is currently used by the FMRISTAT (Worsley *et al.*, 2002) and SPM software (Frackowiak *et al.*, 2003) for the statistical analysis of brain imaging data. The purpose of this paper is to look for something better for the mid range of $FWHM$. Extensions to non-Gaussian random fields are given in Worsley (2005).

2 Improved Bonferroni bounds

The problem of determining P in (1) can be boiled down to this. We have a set of events $A_x = \{Z(x) > t\}$ and we seek a good approximation to the probability of their (finite) union

$$P = \mathbb{P}\left(\cup_{x \in S} A_x\right). \tag{8}$$

We are interested in the case where S is large (typically 50,000) and P is small (typically 0.05). The Bonferroni inequality gives (2), but to do better, we should try to incorporate more information. There is a long history of trying to improve the Bonferroni bound by adding or subtracting more higher order terms of the form

$$P = \mathbb{P}\left(\cup_{x \in S} A_x\right) \leq \sum_{J \subset S} (-1)^{|J|-1} f(J) \mathbb{P}\left(\cap_{x \in J} A_x\right) \tag{9}$$

for some positive function f on subsets of S. Of course with $f(J) = 1$ (9) is an equality, but the idea is to set f to zero for most high order intersections, and still retain a sharp inequality. The remaining terms, involving low-order intersections and hence low-dimensional integrals, can then be evaluated numerically. A good example which works well when $D = 1$, $S = \{1, \ldots, n\}$ and neighbouring values of $Z(x)$ are highly correlated is

$$\mathbb{P}\left(\cup_{x \in S}\right) \leq \sum_{x=1}^{n} \mathbb{P}(A_x) - \sum_{x=1}^{n-1} \mathbb{P}(A_x \cap A_{x+1}) \tag{10}$$

(Hunter, 1976; Worsley, 1982; Efron, 1997). It can easily be generalised by replacing the second summation over adjacent lattice points by a summation over edges of a tree on S. Tomescu (1986) generalised it to hyper-trees. Naiman & Wynn (1992, 1997) considered the conditions on $f(J)$ under which (9) is an equality or inequality, with applications to importance sampling. There has been recent work by Dohmen (2000, 2003) and Dohmen & Tittmann (2004) improving the results of Galambos & Simonelli (1996) and Kwerel (1975) who considered the case where $f(J)$ takes the same value for all subsets J of the same size. Further work along these lines has been done by Kounias (1968), Grable (1993) and Hoppe (1993). A related problem, with links to False Discovery Rate, is addressed by Simes (1986) and Sarkar (1998). Some of these methods have found applications in maximally selected rank statistics (Hothorn & Lausen, 2003).

However none of these methods work well for the problem of a smooth random field sampled on a rectilinear grid. The only exception is (10) in the 1D case, which turns out, as the mesh size decreases, to approach the XEC result (4) which we know to be very accurate for high thresholds, the ones of main interest. The reason is that none of the above improved Bonferroni inequalities take direct advantage of the spatial correlation structure of $Z(x)$.

Our proposed DLM method, introduced in the next section, does take into account the spatial properties of $Z(x)$, and it is very accurate as the mesh size decreases, approaching the XEC result for large thresholds. However it is not strictly speaking an improved Bonferroni inequality because it is not of the form (9). However it does rely on the Bonferroni inequality applied to events other than A_x.

3 An improved Bonferroni-type bound based on discrete local maxima

We now derive an improved Bonferroni-type bound that bridges the gap between small $FWHM$, where BON is accurate, and large $FWHM$, where XEC is accurate. The improved Bonferroni-type bound is the expected number of discrete local maxima (DLM) above threshold:

$$P \leq P_{\mathrm{DLM}} \triangleq \sum_{x \in S} \mathbb{P}(Z(x) > t \text{ and neighbouring } Z\text{'s} < Z(x)). \tag{11}$$

The $2D$ neighbours are those that differ by just one step in each lattice direction. In contrast, P_{BON} is the expected number of lattice points above threshold, and P_{XEC} is the expected Euler characteristic of the search region above threshold. We shall show that, like BON, DLM is conservative, which is reassuring for practical applications, but unlike BON it is very accurate for all $FWHM$; for large $FWHM$ and thresholds DLM converges to XEC.

The fact that DLM is an upper bound follows by noting that the event that $\{Z(x) > t\}$ somewhere in the search region is the union of the events $\{Z(x) > t$ and neighbouring Z's $< Z(x)\}$ over all lattice points, then applying the Bonferroni inequality. Formally,

$$\left\{\max_{x \in S} Z(x) > t\right\} = \bigcup_{x \in S} \{Z(x) > t\} = \bigcup_{x \in S} \left(\{Z(x) > t\} \cap_{y \in \mathcal{N}} \{Z(y) < Z(x)\}\right) \tag{12}$$

where $\mathcal{N} = \mathcal{N}_x$ is the set of neighbouring lattice points about x inside S. Then, by Bonferroni,

$$P = \mathbb{P}\left(\bigcup_{x \in S} \{Z(x) > t\}\right) \leq \sum_{x \in S} \mathbb{P}(\{Z(x) > t\} \cap_{y \in \mathcal{N}} \{Z(y) < Z(x)\}) = P_{\mathrm{DLM}}. \tag{13}$$

We should note that (11) is not strictly speaking an improved Bonferroni inequality, since it is not a function of the probabilities of intersections of events $\{Z(x) > t\}$. This will be discussed more in Section 5.

4 Evaluating the DLM P-value

It remains to calculate

$$\mathbb{P}(\{Z(x) > t\} \cap_{y \in \mathcal{N}} \{Z(y) < Z(x)\}).$$

where $\mathcal{N}$ is the set of $2D$ neighbouring lattice points about x that differ by just one step in each lattice direction. This looks like an integral in $2D + 1$ Gaussian random variables, but the amount of integration can be substantially reduced if we make more assumptions about the spatial correlation structure of the data.

We shall assume that the spatial correlation function is locally *separable*, that is, the product of the correlation functions along the lattice axes:

$$\rho(y, w) = \mathrm{Cor}(Z(y), Z(w)) = \prod_{d=1}^{D} \rho_d(y_d, w_d), \tag{14}$$

where $y = (y_1, \ldots, y_D)$ and $w = (w_1, \ldots, w_D)$ are neighbours of x, and ρ_d is the spatial correlation function along axis d. This correlation structure would arise, for instance, if a lattice of

i.i.d. Gaussian random variables was convolved with a separable filter. A special case is the Gaussian correlation function

$$\rho(y, w) = \exp(-(y_d - w_d)\Lambda(y_d - w_d)'/2), \tag{15}$$

where $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_D)$.

