(12)

# TESTING FOR A SIGNAL WITH UNKNOWN LOCATION AND SCALE IN A STATIONARY GAUSSIAN RANDOM FIELD

David O. Siegmund
Keith J. Worsley

*TECHNICAL REPORT No. 477*
*JANUARY 7, 1994*

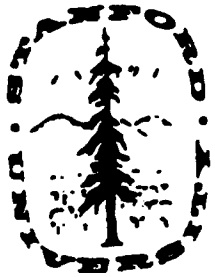
**Prepared Under Contract**
**N00014-92-J-1264 (NR-042-267)**
**FOR THE OFFICE OF NAVAL RESEARCH**





**Approved for public release; distribution unlimited**

**DEPARTMENT OF STATISTICS**
**STANFORD UNIVERSITY**
**STANFORD, CALIFORNIA 94305-4065**



# TESTING FOR A SIGNAL WITH UNKNOWN LOCATION AND SCALE IN A STATIONARY GAUSSIAN RANDOM FIELD

**David O. Siegmund**
**Keith J. Worsley**

*TECHNICAL REPORT No. 477*
*JANUARY 7, 1994*

**Prepared Under Contract**
**N00014-92-J-1264 (NR-042-267)**
**FOR THE OFFICE OF NAVAL RESEARCH**

**And The**
**NATIONAL SCIENCE FOUNDATION**
**Grant No. DMS91-044?2**

**Professor Herbert Solomon, Project Director**



**DEPARTMENT OF STATISTICS**
**STANFORD UNIVERSITY**
**STANFORD, CALIFORNIA 94305-4065**

# Testing for a signal with unknown location and scale in a stationary Gaussian random field

David O. Siegmund * Stanford University

Keith J. Worsley † McGill University

*AMS* 1980 *subject classifications.* Primary 60G60, 62M09; Secondary 60D05, 52A22.

*Key words and phrases.* Euler characteristic, integral geometry, image analysis, Gaussian fields, volume of tubes.

## SUMMARY

We suppose that our observations can be decomposed into a fixed signal plus random noise, where the noise is modelled as a particular stationary Gaussian random field in $N$-dimensional Euclidean space. The signal has the form of a known function centered at an unknown location and multiplied by an unknown amplitude, and we are primarily interested in a test to detect such a signal. There are many examples where the signal scale or width is assumed known, and the test is based on maximising a Gaussian random field over all locations in a subset of $N$-dimensional Euclidean space. The novel feature of this work is that the width of the signal is also unknown and the test is based on maximising a Gaussian random field in $N+1$-dimensions, $N$ dimensions for the location plus one dimension for the width. Two convergent approaches are used to approximate the null distribution: one based on the method of Knowles and Siegmund (1989), which uses a version of Weyl's (1939) tube formula for manifolds with boundaries, and the other based on some recent work by Worsley (1993b), which uses the Hadwiger characteristic of excursion sets as introduced by Adler (1981). Finally we compare the power of our method with one based on a fixed but perhaps incorrect signal width.

*Partially supported by NSF Grant DMS 91-04432 and ONR Contract N00014-92-J-1264 (NR-042-267)

†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

Given a fixed value $\mathbf{t}_0 \in C$, non-negative $\xi$, and positive $\sigma_0$, assume that the random field $\{Z(\mathbf{t}), \mathbf{t} \in \mathbb{R}^N\}$ satisfies

$$dZ(\mathbf{t}) = \xi \sigma_0^{-N/2} f[\sigma_0^{-1}(\mathbf{t} - \mathbf{t}_0)]dt + dW(\mathbf{t}). \tag{1.1}$$

Here $C$ is a subset of $N$ dimensional Euclidean space, e.g., a rectangle, $f$ is a square integrable function, which without loss of generality can be assumed to satisfy

$$\int f(\mathbf{t})^2 d\mathbf{t} = 1, \tag{1.2}$$

and $W$ is Gaussian white noise. The unknown parameter $(\xi, \mathbf{t}_0, \sigma_0)$ represents the amplitude, location and scale of the signal, which is usually positive, symmetric and unimodal. We shall be primarily interested in testing the hypothesis of no signal, i.e., that $\xi = 0$, and to a lesser extent in estimating the value of $\mathbf{t}_0$ when $\xi > 0$. A case of special interest is the Gaussian case, where

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

Although it is unrealistic to suppose that $Z$ is defined throughout $\mathbb{R}^N$, this assumption allows us to avoid problems of edge effects and may be a reasonable approximation when $\sigma_0$ is small compared to the distance between $\mathbf{t}_0$ and the boundary of the region where $Z$ is defined.

Models of at least approximately this form have been used in a number of scientific contexts. One is the model of Worsley *et al.* (1992) of blood flow changes in the human brain observed via positron emission tomography; here $N = 3$ and $C$ is the brain. Another is the model of Rabinowitz (1993) for the geographical clustering of disease incidence, where $N = 2$. Focus for the current research came originally from questions raised by John H. Cobb of the Physics Institute of the University of Oxford, who is involved in searching a portion of the sky for evidence of a point source of high activity against a background of Poisson, hence approximately Gaussian, white noise; here again $N = 2$.

Given an $N$ dimensional kernel $k$, we define the Gaussian process

$$X(\mathbf{t}, \sigma) = \sigma^{-N/2} \int k[\sigma^{-1}(\mathbf{h} - \mathbf{t})]dZ(\mathbf{h}). \tag{1.4}$$

Assuming that the value of $\sigma_0$ is known to lie in the interval $[\sigma_1, \sigma_2]$, we consider the test of $\xi = 0$ that rejects for large (positive) values of

$$X_{\max} = \max_{\mathbf{t},\sigma} X(\mathbf{t}, \sigma), \tag{1.5}$$

where the max extends over $\mathbf{t} \in C$ and $\sigma_1 \leq \sigma \leq \sigma_2$. If the shape of the signal, i.e., $f$, is known, we can take $k = f$. It can be shown (see below) that the log likelihood function is

$$\xi X(\mathbf{t}_0, \sigma_0) - \xi^2/2, \tag{1.6}$$

so the test defined by (1.5) is the likelihood ratio test. This is more or less obvious in the case that one actually observes the process $Z$ defined by (1.1). However, in many cases one can only observe the smoothed process (1.4). Then this statement is a consequence of the following general result. Let $\{Y(\mathbf{r}), \mathbf{r} \in S\}$ be a Gaussian random field with covariance function $R(\mathbf{r}_1, \mathbf{r}_2)$ and mean value function of the form $\mathrm{E}[Y(\mathbf{r})] = \xi R(\mathbf{r}, \mathbf{r}_0)$, parameterized by $\mathbf{r}_0 \in S$ and $\xi \in \mathbb{R}$. Let $Q_{\xi,\mathbf{r}_0}$ denote the distribution of $Y$. Then the log likelihood ratio is

$$\log dQ_{\xi,\mathbf{r}_0}/dQ_0 = \xi Y(\mathbf{r}_0) - \xi^2 R(\mathbf{r}_0, \mathbf{r}_0)/2.$$

If $S$ is finite, this is an easy direct calculation involving the multivariate normal density or an exercise in linear models (see Worsley, 1993a). For $S$ a complete, separable metric space the theory of reproducing kernel Hilbert spaces delivers the goods (cf. Parzen, 1961, Theorem 7A).

Sections 3 and 4 of this paper are concerned with approximate evaluation of the significance level of the test defined by (1.5), i.e., the probability when $\xi = 0$ that $X_{\max}$ exceeds a constant threshold, say $b$. First order approximations for this can easily be derived from the results going back to Belyaev and Pitaberg (1972) (see Adler, 1981, Theorem 6.9.1, p. 160) who give the the following. Suppose $Y(\mathbf{r})$ is a zero mean, unit variance, stationary random field defined on an interval $S \subset \mathbb{R}^n$, and let $Y_{\max} = \max_{\mathbf{r} \in S} Y(\mathbf{r})$ and

$$F(b) = |S| \det[\mathrm{Var}(\partial Y/\partial \mathbf{r})]^{1/2} b^{n-1} \phi(b)/(2\pi)^{n/2}, \tag{1.7}$$

where $|S|$ is the Lebesgue measure of $S$ and $\phi(b) = (2\pi)^{-1/2} \exp(-b^2/2)$. Then

$$\lim_{b \to \infty} \mathrm{P}\{Y_{\max} \geq b\}/F(b) = 1.$$

In our case $n = N + 1$, $Y(\mathbf{r}) = X(\mathbf{t}, \sigma)$, $S$ is the cartesian product of the the region $C$ and the interval $[\sigma_1, \sigma_2]$. It only remains to find the variance matrix of $\partial Y/\partial \mathbf{r}$, which in our case depends on $\sigma$, so the root determinant in (1.7) should be replaced by its average over $S$.

In practice, however, $F(b)$ can be a poor approximation to the exceedence probability of $X_{\max}$. In typical applications to medical images the range of $\sigma$ is small, and since $F(b)$ is proportional to the volume of $S$, then $F(b)$ approaches zero as $\sigma_2 - \sigma_1$ approaches zero. This is obviously unsatisfactory; a better approximation can be obtained by fixing $\sigma$ at say $\sigma_1$ and repeating the above argument for $n = N$. The same phenomenon occurs if we shrink the volume of $S$ to zero; the approximation $F(b)$ approaches zero and yet a better approximation can be obtained by fixing $\mathbf{t}$ and applying $n = 1$ dimensional theory (see Leadbetter, Lindgren and Rootzen, 1983). Clearly what is required is a higher order approximation, and we offer two approaches.

Our first approach in section 3 to the distribution of $X_{\max}$ is based on the Euler, or more specifically the Hadwiger, characteristic of excursion sets. The second approach in

section 4 is concerned with an extension of the method of Knowles and Siegmund (1989), which involves a version of Weyl's (1939) tube formula for manifolds with boundaries. These methods involve a substantial amount of geometric machinery to calculate a quantity related, but not equal, to the probability of interest. Thus it is reassuring that the two methods give essentially the same result and illuminate complementary aspects of the calculation. An interesting aspect of the method of tubes is that for the relevant Gaussian fields an important associated manifold has constant negative curvature.

Section 5 is concerned with the power of the likelihood ratio test We also compare the power of the likelihood ratio test, which involves a data dependent estimate of $\sigma_0$, as indicated in (1.5), with the analogous test based on an arbitrary fixed choice of $\sigma$, which in general differs from $\sigma_0$. Section 6 contains a simple example and an application to positron emission tomography images, and Section 7 contains some additional discussion.

# 2 Preliminaries

Under the null hypothesis of no signal we can write

$$X(\mathbf{t},\sigma) = \sigma^{-N/2} \int k[\sigma^{-1}(\mathbf{h}-\mathbf{t})]dW(\mathbf{h})$$

where $k(\mathbf{h})$, $\mathbf{h} \in \mathbb{R}^N$, is a smooth kernel with $\int k(\mathbf{h})^2 d\mathbf{h} = 1$; sufficient smoothness is discussed below. Then $X(\mathbf{t},\sigma)$ is a zero mean, unit variance Gaussian random field with correlation function

$$\mathrm{Cor}[X(\mathbf{t}_1,\sigma_1), X(\mathbf{t}_2,\sigma_2)] = (\sigma_1\sigma_2)^{-N/2} \int k[\sigma_1^{-1}(\mathbf{h}-\mathbf{t}_1)]k[\sigma_2^{-1}(\mathbf{h}-\mathbf{t}_2)]d\mathbf{h}.$$

Note that for fixed $\sigma$, $X(\mathbf{t},\sigma)$ is stationary in $\mathbf{t}$ and for fixed $\mathbf{t}$, $X(\mathbf{t},\sigma)$ is stationary in $\log\sigma$. Introduce $s = -\log\sigma$ and with a slight abuse of notation let $X(\mathbf{t},s) = X(\mathbf{t},\sigma)$. Then

$$\begin{aligned} \mathrm{Cor}[X(\mathbf{t}_1,s_1), X(\mathbf{t}_2,s_2)] &= e^{N(s_1+s_2)/2} \int k[(\mathbf{h}-\mathbf{t}_1)e^{s_1}]k[(\mathbf{h}-\mathbf{t}_2)e^{s_2}]d\mathbf{h} \qquad (2.1) \\ &= e^{N(s_2-s_1)/2} \int k[\mathbf{h}+(\mathbf{t}_2-\mathbf{t}_1)e^{s_1}]k[\mathbf{h}e^{s_2-s_1}]d\mathbf{h} \qquad (2.2) \end{aligned}$$

but note that $X(\mathbf{t},s)$ is not stationary in $(\mathbf{t},s)$.

