TESTING FOR SIGNALS WITH UNKNOWN LOCATION AND SCALE IN A χ^2 RANDOM FIELD, WITH AN APPLICATION TO fMRI

Keith J. Worsley, * *McGill University*

November 26, 2001

Abstract

Siegmund & Worsley (1995) considered the problem of testing for signals with unknown location and scale in a Gaussian random field defined on $\Re^N$. The test statistic was the maximum of a Gaussian random field in an $N+1$ dimensional 'scale space', N dimensions for location and 1 dimension for the scale of a smoothing filter. Scale space is identical to a continuous wavelet transform with a kernel smoother as the wavelet, though the emphasis here is on signal detection rather than image compression or enhancement. Two methods were used to derive an approximate null distribution for $N = 2$ and $N = 3$: one based on the method of volumes of tubes, the other based on the expected Euler characteristic of the excursion set. The purpose of this paper is two-fold: to show how the latter method can be extended to higher dimensions, and to apply this more general result to χ^2 fields. The result of Siegmund & Worsley (1995) then follows as a special case. In this paper the results are applied to the problem of searching for activation in brain images obtained by functional magnetic resonance imaging (fMRI).

EULER CHARACTERISTIC; DIFFERENTIAL TOPOLOGY; INTEGRAL GEOMETRY; IMAGE ANALYSIS; WAVELET THRESHOLDING; ADAPTIVE FILTERING; MULTI-SCALE; MULTI-RESOLUTION

AMS 1991 SUBJECT CLASSIFICATION: PRIMARY 60G60, 62M09; SECONDARY 60D05, 52A22

*Postal address: Department of Mathematics and Statistics, McGill University, 805 ouest, rue Sherbrooke, Montréal, Qu'ebec, Canada H3A 2K6; E-mail: worsley@math.mcgill.ca. Research supported by the Natural Sciences and Engineering Research Council of Canada, and the Fonds pour la Formation des Chercheurs et l'Aide à la Recherche de Québec.

1 Introduction

1.1 Motivation: functional brain mapping

The main motivation for this paper comes from new methods for functional brain mapping by positron emission tomography (PET), and more recently, functional magnetic resonance imaging (fMRI). This latter imaging method generates 3-D images of BOLD (blood oxygenation level dependent) response, a measure of brain activity, while a subject is performing a task or receiving a stimulus (see Figure 5 for an example in 2-D). Our interest is to test for a correlation between the stimulus and the BOLD response at each point (pixel in 2-D, voxel in 3-D) in the image, then to decide which regions are truly activated and which are just background noise. This is done by thresholding an image of test statistics, often modeled as a Gaussian random field, to reveal only those regions where the changes are 'significant' (Friston *et al.*, 1994; Worsley & Friston, 1995). Usually this threshold is set so that the type I error of finding any activation in the unactivated parts of the image is controlled to be 0.05, say. Since we expect activation to be confined to only a small part of the image, a conservative procedure is to set the threshold based on a null hypothesis of no activation anywhere in the image. Our main interest is to find such a threshold, or equivalently, to find an approximation to the tail probability or P-value of the maximum of the image of test statistics inside a fixed region (called the search region) when the entire image is just bacground noise (see Worsley, 1996, for a non-technical overview).

Often the N-dimensional images are spatially smoothed before analysis by convolution with a filter f to enhance the signal to noise ratio. The most common filter is a Gaussian shaped filter

$$f(\mathbf{t}) = \pi^{-N/4} \exp(-||\mathbf{t}||^2/2), \quad \mathbf{t} \in \Re^N, \tag{1.1}$$

normalized so that $\int f(\mathbf{t})^2 d\mathbf{t} = 1$. The motivation for this comes from the Matched Filter Theorem of signal processing (see e.g. Rosenfeld & Kac, 1982), which states that signal added to white noise is best detected by smoothing with a filter whose shape matches that of the signal. Thus to best detect a 20mm wide signal, a 20mm wide filter should be used. The problem is, of course, that the extent of the signal is usually unknown. In practice, PET and fMRI data are often smoothed with a 10-20mm wide filter as a first guess.

1.2 Scale space and the continuous wavelet transform

The signal width is usually unknown, so it is natural to consider searching over filter width or scale as well as location, that is, to use a filter

$$s^{-N/2} f(\mathbf{t}/s) \tag{1.2}$$

with scale s varying over a predetermined interval $[s_1, s_2]$, again normalized so that the integral of the squared filter is one. This adds an extra dimension to the search space, called scale space. It was first introduced in the brain mapping literature by Poline & Mazoyer (1994ab). A hierarchical multi-scale detection method, similar in spirit to a wavelet decomposition, was proposed by Crivello *et al.* (1995). Siegmund & Worsley (1995) provided a theoretical statistical treatment of scale space analysis, showing that it is closely related

to the problem of detecting a signal with unknown scale, location and amplitude added to a white noise background. They showed that the scale space maximum is the likelihood ratio test statistic for the presence of such a signal, and they gave results for the P-value of this test statistic in 2-D and 3-D. One of the objectives of this paper is to extend these results to any number of dimensions.

The concept of scale space has been explored in many image processing tasks (see Lindeberg, 1994). The difference with what is proposed here is that the filter used in image processing is usually normalized to preserve the *mean*, that is, the filter at scale s is

$$s^{-N} f(\mathbf{t}/s).$$

In our case, likelihood principles dictate that the filter should be normalized to preserve the *variance* of the filtered white noise background, which leads to (1.2) instead. The effect of this on a Gaussian shaped signal smoothed with a Gaussian shaped filter is as follows. In the image processing context, the peak amplitude of the signal always decreases as the filter width is increased; in the case here, the peak amplitude of the signal first increases, reaching a maximum at the width of the signal, then decreases when the filter width is increased beyond that of the signal (a simple illustration of the Matched Filter Theorem). This is because the normalization used here is designed to best detect a signal in a white noise background. In image processing there is usually no random component, and scale space is used to characterize features in the image that are observed without noise.

The normalization used here is the same as that employed by the continuous wavelet transform (Daubechies, 1992). The main difference is that the filter or mother wavelet is usually chosen so that $\int f(\mathbf{t}) d\mathbf{t} = 0$, as well $\int f(\mathbf{t})^2 d\mathbf{t} = 1$, for example the multi-dimensional Marr or 'Mexican hat' wavelet

$$f(\mathbf{t}) = \pi^{-N/4} [4N/(N+2)]^{1/2} (1 - ||\mathbf{t}||^2/N) \exp(-||\mathbf{t}||^2/2). \tag{1.3}$$

The main difference is that the wavelet transform seeks to represent the image at different scales; here the aim is to detect localized image components at an unknown scale. Thresholding the scale space image is analogous to wavelet thresholding in the discrete case. Thus the methods in this paper could be used to set the threshold for the continuous wavelet transform so that the probability of detecting wavelets coefficients above the threshold is controlled at say $\alpha = 0.05$ when the data is white noise.

Finally, Shafie (1998) has considered the case of rotating as well as scaling an elliptically contoured filter of the form

$$\det(\mathbf{\Sigma})^{-\frac{1}{2}} f(\mathbf{\Sigma}^{-1}\mathbf{t})$$

where $\mathbf{\Sigma}$ is an arbitrary symmetric $N \times N$ matrix. The maximum of this 'rotation space' field is sensitive at detecting elliptically contoured signals. Shafie (1998) found an approximation to its P-value in $N = 2$ dimensions using methods similar to those in this paper.

1.3 χ^2 fields

The χ^2 field is a natural extension when several independent Gaussian images are available, each one containing the same width signal at the same location but with possibly differing

amplitudes. The χ^2 field with ν degrees of freedom is then defined as the sum of squares of ν i.i.d. stationary Gaussian random fields with zero mean and unit variance (Adler, 1981). For example, in a typical PET study, k different stimuli are used in an experiment, and the aim is to detect whether some or all of these stimuli produce different localized BOLD response (see for example Fletcher *et al.*, 1996). A suitable test statistic is the between stimuli sum of squares, which under appropriate assumptions is a multiple of a χ^2 field with $k-1$ degrees of freedom under the null hypothesis of equal BOLD response. In another study, Büchel *et al.* (1996) have proposed a quadratic BOLD response to stimulus intensity. Under appropriate assumptions the regression sum of squares is a multiple of a χ^2 field with two degrees of freedom under the null hypothesis of constant BOLD response. In a further study, Bullmore *et al.* (1996) used the power spectrum of fMRI time series at the same frequency as the periodic stimulus to detect the stimulus when the delay of the hemodynamic response is unknown. Again under under appropriate assumptions the power spectrum is a multiple of a χ^2 field with 2 degrees of freedom under the null hypothesis of no BOLD response. Here the two Gaussian fields are the real and imaginary parts of the Fourier transform of the fMRI time series at that frequency. This method will be developed further in section 5.

In all these examples, each Gaussian image was smoothed with the same width filter to enhance the signal to noise ratio. Extending the ideas above, it is natural to search over filter widths as well as over locations, to produce a scale space χ^2 field. Section 4.1 shows that the maximum of the scale space χ^2 field is the likelihood ratio test statistic for detecting the signal when the location, scale and amplitudes are unknown. Note that maximum likelihood principles dictate that the Gaussian images should be smoothed before taking the sum of squares, rather than smoothing the χ^2 field itself (see section 4.2).

1.4 Main objectives

The main purpose of this paper is to find an accurate approximation to the P-value of the maximum of a scale space χ^2 random field. The main tool used is the Euler characteristic (EC) of the excursion set of a random field. The relationship to the P-value of the maximum is as follows. The excursion set is the set of points in the image which exceed a fixed threshold, and the EC of the excursion set counts the number of connected components of the excursion set, minus the number of 'holes'. For high thresholds the holes are a rare occurrence and the EC approximates the number of local maxima that exceed the threshold. For even larger thresholds near the global maximum the EC takes the value 0 if the global maximum is less than the threshold and 1 if it exceeds it, so that the expected EC approximates the exceedence probability or P-value of the global maximum (Hasofer, 1978).

There are only asymptotic results for the P-value of the global maximum and the expected number of local maxima above a high threshold, but the EC has a simple *exact* expectation at any threshold when the image is stationary (no signal). This has been worked out for Gaussian fields (Adler, 1981) and extended to χ^2, t and F fields (Worsley, 1994) and Hotelling's T^2 fields (Cao & Worsley, 1999). Worsley (1995ab) give a boundary correction for when the excursion set touches the boundary of the search region. Adler (2000) has recently shown that this correction gives an even more accurate approximation to the P-value. Moreover there is a very simple algorithm, given in Adler (1981), for calculating the EC from sample data so that the observed EC and the expected EC can be plotted against the threshold

level to detect departures at any level. Such methods have been used for PET data (Worsley *et al.*, 1992) and in astrophysics for galaxy density (Vogeley *et al.*, 1994) and anomalies in the cosmic microwave background radiation, thought to originate from the creation of the universe just after the 'big bang' (Torres, 1994; Worsley, 1995b).

The theory of this paper will draw on results from all the above-mentioned sources, proceeding as follows:

- **Theorem 1** Use a differential topology approach (Worsley, 1995b) to obtain a point-set representation for the EC of the excursion set inside an arbitrary subset C of $\Re^N$ with a smooth boundary, then use an integral geometry approach (Worsley, 1995a) to extend this to an $(N+1)$-dimensional 'prism' which is the Cartesian product of C with an interval I in the extra dimension;
- **Theorem 2** Find the expectation of the EC for a stationary field inside the prism;
- **Theorem 3** Simplify the expected EC for an isotropic field to give a sum of j-dimensional EC intensities multiplied by measures of the j-dimensional curvature of the boundary of C, $j = 0, \ldots, N$;
- **Theorem 4** Apply this to the χ^2 field in which C is the search region for the location of the filter, and the extra dimension is the scale of the filter. The resulting expectations are evaluated using a lemma which gives representations of a χ^2 field and its derivatives, found using methods from Worsley (1994).