If we first condition on the central $Z(x)$, then the assumed separable correlation structure (14) invokes a 'Markovian' property: it can be checked that neighbouring $Z(y)$'s on one lattice axis are conditionally independent of neighbouring $Z(w)$'s on any other lattice axis. We can now evaluate the conditional probability separately for each lattice axis, then multiply them together, then finally integrate over the conditioned central $Z(x)$. Formally, let e_d be a D-vector of zeros with dth component equal to v_d, the step size in the dth direction. Define

$$Q_{xd}(z) \stackrel{\Delta}{=} \mathbb{P}(Z(y) < z, \text{for all } y = x \pm e_d \in S \mid Z(x) = z) \tag{16}$$

and

$$\tilde{P}_{\mathrm{DLM}} \stackrel{\Delta}{=} \sum_{x \in S} \int_t^\infty \left(\prod_{d=1}^{D} Q_{xd}(z) \right) \phi(z)\, dz. \tag{17}$$

If the spatial correlation is separable then

$$P_{\mathrm{DLM}} = \tilde{P}_{\mathrm{DLM}}.$$

To find a simple expression for Q, we shall also assume that $Z(x)$ is locally stationary, so that we need just the two neighbour correlations on the same axis, given by

$$\rho_{d1} = \mathrm{Cor}(Z(x), Z(x + e_d)), \qquad \rho_{d2} = \mathrm{Cor}(Z(x - e_d), Z(x + e_d)). \tag{18}$$

Conditional on $Z(x) = z$, the neighbours have a bivariate Gaussian distribution

$$\left[\begin{array}{c} Z(x + e_d) \\ Z(x - e_d) \end{array} \right] \Bigg| \; Z(x) \sim \mathrm{N}_2 \left(\left[\begin{array}{c} \rho_{d1} z \\ \rho_{d1} z \end{array} \right], \left[\begin{array}{cc} 1 - \rho_{d1}^2 & \rho_{d2} - \rho_{d1}^2 \\ \rho_{d2} - \rho_{d1}^2 & 1 - \rho_{d1}^2 \end{array} \right] \right). \tag{19}$$

Letting X, Y be independent standard Gaussian random variables, $Z(x \pm e_d)$ can be written as

$$\rho_{d1} z + X \sqrt{(1 - 2\rho_{d1}^2 + \rho_{d2})/2} \pm Y \sqrt{(1 - \rho_{d2})/2}. \tag{20}$$

If both neighbours are in S and $|\rho_{d1}| < 1$,

$$Q_{xd}(z) = \mathbb{P}(X \sin \alpha_d \pm Y \cos \alpha_d < h_d z), \tag{21}$$

where

$$\alpha_d = \sin^{-1} \left(\sqrt{\frac{1 - 2\rho_{d1}^2 + \rho_{d2}}{2(1 - \rho_{d1}^2)}} \right), \qquad h_d = \sqrt{\frac{1 - \rho_{d1}}{1 + \rho_{d1}}}.$$

The necessary bivariate integral can be reduced to a single integral by changing to polar coordinates $x = r \cos\theta$, $y = r \sin\theta$ and integrating the radius r analytically. The remaining integral over the angle θ is, for $z > 0$,

$$Q_{xd}(z) = 1 - \frac{1}{\pi} \int_{\alpha_d}^{\pi} \exp(-\tfrac{1}{2} h_d^2 z^2 / \sin^2\theta) d\theta, \tag{22}$$

with a similar expression when $z < 0$. Recognizing that (22) is $1 - 2\overline{\Phi}(h_d z)$ if $\alpha_d = 0$, and changing the limits accordingly, gives, for all z,

$$Q_{xd}(z) = 1 - 2\overline{\Phi}(h_d z^+) + \frac{1}{\pi}\int_0^{\alpha_d} \exp(-\tfrac{1}{2}h_d^2 z^2/\sin^2\theta)d\theta, \tag{23}$$

and $z^+ = z$ if $z > 0$ and 0 otherwise. Let $\phi(z) = \exp(-z^2/2)/\sqrt{2\pi}$. If x is on the boundary of S with just one neighbour in axis direction d, then $Q_{xd}(z) = 1 - \overline{\Phi}(h_d z)$, and is equal to 1 if x has no neighbours. This invokes a boundary correction, similar in purpose to the first D terms in the summation in XEC (4).

5 Relationship between BON, XEC and DLM

When the voxels (lattice points) are independent, P_{DLM} is very slightly smaller than P_{BON} but slightly larger than the true P, specifically,

$$P_{\text{DLM}} = \left(1 - (1-P)^{(2D+1)/N}\right) N/(2D+1). \tag{24}$$

In fact the difference is hardly noticeable in Figure 2 at $FWHM = 0$.

For large $FWHM$ relative to v_d, we now show that P_{DLM} converges to P_{XEC} for large thresholds when the correlation function is separable and Gaussian (15), specifically,

$$P_{\text{DLM}} \approx P_{\text{XEC}} \cdot (1 + O(1/t)). \tag{25}$$

First change variables to $u = \exp(-\frac{1}{2}h_d^2 z^2/\sin^2\theta)$ so that the integral in (23) becomes

$$h_d z \int_0^{u_d(z)} \frac{1}{-2\log u\sqrt{-2\log u - h_d^2 z^2}}\, du, \tag{26}$$

where $u_d(z) = \exp(-\frac{1}{2}h_d^2 z^2/\sin^2\alpha_d)$. This is in fact a more convenient form for numerical integration. When the mesh size v_d approaches 0, $\rho_{d1} \approx 1 - \lambda_d v_d^2/2 \to 1$, $\rho_{d2} \approx 1 - 2\lambda_d v_d^2 \to 1$ and $h_d \approx v_d\sqrt{\lambda}/2 \to 0$. From (23) and (26)

$$Q_{xd}(z) \approx \frac{v_d z\sqrt{\lambda_d}}{\sqrt{2\pi}}(1+B) \tag{27}$$

where $B \le (2/\sqrt{\pi})\exp(-z^2/4)/z^3$, which is negligible for large thresholds. Integrating over z in (17) we obtain

$$P_{\text{DLM}} \approx \left(\prod_{d=1}^{D}\lambda_d^{1/2}\right)(|S|/N)(2\pi)^{-(D+1)/2}t^{D-1}\exp(-t^2/2)\cdot(1+O(1/t^2)) \tag{28}$$

which together with (6) gives (25).

Can we extend DLM to give the remaining terms? To do this, we might be tempted to replace DLM with a 'discrete EC', by analogy with XEC. Another way of counting the EC of the (continuously sampled) excursion set in 2D is to count the number of continuous local maxima, minus saddle points, plus minima of $Z(x)$ inside the excursion set. This suggests that instead of counting discrete local maxima, we can subtract discrete local saddle points and add

discrete local minima. This turns out to be remarkably easy - the resulting adjustment to DLM is to replace Q by

$$Q_{xd}(z) = 1 - \sum_{y=x\pm e_d \in S} \mathbb{P}(Z(y) > z \mid Z(x) = z), \tag{29}$$

a Bonferroni approximation on the two or less events $\{Z(y) > z, y = x \pm e_d \in S \mid Z(x) = z\}$. For separable correlations this removes the troublesome integral from (23) and replaces it with simply

$$Q_{xd}(z) = 1 - \#\{y = x \pm e_d \in S\}\overline{\Phi}(h_d z). \tag{30}$$

Unfortunately this heuristic does not work. The number of discrete local maxima minus saddle points plus minima is *not* the EC of the discretely sampled excursion set, as can be seen from a simple example of a diagonal "ridge" of local maxima, each contributing +1, with no discrete saddle points or local minima, yet the excursion set can be simply connected with an EC of +1.

There is one exception: in the 1D case the number of discrete local maxima minus minima *is* the EC of the discretely sampled excursion set. The reason is that excursion sets in 1D are intervals, and each interval must contain at least one local maximum and one less local minimum. So in this case DLM with (29) counts the number of such intervals, and so it is an upper bound on the P-value P, slightly tighter than the DLM with (23).