We shall need the joint distribution of the derivatives of the field up to second order. Suppose $k(\mathbf{h}) = k(-\mathbf{h})$. Let $X = X(\mathbf{t},s)$, $X_s = \partial X/\partial s$, $\mathbf{X} = \partial X/\partial \mathbf{t}$ and $\mathbf{D} = \partial^2 X/\partial\mathbf{t}\partial\mathbf{t}'$, where prime denotes transpose. At a fixed point $(X, X_s, \mathbf{X}, \mathbf{D})$ is multivariate Gaussian with zero mean and variance matrix

$$\mathrm{Var}\begin{pmatrix} X \\ X_s \\ \mathbf{X} \\ \mathbf{D} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & -e^{2s}\mathbf{\Lambda} \\ 0 & \kappa & 0 & -e^{2s}\mathbf{\Lambda} \\ 0 & 0 & e^{2s}\mathbf{\Lambda} & 0 \\ -e^{2s}\mathbf{\Lambda} & -e^{2s}\mathbf{\Lambda} & 0 & \varepsilon \end{pmatrix}$$

where

$$\Lambda = \int \dot{\mathbf{k}}(\mathbf{h})\dot{\mathbf{k}}'(\mathbf{h})d\mathbf{h}. \qquad \kappa = \int [\mathbf{h}'\dot{\mathbf{k}}(\mathbf{h}) + (N/2)k]^2 d\mathbf{h}.$$

$\dot{\mathbf{k}}(\mathbf{h})$ denotes $\partial k(\mathbf{h})/\partial \mathbf{h}$ and $\varepsilon$ is such that the covariance between the $(i,j)$ element of $\mathbf{D}$ and the $(k,l)$ element of $\mathbf{D}$ is symmetric in $i,j,k,l$, $1 \le i,j,k,l \le N$ (see Adler, 1981, p. 114). For the special case of a Gaussian kernel $k(\mathbf{h}) = f(\mathbf{h})$ from (1.3) we have $\Lambda = \mathbf{I}/2$, where $\mathbf{I}$ is the identity matrix, and $\kappa = N/2$.

Deriving these covariances takes some trial and error; we illustrate the method for $\mathrm{Cov}(\mathbf{D}, X_s)$. From (2.1) we have

$$\begin{aligned}
\frac{\partial}{\partial \mathbf{t}_1}\mathrm{Cor}[X(\mathbf{t}_1,s_1),X(\mathbf{t}_2,s_2)] &= -e^{N(s_1+s_2)/2+s_1}\int \dot{\mathbf{k}}[(\mathbf{h}-\mathbf{t}_1)e^{s_1}]k[(\mathbf{h}-\mathbf{t}_2)e^{s_2}]d\mathbf{h} \\
&= -e^{N(s_1+s_2)/2+s_1}\int \dot{\mathbf{k}}[\mathbf{h}e^{s_1}]k[(\mathbf{h}+\mathbf{t}_1-\mathbf{t}_2)e^{s_2}]d\mathbf{h}, \\
\frac{\partial^2}{\partial \mathbf{t}_1\partial \mathbf{t}_1'}\mathrm{Cor}[X(\mathbf{t}_1,s_1),X(\mathbf{t}_2,s_2)] &= -e^{(N/2+1)(s_1+s_2)}\int \dot{\mathbf{k}}[\mathbf{h}e^{s_1}]\dot{\mathbf{k}}'[(\mathbf{h}+\mathbf{t}_1-\mathbf{t}_2)e^{s_2}]d\mathbf{h} \quad (2.3)
\end{aligned}$$

Differentiating (2.3) with respect to $s_2$, setting $\mathbf{t}_1 = \mathbf{t}_2$ and $s_1 = s_2 = s$, and changing the variable of integration we get

$$\mathrm{Cov}(\mathbf{D}, X_s) = -e^{2s}\int \dot{\mathbf{k}}(\mathbf{h})[\mathbf{h}'\ddot{\mathbf{k}}(\mathbf{h}) + (N/2+1)\dot{\mathbf{k}}'(\mathbf{h})]d\mathbf{h}. \qquad (2.4)$$

Changing the variable of integration in (2.3) first gives

$$\frac{\partial^2}{\partial \mathbf{t}_1\partial \mathbf{t}_1'}\mathrm{Cor}[X(\mathbf{t}_1,s_1),X(\mathbf{t}_2,s_2)] = -e^{(N/2+1)s_1-(N/2-1)s_2}\int \dot{\mathbf{k}}[\mathbf{h}e^{s_1-s_2}]\dot{\mathbf{k}}'[\mathbf{h}+(\mathbf{t}_1-\mathbf{t}_2)e^{s_2}]d\mathbf{h},$$

then differentiating with respect to $s_2$, setting $\mathbf{t}_1 = \mathbf{t}_2$ and $s_1 = s_2 = s$, and changing the variable of integration we get

$$\mathrm{Cov}(\mathbf{D}, X_s) = e^{2s}\int [\ddot{\mathbf{k}}(\mathbf{h})\mathbf{h} + (N/2-1)\dot{\mathbf{k}}(\mathbf{h})]\dot{\mathbf{k}}'(\mathbf{h})d\mathbf{h}. \qquad (2.5)$$

Taking the average of both sides of (2.4) and (2.5), and noting that $\mathbf{D}$ is symmetric, gives

$$\mathrm{Cov}(\mathbf{D}, X_s) = -e^{2s}\int \dot{\mathbf{k}}(\mathbf{h})\dot{\mathbf{k}}'(\mathbf{h})d\mathbf{h} = -e^{2s}\Lambda.$$

The conditions we shall require on the smoothness of the kernel $k(\mathbf{h})$ will ensure that realisations of $Y(\mathbf{r}) = X(\mathbf{t},s)$ have almost surely continuous derivatives up to second order. Exact conditions on the moduli of continuity of $Y(\mathbf{r})$ are given by Adler (1981), Theorem 5.2.2, page 106. A simpler sufficient condition for Gaussian fields, using Theorem 3.4.1 of Adler (1981), page 60. is the following. Let $Y_{ij}(\mathbf{r})$ be the second derivative of $Y(\mathbf{r})$ with respect to components $i$ and $j$ of $\mathbf{r}$. Then we shall require that for some $\epsilon > 0$ and for all $\mathbf{r}_1, \mathbf{r}_2 \in S$

$$\mathrm{E}\{[Y_{ij}(\mathbf{r}_1) - Y_{ij}(\mathbf{r}_2)]^2\} = O(|\log||\mathbf{r}_1 - \mathbf{r}_2|||^{-(1+\epsilon)})$$

for all $\mathbf{r}_1, \mathbf{r}_2 \in S$, and all $i, j = 1, \ldots, n$. This condition is satisfied if. for example, all third derivatives of $Y(\mathbf{r})$ have finite variance, which in turn is assured if the integral of the product of any pair of third derivtives of $k(\mathbf{h})$ times sixth powers of components of $\mathbf{h}$ is finite. This latter condition is met. for example, by the Gaussian kernel (1.3).

# 3 Hadwiger characteristic approach

## 3.1 Motivation

Let $A_b = \{\mathbf{r} : Y(\mathbf{r}) \geq b\}$ be the excursion set of $Y(\mathbf{r})$ above $b$, and let $\chi(A_b)$ be the Euler or Euler-Poincaré characteristic of $S \cap A_b$. The Euler characteristic counts the number of connected components of a set, minus the number of "holes." As the threshold level $b$ increases Adler (1981) Theorem 6.4.1, p. 136. shows that the "holes" in $A_b$ tend to disappear and that we are left with isolated regions each of which contains just one local maximum. Thus for large $b$ the presence of holes is a rare occurrence and the Euler characteristic approaches the number of local maxima above $b$. For even larger $b$ near the global maximum $Y_{\max}$ the Euler characteristic takes the value 0 if $Y_{\max} < b$ and 1 if $Y_{\max} \geq b$. Hasofer (1978) shows that

$$\mathrm{P}\{Y_{\max} \geq b\} \approx \mathrm{P}\{\chi(A_b) \geq 1\} \approx \mathrm{E}[\chi(A_b)] \tag{3.1}$$

as $\mathrm{P}\{\chi(A_b) > 1\} \to 0$ for $b \to \infty$, and so the expected Euler characteristic approximates the exceedence probability of $Y_{\max}$. The advantage of the Euler characteristic is that its expectation can be found exactly; the only remaining point is how well this approximates $\mathrm{P}\{Y_{\max} \geq b\}$.

Some justification for this is as follows. Let $I_b = (Y_{\max} \geq b)$ be the indicator function for $Y_{\max}$, where a logical expression in parentheses takes the value one if the expression is true and zero otherwise (Knuth, 1992), so that $\mathrm{P}\{Y_{\max} \geq b\} = \mathrm{E}(I_b)$. Provided $S$ is connected and itself has no "holes" then the Euler characteristic of the excursion set $\chi(A_b)$ approximates the indicator function $I_b$ both for large $b$ (as shown above), and for small $b$ (since $A_b = S$ for $b$ small enough, and $\chi(S) = 1$). Moreover this property holds for any value of $n$, so that if the region $S$ is "squashed" to form a lower-dimensional manifold embedded in $\mathbb{R}^n$, then $\chi(A_b)$ continues to approximate $I_b$ in the same way. In particular, if $n = 1$, then the two are identical for all $b$. Thus the Euler characteristic has the correct behaviour for all shapes of $S$.

## 3.2 Definition

We now define a more convenient characteristic, the Hadwiger charactersitc, which is defined on different classes of sets than the Euler characteristic, but which equals the Euler

characterisitic within the domain of definition of both. The advantage for us is that it has an iterative definition, given below, which is more amenable to statistical analysis.

Let $C$ be a compact subset of $\mathbb{R}^N$ with a smooth boundary, let $I = [s_1, s_2]$ be the closed interval between $s_1$ and $s_2$ and let $S = C \times I$. Once again we define the excursion set $A_b$ of $X(\mathbf{t}, s)$ above a threshold $b$ to be the set of points in $S$ where $X(\mathbf{t}, s)$ exceeds $b$:

$$A_b = \{(\mathbf{t}, s) \in S : X(\mathbf{t}, s) \geq b\}.$$

Define the *Hadwiger characteristic* $\psi(A)$ of a *basic complex* $A \subset S$ iteratively as follows. For $N = 0$, let $\psi(A)$ be the number of disjoint intervals in $A$. For $N > 0$, let

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

where $\mathcal{E}_u = C \times \{u\}$ and

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

The Hadwiger characteristic is the only characteristic which satisfies the following additivity property: if $A$, $B$, $A \cup B$ and $A \cap B$ are basic complexes then

$$\psi(A \cup B) = \psi(A) + \psi(B) - \psi(A \cap B).$$

If $X(\mathbf{t}, s)$ is sufficiently regular, as defined by Adler (1981), chapter 3, then the excursion set $A_b$ is almost surely a basic complex.