2 The excursion set characteristics

2.1 Euler characteristic

Let $X(\mathbf{t})$, $\mathbf{t} = (t_1, \ldots, t_N)' \in C \subset \Re^N$ be a random field and let $A_b = \{\mathbf{t} \in C \ : \ X(\mathbf{t}) \geq b\}$ be the excursion set above the threshold b. Let $\chi(A)$ be the Euler or Euler-Poincaré characteristic (EC) of a set A; the aim is to find the expectation of $\chi(A_b)$.

A crucial step in deriving statistical properties of excursion set characteristics is to obtain a point-set representation which expresses the characteristic in terms of *local* properties of the excursion set rather than *global* properties such as connectedness. Worsley (1995b) uses differential topology and Morse theory to derive the following point-set representation.

Let $X = X(\mathbf{t})$, let $\dot{\mathbf{X}} = \partial X/\partial \mathbf{t}$ be the gradient N-vector of X, and let $\ddot{\mathbf{X}} = \partial^2 X/\partial \mathbf{t} \partial \mathbf{t}'$ be the $N \times N$ Hessian matrix of X. The boundary ∂C of C must be twice differentiable, and let $\mathbf{c}$ be the $(N-1) \times (N-1)$ inside curvature matrix of ∂C. At a point $\mathbf{t} \in \partial C$, let $\dot{X}_\perp$ be the gradient of X in the direction of the inside normal to ∂C, let $\dot{\mathbf{X}}_{\mathrm{T}}$ be the gradient $(N-1)$-vector in the tangent plane to ∂C, let $\ddot{\mathbf{X}}_{\mathrm{T}}$ be the $(N-1) \times (N-1)$ Hessian matrix in the tangent plane to ∂C. Let $\mathrm{sign}(z) = z/|z|$ if $z \neq 0$ and zero otherwise. A logical expression in parentheses takes the value one if it is true and zero if it is false (Knuth, 1992). Then under suitable regularity conditions given in Adler (1981, Theorem 5.2.2), Worsley (1995b) shows that, with probability one,

$$\chi(A_b) \ = \ \sum_{\mathbf{t} \in C} (X \geq b) \mathrm{sign}[\det(-\ddot{\mathbf{X}})] (\dot{\mathbf{X}} = \mathbf{0})$$

$$+ \sum_{\mathbf{t}\in\partial C} (X \geq b)(\dot{X}_\perp < 0)\text{sign}[\det(-\ddot{\mathbf{X}}_\mathrm{T} - \dot{X}_\perp \mathbf{c})](\dot{\mathbf{X}}_\mathrm{T} = \mathbf{0}). \tag{2.1}$$

2.2 Hadwiger characteristic

A related characteristic, the Hadwiger characteristic (HC), is defined on different classes of sets than the EC, but it equals the EC within the domain of definition of both. The advantage for us is that the HC has an iterative definition, given below, which is very convenient for extending the point-set representation into one higher dimension.

A compact subset B of $\Re^N$ is called a *basic* if the intersections of all k-planes with B is simply connected, $k = 1, \ldots, N$. A set A is a *basic complex* if it can be represented as the union of a finite number of basics such that the intersection of any of these is again a basic (see Adler, 1981, page 71). If $X(\mathbf{t})$ is sufficiently regular, as defined by Adler (1981, Theorem 5.2.2), then Adler (1981, Theorem 4.3.1) shows that the excursion set A_b is almost surely a basic complex.

The *Hadwiger characteristic* $\psi(A)$ of a basic complex $A \subset C$ is defined iteratively as follows (see Adler, 1981, Theorem 4.2.2). For $N = 1$, let $\psi(A)$ be the number of disjoint intervals in A. For $N > 1$, let

$$\psi(A) = \sum_u [\psi(A \cap \mathcal{E}_u) - \psi(A \cap \mathcal{E}_{u^-})],$$

where $\mathcal{E}_u = \{\mathbf{t} \in C \ : \ t_N = u\}$ and

$$\psi(A \cap \mathcal{E}_{u^-}) = \lim_{v \uparrow u} \psi(A \cap \mathcal{E}_v).$$

2.3 Combining the characteristics

The definition of the HC is now used to obtain a point set representation for the excursion set of a random field in one higher dimension inside the 'prism' $C \times I \subset \Re^{N+1}$, where $I = [s_1, s_2] \subset \Re$ is an interval. From now on let $X(\mathbf{t}, s)$, $\mathbf{t} \in C$, $s \in I$ be a random field in $\Re^{N+1}$, and let $B_b = \{(\mathbf{t}, s) \in C \times I \ : \ X(\mathbf{t}, s) \geq b\}$ be the excursion set above b. Define the derivatives $\dot{\mathbf{X}}$, $\ddot{\mathbf{X}}$, $\dot{X}_\perp$, $\dot{\mathbf{X}}_\mathrm{T}$ and $\ddot{\mathbf{X}}_\mathrm{T}$ as before. In addition, let $\dot{X}_s = \partial X / \partial s$.

Theorem 1 *If X satisfies the regularity conditions given by Adler (1981, Theorem 5.2.2), then with probability one,*

$$\begin{aligned}
\chi(B_b) \ = \ & \sum_{\mathbf{t}\in C, s\in I} \Big\{ [(s = s_1)(X \geq b) + (\dot{X}_s > 0)(X = b)] \\
& \times \ \text{sign}[\det(-\ddot{\mathbf{X}})](\dot{\mathbf{X}} = \mathbf{0}) \Big\} \\
& + \ \sum_{\mathbf{t}\in\partial C, s\in I} \Big\{ [(s = s_1)(X \geq b) + (\dot{X}_s > 0)(X = b)] \\
& \times \ (\dot{X}_\perp < 0)\text{sign}[\det(-\ddot{\mathbf{X}}_\mathrm{T} - \dot{X}_\perp \mathbf{c})](\dot{\mathbf{X}}_\mathrm{T} = \mathbf{0}) \Big\}.
\end{aligned} \tag{2.2}$$

Proof. Under the regularity conditions, B_b is almost surely a basic complex and the EC and the HC of B_b almost surely agree. Applying the iterative definition of the HC to B_b gives, with probability one,

$$\chi(B_b) = \sum_s [\chi(B_b \cap \mathcal{E}_s) - \chi(B_b \cap \mathcal{E}_{s^-})],$$

where $\mathcal{E}_s = \{(\mathbf{t}, s) \ : \ \mathbf{t} \in C\}$ and

$$\psi(B_b \cap \mathcal{E}_{s^-}) = \lim_{v \uparrow s} \psi(B_b \cap \mathcal{E}_v).$$

Now $\chi(B_b \cap \mathcal{E}_s)$ is just the EC of $X(\mathbf{t}, s)$ restricted to fixed s, in other words, the EC of $X(\mathbf{t})$ as defined above. This differs from $\chi(B_b \cap \mathcal{E}_{s^-})$ in two ways: when $s = s_1$, the 'base' of the prism, and when $s > s_1$, the interior of the prism.

When $s = s_1$, $\chi(B_b \cap \mathcal{E}_{s^-})$ is zero and so the contributions to $\chi(B_b)$ from $s = s_1$ are the same as the EC of the base, which is given by (2.1) at $s = s_1$. This accounts for the $(s = s_1)(X \geq b)$ terms in (2.2).

When $s > s_1$, $\chi(B_b \cap \mathcal{E}_s)$ differs from $\chi(B_b \cap \mathcal{E}_{s^-})$ only when $X(\mathbf{t}, s) = b$, the boundary of the excursion set, and when $\mathbf{t}$ is a turning point of X, that is $\dot{\mathbf{X}} = \mathbf{0}$. If the EC changes just below s then the interior of the excursion set must be above s, that is, $\dot{X}_s > 0$. Extending the arguments of Adler (1981, page 84), the change in the EC is $\text{sign}[\det(-\ddot{\mathbf{X}})]$. Similar arguments apply to the case when $\mathbf{t}$ is on the boundary of C. These two cases together contribute to the $(\dot{X}_s > 0)(X = b)$ terms in (2.2). ■

2.4 Expectations

The expectations of the point set representation given by Theorem 1 can be found using the methods used to prove Theorem 5.1.1 of Adler (1981, p. 95); see also Worsley (1995ab). Let $x^+ = x$ if $x > 0$ and 0 otherwise. Then the result is given by the following theorem.

Theorem 2 *If $X(\mathbf{t}, s)$ is stationary in $\mathbf{t}$ for fixed s then*

$$\begin{aligned}
\mathrm{E}[\chi(B_b)] &= |C| \Big\{ \Big[\mathrm{E}[(X \geq b) \det(-\ddot{\mathbf{X}}) \mid \dot{\mathbf{X}} = \mathbf{0}] \theta(\mathbf{0}) \Big]_{s=s_1} && (2.3) \\
&+ \int_{s_1}^{s_2} \mathrm{E}[\dot{X}_s^+ \det(-\ddot{\mathbf{X}}) \mid \dot{\mathbf{X}} = \mathbf{0}, X = b] \phi(\mathbf{0}, b) ds \Big\} && (2.4) \\
&+ \int_{\partial C} \Big\{ \Big[\mathrm{E}[(X \geq b)(\dot{X}_\perp < 0) \det(-\ddot{\mathbf{X}}_\mathrm{T} - \dot{X}_\perp \mathbf{c}) \mid \dot{\mathbf{X}}_\mathrm{T} = \mathbf{0}] \theta_\mathrm{T}(\mathbf{0}) \Big]_{s=s_1} && (2.5) \\
&+ \int_{s_1}^{s_2} \mathrm{E}[\dot{X}_s^+ (\dot{X}_\perp < 0) \det(-\ddot{\mathbf{X}}_\mathrm{T} - \dot{X}_\perp \mathbf{c}) \mid \dot{\mathbf{X}}_\mathrm{T} = \mathbf{0}, X = b] \phi_\mathrm{T}(\mathbf{0}, b) ds \Big\} d\mathbf{t}, && (2.6)
\end{aligned}$$

where $\phi(\cdot, \cdot)$, $\theta(\cdot)$, $\phi_\mathrm{T}(\cdot, \cdot)$ and $\theta_\mathrm{T}(\cdot)$ are the densities of $(\dot{\mathbf{X}}, X)$, $\dot{\mathbf{X}}$, $(\dot{\mathbf{X}}_\mathrm{T}, X)$ and $\dot{\mathbf{X}}_\mathrm{T}$, respectively.

2.5 Isotropic fields

It is possible to find a much simplified result for the expectations of Theorem 2 when the field is isotropic in $\mathbf{t}$ for fixed s. First let $X_{|j}(t_1, \ldots, t_j) = X(t_1, \ldots, t_j, 0, \ldots, 0)$ be the restriction

of X to a j-dimensional subspace of $\mathbf{t}$, $j = 0, \ldots, N$. Then $\dot{\mathbf{X}}_{|j}$ is the vector of the first j components of $\dot{\mathbf{X}}$, and $\ddot{\mathbf{X}}_{|j}$ is the $j \times j$ matrix of the first j rows and columns of $\ddot{\mathbf{X}}$, by an obvious extension of the notation. Then define the j-dimensional *EC intensity* as

$$\rho_j(b) = \left[\mathrm{E}[(X \geq b)\det(-\ddot{\mathbf{X}}_{|j}) \mid \dot{\mathbf{X}}_{|j} = \mathbf{0}]\theta_{|j}(\mathbf{0})\right]_{s=s_1} \tag{2.7}$$

$$+ \int_{s_1}^{s_2} \mathrm{E}[\dot{X}_s^+ \det(-\ddot{\mathbf{X}}_{|j}) \mid \dot{\mathbf{X}}_{|j} = \mathbf{0}, X = b]\phi_{|j}(\mathbf{0}, b)ds, \tag{2.8}$$

where $\phi_{|j}(\cdot,\cdot)$ and $\theta_{|j}(\cdot)$ are the joint densities of $(\dot{\mathbf{X}}_{|j}, X)$ and $\dot{\mathbf{X}}_{|j}$, respectively. Using this notation, the two terms (2.3) and (2.4) of $\mathrm{E}[\chi(B_b)]$ equal $|C|\rho_N(b)$. Thus $\rho_j(b)$ can be regarded as the intensity of the EC point-set process per unit volume of $\Re^j$. In different contexts, Worsley (1994,1995ab) and Siegmund & Worsley (1995) refer to $\rho_j(b)$ as the EC *density*; since it is the expectation of a point process taking the values ± 1 it is perhaps better to refer to $\rho_j(b)$ as an intensity.