Another way of counting the intervals in the excursion set is to count the upcrossings of the threshold t by the process $Z(x)$. In other words, for a 1D lattice, we count the events $\{Z(x) < t \cap Z(x + e_1) > t\}$. The expectation of this is in fact identical to the simplest improved Bonferroni inequality (10), so we are back where we started. Incidentally, this shows why (10) works so well for 1D random fields. We might be tempted to extend this idea to higher dimensions and count multidimensional upcrossings in a way similar to the Hadwiger characteristic (Adler, 1981; Worsley, 1995b). However the extra complication, plus the tricky question of how to add a boundary correction, does not seem to be worth the very slight tightening of the bound on P.

Finally, of course, we could attack the discrete EC directly. It can be defined in 2D as the number of lattice points, minus the number of lattice edges, plus the number of lattice squares (or faces) inside the excursion set (Adler, 1981). We can then find its expectation, but unfortunately the Markovian property for separable correlations does not help us to cut down the number of integrals, which grows exponentially in D.

To pursue this properly, we could possibly resort to a specially developed discrete Morse Theory (Forman, 1998; Lewiner *et al.*, 2004) which incorporates events involving all 2^D voxels in a *cell* (cube of adjacent voxels) rather than the $2D + 1$ neighbouring voxels on the axes. There is nothing in principle to prevent us from calculating XEC defined in this way, but again in practice there is no Markovian property and so no nice reduction in the number of multiple integrals. Moreover the expected discrete EC is still only an approximation to P, not a bound.

6 Non-separable spatial correlation

What happens if we use $\tilde{P}_{\mathrm{DLM}}$ to evaluate P_{DLM} when in fact the correlation structure is not separable? We now give an argument that $\tilde{P}_{\mathrm{DLM}}$ is still conservative if the FWHM is large relative to the lattice size. If this is so then the local correlation structure is approximately Gaussian (15) with say $\lambda_1, \ldots, \lambda_D$ on the diagonal of Λ, but with non-zero off-diagonal elements.

Repeating the calculations that lead to (25) we get

$$\tilde{P}_{\mathrm{DLM}} \approx \left[\left(\prod_{d=1}^{D} \lambda_d^{1/2} \right) / |\Lambda|^{1/2} \right] P_{\mathrm{XEC}} \cdot (1 + O(1/t)). \tag{31}$$

It can be shown using Hadamard's inequality that for any positive definite matrix Λ the product of the diagonal elements is greater than the determinant, so that the term in square brackets in (31) is greater than or equal to 1. This implies that $\tilde{P}_{\mathrm{DLM}} \geq P_{\mathrm{XEC}}$ for large $FWHM$. Further, it is known that P_{XEC} is very accurate (Taylor *et al.*, 2005) for large thresholds and large $FWHM$. This implies that $\tilde{P}_{\mathrm{DLM}}$ is conservative for large $FWHM$ whenever P_{XEC} is accurate, i.e. when the underlying correlation function is smooth. It is also slightly larger than the true P-value when $FWHM = 0$, so this suggests that DLM is conservative for all $FWHM$ even if the local spatial correlation is not separable.

However it is not hard to construct examples where $\tilde{P}_{\mathrm{DLM}} < P_{\mathrm{DLM}}$. A simple example is a "checker board" with Z_1 on white squares and Z_2 on black squares, $Z_1, Z_2 \sim \mathrm{N}(0,1)$ independently. Nevertheless we can show that $\tilde{P}_{\mathrm{DLM}}$ is still conservative, using Slepian's inequality as follows. Consider a second example with Z_j on every second square in both lattice directions, $Z_1, Z_2, Z_3, Z_4 \sim \mathrm{N}(0,1)$ independently. It can be checked that the correlation of the second process is separable, with the same correlations as the checker board example on the two axes but lower correlations off the two axes. By Slepian's inequality (or by direct calculation) the P-value of the maximum of the checker-board example is smaller than that of the second example, which in turn is smaller than $\tilde{P}_{\mathrm{DLM}}$ (since it equals the DLM for the second example).

We conjecture that $\tilde{P}_{\mathrm{DLM}}$ is always conservative for any stationary process whether it is separable or not. Some evidence for this from simulations is presented in Section 7 and Worsley (2005).

7 Simulations

The methods were compared on simulated data that matched the 'bubbles' data in Section 8. A 256^2 image of independent zero mean, unit variance Gaussian random variables was smoothed with a separable Gaussian shaped filter to generate a smooth random field $Z(x)$. The filter was normalised to preserve the unit variance of the smoothed image. Fourier methods were used for the convolution, so the smoothed image was periodic. A search region S of size 224^2 was chosen that was far enough from the boundary of the 256^2 image that the periodicity could be ignored. The maximum of $Z(x)$ inside S was recorded.

This was repeated $M = 9999$ times. The $(M+1)P$th largest maximum of $Z(x)$ inside S estimates the true threshold (unbiasedly if the simulated values were uniform). The standard deviation of this estimate was itself estimated by specifying a small width $\delta = 0.02$, then by

$$\frac{(M+1)(P+\delta)\mathrm{th} - (M+1)(P-\delta)\mathrm{th}}{2\delta} \sqrt{\frac{P(1-P)}{M+2}}. \tag{32}$$

This is based on the usual linear approximation to the variance of a function of a random variable. The first term is an estimate of the inverse of the probability density, the second square root is the standard deviation of the sample P-value. Results for a $P = 0.05$ threshold are shown in Figure 2 for $FWHM$ ranging from 0 (no smoothing) to 10 lattice steps.

The DLM P-value is always an accurate upper (conservative) bound on the true P-value, which almost equals BON when $FWHM = 0$ and slightly overestimates XEC when $FWHM > 6$.

In between DLM is better than either of them. The greatest discrepancy occurs at $FWHM = 3$, where the DLM P-value is about half either of the others.

To investigate the conjecture that $\tilde{P}_{\text{DLM}}$ is always conservative, we carried out a small-scale simulation by smoothing a 5^2 periodic 2D lattice of iid zero mean Gaussians with a kernel

$$\begin{bmatrix} b & a & b \\ a & 1 & a \\ b & a & b \end{bmatrix}.$$

The $P = 0.05$ threshold was determined from 199,999 simulations for every value of a and b between -1 and 1 in steps of 0.1. P_{DLM} was estimated by the proportion of local maxima that were above threshold, and $\tilde{P}_{\text{DLM}}$ was calculated from (18)–(23) assuming that the rest of the spatial correlation was separable. The smoothed approximate P-values, with an approximate standard deviation of 0.0001, are shown in Figure 3. For most of the values of a and b, $P < P_{\text{DLM}} < \tilde{P}_{\text{DLM}}$ except a region where $0 < a < b < 1$, but even here $\tilde{P}_{\text{DLM}}$ is still conservative. Inside the shaded region

$$\rho(x + s_1 e_1, x + s_2 e_2) > \rho(x + s_1 e_1, x)\ \rho(x, x + s_2 e_2)$$

for all x and scalar s_1, s_2, so that here the process is "more correlated" than the product of the correlations along the axes. By Slepian's inequality its P-value is lower than that of this associated separable process, for which $P_{\text{DLM}} = \tilde{P}_{\text{DLM}}$. Hence we know that $\tilde{P}_{\text{DLM}}$ must be conservative in the shaded region. A thorough simulation study of this conjecture is beyond the scope of this paper.