A crucial step in deriving statistical properties of excursion 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. To do this, let $\psi_{\mathbf{V}}$ be the contribution to the point set representation from the interior of $S$, let $\psi_{\mathbf{E}}$ be the contribution from $\partial C \times I$, the "edges" in scale space, and let $\psi_{\mathbf{B}}$ be the contribution from $\mathcal{E}_{s_1}$, the "base" of $S$. Then with probability one

$$\psi(A_b) = \psi_{\mathbf{V}} + \psi_{\mathbf{E}} + \psi_{\mathbf{B}}.$$

## 3.3 $N = 1$

Here $\mathbf{t}$, $\mathbf{X}$, $\mathbf{D}$ and $\mathbf{\Lambda}$ are scalars, and set $\mathbf{\Lambda} = \lambda$, say. For $\mathbf{t} \in \partial C$, let $X_\perp$ be the derivative of $X$ with respect to $\mathbf{t}$ in the inwards direction to $C$. Then

$$\begin{aligned} \psi_{\mathbf{V}} &= \sum_S (X_s > 0)[(\mathbf{D} < 0) - (\mathbf{D} > 0)](\mathbf{X} = 0)(X = b), \\ \psi_{\mathbf{E}} &= \sum_{\partial C \times I} (X_s > 0)(X_\perp < 0)(X = b). \end{aligned}$$

We can evaluate the expectation of the point set representations following the methods used to prove Theorem 5.1.1 of Adler (1981, p. 95) to get

$$\begin{aligned} \mathrm{E}(\psi_{\mathrm{V}}) &= -|C| \int_{s_1}^{s_2} \mathrm{E}(X_s^+ \mathbf{D} | \mathbf{X} = 0, X = b) \phi_1(0, b) ds. \\ \mathrm{E}(\psi_{\mathrm{E}}) &= \sum_{\partial C} \int_{s_1}^{s_2} \mathrm{E}[X_s^+ (X_\perp < 0) | X = b] \phi(b) ds, \end{aligned}$$

where $\phi_1(\mathbf{x}, x)$ is the density of $(\mathbf{X}, X)$. Taking expectations first over $\mathbf{D}$ conditional on $X_s$, $\mathbf{X}$ and $X$ we get

$$-\mathrm{E}(X_s^+ \mathbf{D} | \mathbf{X} = 0. X = b) = e^{2s} \lambda \mathrm{E}[X_s^+ (b + X_s/\kappa)] = e^{2s} \lambda [b\kappa^{1/2}/(2\pi)^{1/2} + 1/2]$$

and so for the interior of $S$ we have

$$\begin{aligned} \mathrm{E}(\psi_{\mathrm{V}}) &= |C| \int_{s_1}^{s_2} e^s ds \lambda^{1/2} [b\kappa^{1/2}/(2\pi)^{1/2} + 1/2] \phi(b)/(2\pi)^{1/2} \\ &= |C|(e^{s_2} - e^{s_1})(\lambda\kappa)^{1/2} b\phi(b)/(2\pi) + (|C|/2)(e^{s_2} - e^{s_1}) \lambda^{1/2} \phi(b)/(2\pi)^{1/2}. \end{aligned}$$

For the "edges," each connected component of $C$ contributes two points to $\partial C$, each with $X_\perp$ taken in opposite directions; since $(X_\perp < 0) + (X_\perp > 0) = 1$ and the number of such points is $2\psi(C)$, then

$$\mathrm{E}(\psi_{\mathrm{E}}) = \psi(C)(s_2 - s_1)\kappa^{1/2} \phi(b)/(2\pi)^{1/2}.$$

For the "base," we have from Worsley (1993b),

$$\mathrm{E}(\psi_{\mathrm{B}}) = |C| e^{s_1} \lambda^{1/2} \phi(b)/(2\pi)^{1/2} + \psi(C)[1 - \Phi(b)],$$

where $\Phi(b) = \mathrm{P}\{X < b\} = \int_{-\infty}^{b} \phi(x) dx$. Putting these together, substuting $\sigma_1 = e^{-s_2}$, $\sigma_2 = e^{-s_1}$ for the limits on $\sigma$, and re-arranging, we have

$$\begin{aligned} \mathrm{E}[\psi(A_b)] &= |C|(\sigma_1^{-1} - \sigma_2^{-1})(\lambda\kappa)^{1/2} b\phi(b)/(2\pi) \\ &+ (|C|/2)(\sigma_1^{-1} + \sigma_2^{-1}) \lambda^{1/2} \phi(b)/(2\pi)^{1/2} \\ &+ \psi(C) \log(\sigma_2/\sigma_1) \kappa^{1/2} \phi(b)/(2\pi)^{1/2} \\ &+ \psi(C)[1 - \Phi(b)]. \end{aligned} \tag{3.2}$$

## 3.4 $N > 1$

At a point $(\mathbf{t}, s) \in \partial C \times I$, let $\mathbf{t}_\perp \in \mathbb{R}^N$ be the inside normal to $\partial C$ and let the columns of $(\mathbf{t}_\perp, \mathbf{A})$, $\mathbf{A}$ an $N \times N-1$ matrix, be an orthonormal basis for $\mathbb{R}^N$. Let $X_\perp = \mathbf{t}_\perp' \mathbf{X}$ be the derivative of $X$ normal to $S$, let $\mathbf{X}_{\mathrm{T}} = \mathbf{A}'\mathbf{X}$ be the $N-1$-vector of derivatives of $X$ tangent to $S$, and let $\mathbf{D}_{\mathrm{T}} = \mathbf{A}'\mathbf{D}\mathbf{A}$ be the $N-1 \times N-1$ second derivative matrix of $X$ tangent to $S$. Let $\mathbf{c}_{\mathrm{T}}$ be the $N-1 \times N-1$ curvature matrix of $\partial C$ at $\mathbf{t}$. Finally for a symmetric matrix

$\mathbf{M}$, let $\chi(\mathbf{M})$ take the value +1 if $\mathbf{M}$ has an even number of negative eigen values, and -1 if $\mathbf{M}$ has an odd number of negative eigen values. including multiplicities. Then generalising the results of Adler(1981. p. 84) we have

$$\begin{aligned}\psi_{\mathrm{V}} &= \sum_{S}(X_s > 0)\chi(\mathbf{D})(\mathbf{X} = \mathbf{0})(X = b),\\ \psi_{\mathrm{E}} &= \sum_{\partial C \times I}(X_s > 0)\chi(\mathbf{D}_{\mathrm{T}} + \mathbf{c}_{\mathrm{T}}X_{\perp})(X_{\perp} < 0)(\mathbf{X}_{\mathrm{T}} = \mathbf{0})(X = b).\end{aligned}$$

We evaluate the expectation of the point set representations following the methods used to prove Theorem 5.1.1 of Adler (1981, p. 95) to give

$$\begin{aligned}\mathrm{E}(\psi_{\mathrm{V}}) &= (-1)^N|C|\int_{s_1}^{s_2}\mathrm{E}[X_s^+\det(\mathbf{D})|\mathbf{X} = \mathbf{0}, X = b]\phi_{\mathbf{I}}(\mathbf{0}, b)ds,\\ \mathrm{E}(\psi_{\mathrm{E}}) &= (-1)^{N-1}\int_{s_1}^{s_2}\oint_{\partial C}\mathrm{E}[X_s^+\det(\mathbf{D}_{\mathrm{T}} + \mathbf{c}_{\mathrm{T}}X_{\perp})(X_{\perp} < 0)|\mathbf{X}_{\mathrm{T}} = \mathbf{0}, X = b]\phi_{\mathbf{A}}(\mathbf{0}, b)d\mathbf{A}'\mathbf{t}ds,\end{aligned}$$

where $\phi_{\mathbf{A}}(\mathbf{x}, x)$ is the density of $(\mathbf{A}'\mathbf{X}, X)$. We shall evaluate $\mathrm{E}(\psi_{\mathrm{V}})$ by first taking expectations over $\mathbf{D}$ conditional on $X_s$, $\mathbf{X}$ and $X$. Let $v = (1 + 1/\kappa)^{1/2}$ and let $\mathrm{He}_N(x)$ be the Hermite polynomial of order $N$ in $x$. Following the arguments of Adler(1981, p. 114) we get

$$\mathrm{E}[\det(\mathbf{D})] = e^{2Ns}\det(\mathbf{\Lambda})(-v)^N\mathrm{E}\{\mathrm{He}_N[(b + X_s/\kappa)/v]\}. \tag{3.3}$$

For $\mathrm{E}(\psi_{\mathrm{E}})$, we shall first take expectations over $X_{\perp}$ conditional on $\mathbf{D}_{\mathrm{T}}$, $X_s$, $\mathbf{X}_{\mathrm{T}}$ and $X$. Let $\mathbf{\Lambda}_{\mathrm{T}} = \mathbf{A}'\mathbf{\Lambda}\mathbf{A}$, so that the distribution of $\mathbf{D}_{\mathrm{T}}^* = \mathbf{\Lambda}_{\mathrm{T}}^{-1/2}\mathbf{D}_{\mathrm{T}}\mathbf{\Lambda}_{\mathrm{T}}^{-1/2}$ is invariant under rotations in the tangent plane. Then letting $\mathbf{c}_{\mathrm{T}}^* = \mathbf{\Lambda}_{\mathrm{T}}^{-1/2}\mathbf{c}_{\mathrm{T}}\mathbf{\Lambda}_{\mathrm{T}}^{-1/2}$ we can write

$$\det(\mathbf{D}_{\mathrm{T}} + \mathbf{c}_{\mathrm{T}}X_{\perp}) = \det(\mathbf{\Lambda}_{\mathrm{T}})\det(\mathbf{D}_{\mathrm{T}}^* + \mathbf{c}_{\mathrm{T}}^*X_{\perp}).$$

For any $m \times m$ matrix $\mathbf{M}$, define $\mathrm{detr}_j(\mathbf{M})$ to be the sum of the determinants of all $j \times j$ principal minors of $\mathbf{M}$, $1 \le j \le m$, and one if $j = 0$. Then since any principal minor of $\mathbf{D}_{\mathrm{T}}^*$ has the same distribution we can expand in powers of $X_{\perp}$ to get

$$\mathrm{E}[\det(\mathbf{D}_{\mathrm{T}}^* + \mathbf{c}_{\mathrm{T}}^*X_{\perp})] = \sum_{j=0}^{N-1}\mathrm{E}[\det(\mathbf{D}_{\mathrm{T}}^{*(N-1-j)})]\mathrm{detr}_j(\mathbf{c}_{\mathrm{T}}^*)X_{\perp}^j.$$

where $\mathbf{D}_{\mathrm{T}}^{*(N-1-j)}$ is any $N-1-j \times N-1-j$ principal minor of $\mathbf{D}_{\mathrm{T}}^*$. Now from (3.3) we have

$$\mathrm{E}[\det(\mathbf{D}_{\mathrm{T}}^{*(N-1-j)})] = e^{2(N-1-j)s}(-v)^{N-1-j}\mathrm{E}\{\mathrm{He}_{N-1-j}[(b + X_s/\kappa)/v]\}.$$

Since

$$\mathrm{E}[X_{\perp}^j(X_{\perp} < 0)] = \mathrm{Var}(X_{\perp})^{j/2}(-1)^j2^{(j-1)/2}\Gamma[(j+1)/2]/(2\pi)^{1/2}, \tag{3.4}$$

we get

$$\begin{aligned}\mathrm{E}[\det(\mathbf{D}_{\mathrm{T}}+\mathbf{c}_{\mathrm{T}}X_{\perp})(X_{\perp}<0)] &= (-v)^{N-1}e^{2(N-1)s}\det(\mathbf{\Lambda}_{\mathrm{T}}) \\ &\times \sum_{j=0}^{N-1} e^{-2js}v^{-j}\mathrm{E}\{\mathrm{He}_{N-1-j}[(b+X_s/\kappa)/v]\}\mathrm{detr}_j(\mathbf{c}_{\mathrm{T}}^{*}) \\ &\times \mathrm{Var}(X_{\perp})^{j/2}2^{(j-1)/2}\Gamma[(j+1)/2]/(2\pi)^{1/2}. \qquad (3.5)\end{aligned}$$

In principle we can expand (3.3) and (3.5) in powers of $X_s$, multiply by $X_s^{+}$, take expectations using (3.4), and find a general expression for $\mathrm{E}[\psi(A_b)]$. The algebra is very tedious so we shall now consider the cases of most interest where $N = 2$ or $N = 3$ and the process is isotropic in $\mathbf{t}$ for fixed $s$.