The next result shows that if $X(\mathbf{t}, s)$ is isotropic in $\mathbf{t}$ for fixed s then the expected EC of Theorem 2 depends on the random field X and the threshold b only through the EC intensities $\rho_j(b)$. To do this, some more definitions are needed. For a square $n \times n$ matrix $\mathbf{M}$ let $\mathrm{detr}_j(\mathbf{M})$ be the sum of all $j \times j$ principal minors of $\mathbf{M}$, so that $\mathrm{detr}_n(\mathbf{M}) = \det(\mathbf{M})$, $\mathrm{detr}_1(\mathbf{M}) = \mathrm{trace}(\mathbf{M})$ and define $\mathrm{detr}_0(\mathbf{M}) = 1$. Finally let $a_j = 2\pi^{j/2}/\Gamma(j/2)$ be the surface area of a unit $(j-1)$-sphere in $\Re^j$. Then the expected EC is given by the following.

Theorem 3 *If the field X is isotropic in $\mathbf{t}$ for fixed s then*

$$\mathrm{E}[\chi(B_b)] = \sum_{j=0}^{N} V_j(C)\rho_j(b), \tag{2.9}$$

where $V_j(C)$ is the j-dimensional intrinsic volume *(Klain & Rota, 1997) of C, given by*

$$V_j(C) = \begin{cases} \int_{\partial C} \mathrm{detr}_{N-1-j}(\mathbf{c})d\mathbf{t}/a_{N-j}, & j = 0, \ldots, N-1, \\ |C|, & j = N. \end{cases}$$

Proof. The two terms (2.3) and (2.4) of Theorem 2 give the $j = N$ term of (2.9) by the definition of $\rho_N(b)$, as already noted above. The term (2.5) of Theorem 2 matches the term (2.7) of $\rho_j(b)$, by Theorem 4 of Worsley (1995b). The match between the term (2.6) of Theorem 2 and the term (2.8) of $\rho_j(b)$ can be found by methods very similar to the proof of Theorem 4 of Worsley (1995b), so the details are omitted to save space. ■

Remark 1. Full isotropy is not needed for this Theorem. Inspection of the proof shows that all that is required is that the distribution of the first and second derivatives at every point should be rotationally invariant.

Remark 2. By the Gauss-Bonnet Theorem, the first intrinsic volume is $V_0(C) = \chi(C)$ so that the first term in (2.9) becomes $\chi(C)\mathrm{P}\{X \geq b\}$. The $(N-1)$-dimensional intrinsic volume $V_{N-1}(C)$ is half the $(N-1)$-dimensional surface area of C, and the N-dimensional intrinsic volume $V_N(C)$ is the volume or Lebesgue measure of C. For C a ball of radius r,

$$\int_{\partial C} \mathrm{detr}_{N-1-j}(\mathbf{c})d\mathbf{t} = \binom{N-1}{j} a_N r^j,$$

and so, after some simplification,

$$V_j(C) = \frac{\Gamma\left(\frac{N+1}{2}\right)\pi^{\frac{j+1}{2}}}{\Gamma\left(\frac{N+1-j}{2}\right)\Gamma\left(\frac{j+1}{2}\right)\Gamma\left(\frac{j+2}{2}\right)}\, r^j. \tag{2.10}$$

For $N = 3$ dimensions Worsley (1995a) gives a method for approximating the intrinsic volumes of an arbitrary search region sampled on a rectilinear lattice.

Remark 3. If $X(\mathbf{t}, s)$ is jointly isotropic in $(\mathbf{t}, s)$, then

$$\rho_j(b) = \rho_j^{\mathrm{DT}}(b) + (s_2 - s_1)\rho_{j+1}^{\mathrm{DT}}(b)$$

where

$$\rho_j^{\mathrm{DT}}(b) = \mathrm{E}[\dot{X}_s^+ \det(-\ddot{\mathbf{X}}_{|j-1}) \mid \dot{\mathbf{X}}_{|j-1} = \mathbf{0}, X = b]\phi_{|j-1}(\mathbf{0}, b),$$

is the intensity of the DT characteristic of Adler(1981, Definition 4.4.1). This simplification follows from Worsley (1995b, Theorem 3), which shows that

$$\rho_j^{\mathrm{DT}}(b) = \mathrm{E}[(X \geq b)\det(-\ddot{\mathbf{X}}_{|j}) \mid \dot{\mathbf{X}}_{|j} = \mathbf{0}]\theta_{|j}(\mathbf{0}).$$

The former expression is usually easier to evaluate since it involves one less integration.

3 The scale space χ^2 field

In this section the quite general results above will be applied to a χ^2 field in 'scale space', that is, $\mathbf{t}$ will represent location in $\Re^N$ and s will represent the scale' of the filter used to smooth white noise to produce the Gaussian components of the χ^2 field.

Let $W_1, \ldots, W_\nu$ be independent and identically distributed Brownian sheets ('white noise') in $\Re^N$. Let $f : \Re^N \to \Re$ satisfy

$$\int f(\mathbf{t})^2 d\mathbf{t} = 1.$$

Then the Gaussian processes

$$Z_i(\mathbf{t}, s) = s^{-N/2} \int f[(\mathbf{h} - \mathbf{t})/s] dW_i(\mathbf{h}) \tag{3.1}$$

all have zero mean and unit variance, $i = 1, \ldots, \nu$. It can be shown that $Z_i(\mathbf{t}, s)$ is stationary in $\mathbf{t}$ for fixed s, and in $\log s$ for fixed $\mathbf{t}$ (but not jointly in $(\mathbf{t}, \log s)$). Then define the χ^2 field X with ν degrees of freedom as

$$X(\mathbf{t}, s) = \sum_{i=1}^{\nu} Z_i(\mathbf{t}, s)^2.$$

Note that X is χ^2 at every point $(\mathbf{t}, s)$, and that the smoothing is applied to the Gaussian components, not the χ^2 field itself. Some motivation for this will be given later in sections 4.1 and 4.2.

To apply the results of Theorems 1-3, some conditions on the filter f are required to ensure that X is sufficiently regular. Following Adler (1981, Theorem 7.1.1), these are essentially

the same as those imposed on $Z_1, \ldots, Z_\nu$ to ensure that they are sufficiently regular. This condition is satisfied if, for example, all third derivatives of Z_i have finite variance, which in turn is assured if the integral of the product of any pair of third derivatives of $f(\mathbf{t})$ times sixth degree polynomials in components of $\mathbf{t}$ is finite. This latter condition is met, for example, by the Gaussian filter (1.1) and the Marr wavelet (1.3).

3.1 Joint distribution of the derivatives

In order to evaluate the expectations in Theorem 2, and their simplifications for isotropic fields given in Theorem 3, the joint distribution of the derivatives of X up to second order is needed. This is best achieved by representing the derivatives as functions of independent random variables, a method used in Worsley (1994). The results are simpler if it is assumed from now on that the filter is symmetric, that is $f(\mathbf{t}) = f(-\mathbf{t})$.

The notation $\text{Normal}_d(\mu, \mathbf{\Sigma})$ represents the multivariate normal distribution on $\Re^d$ with mean μ and variance $\mathbf{\Sigma}$, χ^2_m represents the χ^2 distribution with m degrees of freedom, and $\text{Wishart}_d(\mathbf{\Sigma}, m)$ represents the Wishart distribution of a $d \times d$ matrix with expectation $m\mathbf{\Sigma}$ and degrees of freedom m. Let

$$\kappa = \int [\mathbf{t}'\dot{\mathbf{f}} + (N/2)f]^2 d\mathbf{t}, \qquad \mathbf{\Lambda} = \int \dot{\mathbf{f}}\dot{\mathbf{f}}' d\mathbf{t},$$

where $\dot{\mathbf{f}} = \partial f / \partial \mathbf{t}$. For the special case of the Gaussian filter (1.1)

$$\kappa = N/2, \qquad \mathbf{\Lambda} = \mathbf{I}_N/2, \tag{3.2}$$

where $\mathbf{I}_N$ is the $N \times N$ identity matrix. For the Marr wavelet (1.3),

$$\kappa = (N+4)/2, \qquad \mathbf{\Lambda} = \mathbf{I}_N(N+4)/(2N). \tag{3.3}$$

Lemma 1 *The first two derivatives of the χ^2 field X can be written in terms of independent random variables as follows, where the equalities are equalities in law:*

$$\begin{aligned}
\dot{X}_s &= 2s^{-1}(X\kappa)^{\frac{1}{2}}Y, \\
\dot{\mathbf{X}} &= 2s^{-1}X^{\frac{1}{2}}\mathbf{Y}, \\
\ddot{\mathbf{X}} &= 2s^{-2}\left\{\mathbf{P} + \mathbf{Y}\mathbf{Y}' - (X + (X/\kappa)^{\frac{1}{2}}Y)\mathbf{\Lambda} + (vX)^{\frac{1}{2}}\mathbf{H}\right\},
\end{aligned}$$

where $v = 1 + 1/\kappa$, and $X \sim \chi^2_\nu$, $Y \sim \text{Normal}_1(0,1)$, $\mathbf{Y} \sim \text{Normal}_N(\mathbf{0}, \mathbf{\Lambda})$, $\mathbf{P} \sim \text{Wishart}_N(\mathbf{\Lambda}, \nu - 1)$ and $\mathbf{H} \sim \text{Normal}_{N\times N}(\mathbf{0}, \mathbf{M}(\mathbf{\Lambda}))$, all independently; the elements of $\mathbf{M}(\mathbf{\Lambda})$ are such that

$$\text{Cov}(H_{ij}, H_{kl}) = \varepsilon(i,j,k,l) - \lambda_{ij}\lambda_{kl},$$

where H_{ij} is the (ij)th element of $\mathbf{H}$, λ_{ij} is the (ij)th element of $\mathbf{\Lambda}$, and $\varepsilon(i,j,k,l)$ is symmetric in its arguments, $i,j,k,l = 1, \ldots, N$.

The proof is given in the Appendix A1.

3.2 The EC intensity

The final result is an expression for the EC intensity, $\rho_N(b)$, when the fields are isotropic, so that $\mathbf{\Lambda} = \lambda \mathbf{I}_N$ where λ is a scalar.

Theorem 4 *For any $N \geq 1$*

$$\rho_N(b) = \frac{b^{\frac{\nu-N}{2}} e^{-\frac{b}{2}} \lambda^{\frac{N}{2}}}{(2\pi)^{\frac{N}{2}} 2^{\frac{\nu-2}{2}} \Gamma\left(\frac{\nu}{2}\right)} \left\{ \frac{s_1^{-N} + s_2^{-N}}{2} P_{N,\nu}(b) + \frac{s_1^{-N} - s_2^{-N}}{N} \sqrt{\frac{\kappa}{2\pi b}} Q_{N,\nu}(b,\kappa) \right\}$$

where $P_{N,\nu}(b)$ is a polynomial of degree $N-1$ in b with integer coefficients, given by

$$P_{N,\nu}(b) = \sum_{j=0}^{\lfloor (N-1)/2 \rfloor} \sum_{k=0}^{N-1-2j} \binom{\nu-1}{N-1-2j-k} \frac{(-1)^{N-1+j+k}(N-1)!}{2^j j! k!} b^{j+k},$$

and $Q_{N,\nu}(b,\kappa)$ is a polynomial of degree N in b and degree $\lfloor N/2 \rfloor$ in $1/\kappa$, given by

$$Q_{N,\nu}(b,\kappa) = \sum_{m=0}^{\lfloor N/2 \rfloor} \frac{(-1)^m N!}{2^m m!(N-2m)!} \left(\frac{b}{\kappa}\right)^m \frac{P_{N+1-2m,\nu}(b)}{1-2m},$$

where division by the factorial of a negative integer is treated as multiplication by zero.

The proof is given in the Appendix A2.