8 Application to the 'bubbles' experiment

We give an application to the 'bubbles' technique for detecting areas of the face used to discriminate fear from happiness. The data come from the control subjects analysed in Adolphs *et al.* (2005). Subjects were shown a 256^2 image of a face that is either fearful or happy. The images were masked apart from a random number of localised regions or 'bubbles' that reveal only selected parts of the face (see Figure 4). The subject was then asked whether the partially revealed face is fearful or happy, and the trial was repeated ≈2,970 times on each of 10 subjects.

The masked image was generated as follows. The original image I_0 was smoothed by an isotropic Gaussian filter with $FWHM_i = 7.05 \times 2^{i-1}$, to produce images I_i, $i = 1, \ldots, 5$. The smoothed images were differenced to produce images $D_i = I_{i-1} - I_i$, $i = 1, \ldots, 5$ that reveal image features at five different scales. Differenced image D_i was then multiplied by a mask consisting of the sum of a random number of isotropic Gaussian 'bubbles', each with width $2FWHM_i$. The bubble centres were chosen at random from the 256^2 pixels. The number of bubbles for each scale was a multinomial random variable with probabilities inversely proportional to bubble area ($FWHM_i^2$), so that equal areas were revealed by each of the 5 bubble sizes, on average. The total number of bubbles was chosen dynamically to maintain a 0.75 probability of correctly identifying the face.

At each pixel x, let $\hat{p}_{\text{C}}$ be the proportion of bubbles centred at x in the n_{C} correctly classified stimuli, and let $\hat{p}_{\text{I}}$ be the proportion in the n_{I} incorrectly classified stimuli. The number of bubbles per stimulus, $m = 16.5$ on average, was adjusted to maintain the success rate $n_{\text{C}}/(n_{\text{C}} + n_{\text{I}}) \approx 0.75$. If face discrimination is unrelated to the bubbles, then these proportions

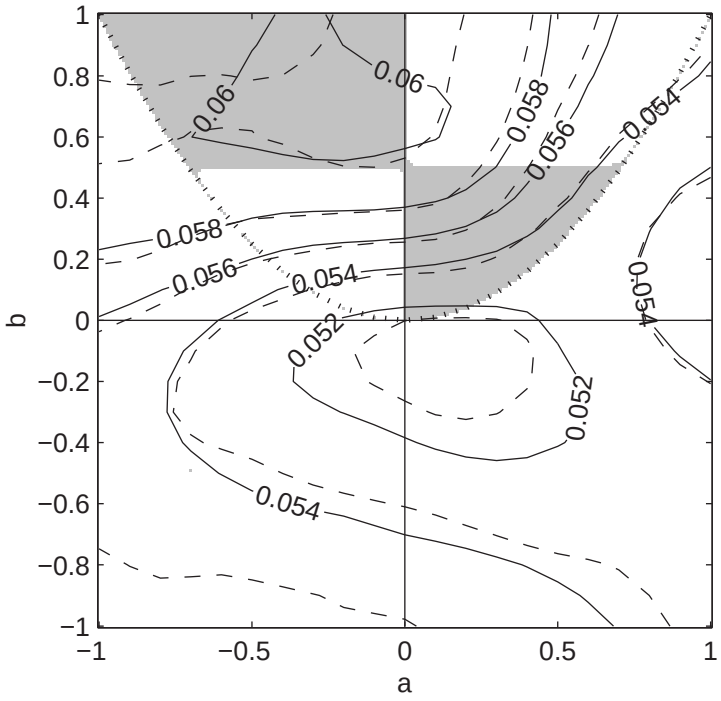

Figure 3: Support for the conjecture that $\tilde{P}_{\mathrm{DLM}}$ is always conservative. A 5^2 array of i.i.d. Gaussians was smoothed with a 3^2 kernel with a on the axes and b on the diagonals. The $P = 0.05$ threshold was determined from 199,999 simulations and P_{DLM} (solid contours) and $\tilde{P}_{\mathrm{DLM}}$ (dashed contours) were determined for this threshold. The dotted line ($b = a^2$) is where the kernel, and hence the correlation, is separable, so that $\tilde{P}_{\mathrm{DLM}} = P_{\mathrm{DLM}}$. Inside the shaded region $\tilde{P}_{\mathrm{DLM}}$ is always conservative by Slepian's inequality.

are expected to be $p = m/256^2$. The test statistic $Z(x)$ is then defined as

$$Z(x) = \frac{\hat{p}_{\mathrm{C}} - \hat{p}_{\mathrm{I}}}{\sqrt{p(1-p)(1/n_{\mathrm{C}} + 1/n_{\mathrm{I}})}}. \tag{33}$$

We only used the information contained in the bubble centres, ignoring the information contained in the bubble $FWHM$s.

The unsmoothed image $Z(x)$ is shown in the top left panel in Figure 5 (the search region S is the interior of the black frame). This was then smoothed by Gaussian filters of $FWHM$=1.5, 3, 6, 12 and 24 pixels (top row) normalised to preserve the variance. Note that $Z(x)$ is a linear function of the ≈29,700 binary images of presence/abscence of bubble centres (1=bubble centred at x, 0=otherwise), so that smoothing each of the binary images, then calculating a Gaussian statistic for comparing their means, is equivalent (up to a constant) to smoothing $Z(x)$.

We now check the conditions for P_{DLM} to be evaluated by the simpler form $\tilde{P}_{\mathrm{DLM}}$. First, the noise component of the unsmoothed image $Z(x)$ is well-approximated by i.i.d. Gaussian random variables, by the Central Limit Theorem. Since the filter is axis-aligned Gaussian then it is separable, which implies that its spatial correlation is also separable, thus satisfying the condition (14) for $P_{\mathrm{DLM}} = \tilde{P}_{\mathrm{DLM}}$.

Three methods for thresholding at $P = 0.05$ were used: BON, XEC and DLM (rows 2 to 4). Note that the third image ($FWHM = 3$) is where DLM outperformed both BON and XEC