## 3.5 $N = 2$

We can evaluate (3.3) by expanding and using (3.4):

$$v^2\mathrm{E}\{X_s^{+}\mathrm{He}_2[(b+X_s/\kappa)/v]\} = \mathrm{E}\{X_s^{+}[(b+X_s/\kappa)^2-(1+1/\kappa)]\} = \kappa^{1/2}(b^2-1+1/\kappa)/(2\pi)^{1/2}+b,$$

so that, substituting and integrating over $s$, we get

$$\mathrm{E}(\psi_{\mathrm{V}}) = (|C|/2)e^{2(s_2-s_1)}\det(\mathbf{\Lambda})^{1/2}[\kappa^{1/2}(b^2-1+1/\kappa)/(2\pi)^{1/2}+b]\phi(b)/(2\pi).$$

It is hard to simplify (3.5) without making the further assumption that the field is isotropic in $\mathbf{t}$ for fixed $s$, so that we can write $\mathbf{\Lambda} = \lambda\mathbf{I}$. Then conditional on $X_s$, $\mathbf{X} = \mathbf{0}$ and $X = b$

$$\mathrm{E}[\det(\mathbf{D}_{\mathrm{T}}+\mathbf{c}_{\mathrm{T}}X_{\perp})(X_{\perp}<0)] = -e^{2s}\lambda(b+X_s/\kappa)/2 - \mathbf{c}_{\mathrm{T}}e^{s}\lambda^{1/2}/(2\pi)^{1/2}$$

so that conditional on $\mathbf{X} = \mathbf{0}$ and $X = b$

$$\mathrm{E}[X_s^{+}\det(\mathbf{D}_{\mathrm{T}}+\mathbf{c}_{\mathrm{T}}X_{\perp})(X_{\perp}<0)] = -e^{2s}\lambda[b\kappa^{1/2}/(2\pi)^{1/2}+1/2]/2 - \mathbf{c}_{\mathrm{T}}e^{s}(\lambda\kappa)^{1/2}/(2\pi).$$

Now since $\oint_{\partial C}\mathbf{c}_{\mathrm{T}}d\mathbf{t}_{\mathrm{T}} = 2\pi\psi(C)$ we get

$$\begin{aligned}\mathrm{E}(\psi_{\mathrm{E}}) &= \int_{s_1}^{s_2}\{|\partial C|e^{s}\lambda^{1/2}[b\kappa^{1/2}/(2\pi)^{1/2}+1/2]/2+\psi(C)\kappa^{1/2}\}\phi(b)/(2\pi)^{1/2}ds \\ &= \{|\partial C|e^{s_2-s_1}\lambda^{1/2}[b\kappa^{1/2}/(2\pi)^{1/2}+1/2]/2+\psi(C)(s_2-s_1)\kappa^{1/2}\}\phi(b)/(2\pi)^{1/2}.\end{aligned}$$

Finally, from Worsley (1993b), we have

$$\mathrm{E}(\psi_{\mathrm{B}}) = |C|e^{2s_1}\lambda b\phi(b)/(2\pi)+(|\partial C|/2)e^{s_1}\lambda^{1/2}\phi(b)/(2\pi)^{1/2}+\psi(C)[1-\Phi(b)].$$

Combining these and substuting $\sigma_1 = e^{-s_2}$, $\sigma_2 = e^{-s_1}$ for the limits on $\sigma$. we have

$$\begin{aligned}
\mathrm{E}[\psi(A_b)] &= (|C|/2)(\sigma_1^{-2} - \sigma_2^{-2})\lambda\kappa^{1/2}(b^2 - 1 + 1/\kappa)\phi(b)/(2\pi)^{3/2} \\
&+ (|C|/2)(\sigma_1^{-2} + \sigma_2^{-2})\lambda b\phi(b)/(2\pi) \\
&+ (|\partial C|/2)(\sigma_1^{-1} - \sigma_2^{-1})(\lambda\kappa)^{1/2} b\phi(b)/(2\pi) \\
&+ (|\partial C|/4)(\sigma_1^{-1} + \sigma_2^{-1})\lambda^{1/2}\phi(b)/(2\pi)^{1/2} \\
&+ \psi(C)\log(\sigma_2/\sigma_1)\kappa^{1/2}\phi(b)/(2\pi)^{1/2} \\
&+ \psi(C)[1 - \Phi(b)]. \qquad (3.6)
\end{aligned}$$

## 3.6 $\mathcal{N} = 3$

We can evaluate (3.3) by expanding and using (3.4):

$$\begin{aligned}
v^3\mathrm{E}\{X_s^+ \mathrm{He}_3[(b + X_s/\kappa)/v]\} &= \mathrm{E}\{X_s^+[(b + X_s/\kappa)^3 - 3(1 + 1/\kappa)(b + X_s/\kappa)]\} \\
&= \kappa^{1/2}(b^3 - 3b + 3b/\kappa)/(2\pi)^{1/2} + 3(b^2 - 1)/2,
\end{aligned}$$

so that. substituting and integrating over $s$. we get

$$\mathrm{E}(\psi_V) = (|C|/3)e^{3(s_2 - s_1)}\det(\mathbf{\Lambda})^{1/2}[\kappa^{1/2}(b^3 - 3b + 3b/\kappa)/(2\pi)^{1/2} + 3(b^2 - 1)/2]\phi(b)/(2\pi)^{3/2}.$$

It is hard to simplify (3.5) without making the further assumption that the field is isotropic in $\mathbf{t}$ for fixed $s$, so that we can write $\mathbf{\Lambda} = \lambda\mathbf{I}$. Then conditional on $X_s$, $\mathbf{X} = \mathbf{0}$ and $X = b$

$$\begin{aligned}
\mathrm{E}[\det(\mathbf{D}_T + \mathbf{c}_T X_\perp)(X_\perp < 0)] &= e^{4s}\lambda^2[(b + X_s/\kappa)^2 - (1 + 1/\kappa)]/2 \\
&+ \mathrm{trace}(\mathbf{c}_T)e^{3s}\lambda^{3/2}(b + X_s/\kappa)/(2\pi)^{1/2} + \det(\mathbf{c}_T)e^{2s}\lambda/2,
\end{aligned}$$

so that conditional on $\mathbf{X} = \mathbf{0}$ and $X = b$

$$\begin{aligned}
\mathrm{E}[X_s^+ \det(\mathbf{D}_T + \mathbf{c}_T X_\perp)(X_\perp < 0)] &= e^{4s}\lambda^2[\kappa^{1/2}(b^2 - 1 + 1/\kappa)/(2\pi)^{1/2} + b]/2 \\
+\mathrm{trace}(\mathbf{c}_T)e^{3s}\lambda^{3/2}[b\kappa^{1/2}/(2\pi)^{1/2} + 1/2]/(2\pi)^{1/2} &+ (\det(\mathbf{c}_T)/2)e^{2s}\lambda\kappa^{1/2}/(2\pi)^{1/2}.
\end{aligned}$$

Define the mean curvature of $\partial C$ (see Santaló, 1976, p. 222) to be

$$H(\partial C) = \oiint_{\partial C} \mathrm{trace}(\mathbf{c}_T)d\mathbf{t}_T/2.$$

Now since $\oint_{\partial C} \det(\mathbf{c}_T)d\mathbf{t}_T = 4\pi\psi(C)$ we get

$$\begin{aligned}
\mathrm{E}(\psi_E) &= \int_{s_1}^{s_2} \{|\partial C|e^{2s}\lambda[\kappa^{1/2}(b^2 - 1 + 1/\kappa)/(2\pi)^{1/2} + b]/2 \\
&+ 2H(\partial C)e^s\lambda^{1/2}[b\kappa^{1/2}/(2\pi)^{1/2} + 1/2]/(2\pi)^{1/2} + \psi(C)(2\pi\kappa)^{1/2}\}\phi(b)/(2\pi)ds \\
&= \{(|\partial C|/2)(e^{2s_2} - e^{2s_1})\lambda[\kappa^{1/2}(b^2 - 1 + 1/\kappa)/(2\pi)^{1/2} + b]/2 \\
&+ 2H(\partial C)(e^{s_2} - e^{s_1})\lambda^{1/2}[b\kappa^{1/2}/(2\pi)^{1/2} + 1/2]/(2\pi)^{1/2} + \psi(C)(s_2 - s_1)(2\pi\kappa)^{1/2}\}\phi(b)/(2\pi).
\end{aligned}$$

Finally, from Worsley (1993b), we have

$$\begin{aligned} \mathrm{E}(\psi_{\mathrm{B}}) &= |C|e^{3s_1}\lambda^{3/2}(b^2-1)\phi(b)/(2\pi)^{3/2} + (|\partial C|/2)e^{2s_1}\lambda b\phi(b)/(2\pi) \\ &+ [H(\partial C)/\pi]e^{s_1}\lambda^{1/2}\phi(b)/(2\pi)^{1/2} + \psi(C)[1-\Phi(b)]. \end{aligned}$$

Combining these and substuting $\sigma_1 = e^{-s_2}$, $\sigma_2 = e^{-s_1}$ for the limits on $\sigma$. we have

$$\begin{aligned} \mathrm{E}[\psi(A_b)] &= |C|(1/3)(\sigma_1^{-3}-\sigma_2^{-3})\lambda^{3/2}\kappa^{1/2}(b^3-3b+3b/\kappa)\phi(b)/(2\pi)^2 \\ &+ (|C|/2)(\sigma_1^{-3}+\sigma_2^{-3})\lambda^{3/2}(b^2-1)\phi(b)/(2\pi)^{3/2} \\ &+ (|\partial C|/4)(\sigma_1^{-2}-\sigma_2^{-2})\lambda\kappa^{1/2}(b^2-1+1/\kappa)\phi(b)/(2\pi)^{3/2} \\ &+ (|\partial C|/4)(\sigma_1^{-2}+\sigma_2^{-2})\lambda b\phi(b)/(2\pi) \\ &+ [H(\partial C)/\pi](\sigma_1^{-1}-\sigma_2^{-1})(\lambda\kappa)^{1/2}b\phi(b)/(2\pi) \\ &+ [H(\partial C)/(2\pi)](\sigma_1^{-1}+\sigma_2^{-1})\lambda^{1/2}\phi(b)/(2\pi)^{1/2} \\ &+ \psi(C)\log(\sigma_2/\sigma_1)\kappa^{1/2}\phi(b)/(2\pi)^{1/2} \\ &+ \psi(C)[1-\Phi(b)]. \end{aligned} \tag{3.7}$$

It is straightforward, though cumbersome, to generalise to piece-wise smooth sets $C$. Let $\partial C_{\mathrm{S}}$ be the smooth part of $\partial C$, and let $\partial C_{\mathrm{E}}$ be the smooth curves or "edges" that bound the smooth parts of $\partial C$. Let $\delta$ be the internal angle between the normals to the two parts of $\partial C_{\mathrm{S}}$ on either side of $\partial C_{\mathrm{E}}$, and let $\mathbf{t}_{\mathrm{E}}$ be a unit vector tangent to $\partial C_{\mathrm{E}}$. Then the above results will hold with $H(\partial C)$ replaced by

$$H(\partial C) = \left(\oint\!\!\!\oint_{\partial C_{\mathrm{S}}} \mathrm{trace}(\mathbf{c}_{\mathrm{T}})d\mathbf{t}_{\mathrm{T}} + \int_{\partial C_{\mathrm{E}}} \delta d\mathbf{t}_{\mathrm{E}}\right) \Big/ 2.$$

If $\partial C$ is smooth everywhere, so that the second term is zero, then $H(\partial C)$ is the mean curvature of $\partial C$ as before. If $C$ is a polyhedron, so that the first term is zero, then $H(\partial C)$ is half the sum of the lengths of the edges of $C$ multiplied by their angular deficiency. If $C$ is convex, let $\Delta(\partial C)$ be the average, over all rotations, of the maximum perpendicular distance between two parallel planes that touch $\partial C$; this is known in stereology as the mean caliper diameter of $C$. Then it can be shown that $\Delta(\partial C) = H(\partial C)/(2\pi)$ (see Santaló, 1976, p. 226). Values of $H(\partial C)$ for some common geometric solids are given by Santaló (1976, p. 229).