Corollary 1 *Up to three dimensions, the EC intensities are:*

$$\begin{aligned}
\rho_0(b) &= \int_b^\infty \frac{y^{\frac{\nu}{2}-1} e^{-\frac{y}{2}}}{2^{\frac{\nu}{2}} \Gamma\left(\frac{\nu}{2}\right)} dy + \frac{b^{\frac{\nu}{2}} e^{-\frac{b}{2}}}{2^{\frac{\nu-2}{2}} \Gamma\left(\frac{\nu}{2}\right)} \log\left(\frac{s_2}{s_1}\right) \sqrt{\frac{\kappa}{2\pi b}} \\
\rho_1(b) &= \frac{b^{\frac{\nu-1}{2}} e^{-\frac{b}{2}} \lambda^{\frac{1}{2}}}{(2\pi)^{\frac{1}{2}} 2^{\frac{\nu-2}{2}} \Gamma\left(\frac{\nu}{2}\right)} \left\{ \frac{s_1^{-1} + s_2^{-1}}{2} + \left(s_1^{-1} - s_2^{-1}\right) \sqrt{\frac{\kappa}{2\pi b}} [b - (\nu-1)] \right\} \\
\rho_2(b) &= \frac{b^{\frac{\nu-2}{2}} e^{-\frac{b}{2}} \lambda}{(2\pi) 2^{\frac{\nu-2}{2}} \Gamma\left(\frac{\nu}{2}\right)} \left\{ \frac{s_1^{-2} + s_2^{-2}}{2} [b - (\nu-1)] \right. \\
&+ \left. \frac{s_1^{-2} - s_2^{-2}}{2} \sqrt{\frac{\kappa}{2\pi b}} \left[b^2 - (2\nu-1)b + (\nu-1)(\nu-2) + \frac{b}{\kappa} \right] \right\} \\
\rho_3(b) &= \frac{b^{\frac{\nu-3}{2}} e^{-\frac{b}{2}} \lambda^{\frac{3}{2}}}{(2\pi)^{\frac{3}{2}} 2^{\frac{\nu-2}{2}} \Gamma\left(\frac{\nu}{2}\right)} \left\{ \frac{s_1^{-3} + s_2^{-3}}{2} \left[b^2 - (2\nu-1)b + (\nu-1)(\nu-2) \right] \right. \\
&+ \left. \frac{s_1^{-3} - s_2^{-3}}{3} \sqrt{\frac{\kappa}{2\pi b}} \left[b^3 - 3\nu b^2 + 3(\nu-1)^2 b - (\nu-1)(\nu-2)(\nu-3) + 3\frac{b}{\kappa}[b - (\nu-1)] \right] \right\}
\end{aligned}$$

It is now possible to find the EC intensity for a Gaussian random field in scale space, thus generalizing the result of Siegmund & Worsley (1995) to any number of dimensions. This is obtained by setting $\nu = 1$ and noting that the expected EC of the excursion set above b^2 of a χ^2 field with one degree of freedom is, by symmetry, twice the expected EC of a Gaussian random field above b:

Corollary 2 *For X a Gaussian random field in scale space, the EC intensity is, for $N \geq 1$,*

$$\rho_N(b) = \frac{e^{-\frac{b^2}{2}}\lambda^{\frac{N}{2}}}{(2\pi)^{\frac{N+1}{2}}}\left\{\frac{s_1^{-N}+s_2^{-N}}{2}\mathrm{He}_{N-1}(b) + \frac{s_1^{-N}-s_2^{-N}}{N}\sqrt{\frac{\kappa}{2\pi}}G_N(b,\kappa)\right\}$$

where $\mathrm{He}_n(b)$ is the Hermite polynomial of degree n in b and $G_N(b,\kappa)$ is a polynomial of degree N in b and degree $\lfloor N/2 \rfloor$ in $1/\kappa$, given by

$$G_N(b,\kappa) = \sum_{m=0}^{\lfloor N/2 \rfloor} \frac{(-1)^m N!}{2^m m!(N-2m)!}\left(\frac{1}{\kappa}\right)^m \frac{\mathrm{He}_{N-2m}(b)}{1-2m}.$$

For $N = 0$, the EC intensity is

$$\rho_0(b) = \int_b^\infty \frac{e^{-\frac{y^2}{2}}}{(2\pi)^{\frac{1}{2}}}dy + \frac{e^{-\frac{b^2}{2}}}{(2\pi)^{\frac{1}{2}}}\log\left(\frac{s_2}{s_1}\right)\sqrt{\frac{\kappa}{2\pi}}$$

Remark 1. One curious feature of these results is the presence of terms in powers of λ_j/κ for $N \geq 2$. This is curious because κ measures the roughness of the field in the scale dimension; EC should decrease if the roughness decreases, but the λ_j/κ term makes the expected EC increase as κ decreases. It would be very interesting to find a lower bound to λ/κ. It is possible to construct a filter $f(\mathbf{t}) \propto (\mathbf{t}'\mathbf{t})^{-N/4}$ for which κ is identically zero, but for this filter λ is infinite. Truncating the filter near zero and extrapolating it linearly in $\mathbf{t}'\mathbf{t}$, and truncating it at a large value of $\mathbf{t}'\mathbf{t}$ and extrapolating it quadratically to zero makes a filter for which κ can be made arbitrarily small, but at the same time λ becomes arbitrarily large. There is no contradiction here because the field is arbitrarily rough in the location dimensions, producing an arbitrarily large expected EC. It would be very interesting to find a filter for which κ can be made arbitrarily small while λ is bounded.

Remark 2. It might be thought worthwhile to generalize still further to t and F fields, as in Worsley (1994), to cover the case of unknown non-stationary standard deviation $\sigma(\mathbf{t})$, say. Here the component random fields in the numerator and denominator might then be all smoothed and used to construct scale space images of t or F statistics. No doubt the methods of this paper can be used to find the expected EC under the assumption of stationary $\sigma(\mathbf{t})$. But a moments reflection will show that this has little practical usefulness since if $\sigma(\mathbf{t})$ is non-stationary then it can never be completely eliminated from scale space t or F fields. It can be eliminated *at a fixed scale* if the underlying model for the observations has stationary Gaussian errors multiplied point-wise by $\sigma(\mathbf{t})$, since point-wise division by an estimator of $\sigma(\mathbf{t})$ derived from an independent χ^2 field with the same error structure will produce a t or F field whose distribution is free of $\sigma(\mathbf{t})$. However smoothing the observations produces errors which cannot be expressed as a stationary Gaussian field multiplied by a non-stationary standard deviation. In other words, for fixed s, $\sigma(\mathbf{t})$ can be eliminated from t or F fields based on

$$Z_i(\mathbf{t},s) = \sigma(\mathbf{t})s^{-N/2}\int f[(\mathbf{h}-\mathbf{t})/s]dW_i(\mathbf{h}),$$

but $\sigma(\mathbf{h})$ cannot be eliminated from from t or F fields based on

$$Z_i(\mathbf{t}, s) = s^{-N/2} \int f[(\mathbf{h} - \mathbf{t})/s]\sigma(\mathbf{h}) dW_i(\mathbf{h}).$$

Hence there is no way that $\sigma(\mathbf{t})$ can be removed by simply dividing, point-wise, by a suitable estimator. To be sure, the marginal distribution at any single point is free of $\sigma(\mathbf{t})$, but the joint distribution of the random field at pairs of points is not free of $\sigma(\mathbf{t})$, so that the distribution of the random field as a whole is not free of $\sigma(\mathbf{t})$. The only possibility of removing $\sigma(\mathbf{t})$ is to smooth the t or F field. However a simple expression for the distribution of such a field, even at a single point, is intractable. In Appendix A3 we present a method that partially overcomes these problems in fMRI data.

4 Practical issues in signal detection

4.1 Likelihood ratio tests

Suppose that a signal of the following form is added to the white noise error:

$$\mathrm{E}[dW_i(\mathbf{t})] = \mu_i f[(\mathbf{t} - \tau)/\sigma] d\mathbf{t}, \tag{4.1}$$

where μ_i is an unknown amplitude, $\tau \in C$ is an unknown location and $\sigma \in [s_1, s_2]$ is an unknown scale, $i = 1, \ldots, \nu$. In other words, the unknown location and scale of the signal are common to each component, but the amplitudes are different. We wish to test the null hypothesis of no signal, that is

$$H_0 \ : \ \mu_i = 0, \ i = 1, \ldots, \nu.$$

It is straightforward to show (see Siegmund & Worsley, 1995) that the -2log(likelihood ratio) test statistic is

$$X_{\max} = \max_{\mathbf{t} \in C, s \in I} X(\mathbf{t}, s),$$

and the arguments where the maximum is attained are the maximum likelihood estimators of τ and σ. In other words, the white noise plus signal should be smoothed with a filter that matches the signal, as in the Matched Filter Theorem.

4.2 Smoothing the χ^2 field

It might be tempting to smooth the χ^2 field, rather than the Gaussian components. Apart from the intractability of the distribution of a smoothed χ^2 field, this is not the maximum likelihood procedure. Since maximum likelihood based methods are often asymptotically most powerful and most efficient, it seems preferable to smooth the Gaussian components than to smooth the χ^2 field. Note however that, like outlier detection and the change-point problem, no uniformly most powerful test exists for detecting a signal at an unknown location, since the test statistic $X(\tau, \sigma)$ at the true location of the signal is always more powerful than any other test statistic (see Worsley, 1986).

4.3 Correlated image data

In practice the image data is spatially correlated, particularly PET data, which means that $Z_i(\mathbf{t}, s)$ cannot be calculated from the underlying white noise data as in (3.1). However if the spatial correlation function is Gaussian with scale s^* say, then the image can still be modeled as (3.1) with an effective filter that is also Gaussian with scale $s = s^*/\sqrt{2}$. This means that smoothing to an effective scale of $\tilde{s}$ can be achieved by using a filter of scale $\sqrt{\tilde{s}^2 - s^2}$ on the spatially correlated images. This and other practical issues are dealt with in Worsley *et al.* (1996).

4.4 Detecting multiple signals

It might be supposed that $X_{\max}$ is only powerful at detecting one localized signal. However it is equally useful at detecting several localized signals provided they are well separated in scale space. This is because if the amplitude of a signal component is sufficiently large then the value of the scale space χ^2 field at the location of the signal will cross the critical threshold of $X_{\max}$ and that signal component will be detected. In the example in section 5 this method does detect multiple signal components.

In fact there is good reason to suppose that the EC is an accurate estimator of the number of isolated unimodal signal components. The argument is as follows. Suppose that the critical threshold b is chosen so that $\mathrm{E}[\chi(B_b)] = \alpha$, where α is small, say 0.05. If there are k strong isolated unimodal signal components, so that X cross the critical threshold in the neighbourhood of each signal with high probability, then each will contribute +1 to the EC with high probability. For the region outside the signals, where the means of the Gaussian fields are zero, the probability that X will cross the critical threshold is less than α (approximately), since the volume of the regions outside the signal components is less than the total search region. Thus the probability that the observed EC equals the number of signal components is at least $1 - \alpha$, approximately.

4.5 Closely spaced signals

For the above methods to work, the signals must be well separated. If two signals of the same scale are close together in location then they might be detected as a single signal at a larger scale and at a location midway between them. Without loss of generality, assume that only one Gaussian component contains signals, and the other $\nu - 1$ have a mean of zero. Then it can be shown (numerically) that for two 3-D Gaussian shaped signals of equal amplitude smoothed with a Gaussian shaped filter, the scale space signal moves from bimodal to unimodal at a separation of about $t = 3.74$ standard deviations. The scale at which this happens is $s = \sqrt{t^2/N + \sqrt{t^4/N^2 + 1}} = 1.64$ standard deviations. In other words, the minimum resolvable separation is 3.74 standard deviations.

The same thing happens in the scale dimension. If the two signals have the same location and the same integrated square but different scales, it can be shown that the minimum resolvable separation is $(\sqrt{N/2 + 1} + 1)/(\sqrt{N/2 + 1} - 1)$ standard deviations. This means that for 3-D data, two superimposed signals which have a standard deviation differing by a

factor of at least 4.44 can be resolved, even though the signals have the same location. This will be illustrated in the example in section 5.