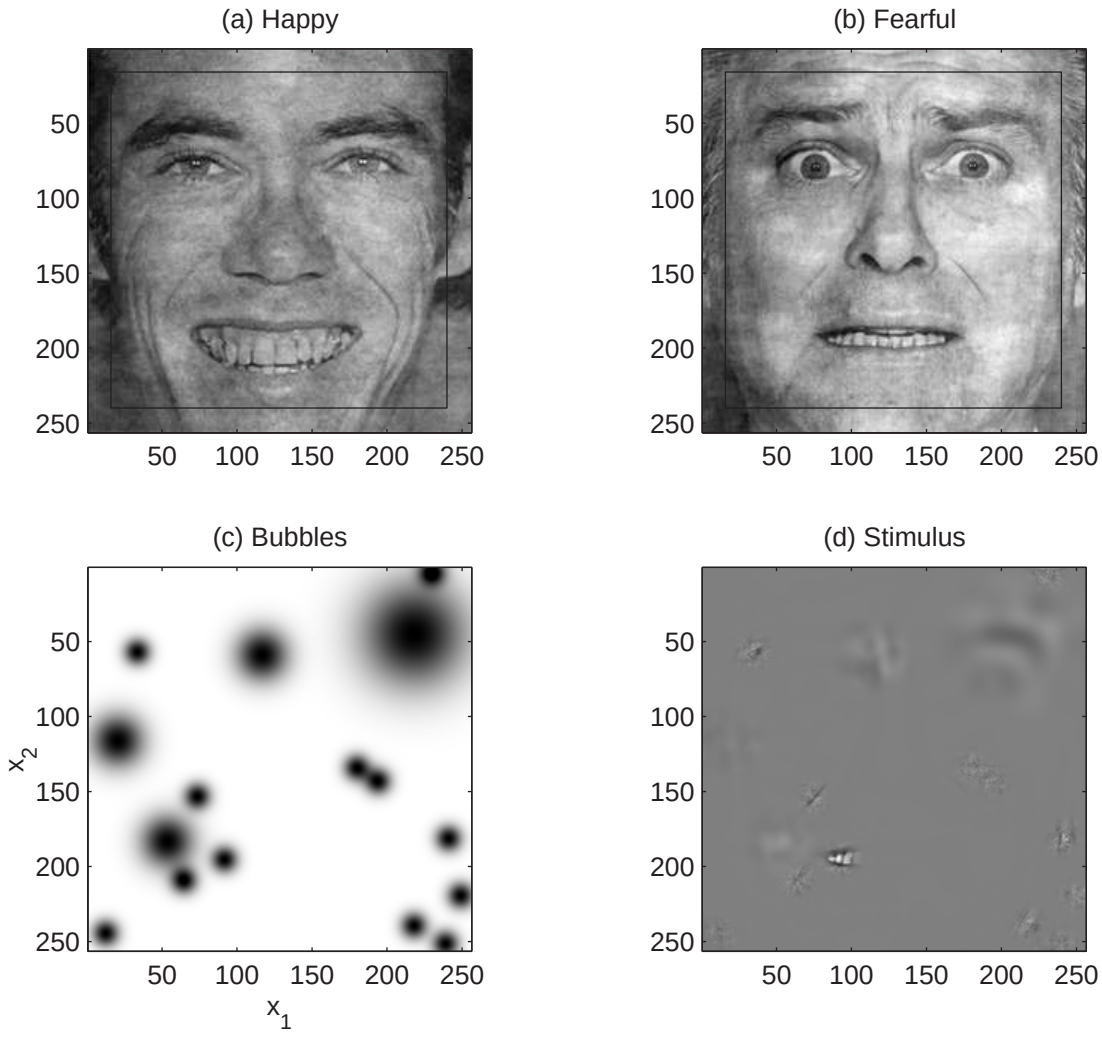

Figure 4: Bubbles experiment. The subject is asked to discriminate between the happy (a) and fearful (b) faces on presentation of the stimulus (d) which is one of the two faces (here the fearful face) partially revealed by random 'bubbles' (c). The 224^2 search region S is inside the black frame in (a) and (b).

in the simulations (Figure 2). The fourth image ($FWHM = 12$) is close to the smoothness of the smallest bubbles ($FWHM = 14.1$). The BON method picks up some discriminatory features in the eyes, but as the smoothing increases, it does not do as well as XEC. XEC on the other hand, fails to pick up any features when there is little or no smoothing. DLM combines the best of both: it does as well as BON at low smoothness, and almost as well as XEC at high smoothness. Recall that DLM, like BON, is an upper bound, so it is always conservative, whereas XEC is merely an approximation.

The analysis we have presented here is preliminary. A natural question is what amount of smoothing to use. Although one can make a good argument that the amount of smoothing should match the size of the bubble (here at least 14), from a purely inferential point of view, the optimal smoothing should match the shape of the underlying signal, by the well-known Matched Filter Theorem of signal processing. In other words, a filter the same size as the eyes might be optimal at detecting bubble centres clustered in the eye region. A formal analysis can

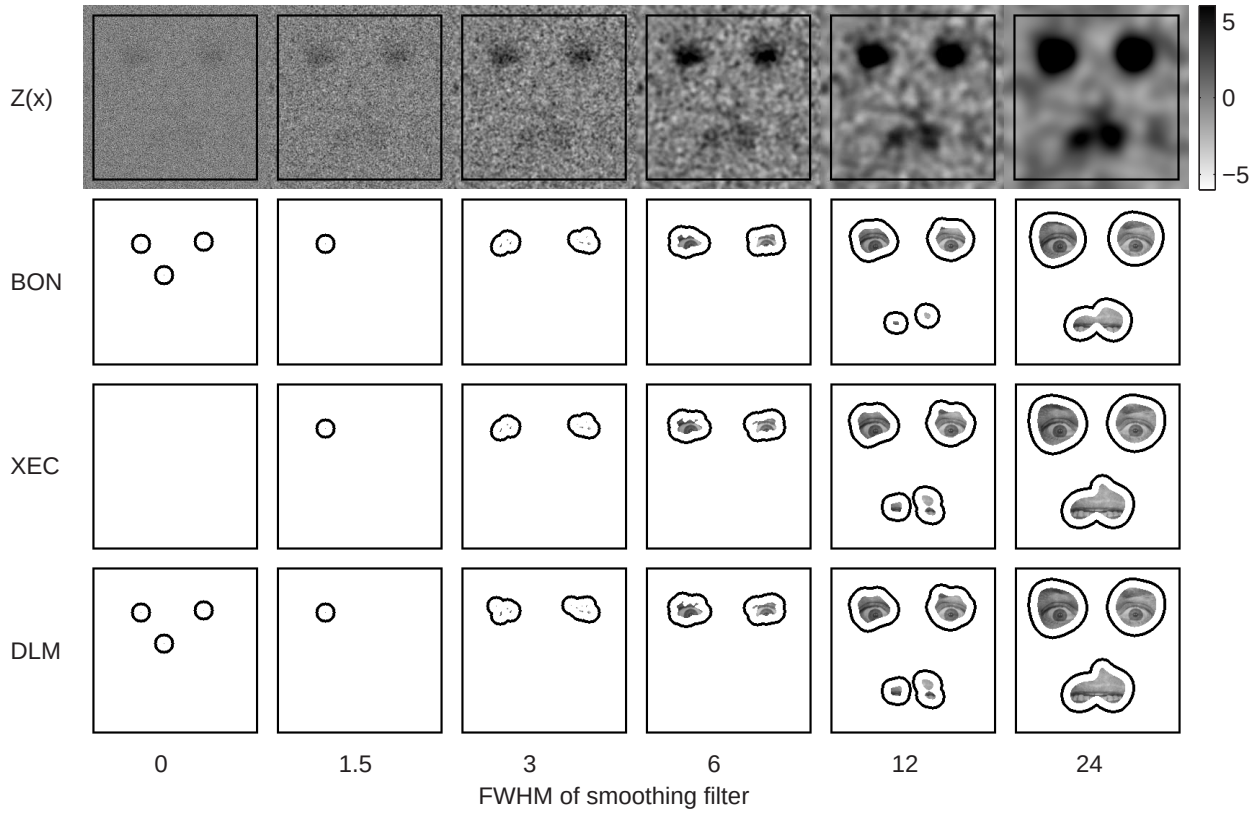

Figure 5: Bubbles analysis. Top row: $Z(x)$ and 224^2 search region S (black frame) smoothed with increasing $FWHM$ Gaussian filters; Rows 2-4: images in first row thresholded at the $P = 0.05$ level using the BON, XEC and DLM methods. Excursion sets, revealing the underlying fearful face, have been outlined in black to make them easier to identify. Note that DLM does as well as either BON or XEC at detecting facial features that discriminate between fear and happiness, the upper eyes and upper lip.

be done by searching over filter width, as well as pixel location, known as scale space (Siegmund & Worsley, 1995). The price to pay for this is an increase in threshold; for searching over a 10 fold scale range from 10 to 100 $FWHM$ the threshold is $t = 4.31$, as opposed to 4.23 – 2.95 at fixed scales from 10 to 100 $FWHM$. Scale space local maxima (in 3D) detected the eyes at $FWHM \approx 25$ pixels ($Z_{\max} \approx 17.2$) and the mouth at a higher smoothness of $FWHM \approx 40$ pixels ($Z_{\max} \approx 8.0$) - there were no other local maxima above threshold. These estimated $FWHM$s are indeed maximum likelihood estimates of the true $FWHM$s of the Gaussian shaped signals added to the unsmoothed $Z(x)$ (Siegmund & Worsley, 1995).