## 3.7 Manifolds in $\mathbb{R}^3$

We can now find the result for a piece-wise smooth two dimensional manifold $C$ embedded in $\mathbb{R}^3$ by thickening it slightly, applying the above, and taking the limit as the thickness tends to zero. ¿From (3.7) we see that the result is identical to (3.6) in two dimensions, obviously a special case when $C$ is flat. Thus no matter how $C$ is folded or even creased, the

expectation of the Hadwiger characteristic is the same, even though the correlation structure on $C$ depends on the folding and is not necessarily stationary unless $C$ is flat.

If $C$ is a piece-wise smooth two dimensional surface homeomorphic to the surface of a sphere, $|\partial C| = 0$ and $\psi(C) = 2$. This result could be useful for directional data, such as the astronomical example, which is modelled as a random field on the surface of a sphere. However the fact that $C$ contains a "hole" means that $\mathrm{E}[\psi(A_b)]$ may not be a good approximation to $\mathrm{P}\{X_{\max} \geq b\}$ for small $b$.

# 4 Volume of tubes

An alternative to the methods of Section 3 is the differential geometric method of Weyl (1939). See also Naiman (1990), Knowles and Siegmund (1989), Sun (1993), and Siegmund and Zhang (1994). As will become apparent below, it is not obvious that we are in general evaluating the same quantity as was calculated in Section 3, so it is reassuring to obtain the same answer in particular cases. However, our main goal is not only to reproduce the answer by different means, but also to try to add some insight. The method of tubes breaks the complicated calculation into a large number of small pieces whose relative importance can to some extent be assessed without performing the entire calculation. Hence the method might be suitable for application when one would prefer a simple approximate result to a more complicated exact one. Alternatively, since each of the pieces has its own geometric meaning, the method illuminates some mysterious expressions obtained by the formal evaluations of Section 3. To this end we show the geometric origin of the somewhat enigmatic term involving $1/\kappa$, which appears in the display (3.6) for the case $N = 2$ and in the corresponding formulas for larger values of $N$, but which does not appear in (3.1) for $N = 1$.

Since there is no particular advantage in reparameterizing $\sigma$ as $\exp(-s)$ as in Sections 2 and 3, we consider the Gaussian $X(\mathbf{t}, \sigma)$ defined in (1.4), which has the Karhunen-Loève expansion (cf. Loève (1963, p. 478)):

$$X(\mathbf{t}, \sigma) = \sum_k \gamma_k(\mathbf{t}, \sigma) Z_k, \tag{4.1}$$

where the $Z$'s are independent standard normal and

$$\mathrm{Cov}[X(\mathbf{t}_1, \sigma_2), X(\mathbf{t}_2, \sigma_2)] = \sum_k \gamma_k(\mathbf{t}_1, \sigma_1)\gamma_k(\mathbf{t}_2, \sigma_2). \tag{4.2}$$

If the series in (4.1) terminates after a finite number of terms, the following discussion yields a very precise approximation to the tail of the distribution of $X_{\max}$ when $\xi = 0$. It presumably applies approximately otherwise, but a delicate interchange of limits is involved. See Sun (1993) for a detailed discussion of the nature of this approximation in a special case.

We omit discussion of these technicalities and henceforth proceed formally as if the series (4.1) contains $m$ terms. It turns out that ultimately all our calculations are intrinsic to the manifold $M$ defined below, so the quantities we evaluate are the same whether or not the series terminates after finitely many terms.

If (4.1) is divided and multiplied by the norm of the vector $\mathbf{Z} = (Z_1, \ldots, Z_m)'$, it takes the form of the product of $\langle \gamma(\mathbf{t}, \sigma), \mathbf{U} \rangle$ and $||\mathbf{Z}||$, where $\gamma(\mathbf{t}, \sigma) = (\gamma_1(\mathbf{t}, \sigma), \ldots, \gamma_m(\mathbf{t}, \sigma))'$ $\mathbf{U}$ is uniformly distributed on the unit sphere in $m$ dimensions and $||\mathbf{Z}||$ is independently distributed as a $\chi$ random variable with $m$ degrees of freedom. By conditioning on $||\mathbf{Z}||$, we can reduce the problem of evaluating the tail of the distribution of (1.5) when $\xi = 0$ to that of evaluating

$$\mathrm{P}\{\max_{\mathbf{t},\sigma} \langle \gamma(\mathbf{t},\sigma), \mathbf{U} \rangle > w\} \tag{4.3}$$

for values of $w$ close to one. Moreover, by (4.2) the vector $\gamma$ defines a parameterized submanifold of the unit sphere. so (4.3) can be interpreted geometrically as the volume of the tube of geodesic radius $\cos^{-1}(w)$ about this manifold, divided by the volume of the unit sphere. We seek to calculate the volume of that tube.

Let $M = \{\gamma(\mathbf{t},\sigma) : \mathbf{t} \in C, \sigma_1 \leq \sigma \leq \sigma_2\}$. The metric tensor (first fundamental form) of the manifold $M$ is given by the $(N+1) \times (N+1)$ matrix with entries $\langle \gamma_i, \gamma_j \rangle$, where the subscripts denote differentiation with respect to the corresponding argument. Assuming $k$ is symmetric about zero in each of its arguments and invariant under permutation of its arguments, we see from the covariance calculations of Section 2 that the matrix is diagonal: the first $N$ diagonal entries equal $\lambda/\sigma^2$, and the final entry is $\kappa/\sigma^2$. It will be useful below to note that in the special case $\lambda = \kappa = 1$, this is the metric tensor of hyperbolic space having constant curvature -1 (e.g., Boothby (1986, p. 404)).

For $N = 1$, Corollary 2 of Knowles and Siegmund (1989) (corrected for minor errors of calculation) yields

$$\mathrm{P}\{\max \langle \gamma(\mathbf{t},\sigma), \mathbf{Z} \rangle > b\} = |M| b\phi(b)/(2\pi) + (|\partial M|/2)\phi(b)/(2\pi)^{1/2} + \chi(M)[1 - \Phi(b)][1 + o(1)] \tag{4.4}$$

as $b \to \infty$. Here $|M|$ is the area of the manifold, $|\partial M|$ is the length of its boundary, and $\chi(M)$ is its Euler-Poincarè characteristic (equal to 1 if, as we assume for simplicity, $C$ is an interval of real numbers). In the present case the area element of the manifold is $(\lambda\kappa)^{1/2} dt d\sigma/\sigma^2$, so $|M| = |C|(\sigma_1^{-1} - \sigma_2^{-1})(\lambda\kappa)^{1/2}$; also $|\partial M| = |C|(\sigma_1^{-1} + \sigma_2^{-1})\lambda^{1/2} + \kappa^{1/2} \log(\sigma_2/\sigma_1)$, so the right hand sides of (4.4) and (3.2) are the same.

It is interesting to note that in the derivation of (4.4) the Euler characteristic $\chi(M)$ arises via an application of the Gauss-Bonnet Theorem as $(2\pi)^{-1}$ times the sum of three terms: (i) the Gaussian curvature of the manifold integrated with respect to the area element, (ii) the geodesic curvature of the boundary of $M$ integrated with respect to arc length, and (iii) the sum of the angles through which the tangent to the boundary of $M$ rotates at the corners.

See, for example, Millman and Parker (1977, p. 185). In the present case it may be shown by standard elementary calculations based on the metric tensor given above that the Gaussian curvature is the constant $-1/\kappa$. For fixed $\mathbf{t}$, as a function of $\sigma$ the curve $\gamma(\mathbf{t}, \sigma)$ is geodesic, so its geodesic curvature is 0. For $\sigma = \sigma_\nu$, as a function of $\mathbf{t}$ the geodesic curvature of $\gamma(\mathbf{t}, \sigma_\nu)$ is $(-1)^{(\nu-1)}/\kappa^{1/2}$, provided the boundary is oriented in the customary counterclockwise order in the $(\mathbf{t}, \sigma)$ plane. We omit the details, since the case $N = 2$ discussed below is similar in principle and more complcated in detail. Hence, before adding the integral over the manifold of the Gaussian curvature and the integral over the boundary of the geodesic curvature, we have obtained two terms involving $1/\kappa^{1/2}$, which then sum to zero.

We now consider the case $N = 2$. As the arguments of Weyl (1939) show, the first and second order terms again involve the volume of the manifold $M$ and the area of $\partial M$, and are easily evaluated. (See Knowles and Siegmund (1989) or Naiman (1990) for details.) For example, in the present case the volume element is $dV = \lambda\kappa^{1/2} dt_1 dt_2 d\sigma/\sigma^3$, so

$$|M| = |C|(\sigma_2^{-2} - \sigma_1^{-2})\lambda\kappa^{1/2}/2. \tag{4.5}$$

The terms in $b^2\phi(b)$ and $b\phi(b)$ are the same as in (3.6). For many applications these terms will provide an adequate approximation by themselves.

The other terms involve the curvature of the manifold, the curvature of the boundary, and lower dimensional boundary corrections. A complete derivation involves more advanced tools of differential geometry. We sketch the possibilities here by deriving the term involving $1/\kappa$ in the first line of (3.6).

By the methods of Knowles and Siegmund (1989) one can see that this term arises from

$$\int_M J dV + 2\int_{\partial M} H_\partial dA_\partial. \tag{4.6}$$

The expressions in (4.6) are as follows. Let $\mathbf{n}_\nu, \nu = 1, \cdots, m - N - 1$ be mutually orthogonal unit normals to the tangent space of $M$. Then $J$ is the sum over $\nu$ of the sum of pairwise products of the eigenvalues of the Weingarten map (Millman and Parker (1977, p. 125)) in the direction $\mathbf{n}_\nu$; $dV$ is the volume element of the manifold; $H_\partial$ is the geodesic mean curvature of the boundary, i.e., the mean curvature in the direction of a vector which is normal to the boundary of $M$, but lying in the tangent space of $M$ pointing into the interior of $M$; and $dA_\partial$ is the area element of the boundary. (The expression (4.6) appears in a brief discussion of the case of a three dimensional manifold in a preliminary version of Knowles and Siegmund (1989), but not in the published version.)

Up to a constant multiple, $J$ is the scalar curvature of the manifold. It can be identified as such and laboriously calculated directly from the metric tensor of the manifold. See Sun (1993) for a discussion and summary of the algorithm. (For a manifold of the simple structure of $M$, we suggest a different method below.) The result is $J = -3/\kappa$, a constant,

so by (4.5)

$$\int_M J dV = -3\lambda\kappa^{-1/2}|C|(\sigma_1^{-2} - \sigma_2^{-2})/2. \tag{4.7}$$

We now consider the geodesic mean curvature of the boundary. It turns out that we are primarily interested in the image of $\partial C \times \{\sigma_1, \sigma_2\}$, since the rest of the boundary gives rise to one of the other terms in (3.6). Hence assume that $\sigma$ is fixed and consider the surface $\gamma(\mathbf{t}, \sigma)$ as a function of $\mathbf{t}$ with unit normal $\sigma\gamma_3/\kappa^{1/2}$. From the covariance calculations in Section 2 it follows easily that $\langle \gamma_{ii}, \gamma_3 \rangle = \lambda/\sigma^3$ for $i = 1, 2$. Hence since the metric tensor is diagonal, the matrix of the Weingarten map (Millman and Parker (1977, p. 125)) has diagonal entries $\langle \gamma_{ii}, \gamma_3 \rangle / (||\gamma_3|| ||\gamma_{ii}||^2) = [\sigma\lambda/(\kappa^{1/2}\sigma^3)]/(\lambda/\sigma^2) = \kappa^{-1/2}$, so $2H = 2\kappa^{-1/2}$. In order to maintain our convention that $H$ be calculated with respect to an inward pointing normal, its sign should be negative at $\sigma = \sigma_2$ and positive at $\sigma = \sigma_1$. Since the area element is $\lambda\sigma^{-2} dt_1 dt_2$, it follows that the integral of twice the geodesic mean curvature over the image of $\partial C \times \{\sigma_1, \sigma_2\}$ is

$$\int 2H_\partial dA_\partial = 2\lambda\kappa^{-1/2}|C|(\sigma_1^{-2} - \sigma_2^{-2}). \tag{4.8}$$

Addition of (4.7) and (4.8) yields the coefficient associated with the $1/\kappa$ term on the first line of (3.6).