4.6 Choosing the scale range

Finally, the question arises of how to select the range of scales $[s_1, s_2]$. Clearly this should cover the anticipated range of the scales of the underlying signals. Siegmund & Worsley (1995) compare the power of the scale space test with the test at a fixed scale, both when the fixed scale matches that of the signal, and when it is misspecified. As expected, the fixed scale test is more powerful than the scale space test, but less powerful when the fixed scale is misspecified.

5 Application

The motivation for this work came from a problem in functional magnetic resonance imaging (fMRI). This new technique is able to produce a time series of images of BOLD response in the human brain, opening up the possibility of finding which areas of the brain are 'activated' by an external stimulus (see Lange & Zeger, 1997). One of the first experiments was to locate the regions of the brain that respond to a simple visual stimulus (see for example Kwong *et al.*, 1992; Ogawa, 1992). In a similar experiment performed at the Montreal Neurological Institute, a subject was given a simple visual stimulus that was switched on and off every 24 seconds, while every 6 seconds a BOLD fMRI image was obtained (Ouyang *et al.*, 1994). This was repeated for 240 seconds, so that a time series of 40 2-D images was obtained, each 128×128 pixels. One of these images is shown on the left of Figure 5, which is an oblique slice through the visual cortex in the lower part of the image. Darker shading represents higher BOLD response. The time series at two pixels, one in the visual cortex and one outside, are also shown. There appears to be strong correlation with the stimulus at the lower pixel in the visual cortex, but none at the upper pixel outside the visual cortex.

The problem then is to locate the regions of the image which show changes in BOLD response that are synchronized with the signal. Friston *et al.* (1994) have modeled the BOLD response as the input stimulus (in this case a square wave with 0=off, 1=on) plus additive white noise convolved with a Poisson-shaped hemodynamic response function with an 8 second mean delay. More recently, Purdon *et al.* (1998ab) have used a physiological model for the hemodynamic response function that accounts for "undershoot" and delay dynamics.

There has been some speculation that the hemodynamic response is not the same at every pixel (Bullmore *et al.*, 1996; Lange & Zeger, 1997; Purdon *et al.*, 1998ab). An alternative method of detecting a signal is to look for pixels where the frequency of the stimulus matches the frequency of the response, ignoring the delay and the shape of the response, using the magnitude of the periodogram at that frequency (Bullmore *et al.*, 1996). This is equivalent to cross-correlating with a sine and cosine function, to give $\nu = 2$ random fields which approximate independent white noise processes $dW_1(\mathbf{t})$ and $dW_2(\mathbf{t})$. The magnitude of the periodogram is then $dW_1(\mathbf{t})^2 + dW_2(\mathbf{t})^2$. Full details are given in Appendix A3.

Figure 5, panels 1 and 2, show images of $dW_1(\mathbf{t})$ and $dW_2(\mathbf{t})$. Some activation is clearly

present in the second image $dW_2(\mathbf{t})$ in the lower portion of the image corresponding to the visual cortex, as expected. To check sensitivity, three simulated Gaussian-shaped signals were created with delays of 0°, 45° and 90°, and widths of 9mm, 15mm and 25mm, respectively. As is customary in the imaging literature, width is measured by the *Full Width at Half Maximum (FWHM)* defined as the width of the function at half its maximum height. For a Gaussian-shaped function, FWHM $= \sqrt{8 \log 2}\, s$, where s is the standard deviation. Since no signal was expected outside the visual cortex (and none was in fact detected), these three signals were located in the upper half of the image (Figure 5, panels 3 and 4) and added to $dW_1(\mathbf{t})$ and $dW_2(\mathbf{t})$. The magnitude of the periodogram $dW_1(\mathbf{t})^2 + dW_2(\mathbf{t})^2$ without and with the simulated signal is shown in Figure 5, panels 5 and 6, respectively. The remaining panels in Figure 5 show the χ^2 fields $X(\mathbf{t}, s)$ resulting from smoothing $dW_1(\mathbf{t})$ and $dW_2(\mathbf{t})$ with the Gaussian filter (1.1) for 10 widths ranging from 6.2mm to 34.4mm FWHM, roughly equally spaced on a log scale.

For the purpose of finding $\mathrm{E}[\chi(B_b)]$, the search region C was approximated by a circle of radius $r = 61.7$mm chosen so that its area matched that of the set of pixels shown in Figures 1 and 2. $\mathrm{E}[\chi(B_b)]$ was then calculated using (2.10) with EC intensities taken from Corollary 1. This was equated to 0.05 and solved for b to give $b = 25.3$, which by the arguments in the previous section is an approximate level 0.05 critical value for $X_{\max}$. Scale space local maxima above this threshold are marked by white crosses on Figure 5. All three simulated signals are detected (panels 8, 11 and 14) at 7.5, 13.9, 24.7mm smoothing, close to their actual widths of 9, 15, 25mm, respectively. Several local maxima appear in the visual cortex, culminating in the global maximum of $X_{\max} = 208$ in panel 15 at 29.3mm FWHM smoothing.

The scale space excursion set B_b at the $b = 25.3$ threshold is shown in Figure 5a. Four components are clearly evident: the three simulated signals in the upper part of the figure and the activation in the visual cortex in the lower part. The EC of the excursion set is $\chi(B_b) = 6$ (there are two other small connected components not easily visible in the figure). The excursion set at a much higher threshold $b = 175$ is shown in Figure 5b, this time tilted to emphasize the scale dimension. Four connected components are clearly visible ($\chi(B_b) = 4$) corresponding to the three simulated signals and the activation in the visual cortex.

Note that each of the simulated signals is optimally detected at a filter width close to the width of the signal, as predicted by the Matched Filter Theorem. There is a tendency for the signals to be detected at a filter width less than that of the signal, but this may be due to random error – note that the image is quite smooth in the scale dimension, as is evident from Figure 5b.

The stimulus is detected at all scales, suggesting a mix of sharp foci of activation embedded in a broader 'penumbra' of lesser activation. This is more clearly seen at 11.3mm smoothing (Figure 5, panel 10). The sharp foci may correspond to arterial blood flow, whereas the penumbra may correspond to BOLD response in the capillaries. The fact that many foci appear at the lowest filter width suggest that these foci are sharper than 6.2mm.

Finally Figure 5 plots the observed and expected EC for a range of thresholds on a log scale. In the range from $25 < b < 200$ the observed EC is fairly stable at 4, well above expectation. Again, these four components correspond to the three simulated signals and the visual cortex activation. For $1 < b < 25$ the observed and expected EC are in rough agreement, though it is difficult to interpret this in practical terms. For $b < 1$, the observed

EC and expected EC differ, probably due to lack of continuity in the discretely sampled random fields.

Remark. Note that at $b = 0$ the expected EC is negative; there is a simple explanation for this which is peculiar to χ^2 fields whose degrees of freedom (here 2) are less than the number of dimensions (here 3). The χ^2 field with 2 degrees of freedom is zero if and only if the two scale space component Gaussian fields $Z_1(\mathbf{t}, s)$ and $Z_2(\mathbf{t}, s)$ are both zero. The sets where this happens are smooth 2-D surfaces embedded in 3-D, and their intersection, which is where the χ^2 field is zero, is a set of smooth 1-D lines in 3-D (Worsley, 1996). Thus the excursion set of the χ^2 field near $b = 0$ consists of a set of linear 'holes' which either form complete rings or which connect to the outside of the search region; there can be no free ends. Worsley (1996) conjectures that these rings can be linked or even knotted with probability greater than zero. Each ring-shaped hole contributes zero to the EC, but each hole that enters and leaves through the boundary of the search region effectively adds a 'handle' to the excursion set and contributes -1 to the EC. Figure 5 shows that 47.6 of these 'handles' are expected; the fact that never more than 20 are observed (at $b = 0.57$) is probably due to insufficient sampling of the continuous random field by a voxel grid. Below this threshold the linear holes break up into isolated voxels that each contribute +1 to the EC, and their connectivity is lost in the gaps between the voxels.

Appendix

A1 Proof of Lemma 1

Let $\dot{\mathbf{Z}}_i = \partial Z_i / \partial \mathbf{t}$ denote the gradient vector of Z_i and let $\ddot{\mathbf{Z}}_i = \partial^2 Z_i / \partial \mathbf{t} \partial \mathbf{t}'$ denote the Hessian matrix with respect to $\mathbf{t}$. Finally, let $\dot{Z}_{is} = \partial Z_i / \partial s$, $i = 1, \ldots, \nu$. Then

$$\begin{aligned} \dot{X}_s &= 2 \sum_{i=1}^{\nu} Z_i \dot{Z}_{is}, \\ \dot{\mathbf{X}} &= 2 \sum_{i=1}^{\nu} Z_i \dot{\mathbf{Z}}_i, \\ \ddot{\mathbf{X}} &= 2 \sum_{i=1}^{\nu} (Z_i \ddot{\mathbf{Z}}_i + \dot{\mathbf{Z}}_i \dot{\mathbf{Z}}_i'). \end{aligned}$$

Now the joint distribution of $(Z_i, \dot{Z}_{is}, \dot{\mathbf{Z}}_i, \ddot{\mathbf{Z}}_i)$ is multivariate normal with mean zero and variance

$$\mathrm{Var} \begin{pmatrix} Z_i \\ \dot{Z}_{is} \\ \dot{\mathbf{Z}}_i \\ \ddot{\mathbf{Z}}_i \end{pmatrix} = \begin{pmatrix} 1 & 0 & \mathbf{0} & -s^{-2}\mathbf{\Lambda} \\ 0 & s^{-2}\kappa & \mathbf{0} & -s^{-3}\mathbf{\Lambda} \\ \mathbf{0} & \mathbf{0} & s^{-2}\mathbf{\Lambda} & \mathbf{0} \\ -s^{-2}\mathbf{\Lambda} & -s^{-3}\mathbf{\Lambda} & \mathbf{0} & s^{-4}\mathbf{M}(\mathbf{\Lambda}) \end{pmatrix} \tag{A.1}$$

(Siegmund & Worsley, 1995).

The first derivatives are straightforward. Since $\dot{Z}_{is} \sim \mathrm{Normal}_1(0, s^{-2}\kappa)$, independent of Z_i, then conditional on Z_i for all $i = 1, \ldots, \nu$,

$$\dot{X}_s = 2 \sum_{i=1}^{\nu} Z_i \dot{Z}_{is} \sim \mathrm{Normal}(0, 4X s^{-2} \kappa).$$

Since this depends on Z_i for all $i = 1, \ldots, \nu$ only through X then it is the distribution of $\dot{X}_s$ conditional on X alone. This leads straight to the representation for $\dot{X}_s$, and that for $\dot{\mathbf{X}}$ in a similar way, with

$$Y = \sum_{i=1}^{\nu} Z_i \dot{Z}_{is} s/(X\kappa)^{1/2}, \qquad \mathbf{Y} = \sum_{i=1}^{\nu} Z_i \dot{\mathbf{Z}}_i / X^{1/2}.$$

The second derivative will be found next. Conditional on Z_i and $\dot{Z}_{is}$

$$\ddot{\mathbf{Z}}_i \sim \text{Normal}_{N\times N}\{-s^{-2}(Z_i + \dot{Z}_{is}s/\kappa)\mathbf{\Lambda}, s^{-4}v\mathbf{M}(\mathbf{\Lambda})\},$$

independent of $\dot{\mathbf{Z}}_i$, where $v = 1 + 1/\kappa$. Hence conditional on Z_i and $\dot{Z}_{is}$ for all $i = 1, \ldots, \nu$,

$$\sum_{i=1}^{\nu} Z_i \ddot{\mathbf{Z}}_i \sim \text{Normal}_{N\times N}\{-s^{-2}(X + (X/\kappa)^{\frac{1}{2}}Y)\mathbf{\Lambda}, s^{-4}vX\mathbf{M}(\mathbf{\Lambda})\}.$$

Since this depends on Z_i and $\dot{Z}_{is}$ only through X and Y, then conditioning on X and Y is all that is needed. Note that X, Y and $\dot{\mathbf{Z}}_i$ for all $i = 1, \ldots, \nu$ are all independent, so conditional on X, Y and $\dot{\mathbf{Z}}_i$ for all $i = 1, \ldots, \nu$

$$\ddot{\mathbf{X}} = 2s^{-2}\left\{\sum_{i=1}^{\nu} \dot{\mathbf{Z}}_i\dot{\mathbf{Z}}_i' - (X + (X/\kappa)^{\frac{1}{2}}Y)\mathbf{\Lambda} + (vX)^{1/2}\mathbf{H}\right\}.$$