9 Discussion

We have derived an accurate upper bound for the maximum of a discrete lattice of (correlated) Gaussian random variables inside a fixed search region of arbitrary shape. The DLM P-value P_{DLM} is simply the Bonferroni bound applied to discrete local maxima, as opposed to P_{BON}, Bonferroni applied to the lattice values themselves. This very simple idea produces a bound that is exponentially sharper than P_{BON}, and whose error is exponentially smaller than the

P-value itself (see Appendix A).

The main attraction of P_{DLM} comes when the lattice values are sampled from a smooth continuous Gaussian field. As the mesh size decreases, P_{DLM} does not become increasingly conservative, despite the fact that the number of lattice points in the search region increases. In contrast, P_{BON} or other Bonferroni approximations discussed in Section 2 and Appendix A become increasingly conservative. This is true even though, for a fixed mesh size, there is no gain in exponential accuracy of P_{DLM} when compared to the second-order Bonferroni correction including only nearest-neighbour edges. However, as the mesh size goes to zero, P_{DLM} approaches the P-value of the maximum of the continuous random field from which it was sampled. In this sense it is the only approximation discussed here that "interpolates" between P_{BON} and P_{XEC}: two approximations which we know are accurate at the extremes of mesh size.

The evaluation of P_{DLM} still requires an integral in as many variables as neighbours used to determine the discrete local maxima, repeated at each lattice point in the search region. Even if the lattice values are stationary, so that essentially only one multidimensional integral is needed (ignoring the boundaries of the search region), this is still prohibitive if the number of dimensions is large. However if we are willing to to assume that the local correlation structure is separable, the calculations can be reduced to a single bivariate integral, giving $\tilde{P}_{\mathrm{DLM}}$ (11).

The only remaining question is: if the spatial correlation is not separable, how close is $\tilde{P}_{\mathrm{DLM}}$ to the true P-value? In Section 6, we showed that $\tilde{P}_{\mathrm{DLM}}$ is conservative if the observations were finely sampled from any smooth stationary Gaussian field. We were able to construct a non-smooth example where $\tilde{P}_{\mathrm{DLM}} < P_{\mathrm{DLM}}$, yet despite this, $\tilde{P}_{\mathrm{DLM}}$ was still conservative. We have so far been unable to come up with a stationary lattice for which $\tilde{P}_{\mathrm{DLM}}$ is anti-conservative. Note that stationarity is not necessary for $\tilde{P}_{\mathrm{DLM}} = P_{\mathrm{DLM}}$, but it is needed to obtain a workable expression for $\tilde{P}_{\mathrm{DLM}}$.

A referee has asked why we cannot simply use simulations if we know the entire distribution of the lattice random variables, as we do for the bubbles application. For the bubbles, generating 9999 smoothed 256^2 Gaussian images is not too time consuming; this was how we obtained the results in Figure 2. But our other main interest is applications to 3D brain imaging data, where the images can be as large as 256^3, which is currently prohibitively expensive to simulate (Worsley, 2005). Moreover the exact correlation structure of the data is not known, neighbour correlations must be estimated from residuals, and the residuals are not stationary. Most importantly, the variances are not known, so the test statistic is a T statistic rather than a Gaussian statistic, requiring an additional simulated image for each degree of freedom (typically 100). These issues are addressed in Worsley (2005).

Appendix

A Accuracy of the DLM P-value

In the continuous setting, P_{XEC} is known to be a very accurate approximation to (1). From (Taylor *et al.*, 2005)

$$P_{\mathrm{XEC}} = P \cdot \left(1 + O_e(e^{-u^2/2\sigma_c^2(Z)})\right)$$

for some critical variance $\sigma_c^2(Z)$ of the process Z where O_e is to be interpreted as "exponentially of order of". In other words, the relative error is exponentially small in u^2.

The expected number of local maxima above a given level is has the same exponential behaviour, so one might ask how accurate P_{DLM} is being the expected number of *discrete* local maxima above a given level. Here, we show that its relative error is also exponentially small. Our proof holds for any covariance structure as well as non-regular lattices, however we can only work out P_{DLM} explicitly only if the covariance function is a mixture of axis-aligned Gaussian kernels, in which case $P_{\mathrm{DLM}} = \tilde{P}_{\mathrm{DLM}}$. It should be emphasized that the results below refer to P_{DLM} and not the approximation $\tilde{P}_{\mathrm{DLM}}$.

THEOREM **A.1** *Let $(Z(x))_{x\in S}$ be a centered, zero mean Gaussian field on the finite set S and P_{DLM} be the expected number of discrete local maxima of Z on S based on a given neighbourhood structure (graph) G. Then,*

$$\liminf_{t\to\infty} -\frac{t^2}{2}\log\left(P_{\mathrm{DLM}} - P\right) = 1 + \frac{1}{\sigma^2_{c,DLM}(Z,G)} \tag{34}$$

where the critical variance is given by

$$\sigma^2_{c,\mathrm{DLM}}(Z,G) \triangleq \max_{x\in S}\max_{y\in S\setminus\widetilde{\mathcal{N}}} \frac{1+\rho(x,y)}{1-\rho(x,y)} \tag{35}$$

Above, ρ is the correlation function of Z and $\widetilde{\mathcal{N}} = \mathcal{N}\cup\{x\}$ is the set of neighbours of x (including x itself) determined by G.

Proof. The proof is based on the following equality, assuming that ties have probability 0

$$\begin{aligned} P &= \sum_{x\in S}\mathbb{P}\left(Z(x) > t, Z(x) > Z(y)\ \forall y\neq x\right) \\ &= \sum_{x\in S}\mathbb{E}\left(1_{\{Z(x)>t\}}\cdot 1_{\{Z(x)>Z(y)\ \forall y\neq x\}}\right). \end{aligned} \tag{36}$$

Therefore,

$$\begin{aligned} P_{\mathrm{DLM}} - P &= \sum_{x\in S}\mathbb{E}\left(1_{\{Z(x)>t\}}\cdot\left(1_{\{Z(x)>Z(y)\ \forall y\in\mathcal{N}\}} - 1_{\{Z(x)>Z(y)\ \forall y\neq x\}}\right)\right). \\ &= \sum_{x\in S}\mathbb{E}\left(1_{\{Z(x)>t\}}1_{\{Z(x)>Z(y)\ \forall y\in\mathcal{N}\}}\cdot\left(1 - 1_{\{Z(x)>Z(y)\ \forall y\notin\widetilde{\mathcal{N}}\}}\right)\right). \\ &= \sum_{x\in S}\mathbb{E}\left(1_{\{Z(x)>t\}}\cdot 1_{\{Z(x)>Z(y)\ \forall y\in\mathcal{N}\}}\cdot 1_{\cup_{y\in S\setminus\widetilde{\mathcal{N}}}\{Z(y)>Z(x)\}}\right) \\ &\leq \sum_{x\in S}\mathbb{E}\left(1_{\{Z(x)>t\}}\cdot 1_{\cup_{y\in S\setminus\widetilde{\mathcal{N}}}\{Z(y)>Z(x)\}}\right). \end{aligned} \tag{37}$$