The geodesic mean curvature at a point $p$ on the image of $\partial C \times (\sigma_1, \sigma_2)$ leads to another term in (3.6). For simplicity we assume that $C$ has a smooth boundary. The point $p$ can be represented parametrically by $p = \alpha(t, \sigma)$, where for each fixed value of $\sigma$, as a function of $t$, $\alpha$ is a unit speed curve. Tangent vectors to the surface $\alpha$ are $\partial\alpha/\partial\sigma = \gamma_3$ and $\partial\alpha/\partial t = \lambda^{-1/2}\sigma[\gamma_1 \cos(\theta) + \gamma_2 \sin(\theta)]$. The relevant (unit) normal is $\mathbf{n} = \lambda^{-1/2}\sigma[-\gamma_1 \sin(\theta) + \gamma_2 \cos(\theta)]$. Obviously $\theta$ is the angle of rotation carrying $\gamma_1, \gamma_2$ into $\partial\alpha/\partial t, \mathbf{n}$. By differentiating the relation $||\gamma_3||^2 = \kappa/\sigma^2$ with respect to $t_i$ and $\langle \gamma_3, \gamma_i \rangle = 0$ with respect to $\sigma$, we see that $\langle \gamma_{33}, \gamma_i \rangle = 0$ for $i = 1, 2$. It follows that one diagonal element of the second fundamental form, and since the metric tensor is diagonal one diagonal element of the matrix of the Weingarten map is zero. The other diagonal entry is $\langle \partial^2\alpha/\partial t^2, \mathbf{n} \rangle$, which is by definition the geodesic curvature of the curve $\alpha$ in the surface $\gamma(\mathbf{t}, \sigma)$ considered as a function of $\mathbf{t}$ with $\sigma$ fixed. Hence the trace of Weingarten map is $2H_\partial = k_g$, the geodesic curvature of $\alpha$. The area element at $p$ is given by $dA_\partial = (\kappa)^{1/2}\sigma^{-1} dt d\sigma$. Since the surface $\gamma(\mathbf{t}, \sigma)$ for fixed $\sigma$ has as metric tensor a multiple of the identity, its Gaussian curvature is zero. Hence the integral of the geodesic curvature of the boundary with respect to arc length is $2\pi$ times the Euler characteristic of the surface, or equivalently the Euler characteristic of $C$. Hence the integral of $2H_\partial$ over this part of $\partial M$ yields $2\pi\psi(C)\kappa^{1/2}\log(\sigma_2/\sigma_1)$, which appears in the third line of (3.6).

The use of differential forms simplifies certain calculations associated with the volume of tubes. It is a particularly attractive alternative in the present case of a diagonal metric tensor, since it allows us to calculate directly and easily the curvature $J$ without necessarily

identifying it as a multiple of the scalar curvature of the manifold. Although these calculations should be regarded as standard, there does not seem to be a convenient reference, and it is not difficult to proceed from first principles, at least in low dimensional problems. A useful general reference is Boothby (1986).

Since our main goal here is insight, we shall consider the case of a tube about a manifold embedded in $m$-dimensional Euclidean space. The added features of the calculation for a tube about a manifold embedded in the unit sphere are discussed in Knowles and Siegmund (1989). Except for a brief remark at the end of this section, we shall also restrict ourselves to the case of present interest, namely a three dimensional manifold. Although the methods work also in higher dimensions, the computational complications can become quite substantial.

We denote by $e_i, i = 1, \cdots, m$ an orthonormal moving frame for $\mathbb{R}^m$, with the first three elements a frame for the tangent space of the manifold $M$ and the rest orthogonal to $M$. Let $\{\omega^i\}$ and $\{\omega_i^j\}$ denote the dual forms and connection forms respectively. As usual we assume the $\omega$'s are restricted to the tangent space of the manifold, so $\omega^4 = \cdots = \omega^m = 0$. A point on the manifold satisfies $dp = \sum_1^3 \omega^i e_i$. Recall also that

$$de_i = \sum_{j=1}^{m} \omega_i^j e_j, \tag{4.9}$$

$$d\omega_i^k = \sum_{j=1}^{m} \omega_i^j \wedge \omega_j^k \tag{4.10}$$

and $\omega_i^k = -\omega_k^i$.

For all sufficiently small $a$, a point in the tube of radius $a$ about $M$ (except for points whose closest point in the manifold is on the boundary of the manifold) can be parameterized as

$$q = p + \sum_{4}^{m} t_i e_i, \tag{4.11}$$

where $\sum t_i^2 \leq a^2$. We want to identify 1-forms $\{\tilde{\omega}^i, i = 1, \cdots, m\}$ such that $dq = \sum_1^m \tilde{\omega}^i e_i$. Then the volume of the tube can be obtained by integrating the volume element

$$dV(q) = \tilde{\omega}^1 \wedge \cdots \wedge \tilde{\omega}^m. \tag{4.12}$$

Differentiating (4.11) and applying the structural equation (4.9), we see after collecting terms that $\tilde{\omega}_i = \omega_i - \sum_{j=4}^m t_j \omega_i^j$ for $i = 1, 2, 3$ and $\tilde{\omega}_i = dt_i - \sum_{j=4}^m t_j \omega_i^j$ for $4 \leq j \leq m$. In taking the wedge product of the $\tilde{\omega}^i$, we note that since the manifold $M$ is three dimensional any wedge product involving more than 3 among the various $\omega^i$ and $\omega_i^j$ vanishes. Hence for $4 \leq i \leq m$ the sums $\sum_{j=4}^m t_j \omega_i^j$ can be neglected. It follows that the volume element (4.12) can be expanded as a sum of wedge products of the following form. Each term contains $dt_4 \wedge \cdots \wedge dt_m$. One term contains $\omega^1 \wedge \cdots \wedge \omega^3$. When integrated over the product of $M$

and the ball of radius $a$. it yields the volume of the manifold multiplied by the volume of the ball. There are also terms containing some of the $t_j$ raised to the first or the third power. These terms vanish when integrated over the ball of radius $a$. Finally there is an expression of the form

$$\sum_{j=4}^{m} t_j^2[\omega^1 \wedge \omega_2^j \wedge \omega_3^j + \omega_1^j \wedge \omega^2 \wedge \omega_3^j + \omega_1^j \wedge \omega_2^j \wedge \omega^3] \wedge dt_4 \wedge \cdots \wedge dt_m. \tag{4.13}$$

Integration of (4.13) over the ball of radius $a$ yields the same constant for all values of $j$, so there remains the problem of evaluating the integral over the manifold of

$$\sum_{j=4}^{m} [\omega^1 \wedge \omega_2^j \wedge \omega_3^j + \omega_1^j \wedge \omega^2 \wedge \omega_3^j + \omega_1^j \wedge \omega_2^j \wedge \omega^3]. \tag{4.14}$$

Since this is a 3-form on a 3 dimensional manifold, it must be a scalar multiple of the volume element $\omega^1 \wedge \omega^2 \wedge \omega^3$. From $\omega^1 \wedge \omega^2 \wedge \omega^3(e_1, e_2, e_3) = 1$ it follows that to evaluate this scalar, it suffices to evaluate (4.14) at $(e_1, e_2, e_3)$, and since $\omega^i(e_k) = \delta_{i,k}$, this evaluation yields a sum over $j$ of a sum of three determinants of $2 \times 2$ matrices whose $i, k$ element is $\omega_i^j(e_k)$. In terms of the curvature forms

$$\Omega_i^k = d\omega_i^k - \sum_{l=1}^{3} \omega_i^l \wedge \omega_l^k \tag{4.15}$$

$(i, k = 1, 2, 3)$ (cf. Boothby (1986, p. 390), which by (4.10) equals $-\sum_{j=4}^{m} \omega_i^j \wedge \omega_k^j$, this evaluation of (4.14) becomes

$$-\sum_{1 \le i \le k \le 3} \Omega_i^k(e_i, e_k). \tag{4.16}$$

Finally, to evaluate (4.16), we follow with minor changes the example in Boothby (1986, p. 408), which deals with the special case of our manifold with $\lambda = \kappa = 1$. We find that $\omega_i^k = \kappa^{-1/2}(\delta_{3,k}\omega^i - \delta_{3,i}\omega^k)$ $(i, k = 1, 2, 3)$, and hence $\Omega_i^k = -\omega_i^3 \wedge \omega_3^k = \kappa^{-1}\omega^i \wedge \omega^k$. Hence (4.16) equals $-3/\kappa$, which is exactly the value of $J$ given above, as it should be.

**Remarks.** Evaluation of the geodesic mean curvature of the boundary by this method is also easy. For example, in the case of fixed $\sigma$, we put $e_3 = \gamma_3/||\gamma_3||$ (subject to a correct selection of the sign). A calculation similar to that given above shows that the desired quantity is $\omega^1 \wedge \omega_2^3 + \omega_1^3 \wedge \omega^2$, evaluated at $e_1, e_2$. This equals $\omega_2^3(e_2) + \omega_1^3(e_1) = 2\kappa^{-1/2} = 2H$, as above.

The case $N = 3$ can be handled similarly, although now there are some additional terms to consider. The most interesting, although in the end it makes no contribution to the final formula, involves the Gauss-Bonnet integrand for the four dimensional manifold. It arises from the 4-form $\Omega_1^2 \wedge \Omega_3^4$. As in the case $N = 1$, where the integral of the Gaussian curvature over the manifold is exactly canceled by the integral of the geodesic curvature of the boundary, the integral over the manifold of this 4-form is canceled by the integral over the boundary faces where $\sigma = \sigma_1$ or $\sigma_2$ of a three form involving the sum of twice

$\omega_1^4 \wedge \omega_2^4 \wedge \omega_3^4$ and $\omega_1^4 \wedge \Omega_2^3 - \omega_2^4 \wedge \Omega_1^3 + \Omega_1^2 \wedge \omega_3^4$. These forms are defined in terms of the frame $e_1, \cdots . e_4$ adapted to the manifold with $e_1, e_2, e_3$ spanning the tangent space of the boundary and $e_4 = \gamma_4/||\gamma_4||$ perpendicular to the boundary.