Let $\mathbf{Z} = (Z_1, \ldots, Z_\nu)'$, $\dot{\mathbf{Z}} = \partial\mathbf{Z}/\partial\mathbf{t}'$, $\mathbf{A} = \mathbf{I}_\nu - \mathbf{Z}\mathbf{Z}'/X$ and $\mathbf{P} = \dot{\mathbf{Z}}'\mathbf{A}\dot{\mathbf{Z}}$. Since $\mathbf{Y} = \dot{\mathbf{Z}}'\mathbf{Z}/X^{1/2}$ then

$$\sum_{i=1}^{\nu} \dot{\mathbf{Z}}_i\dot{\mathbf{Z}}_i' = \dot{\mathbf{Z}}'\dot{\mathbf{Z}} = \mathbf{P} + \mathbf{Y}\mathbf{Y}'.$$

Assume for the moment that $\mathbf{Z}$ is fixed. Then $\mathbf{A}$ is fixed and $\mathbf{P} \sim \text{Wishart}_N(\mathbf{\Lambda}, \nu - 1)$ since $\mathbf{A}$ is idempotent of rank $\nu - 1$. Since $\mathbf{A}\mathbf{Z} = \mathbf{0}$ and $\mathbf{Y}$ is a linear combination of $\dot{\mathbf{Z}}$ then $\mathbf{Y}$ is independent of $\mathbf{P}$ conditional on $\mathbf{Z}$. Since the distributions of $\mathbf{P}$ and $\mathbf{Y}$ do not depend on $\mathbf{Z}$, then $\mathbf{P}$ and $\mathbf{Y}$ are independent unconditionally, which completes the proof.

A2 Proof of Theorem 4

The term (2.7), corresponding to the 'base', is

$$\left[\text{E}[(X \geq b)\det(-\ddot{\mathbf{X}}) \mid \dot{\mathbf{X}} = \mathbf{0}]\theta(\mathbf{0})\right]_{s=s_1} = \frac{b^{\frac{\nu-N}{2}}e^{-\frac{b}{2}\lambda^{\frac{N}{2}}}}{(2\pi)^{\frac{N}{2}}2^{\frac{\nu-2}{2}}\Gamma\left(\frac{\nu}{2}\right)}s_1^{-N}P_{N,\nu}(b), \tag{A.2}$$

found by combining Theorem 3.7 of Worsley (1994) with Theorem 3 of Worsley (1995b). We now concentrate on (2.8), starting with the expectation in the integrand. Conditional on $\dot{\mathbf{X}} = \mathbf{0}$ and $X = b$, Lemma 1 can be used to write

$$\ddot{\mathbf{X}} = 2s^{-2}\left\{\mathbf{P} - (b + (b/\kappa)^{\frac{1}{2}}Y)\mathbf{\Lambda} + (vb)^{\frac{1}{2}}\mathbf{H}\right\},$$

where $v = 1+1/\kappa$, $Y \sim \text{Normal}_1(0,1)$, $\mathbf{P} \sim \text{Wishart}_N(\mathbf{\Lambda}, \nu-1)$ and $\mathbf{H} \sim \text{Normal}_{N\times N}(\mathbf{0}, \mathbf{M}(\mathbf{\Lambda}))$, all independently. Conditional on Y as well as $\dot{\mathbf{X}} = \mathbf{0}$ and $X = b$, Lemma A.6 of Worsley (1994) (with $\eta = \infty$) gives:

$$\begin{aligned} & \mathrm{E}[\dot{X}_s^+ \det(-\ddot{\mathbf{X}}) \mid Y, \dot{\mathbf{X}} = \mathbf{0}, X = b] \\ = \; & 2s^{-1}(b\kappa)^{\frac{1}{2}} Y^+ s^{-2N} \det(\mathbf{\Lambda}) 2^N \sum_{j=0}^{\lfloor N/2 \rfloor} \sum_{k=0}^{N-2j} \binom{\nu-1}{N-2j-k} \frac{(-1)^{N+j+k} N!}{2^j j! k!} v^j b^j (b + (b/\kappa)^{\frac{1}{2}} Y)^k. \end{aligned}$$

Expanding in terms of Y and taking expectations over Y:

$$\mathrm{E}[Y^+ (b + (b/\kappa)^{\frac{1}{2}} Y)^k] = \sum_{l=0}^{k} \binom{k}{l} b^{k-\frac{l}{2}} \kappa^{-\frac{l}{2}} 2^{\frac{l}{2}} \Gamma\left(\frac{l}{2}+1\right) (2\pi)^{-\frac{1}{2}}.$$

The joint density of $(\dot{\mathbf{X}}, X)$ is the product of the density of $\dot{\mathbf{X}}$ conditional on X times the marginal density of X. Evaluated at $\dot{\mathbf{X}} = \mathbf{0}$ and $X = b$ this is:

$$\phi(\mathbf{0}, b) = (8\pi b)^{-\frac{N}{2}} s^N \det(\mathbf{\Lambda})^{-\frac{1}{2}} \frac{b^{\frac{\nu}{2}-1} e^{-\frac{b}{2}}}{2^{\frac{\nu}{2}} \Gamma\left(\frac{\nu}{2}\right)}.$$

Combining these results, and integrating from $s = s_1$ to $s = s_2$, gives the second term (2.8):

$$\int_{s_1}^{s_2} \mathrm{E}[\dot{X}_s^+ \det(-\ddot{\mathbf{X}}) \mid \dot{\mathbf{X}} = \mathbf{0}, X = b] \phi(\mathbf{0}, b) ds = \frac{b^{\frac{\nu-N}{2}} e^{-\frac{b}{2}} \lambda^{\frac{N}{2}}}{(2\pi)^{\frac{N}{2}} 2^{\frac{\nu-2}{2}} \Gamma\left(\frac{\nu}{2}\right)} \frac{s_1^{-N} - s_2^{-N}}{N} \sqrt{\frac{\kappa}{2\pi b}} Q^*_{N,\nu}(b, \kappa) \tag{A.3}$$

where

$$Q^*_{N,\nu}(b,\kappa) = \sum_{j=0}^{\lfloor N/2 \rfloor} \sum_{k=0}^{N-2j} \binom{\nu-1}{N-2j-k} \frac{(-1)^{N+j+k} N!}{2^j j! k!} v^j b^j \sum_{l=0}^{k} \binom{k}{l} b^{k-\frac{l}{2}} \kappa^{-\frac{l}{2}} 2^{\frac{l}{2}} \Gamma\left(\frac{l}{2}+1\right).$$

Adding the two terms (A.2) and (A.3) gives the answer we seek. However it is not in quite the right form. Inspection of (A.2) plus (A.3) shows that it is equal to $b^{\frac{\nu-N}{2}}$ times a polynomial in half powers of b, whereas we know from Siegmund & Worsley (1995) that the Gaussian case ($\nu = 1$) is a polynomial in b. This suggests that the half odd powers should cancel. We now proceed to show this. Writing

$$v^j = \sum_{i=0}^{j} \binom{j}{i} \kappa^{-i},$$

replacing j by $j' = j - i$, replacing l by $m = l + 2i$, replacing k by $k' = k - m + 2i$, and re-arranging the order of summation, this can be factored into

$$Q^*_{N,\nu}(b,\kappa) = \sum_{m=0}^{N} \binom{N}{m} \left(\frac{b}{\kappa}\right)^{\frac{m}{2}} P_{N+1-m,\nu}(b) c_m$$

where

$$c_m = \sum_{i=0}^{\lfloor m/2 \rfloor} \frac{(-1)^i 2^{\frac{m}{2}} m!}{4^i i! (m-2i)!} \Gamma\left(\frac{m}{2}+1-i\right).$$

Replacing the gamma function by the integral

$$\Gamma\left(\frac{m}{2}+1-i\right)=\int_0^\infty 2^{i-\frac{m}{2}}x^{m-2i+1}e^{-x^2/2}dx,$$

and interchanging summation and integration gives

$$c_m=\int_0^\infty x\mathrm{He}_m(x)e^{-x^2/2}dx,$$

where $\mathrm{He}_m(x)$ is the Hermite polynomial of degree m in x. Then $c_0=1$ and $c_1=\sqrt{\pi/2}$; for $m\geq 2$ integration by parts gives

$$c_m=\mathrm{He}_{m-2}(0)=\begin{cases}0, & m \text{ odd}\\ \frac{(-1)^{\frac{m}{2}-1}m!}{2^{\frac{m}{2}}(m/2)!(m-1)}, & m \text{ even}\end{cases}$$

This eliminates half the terms in $Q^*_{N,\nu}(b,\kappa)$; the remainder can be written as

$$Q^*_{N,\nu}(b,\kappa)=Q_{N,\nu}(b,\kappa)+\frac{N}{2}\sqrt{\frac{2\pi b}{\kappa}}P_{N,\nu}(b).$$

Using this in (A.3), and adding (A.2) and (A.3) gives the final result. Since the algebra is quite messy, this was checked using the computer algebra package MAPLE.

A3 Details of the analysis of the fMRI data

Let $Y(\mathbf{t},t)$ represent the BOLD response at location $\mathbf{t}$ in $D=2$ dimensions and time t. The standard deviation of fMRI data is not constant across the image, so the first step was to standardise the data to unit standard deviation at each pixel. To do this, a simple sine wave with period $T=48$ seconds

$$\hat{Y}(\mathbf{t},t)=m(\mathbf{t})+b_1(\mathbf{t})\sin(2\pi t/T)+b_2(\mathbf{t})\cos(2\pi t/T)$$

was fitted to the BOLD response with coefficients estimated by

$$\begin{aligned}m(\mathbf{t}) &= \sum_t Y(\mathbf{t},t)/n,\\ b_1(\mathbf{t}) &= \sum_t \sin(2\pi t/T)Y(\mathbf{t},t)/(n/2),\\ b_2(\mathbf{t}) &= \sum_t \cos(2\pi t/T)Y(\mathbf{t},t)/(n/2),\end{aligned}$$

where summation is over $n=40$ scans. The pixel variance was then estimated by

$$\hat{\sigma}(\mathbf{t})^2=\sum_t[Y(\mathbf{t},t)-\hat{Y}(\mathbf{t},t)]^2/(n-3),$$

and the standardised data is

$$Y^*(\mathbf{t},t)=[Y(\mathbf{t},t)-\mu(\mathbf{t})]/\hat{\sigma}(\mathbf{t}),$$

which should now have equal standard deviation across all pixels.

The sine and cosine coefficients were re-estimated using the standardised data, to give

$$\begin{aligned} b_1^*(\mathbf{t}) &= \sum_t \sin(2\pi t/T) Y^*(\mathbf{t}, t)/(n/2), \\ b_2^*(\mathbf{t}) &= \sum_t \cos(2\pi t/T) Y^*(\mathbf{t}, t)/(n/2). \end{aligned}$$

These were then smoothed using a Gaussian filter with FWHM $w = \sqrt{8 \log 2}\ s$ =5.0, 6.0, 7.1, 8.5, 10.2, 12.2, 14.6, 17.5, 20.9, 25.0mm equally spaced on a log scale to give uniform coverage of scale space (see Siegmund & Worsley, 1995):

$$b_i^*(\mathbf{t}, s) = K^*(s) \sum_{\mathbf{h}} f[(\mathbf{h} - \mathbf{t})/s] b_i^*(\mathbf{h}), \quad i = 1, 2,$$

where, because we have replaced an integral by a summation,

$$K^*(s) = \left(\sum_{\mathbf{h}} f[\mathbf{h}/s]^2 \right)^{-\frac{1}{2}}.$$

Note that although $b_i^*(\mathbf{t})$ is not exactly Gaussian, since it was divided by an estimated (random) standard deviation, the Central Limit Theorem should ensure that $b_i^*(\mathbf{t}, s)$ is closer to Gaussian, if s is reasonably large, since it is a (weighted) sum of $b_i^*(\mathbf{t})$.