In other words,

$$P_{\mathrm{DLM}} - P = \sum_{x\in S}\mathbb{P}\left(Z(x) > t, \exists\, y\in S\setminus\widetilde{\mathcal{N}} \text{ s.t. } Z(y) > Z(x)\right).$$

Now,

$$\{Z(y) > Z(x)\} = \left\{\frac{Z(y) - \rho(x,y)Z(x)}{1-\rho(x,y)} > Z(x)\right\}$$

and for each $y \neq x$

$$Z^x(y) \triangleq \frac{Z(y) - \rho(x,y)Z(x)}{1-\rho(x,y)}$$

is a centered Gaussian random variable independent of $Z(x)$ with variance

$$\sigma_r^2(x,y) = \frac{1+\rho(x,y)}{1-\rho(x,y)}.$$

Therefore, the random field $(Z^x(y))_{y\in S\setminus\widetilde{\mathcal{N}}}$ is independent of $Z(x)$ and

$$\begin{aligned} P_{\mathrm{DLM}} - P &= \sum_{x\in S} \mathbb{P}\left(Z(x) > t, \exists\, y \in S\setminus\widetilde{\mathcal{N}} \text{ s.t. } Z(y) > Z(x)\right) \\ &\leq \sum_{x\in S} \mathbb{P}\left(Z(x) > t, \exists\, y \in S\setminus\widetilde{\mathcal{N}} \text{ s.t. } Z(y) > t\right) \\ &= \sum_{x\in S}\sum_{y\in S\setminus\widetilde{\mathcal{N}}} \overline{\Phi}(t)\cdot\overline{\Phi}\left(t/\sigma_r(x,y)\right) \\ &\leq N^2 \frac{1}{t^2}\exp\left(-\left(1+\frac{1}{\sigma^2_{c,\mathrm{DLM}}(Z,G)}\right)\frac{t^2}{2}\right). \end{aligned} \tag{38}$$

The last inequality is quite conservative as the number of neighbours whose variance achieves

$$\sigma^2_{c,\mathrm{DLM}}(Z,G) = \max_{x\in S}\max_{y\in S\setminus\widetilde{\mathcal{N}}} \sigma_r^2(x,y)$$

is typically much smaller than N. □

The usefulness of this bound depends on a well-chosen G. For instance if G_{BON} is a graph with N nodes and no edges, then $P_{\mathrm{DLM}} = P_{\mathrm{BON}}$, and the theorem says exponential rate of decay is given by the maximal off-diagonal entry in the correlation matrix of $(Z(x))_{x\in S}$. When the parameter space S consists of points of a regularly sampled lattice in $\mathbb{R}^D$ with isotropic covariance function and edge lengths v then

$$\sigma^2_{c,\mathrm{DLM}}(Z, G_{\mathrm{BON}}) = \frac{1+\rho(v)}{1-\rho(v)}.$$

In order to beat Bonferroni then, one should choose G so that

$$\sigma^2_{c,\mathrm{DLM}}(Z,G) < \sigma^2_{c,\mathrm{DLM}}(Z, G_{\mathrm{BON}}).$$

If S is as above and G_{NN} is the nearest neighbours graph with $2D$ neighbours then the critical variance is achieved at the 2nd nearest neighbours and

$$\sigma^2_{c,\mathrm{DLM}}(Z, G_{\mathrm{NN}}) = \frac{1+\rho(\sqrt{2}v)}{1-\rho(\sqrt{2}v)}$$

which is generally smaller than $\sigma^2_{c,DLM}(Z, G_{\mathrm{BON}})$ particularly when ρ is Gaussian as will be the case below.

In summary, P_{DLM} is exponentially sharper than P_{BON}. However, unlike P_{XEC} in the continuous setting, not all terms of P_{DLM} are exponentially sharp in general. By terms, we are referring to the terms in the inclusion-exclusion expansion

$$\begin{aligned}
&\mathbb{P}\left(Z(x) > t, Z(x) > Z(y),\, \forall y \in \mathcal{N}\right) \\
&\quad = \mathbb{E}\left(1_{\{Z(x)>t\}}\left(1 - 1_{\cup_{y\in\mathcal{N}}\{Z(y)>Z(x)\}}\right)\right) \\
&\quad = \mathbb{P}(Z(x) > t) - \sum_{y\in\mathcal{N}} \mathbb{P}(Z(x) > t, Z(y) > Z(x)) + \ldots
\end{aligned} \tag{39}$$

The third order terms are of the form

$$\mathbb{P}(Z(x) > t, Z(x) > Z(y), Z(x) > Z(w)).$$

If the covariance function is Gaussian and w and y are chosen along different axes then the above probability is exponentially of order

$$\exp\left(-\left(1 + \frac{2(1-\rho(v))}{1+\rho(v)}\right) t^2/2\right)$$

and it can be easily verified that

$$\frac{1+\rho(v)}{2(1-\rho(v))} < \sigma^2_{c,DLM}(Z, G_{\text{NN}}).$$

Therefore, for v fixed, the third (and higher order) terms are exponentially smaller than the error $P_{\text{DLM}} - P$.

B Non-lattice sampled data

In some cases the random field Z is not sampled on a lattice, but rather a triangulated surface S. In this case, each point can have differing numbers of neighbours and there is no natural correspondence between opposite pairs of edges, as in the lattice. However, the results of Section A still hold: if we can evaluate P_{DLM} for the graph based in the triangulation, we will generally have a bound that is exponentially sharper than Bonferroni. This implicitly assumes that connectivity of nodes in the triangulated surface is at least approximately related to the covariance of Z, which would be the case of Z is thought of as the restriction of an isotropic random field to the triangulated surface S. The results of Section A show that, even when it is possible to evaluate P_{DLM} not all terms in the expansion (39) are exponentially sharp. In fact, the only terms that were exponentially sharp were the terms

$$\begin{aligned}
&\mathbb{P}(Z(x) > t, Z(y) > Z(x), y \in \mathcal{N}) \\
&\quad = \mathbb{P}(Z(x) > t, Z^x(y) > Z(x), y \in \mathcal{N}) \\
&\quad = \int_t^\infty \overline{\Phi}(z/\sigma_r(x,y))\phi(z)\, dz
\end{aligned} \tag{40}$$

This suggests an approximation of P_{DLM} including only the terms (40), i.e.

$$P_{\text{DLM}} \approx N\overline{\Phi}(t) - \sum_{x\in S}\sum_{y\in\mathcal{N}} \int_t^\infty \overline{\Phi}(z/\sigma_r(x,y))\phi(z)\, dz. \tag{41}$$