# 5 Power

Suppose now that a signal is present, so we observe the field satisfying

$$dZ(\mathrm{t}) = \sigma_0^{-N/2} \xi f[\sigma_0^{-1}(\mathrm{t} - \mathrm{t}_0)]dt + dW(\mathrm{t}),$$

where $f$ is a positive, smooth function, $\mathrm{t}_0$ is the unknown location of the signal, $\xi, \sigma_0$ are nuisance parameters and $dW$ is white noise. After smoothing with the kernel $\sigma^{-N/2}k[\sigma^{-1}(\mathrm{h} - \mathrm{t})]$, where $k$ is chosen to equal $f$ and $\sigma$ to equal $\sigma_0$ insofar as possible. we are interested in the Gaussian field

$$X(\mathrm{t}, \sigma) = \sigma^{-N/2} \int k[\sigma^{-1}(\mathrm{h} - \mathrm{t})]dZ(\mathrm{h}), \tag{5.1}$$

which has mean value

$$\mu(\mathrm{t}, \sigma) = \xi(\sigma\sigma_0)^{-N/2} \int f[\sigma_0^{-1}(\mathrm{h} - \mathrm{t}_0)]k[\sigma^{-1}(\mathrm{h} - \mathrm{t})]d\mathrm{h} \tag{5.2}$$

and the covariance behavior described in Section 2. The power of the test suggested in Section 1 is

$$\mathrm{P}\{X_{\max} > b\}. \tag{5.3}$$

Both (5.2) and (5.3) depend on the unknown parameters $\xi, \mathrm{t}_0, \sigma_0$, although this dependency is suppressed in the notation. If the range of $\sigma$ includes the true value $\sigma_0$, we expect most of the probability (5.3) to arise from the probability that $X(\mathrm{t}_0, \sigma_0)$ exceeds $b$, which equals $1 - \Phi(b - \mu_0)$, where $\mu_0 = \mu(\mathrm{t}_0, \sigma_0)$ and $\Phi$ is the standard normal distribution function. Hence we begin with the decomposition of (5.3) as

$$1 - \Phi(b - \mu_0) + \int_{\{X(\mathrm{t}_0, \sigma_0) < b\}} \mathrm{P}\{X_{\max} > b | X(\mathrm{t}_0, \sigma_0)\}dP. \tag{5.4}$$

We consider below two different situations regarding the choices of $k$ and $\sigma$: (i) $k = f$ and $\sigma$ is chosen adaptively as discussed in Sections 2-4, and (ii) $k$ and $\sigma$ are arbitrary, but fixed, and in general differ from $f$ and $\sigma_0$. In the former case we assume that $\sigma_1 < \sigma_0 < \sigma_2$.

We begin with consideration of case (i). Note that in this case $\mu_0 = \xi$. Also, from the form (1.6) of the likelihood function it follows that $X(\mathrm{t}_0, \sigma_0)$ is sufficient for $\xi$ and hence the conditional probability in (5.4) can be evaluated assuming that $\xi = 0$. To emphasize this situation we write $P_0$ in place of $P$. Our approximation is based on the following considerations that have proved useful in related contexts (e.g., Siegmund (1985, pp. 202-203), James, James, and Siegmund (1987)). If $b$ and $\xi$ are large, then large values of the

sample paths of $X$ are likely to occur near to $\mathbf{t}_0, \sigma_0$. The conditional probability in (5.4) will be small unless $X(\mathbf{t}_0, \sigma_0)$ is close to $b$, say $X(\mathbf{t}_0, \sigma_0) = b - y$, with $y$ close to 0. Hence if the sample paths of the field are to cross the level $b$. they must do so in a small neighborhood of the value $\mathbf{t}_0, \sigma_0$. Moreover, as noted by various authors (e.g., Adler (1981, p. 157), Aldous (1989, p. 65), Leadbetter, Lindgren Rootzén (1983. pp. 201 ff.), if $X$ takes on a large value at $\mathbf{t}_0, \sigma_0$, then in a small neighborhood of that point it will behave very much like a quadratic function whose random part will come from the linear term. This suggests the local approximation

$$X(\mathbf{t}, \sigma) \approx (b - y) + \mathbf{u}'\dot{X}(\mathbf{t}_0, \sigma_0) + \mathbf{u}'\mathrm{E}_0[\ddot{X}(\mathbf{t}_0, \sigma_0)|X(\mathbf{t}_0, \sigma_0) = b - y]\mathbf{u}/2, \tag{5.5}$$

where $\mathbf{u} = ((\mathbf{t} - \mathbf{t}_0)', \sigma - \sigma_0)'$. The right hand side of (5.5) will exceed $b$ for some value of $\mathbf{u}$ if and only if its maximum does. A straightforward maximization of the right hand side of (5.5) yields

$$b - y - \dot{X}(\mathbf{t}_0, \sigma_0)'\{\mathrm{E}_0[\ddot{X}(\mathbf{t}_0, \sigma_0)|X(\mathbf{t}_0, \sigma_0) = b - y]\}^{-1}\dot{X}(\mathbf{t}_0, \sigma_0)/2, \tag{5.6}$$

which after an evaluation of the indicated conditional expectation of the Hessian of $X$ yields

$$b - y + \dot{X}(\mathbf{t}_0, \sigma_0)'\Sigma^{-1}\dot{X}(\mathbf{t}_0, \sigma_0)/[2(b - y)], \tag{5.7}$$

where

$$\Sigma = -\mathrm{E}_0[X(\mathbf{t}_0, \sigma_0)\ddot{X}(\mathbf{t}_0, \sigma_0)] = \mathrm{E}_0[\dot{X}(\mathbf{t}_0, \sigma_0)\dot{X}(\mathbf{t}_0, \sigma_0)']. \tag{5.8}$$

The quadratic form in (5.7) is independent of $X(\mathbf{t}_0, \sigma_0)$ and is distributed as a $\chi^2$ random variable with $n = N + 1$ degrees of freedom. Hence the conditional probability in (5.4) is approximately, in the obvious notation,

$$\mathrm{P}\{\chi_n^2 > 2y(b - y)\}. \tag{5.9}$$

Since we are assuming that $b$ is large and expect that the important values of $y$ are close to 0, we propose to replace $b - y$ by $b$ in (5.9) when we substitute into (5.4). Likewise we neglect the term involving $y^2$ in $\phi(b - \xi - y)$. Then the integral in (5.4) takes the form of a Laplace transform, which after an integration by parts is easily shown to equal

$$1 - \Phi(b - \xi) + \phi(b - \xi)[1 - (b/\xi)^{d/2}]/(\xi - b). \tag{5.10}$$

**Remarks.** Except for the smoothness and behavior for large $\mathbf{t}$ necessary to permit the expansion (5.5) and the integration by parts implicit in (5.8), there are essentially no conditions imposed on $f$. It appears that with a more careful analysis the preceding argument requires that $X$ be only once continuously differentiable, not twice differentiable, although the stronger assumption is satisfied in the applications discussed below.

We now turn to case (ii). Although the conditional probability in (5.4) now depends on the underlying parameters. the argument given above continues to hold with suitable modifications. In particular, in the following analogue of (5.5) the gradient vector $\dot{X}$ and Hessian matrix $\ddot{X}$ involve differentiation with respect to the $N$ spatial variables, but not the scale factor $\sigma$, which is constant. Instead of (5.5) we have

$$X(\mathbf{t},\sigma) \approx b - y + (\mathbf{t}-\mathbf{t}_0)'\dot{X}(\mathbf{t}_0,\sigma) + (\mathbf{t}-\mathbf{t}_0)'\mathrm{E}[\ddot{X}(\mathbf{t}_0,\sigma)|X(\mathbf{t}_0,\sigma) = b-y](\mathbf{t}-\mathbf{t}_0)/2. \quad (5.11)$$

Also

$$\mathrm{E}[\ddot{X}|X] = \mathrm{E}[\ddot{X}] - \mathrm{Cov}[X,\ddot{X}]\mathrm{E}(X) + \mathrm{Cov}[X,\ddot{X}]X, \quad (5.12)$$

but unlike case (i) above the sum of the first two terms does not vanish. We shall assume that $f$ and $k$ are symmetric about zero in each argument and invariant under permutations of the arguments, so that the right hand side of (5.12) is the product of the identity matrix and a scalar of the form

$$\xi\eta + \lambda\sigma^{-2}X(\mathbf{t}_0,\sigma). \quad (5.13)$$

Here $\xi$ and $\lambda$ are as defined above, and $\eta$ is easily evaluated in terms of integrals involving $k, f$, and their partial derivatives. We omit the general expression. In the special case that both $k$ and $f$ are the Gaussian kernels (1.3)

$$\eta = (2\sigma^2)^{-1}[2\sigma\sigma_0/(\sigma^2+\sigma_0^2)]^{N/2}[(\sigma^2-\sigma_0^2)/(\sigma^2+\sigma_0^2)]. \quad (5.14)$$

The right hand side of (5.11) can be maximized as above, and the resulting quadratic form is a multiple of a $\chi^2$ random variable with $N$ degrees of freedom. This can be substituted into (5.4) and integrated approximately as above to yield as an approximation to the power in case (ii)

$$1 - \Phi(b-\mu_0) + (\mu_0 - b)^{-1}\phi(b-\mu_0)\{1 - [(b+\sigma^2\xi\eta/\lambda)/(\mu_0+\sigma^2\xi\eta/\lambda)]^{N/2}\}. \quad (5.15)$$

For the special case of Gaussian kernels, $\eta$ is given by (5.14), $\lambda = 1/2$, and

$$\mu_0 = \xi[2\sigma\sigma_0/(\sigma^2+\sigma_0^2)]^{N/2}. \quad (5.16)$$

Note that (5.10) and (5.15) are the same if $\sigma = \sigma_0$ and $N = n$.

We have investigated the numerical accuracy of the approximation (5.15) for the special case $N = 1, \sigma = \sigma_0$. In one dimension it is possible to use the expected number of upcrossings of the process $X$ to give a tight upper bound on the conditional probability in (5.4), which can then be integrated numerically to give a tight upper bound on (5.4) (cf. Davies (1977), Knowles, Siegmund, and Zhang (1991)). From such a comparison, it appears that the $\chi^2$ approximation for the conditional probability is very good for small $y$, although it deteriorates for larger values. Over a wide range of values for $b$ and $\xi$ the upper bound and the

approximation (5.15) are in quite good agreement– usually to the first two significant figures. We also tried substituting the $\chi^2$ approximation for the conditional probability directly into (5.4) and integrating numerically with essentially the same results.

An interesting application of the preceding calculations is to compare the efficiency of adaptive choice of the scale parameter $\sigma$, as suggested above, with an arbitrary fixed (but frequently incorrect) value. For an example we suppose that $N = 2$. We also assume that both $f$ and $k$ are Gaussian and that $C$ is a $T \times T$ square with $T = 100$. (This is roughly the size of the region corresponding to the astronomical example mentioned in the introduction, although in that case the shape and hence the boundary of the region $C$ are different.) Under these assumptions, when $\xi = 0$, an approximation to the probability that $\max_t X(t, \sigma) > b$ is

$$[|C|b/(4\pi\sigma^2) + |\partial C|/(4\pi^{1/2}\sigma)]\phi(b). \tag{5.17}$$

Assume that $\sigma_0$ is thought to equal 1, but this value may be in error to the extent that the true value may be as small as 0.5 or as large as 2.0. For the test based on a fixed choice of $\sigma$, it is easy to see from (5.15) that one makes a slightly more serious error by choosing $\sigma$ larger than $\sigma_0$ than by choosing it smaller. For significance level 0.05 and power 0.9 some experimentation shows that the choice $\sigma \approx 0.9$ minimizes the maximum value of $\xi$ necessary to obtain the desired power at both $\sigma_0 = 0.5$ and 2.0. In particular $b = 4.58$ and $\xi = 6.8$ satisfy these requirements.

Suppose, on the other hand we search adaptively for $\sigma_0$ over the range (0.33, 3.0). By (3.5) the value of $b$ yielding a significance level of 0.05 is $b \approx 5.1$. Assuming that the value of $\xi$ is proportional to the square root of the sample size, we might reasonably measure the relative efficiency of these two procedures by the square of the ratio of the values of $\xi$ necessary to obtain a given power. For a power of 0.9, at $\sigma_0 = 1.0$, the non-adaptive procedure is about 16% more efficient, while at $\sigma_0 = 0.5$ or 2.0 the adaptive procedure is about 22% more efficient. The range of values of $\sigma_0$ over which the non-adaptive procedure is more efficient is from about .63 to about 1.5. Similar results hold at power of 0.80 and 0.95.

As one might expect from the form of (5.16), these relative efficiencies are larger in higher dimensions. For a similar scenario with $N = 3$, the adaptive procedure is about 33% more efficient when $\sigma_0$ is misspecified by a factor of 2, whereas the nonadaptive procedure is about 25% more efficient when $\sigma_0 = 1$.