However the above method of constructing Gaussian scale space images may not ensure that $b_i^*(\mathbf{t}, s)$ has unit standard deviation. This is because, although $b_i^*(\mathbf{t})$ usually has very low spatial correlation for fMRI data, the spatial correlation is not quite zero and so $b_i^*(\mathbf{t})$ is not exactly white noise. Positive correlation, for example, will increase the standard deviation of $b_i^*(\mathbf{t}, s)$, and this was in fact observed. Since $b_i^*(\mathbf{t})$ and hence $b_i^*(\mathbf{t}, s)$ should have constant standard deviation across the image, this suggests that a global correction for the standard deviation is all that is necessary. To do this, the global standard deviation of the smoothed data was estimated separately using the coefficients at two periodicities, $A = 240/6$ and $C = 240/4$ seconds, either side of the periodicity of the signal ($T = 240/5$ seconds), that is

$$\begin{aligned} a_1^*(\mathbf{t}) &= \sum_t \sin(2\pi t/A) Y^*(\mathbf{t}, t)/(n/2), \\ a_2^*(\mathbf{t}) &= \sum_t \cos(2\pi t/A) Y^*(\mathbf{t}, t)/(n/2). \\ c_1^*(\mathbf{t}) &= \sum_t \sin(2\pi t/C) Y^*(\mathbf{t}, t)/(n/2), \\ c_2^*(\mathbf{t}) &= \sum_t \cos(2\pi t/C) Y^*(\mathbf{t}, t)/(n/2). \end{aligned}$$

These were smoothed as above:

$$\begin{aligned} a_i^*(\mathbf{t}, s) &= K^*(s) \sum_{\mathbf{h}} f[(\mathbf{h} - \mathbf{t})/s] a_i^*(\mathbf{h}), \\ c_i^*(\mathbf{t}, s) &= K^*(s) \sum_{\mathbf{h}} f[(\mathbf{h} - \mathbf{t})/s] c_i^*(\mathbf{h}), \quad i = 1, 2. \end{aligned}$$

These four images should contain no signal, and if the global variance (averaged over all pixels) is smooth in the frequency, then the average of these global variances, which linearly interpolates the variance across frequency, should be a good estimator of the global variance of $b_i^*(\mathbf{t}, s)$. Thus the estimator of the global variance is

$$v^2(s) = \sum_{\mathbf{t}} [a_1^*(\mathbf{t}, s)^2 + a_2^*(\mathbf{t}, s)^2 + c_1^*(\mathbf{t}, s)^2 + c_2^*(\mathbf{t}, s)^2]/(4M),$$

where M is the number of pixels in the image. This was validated by repeating the operations for the upper part of the image, outside the visual cortex, where no signal should be present, and there was good agreement between $v^2(s)$ and the mean sum of squares of $b_i^*(\mathbf{t}, s)$. Our final definition of the scale space Gaussian images is therefore

$$Z_i(\mathbf{t}, s) = b_i^*(\mathbf{t}, s)/v(s), \quad i = 1, 2,$$

and the χ^2 field with $\nu = 2$ degrees of freedom is

$$X(\mathbf{t}, s) = Z_1(\mathbf{t}, s)^2 + Z_2(\mathbf{t}, s)^2.$$

The same considerations as mentioned above also affect the smoothness of the scale space data. The effective scale was estimated from (A.1) by

$$\tilde{s} = [\det(\mathbf{\Lambda})/\det(\mathrm{Var}(\dot{\mathbf{Z}}_i))]^{\frac{1}{2N}}.$$

$\mathbf{\Lambda} = \mathbf{I}/2$ from (3.2), and $\mathrm{Var}(\dot{\mathbf{Z}}_i)$ was estimated from the pure noise images $a_i^*(\mathbf{t}, s)$ and $c_i^*(\mathbf{t}, s)$ by replacing derivatives by differences at adjacent pixels (Worsley, 1995a), then pooling the variance over all pixels and images. The results were $\tilde{s}$ =2.6, 3.2, 3.9, 4.8, 5.9, 7.3, 8.8, 10.5, 12.4, 14.6mm ($\tilde{w} = \sqrt{8 \log 2}\ \tilde{s}$ = 6.2, 7.5, 9.2, 11.3, 13.9, 17.1, 20.7, 24.7, 29.3, 34.4mm FWHM) respectively, approximately 35% higher than the nominal FWHMs. Thus s_1 = 2.6mm and s_2 = 14.6mm were used in Corollary 1 in place of the nominal values of $5/\sqrt{8 \log 2}$ = 2.1mm and $25/\sqrt{8 \log 2}$ = 10.6mm, respectively.

Acknowledgment

The author would like to thank Alan Evans and Mark Wolforth of the Montreal Neurological Institute for permission to use the fMRI data, and an anonymous referee of an earlier version for very useful comments.

References

Adler, R.J. (1981). *The Geometry of Random Fields.* Wiley, New York.

Adler, R.J. (2000). On excursion sets, tube formulae, and maxima of random fields. *Annals of Applied Probability,* in press.

Büchel, C., Wise, R. J. S., Mummery, C. J., Poline, J-B. & Friston, K. J. (1996). Non-linear regression in parametric activation studies. *NeuroImage*, **4**, 60-66.

Bullmore, E.T., Brammer, M.J., Williams, S.C.R., Rabe-Hesketh, S., Janot, N., David, A.S., Mellers, J.D.C., Howard, R. & Sham, P. (1996). Statistical methods of estimation and inference for functional MR image analysis. *Magnetic Resonance in Medicine*, **35**, 261-277.

Cao, J. & Worsley, K.J. (1999). The detection of local shape changes via the geometry of Hotelling's T^2 fields. *Annals of Statistics*, **27**, 925-942.

Crivello, F., Tzourio, N., Poline, J.B., Woods, R.P., Mazziotta, J.C. and Mazoyer, B. (1995). Intersubject variability in functional neuroanatomy of silent verb generation: Assessment by a new activation detection algorithm based on amplitude and size information. *NeuroImage*, **2**, 253-263.

Daubechies, I. (1992). *Ten Lectures on Wavelets.* CBMS-NSF Regional Conference Series in Applied Mathematics, **61**, Philadelphia.

Fletcher, P.C., Dolan, R.J., Shallice, T., Frith, C.D., Frackowiak, R.S.J. & Friston, K.J. (1996). Is multivariate analysis of PET data more revealing than the univariate approach? Evidence from a study of episodic memory retrieval. *NeuroImage*, **3**, 209-215.

Friston, K.J., Jezzard, P. & Turner, R. (1994). Analysis of functional MRI time series. *Human Brain Mapping*, **1**,153-171.

Hasofer, A.M. (1978). Upcrossings of random fields. *Supplement to Advances in Applied Probability,* **10**, 14-21.

Klain, G. & Rota, G.-C. (1997). *Introduction to Geometric Probability.* Cambridge, UK: Cambridge University Press.

Knuth, D.E. (1992). Two notes on notation. *The American Mathematical Monthly*, **99**, 403-422.

Kwong, K.K., Belliveau, J.W., Chesler, D.A., Goldberg, I.E., Weisskoff, R.M., Poncelet, B.P., Kennedy, D.N., Hoppel, B.E., Cohen, M.S., Turner, R., Cheng, H-M., Brady, T.J. & Rosen, B.R. (1992). Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proceedings of the National Academy of Science*, **89**, 5675-5679.

Lange, N. & Zeger, S.L. (1997). Non-linear Fourier time series analysis for human brain mapping by functional magnetic resonance imaging (with Discussion). *Applied Statistics*, **46**, 1-29.

Lindeberg, T. (1994). *Scale-space Theory in Computer Vision*, Kluwer Academic Publishers, Boston.

Ogawa, S., Tank, D.W., Menon, R., Ellerman, J.M., Kim, S-G., Merkle, H. & Ugurbil, K. (1992). Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. *Proceedings of the National Academy of Science*, **89**, 5951-5955.

Ouyang, X., Pike, G.B. & Evans, A.C. (1994). fMRI of human visual cortex using temporal correlation and spatial coherence analysis. *13th Annual Symposium of the Society of Magnetic Resonance in Medicine.*

Poline, J-B. & Mazoyer, B.M. (1994a). Enhanced detection in activation maps using a multifiltering approach. *Journal of Cerebral Blood Flow and Metabolism*, **14**, 690-699.

Poline, J-B. & Mazoyer, B.M. (1994b). Analysis of individual brain activation maps using hierarchical description and multiscale detection. IEEE Transactions on Medical Imaging, **13(4)**, 702-710.

Purdon, P.L., Solo, V., Brown, E., Buckner, R., Rotte, M., and Weisskoff, R.M. (1998a). fMRI noise variability across subjects and trials: insights for noise estimation methods. *NeuroImage*, **7**:S617.

Purdon, P.L., Solo, V., Brown, E., and Weisskoff, R.M. (1998b). Signal processing in fMRI: Noise estimation with regularization and hemodynamic response modeling. *NeuroImage*, **7**:S618.

Rosenfeld, A. & Kak, A.C. (1982). *Digital Picture Processing, Volume 2.* Academic Press, Orlando.

Shafie, K. (1998). *The geometry of Gaussian rotation space random fields.* Ph.D. thesis, Department of Mathematics and Statistics, McGill University.

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**, 608-639.

Torres, S. (1994). Topological analysis of *COBE*-DMR cosmic microwave background maps. *Astrophysical Journal*, **423**, L9-L12.

Vogeley, M.S., Park, C., Geller, M.J., Huchra, J.P. & Gott, J.R. (1994). Topological analysis of the CfA redshift survey. *Astrophysical Journal*, **420**, 525-544.

Worsley, K.J. (1986). Confidence regions and tests for a change-point in a sequence of exponential family random variables. *Biometrika*, **73**, 91-104.

Worsley, K.J. (1994). Local maxima and the expected Euler characteristic of excursion sets of χ^2, F and t fields. *Advances in Applied Probability*, **26**, 13-42.

Worsley, K.J. (1995a). 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**, 640-669.

Worsley, K.J. (1995b). Boundary corrections for the expected Euler characteristic of excursion sets of random fields, with an application to astrophysics. *Advances in Applied Probability*, **27**, 943-959.

Worsley, K.J. (1996). The geometry of random images. *Chance*, **9(1)**, 27-40.

Worsley, K.J., Marrett, S., Neelin, P. & Evans, A.C. (1996). Searching scale space for activation in PET images. *Human Brain Mapping*, **4**, 74-90.

Worsley, K.J., Evans, A.C., Marrett, S. & Neelin, P. (1992). A three dimensional statistical analysis for CBF activation studies in human brain. *Journal of Cerebral Blood Flow and Metabolism*, **12**, 900-918.

Worsley, K.J. & Friston, K.J. (1995). Analysis of fMRI time-series revisited - again. *NeuroImage*, **2**, 173-181.