However, this would not be a bound on P. To remedy this we could include the next term in the inclusion exclusion expansion

$$\sum_{x\in S}\sum_{y,w\in\mathcal{N},y\neq w}\mathbb{P}\left(Z(x)>t,Z(y)>Z(x),Z(w)>Z(x)\right) \tag{42}$$

yielding a final upper bound

$$\begin{aligned}P_{\mathrm{DLM}}\leq N\overline{\Phi}(t)-\sum_{x\in S}\sum_{y\in\mathcal{N}}\int_t^\infty\overline{\Phi}(z/\sigma_r(x,y))\phi(z)\,dz\\ +\sum_{x\in S}\sum_{y,w\in\mathcal{N},y\neq w}\int_t^\infty\left(\int_{[z,\infty)^2}\frac{e^{-v\Sigma^{-1}_{x,y,w}v'/2}}{2\pi|\Sigma_{x,y,w}|^{1/2}}\,dv\right)\phi(z)\,dz,\end{aligned} \tag{43}$$

where

$$\Sigma_{x,y,w}=\begin{pmatrix}\sigma_r(x,y) & \frac{\rho(y,w)-\rho(y,x)\rho(w,x)}{(1-\rho(y,x))(1-\rho(w,x))}\\ \frac{\rho(y,w)-\rho(y,x)\rho(w,x)}{(1-\rho(y,x))(1-\rho(w,x))} & \sigma_r(x,w)\end{pmatrix}. \tag{44}$$

While this is indeed an upper bound for P_{DLM} and hence for P itself, the results of Section A suggest the added accuracy obtained by adding these terms is generally undetectable on an exponential scale. Unfortunately, as the mesh of the surface goes to 0, the approximation (43) does not approach P_{XEC} if $D>2$.

With highly convoluted triangulations such as the cortical surface, it may be the case that the nearest point to a given node is not a neighbour of the given node. In the results of (Taylor *et al.*, 2005), in the continuous setting, this is analogous to a "global overlap". To improve the accuracy of (43) it is likely a good idea to add such points to $\mathcal{N}$.

Acknowledgement

This work was partially funded by NSF grant DMS-0405970 and the Natural Sciences and Engineering Research Council of Canada.

References

Adler, R.J. (1981). *The geometry of random fields.* New York:Wiley.

Adler, R.J. & Taylor J. E. (2005). *Random fields and geometry.* In preparation. Preliminary versions of Chapters 1-13 available at `http://www-stat.stanford.edu/~jtaylo/papers/`

Adolphs, R., Gosselin, F., Buchanan, T.W., Tranel, D., Schyns, P. & Damasio, A.R. (2005). A mechanism for impaired fear recognition after amygdala damage. *Nature* **433**, 68-72.

Chauvin, A., Worsley, K.J., Schyns, P.G., Arguin, M. & Gosselin, F. (2004). A sensitive statistical test for smooth classification images. *Journal of Vision* **5**, 659-67.

Dohmen, K. (2000). Improved Bonferroni inequalities via union-closed set systems. *Journal of Combinatorial Theory Series A* **92**, 61-7.

Dohmen, K. (2003). Improved inclusion-exclusion identities and Bonferroni inequalities with reliability applications. *SIAM Journal on Discrete Mathematics* **16**, 156-71.

Dohmen, K. & Tittmann, P. (2004). Bonferroni-Galambos inequalities for partition lattices, *Electronic Journal of Combinatorics* **11**, R85.

Efron, B. (1997). The length heuristic for simultaneous hypothesis tests. *Biometrika* **84**, 143-157.

Forman, R (1998). Morse Theory for cell complexes. *Advances in Mathematics* **134**, 90-145.

Frackowiak, R.S.J., Friston, K.J., Frith, C., Dolan, R., Price, C.J., Zeki, S., Ashburner, J. & Penny, W.D. (2003). *Human Brain Function*. New York:Academic Press.

Galambos, J. & Simonelli, I. (1996). Bonferroni-type inequalities with applications. *Springer Series in Statistics, Probability and Its Applications* New York: Springer-Verlag.

Gosselin, F. & Schyns, P.G. (2001). Bubbles, a technique to reveal the use of information in recognition. *Vision Research* **41**, 2261-71.

Grable, D.A. (1993). Sharpened Bonferroni inequalities. *Journal of Combinatorial Theory, Series B* **57**, 131-7.

Hoppe, F.M. (1993). Beyond inclusion-and-exclusion: natural identities for P[exactly t events] and P[at least t events] and resulting inequalities. *International Statistical Review* **61**, 435–446.

Hothorn, T. & Lausen, B. (2003). On the exact distribution of maximally selected rank statistics. *Computational Statistics and Data Analysis* **43**, 121-37.

Hunter, D. (1976). An upper bound for the probability of a union. *Journal of Applied Probability* **13**, 597-603.

Kounias, E.G. (1968). Bounds for the probability of a union of events, with applications. *Annals of Mathematical Statistics* **39**, 2154-8.

Kwerel, S.M. (1975). Most stringent bounds on aggregated probabilities of partially specified dependent probability systems. *Journal of the American Statistical Association* **70**, 472-9.

Lewiner, T., Lopes, H. & Tavares, G. (2004). Applications of Forman's discrete Morse Theory to topology visualization and mesh compression. *IEEE Transactions on Visualization and Computer Graphics* **10**, 499–508.

Naiman, D.Q. & Wynn, H.P. (1992). Inclusion-exclusion-Bonferroni identities and inequalities for discrete tube-like problems via Euler characteristics. *Annals of Statistics* **20**, 43–76.

Naiman, D.Q. & Wynn, H.P. (1997). Abstract tubes, improved inclusion-exclusion identities and inequalities and importance sampling. *Annals of Statistics* **25**, 1954–83.

Sarkar, S.K. (1998). Some probability inequalities for ordered MTP_2 random variables, a proof of the Simes conjecture. *Annals of Statistics* **26**, 494-504.

Siegmund, D.O & Worsley, K.J. (1995). Testing for a signal with unknown location and scale in a stationary Gaussian random field. *Annals of Statistics* **23**, 608–39.

Simes, R.J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika* **73**, 751–4.

Schyns, P.G., Bonnar, L. & Gosselin, F. (2002). Show me the features! Understanding recognition from the use of visual information. *Psychological Science* **13,** 402-9.

Taylor, J.E. & Adler, R.J. (2003). Euler characteristics for Gaussian fields on manifolds. *Annals of Probability* **31**, 533–63.

Taylor, J.E., Takemura, A. and Adler, R.J. (2005). Validity of the expected Euler characteristic heuristic. *Annals of Probability*, **33**, 1362–1396.

Tomescu, I. (1986). Hypertrees and Bonferroni inequalities, *Journal of Combinatorial Theory Series B* **41**, 209-17.

Worsley, K.J. (1982). An improved Bonferroni inequality and applications. *Biometrika* **69**, 297–302.

Worsley, K.J. (1995a). Boundary corrections for the expected Euler characteristic of excursion sets of random fields, with an application to astrophysics. *Advances in Applied Probability* **27**, 943–59.

Worsley, K.J. (1995b). Estimating the number of peaks in a random field using the Hadwiger characteristic of excursion sets, with applications to medical images. *Annals of Statistics* **23**, 640–69.

Worsley, K.J., Liao, C., Aston, J.A.D., Petre, V., Duncan, G.H., Morales, F. & Evans, A.C. (2002). A general statistical analysis for fMRI data. *NeuroImage* **15**, 1–15.

Worsley, K.J. (2003). Detecting activation in fMRI data. *Statistical Methods in Medical Research* **12**, 401–18.

Worsley, K.J. (2005). An improved theoretical P-value for SPMs based on discrete local maxima. *NeuroImage*, available on-line.