Figure 1. Blood flow (jagged lines) at two pixels chosen to illustrate
correlation (lower) and no correlation (upper) with the stimulus (dashed line)

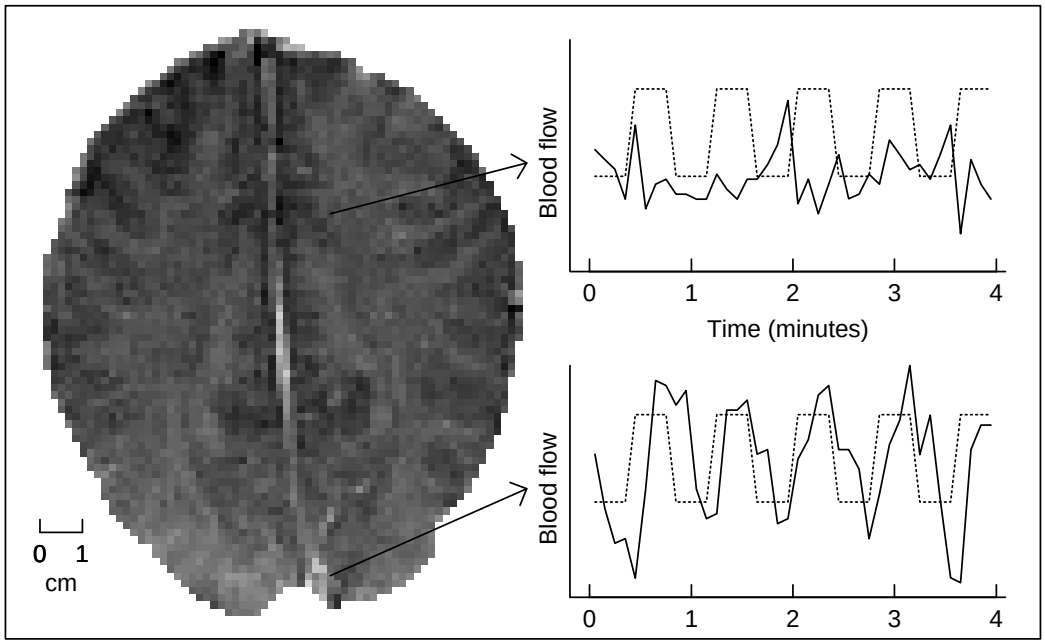

Figure 1: One of 40 fMRI images of the human brain while the subject was responding to an on-off visual stimulus repeated five times. The image shows an oblique slice through the visual cortex in the lower part of the image. Darker shading represents higher BOLD (blood oxygenation level dependent) response. The time series $Y(\mathbf{t}, t)$ (solid lines) at two pixels, one in the visual cortex and one outside, are also shown. There appears to be strong correlation with the stimulus (dashed lines) at the lower pixel in the visual cortex, but none at the upper pixel outside the visual cortex.

Figure 2. Scale space analysis of data + simulated signal

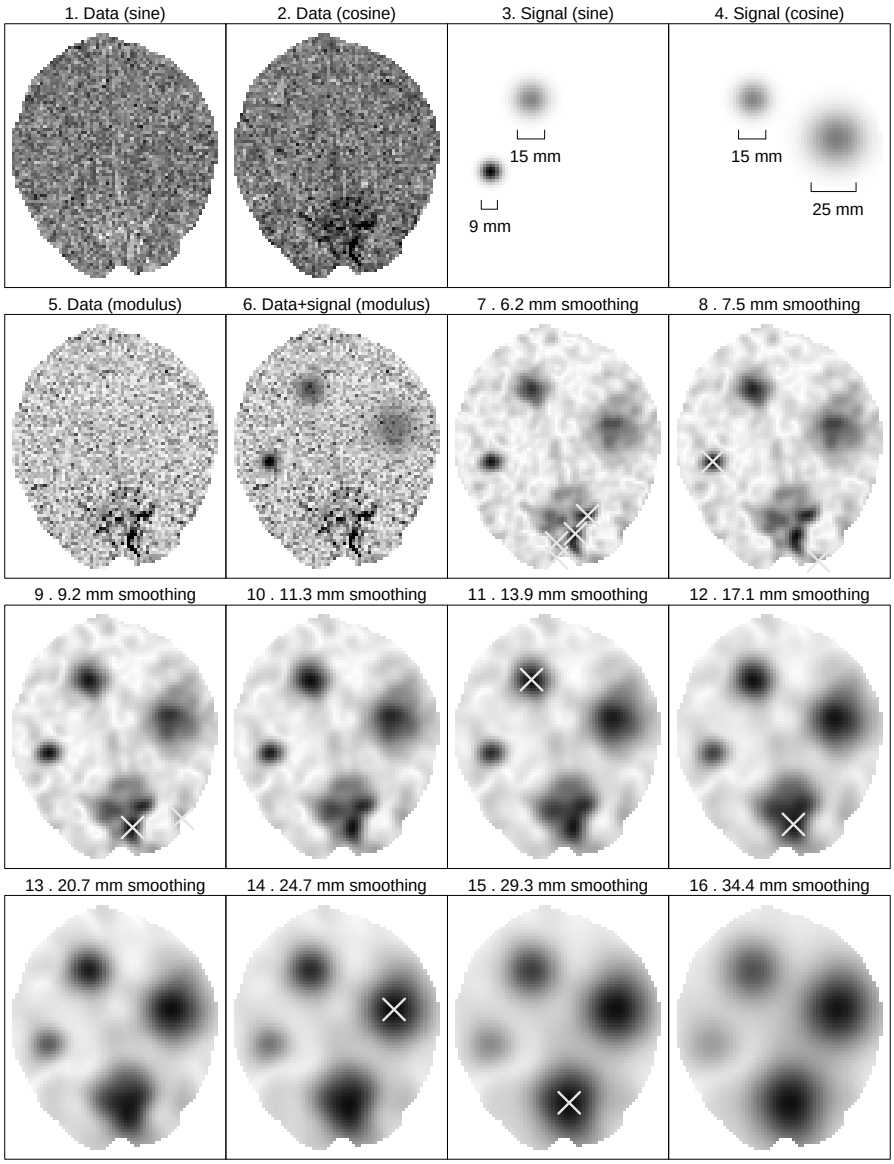

Figure 2: Unsmoothed cross-covariance of the 40 images with sine and cosine functions whose frequency matches that of the stimulus to give $dW_1(\mathbf{t})$ (panel 1) and $dW_2(\mathbf{t})$ (panel 2), respectively. Three simulated Gaussian-shaped signals with delays of 0°, 45° and 90°, and 9mm, 15mm and 25mm FWHM, respectively, were added to the sine (panel 3) and cosine (panel 4) components. The sum of squares of the sine and cosine components, $dW_1(\mathbf{t})^2 + dW_2(\mathbf{t})^2$, is shown without (panel 5) and with (panel 6) the simulated signal added. Panels 7–16 show the sum of squares of the sine and cosine components $X(\mathbf{t}, s)$, separately smoothed by a Gaussian-shaped filter varying from 6.2mm to 34.4mm FWHM. Scale space local maxima above the approximate 0.05 critical threshold for $X_{\max}$ ($b = 25.3$) are marked by white crosses. All three simulated signals are detected (panels 8, 11 and 14) at 7.5, 13.9, 24.7mm smoothing, close to their actual widths of 9, 15, 25mm, respectively. Several local maxima appear in the visual cortex, culminating in the global maximum of $X_{\max} = 208$ in panel 15 at 29.3mm FWHM smoothing.

Figure 3a. Excursion set at b=25.3, EC=6

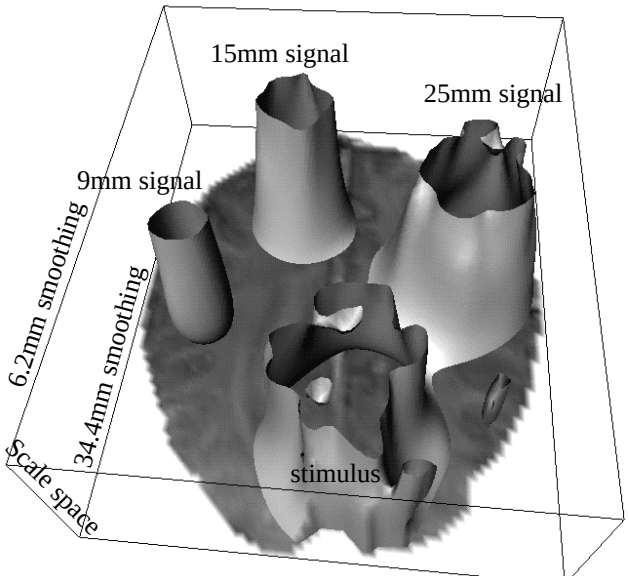

Figure 3b. Excursion set at b=175, EC=4

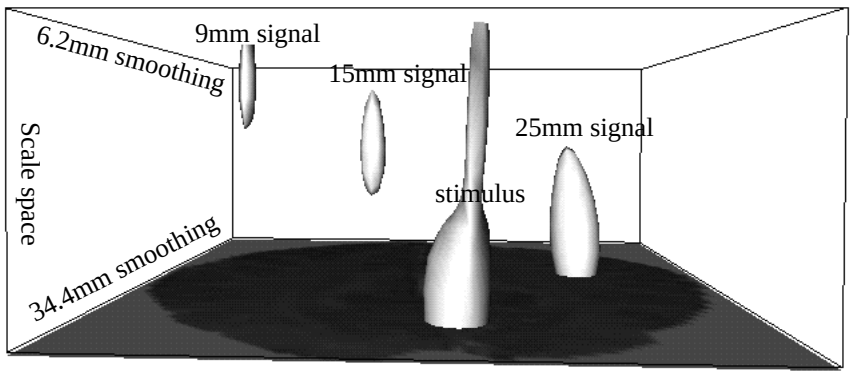

Figure 3: The scale space excursion set B_b above (a) $b = 25.3$, the approximate 0.05 critical threshold for $X_{\max}$, and (b) a higher threshold $b = 175$. The EC of the excursion set $\chi(B_b)$ is 6 in (a) and 4 in (b). The four components are clearly evident: the three simulated signals in the upper part of the figure and the activation in the visual cortex in the lower part. Note in (b) that each of the simulated signals is optimally detected at a filter width close to the width of the signal, as predicted by the Matched Filter Theorem. The stimulus is detected at all scales, suggesting a mix of sharp foci of activation embedded in a broader 'penumbra' of lesser activation (visible in Figure 5, panel 10).

Figure 4. Euler characteristic plot

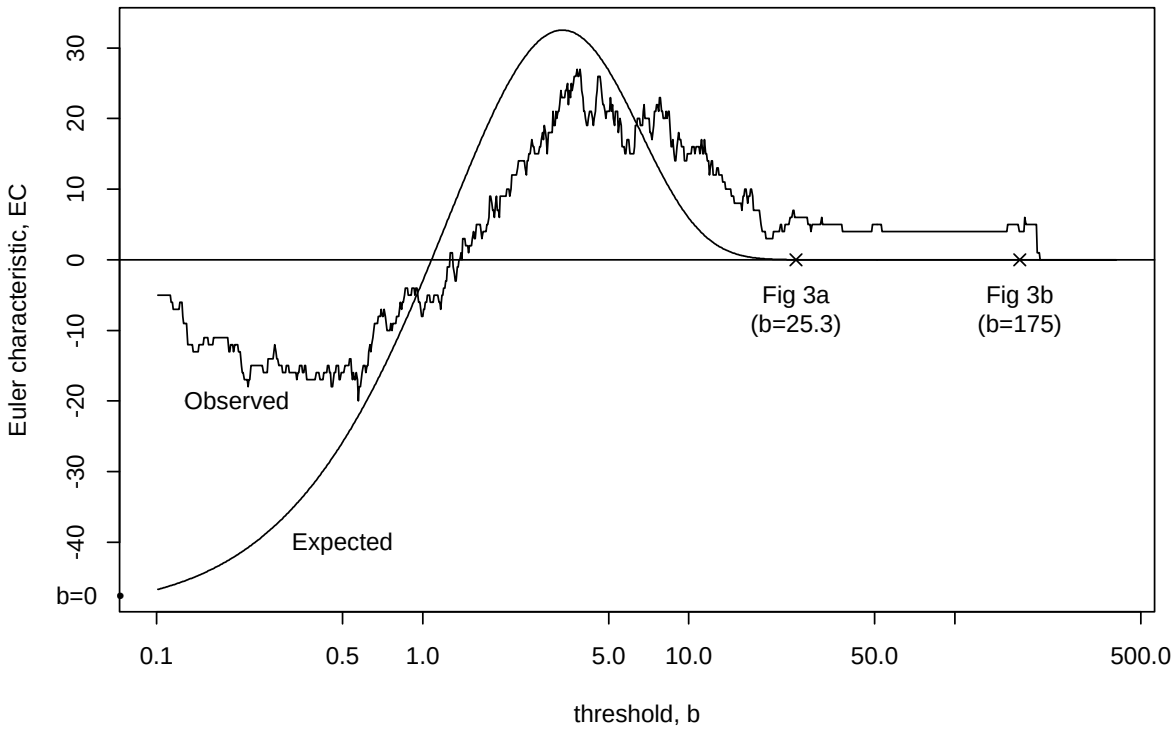

Figure 4: Plot of the observed and expected EC against threshold on a log scale. At high thresholds the EC is fairly stable at 4, well above expectation, corresponding to the three simulated signals and the visual cortex activation.