# 6 Example and Application

## 6.1 Example

We illustrate the methods in this paper on some simulated data for $N = 1$. Gaussian white noise was simulated on the interval [-40.40] and smoothed for $\mathbf{t} \in C = [-10.10]$ using the Gaussian kernel (1.3) for $\sigma \in I = [0.2.5.0]$. This was repeated for the same Gaussian white noise plus a Gaussian signal located at $\mathbf{t_0} = 0$ with scale $\sigma_0 = 1$ and height $\xi = 6$. The results are shown in Figure 1 for the lower limit $\sigma_1 = 0.2$, the upper limit $\sigma_2 = 5$, and the true scale $\sigma_0 = 1$. Figure 2 shows the same data for all $\sigma$ plotted on the log scale for $\sigma$, so that the field is stationary in separate horizontal and vertical directions (but not jointly). The test statistic is $X_{\mathbf{max}} = 2.24$ for noise only, and $X_{\mathbf{max}} = 6.92$ for signal plus noise; in the latter case the maximum is located at $\mathbf{t} = -0.16$ and $\sigma = 1.18$, close to the location and scale of the signal. Adler (1981, p. 117 ff.) gives a simple and fast way of approximating the Hadwiger characteristic from equally spaced sample data, and we plot this against the threshold $b$ in Figure 3. It is in reasonable agreement with the expected value (3.2) for the noise only data. The approximate level 0.05 critical value for $X_{\mathbf{max}}$ is 3.40, found by equating (3.2) to 0.05 and solving for $b$. No signal is detected in the noise only data, but the signal is detected in the signal plus noise data. The contributions of the individual terms in (3.2) at $b = 3.40$ are 0.032 for the first "interior" term, 0.018 for the second 'location' term, 0.001 for the third "scale" term, and $\mathrm{P}\{X \geq 3.40\} = 0.0003$ for the last term. Thus in this example the "scale" corrections to the first term are an important part (38%) of the approximate probability. However the contribution from the sides is relatively unimportant, despite the 25-fold range of kernel scale; this can be seen from Figure 2, where the image appears a lot smoother in the vertical direction than along the base, so that the excursion set touches the sides a lot less than the base. If we use the data along the base only, that is we restrict $\sigma$ to $\sigma_1 = 0.2$, then the maximum $X(\mathbf{t}, \sigma_1)$ is unchanged for noise only and still high (5.07) for signal plus noise; the approximate 0.05 critical value is slightly lower at 3.30 and we reach the same conclusion as before. Thus in this case searching over kernel scales does not appreciably alter the analysis.

## 6.2 Application

We apply our results to $N = 3$ dimensional medical images. Talbot et al. (1991) carried out an experiment in which PET cerebral blood flow images were obtained for 9 subjects while a thermistor was applied to the forearm at both warm ($37°C$) and hot ($48°C$) temperaratures, each condition being studied twice on each subject. The purpose of the experiment was to find regions of the brain that were activated by the hot stimulus, compared to the warm

stimulus. Individual images were aligned and sampled on a $128 \times 128 \times 80$ lattice of voxels, separated at approximately 1.4mm, 1.7mm and 1.5mm on the front-back, left-right and vertical axes, respectively. For the present work, we analysed the difference images of the two warm conditions as a dataset which should have an expectation of zero throughout. We also analysed the difference between the average of the two hot conditions and the average of the two warm conditions to search for activation due to the painful heat stimulus. These difference images were averaged over subjects to increase the signal to noise ratio. Worsley et al. (1992) estimated the voxel standard deviation by pooling the voxel sample variance (with 8 degrees of freedom) over all voxels. A normalized image was produced by dividing each voxel by the pooled standard deviation and multiplying by the square root of the number of subjects.

In PET imaging the resolution of the image is determined by the "point response function," which is measured by placing a point source of isotope in the PET camera and measuring the response. It can be reasonably approximated by a Gaussian kernel with $\sigma = 2.87$mm. Worsley et al. (1992) found that the distribution of the noise component of the normalized image is well approximated by that of a stationary white noise Gaussian random field smoothed with the point response function. We are thus able to observe the smoothed process $X$ for $\sigma = 2.87$mm but not the unsmoothed process $Z$. If the value $\sigma_0$ in the unobserved model (1.1) exceeds $\sigma_1 = 2.87$ (and the signal shape is Gaussian), it would be appropriate to introduce additional smoothing by a Gaussian kernel with scale $\sigma = (\sigma_0^2 - \sigma_1^2)^{1/2}$. Since $\sigma_0$ is unknown, the normalized image was smoothed with seven Gaussian kernels ranging in scale from zero to 14.0mm so that the scale ranged from $\sigma_1 = 2.87$mm to $\sigma_2 = (\sigma_1^2 + 14.0^2)^{1/2} = 14.3$mm. The Gaussian kernels were chosen so that the the resulting $\log(\sigma)$ values were uniformly spaced on $[\log(\sigma_1), \log(\sigma_2)]$. The choice of seven kernels gave a sampling interval to resolution ratio on the $\log(\sigma)$ scale that approximated the sampling interval to resolution ratio of locations in the highest resolution ($\sigma = \sigma_1$) image.

The region of the brain $C$ of interest was chosen to be a hemisphere with radius r=7cm covering the top part of the brain. The test statistic was $X_{\max} = 4.05$ for noise only, and $X_{\max} = 7.47$ for signal plus noise; in the latter case the maximum was located at $\sigma = \sigma_2 = 14.3$mm in the anterior cingulate/supplementary motor area, plus three other local maxima all with $X(\mathbf{t}, \sigma) > 5$. The volume of $C$ was $|C| = (2/3)\pi r^3 = 718\text{cm}^3$, its area was $|\partial C| = 2\pi r^2 + \pi r^2 = 462\text{cm}^2$, and its mean caliper diameter was $\Delta(\partial C) = r + \pi r/4 = 12.5$cm giving $H(\partial C) = 79$cm. The observed Hadwiger characteristic is plotted against the threshold $b$ in Figure 4, together with the expected value (3.7). There is reasonable agreement for the noise only data, but when the signal is present there are some discrepancies in the upper tail. The approximate level 0.05 critical value for $X_{\max}$ is 4.92, found by equating (3.7) to 0.05 and solving for $b$, so that no signal is detected in the noise only data, but the signal

is detected in the signal plus noise data. The contributions of the individual terms in (3.2) at $b = 4.92$ are 0.028 for the first "interior" term, 0.018 for the second 'location' term, and 0.0028 and 0.0013 for the third and fourth 'scale' terms; the rest are below 0.0001. As in the $N = 1$ example the "edge" corrections to the first term are an important part (43%) of the approximate probability, but the contribution from the scale is relatively unimportant. If we use the data along the base only, that is we restrict $\sigma$ to $\sigma_1 = 2.87$mm, then the maximum $X(\mathbf{t}, \sigma_1)$ is unchanged for noise only but much lower (4.89) for signal plus noise; the approximate 0.05 critical value is slightly lower at 4.86 and so we fail to detect a signal in the noise only data set and we just detect a signal in the signal plus noise data set. This nicely illustrates the advantages of our approach of searching over kernel scales as well as locations.

# 7 Discussion and Open Problems

There remain a number of related problems, which we discuss briefly below.

## 7.1 Confidence Regions

When a signal is present, i.e., $\xi > 0$, we may want to estimate its location and size, say by a confidence region. This problem is related to finding confidence sets for a change-point, and we can make use of ideas appearing in that literature. For a review of a number of possibilities see Siegmund (1988a) and references cited there. For example, suppose that we know the signal shape $f$. Then from the form (1.6) for the likelihood function, it follows that a $1 - \alpha$ conditional likelihood ratio confidence region for $\mathbf{t}_0, \sigma_0$ is given by

$$\{(\mathbf{t}_0, \sigma_0) : X(\mathbf{t}_0, \sigma_0) \geq X_{\max} - c\},$$

where $c = c[X(\mathbf{t}_0, \sigma_0)]$ is such that

$$\mathrm{P}\{X_{\max} \geq X(\mathbf{t}_0, \sigma_0) + c | X(\mathbf{t}_0, \sigma_0)\} = \alpha.$$

Since $X(\mathbf{t}_0, \sigma_0)$ is sufficient for $\xi$, the conditional probability does not depend on the unknown nuisance parameter.

The same conditional probability appears in (5.4), so the method suggested in Section 5 for evaluating this probability is applicable. The only difference is that when we come to assess the accuracy of the approximation given in (5.9), we are now interested in the conditional probability itself, and particularly in values of $b$ and $y$ which make the conditional probability small. Since the power calculations in Section 5 and hence the argument given there involve values of $b$ and $y$ for which the conditional probability is large, and since the

approximation (5.9) is not even monotonic in $y$ when $y$ is a substantial fraction of $b$, it is not clear whether the same approximation will work here. We have performed numerical calculations assuming that $N = 1$ and $\sigma_0$ is known to compare conditional likelihood ratio confidence regions based on the approximation (5.9) with the presumed gold standard: numerical calculation of the expected number of upcrossings (cf. Knowles, Siegmund, and Zhang (1991)). For $X_{\max} = 5.0$, 4.0, and 3.5. we find that the two approximations give essentially the same 0.95 confidence regions. At 5.0 they also give essentially the same 0.99 regions, but at 4.5 the regions differ somewhat and at 3.5 the use of the approximation (5.9) breaks down.

It would be interesting to study this issue in more detail and give an approximation to the conditional probability in (5.4) that is more generally accurate when that probability is small.

## 7.2 Box Shaped Kernels and Signals

In some cases it might be thought appropriate to use a discontinuous kernel, either because it appears that the signal itself is discontinuous or as a convenient approximation. An example would be the box shaped kernel $k(\mathbf{x}) = 1$ if $\max|x_i| \le 1/2$ and 0 otherwise. A systematic study of this case is beyond the scope of the present paper. To illustrate briefly some of the essential features, we suppose that $\sigma_0$ is known and for definiteness that $N = 2$. Then if $C$ is large relative to $\sigma$, so edge effects can be neglected, we find by the methods of Siegmund (1988b) or Loader (1991) that when $\xi = 0$,

$$\mathrm{P}\{\max_{\mathbf{t}} X(\mathbf{t}, \sigma) > b\} \approx (|C|/\sigma^2) b^3 \phi(b). \tag{7.1}$$

Because of the extra roughness in the sample paths of $X(\mathbf{t}, \sigma)$ in this case. the tail probability involves two extra powers of $b$ in comparison with (5.17) and even one power more than (3.5), where we also search for the correct $\sigma_0$. (Another consequence of this sample path roughness is that in most cases one will also want to make a correction for the inevitable discrete grid of observations, for which the model of continuously observed white noise usually does not provide an adequate approximation. We omit discussion of this aspect of the problem. See, for example, Siegmund (1988b).)

A substantial elaboration of the argument of James, James, and Siegmund (1987) shows that if in this case $f = k$ and $\sigma = \sigma_0$, the power of the likelihood ratio test is given approximately by

$$1 - \Phi(b - \xi) + \phi(b - \xi)\{[4b^2(b + \xi) + \xi^2(3b + \xi)]/[\xi(b + \xi)]^2\}. \tag{7.2}$$

It is interesting to consider the possibility that we use a box shaped kernel when the signal is smooth or conversely a smooth kernel when the signal is box shaped. Suppose,

for example, that $f$ is Gaussian but one uses a box shaped kernel. It is easy to see that the expectation of $X(t_0, \sigma)$ is maximized by taking $\sigma$ about equal to $2.8\sigma_0$ and that the maximum value is about $\mu_0 = 0.94355^N \times \xi = 0.89\xi$. Solely on the basis of this effect on the non-centrality parameter, one would expect the relative efficiency of the box shaped kernel to the Gaussian kernel to be about 80%. However, this analysis neglects the effect of the different kernels on the threshold $b$ and the contribution of the second term in (5.4) to the power of the test. For a 0.05 test for the same $100 \times 100$ square in the example at the end of Section 5, the threshold with a box shaped kernel and $\sigma = 2.8, \sigma_0 = 1$ is about $b = 5.34$ compared to $b = 4.53$ when using a Gaussian kernel with $\sigma = \sigma_0 = 1$ (cf. (5.17)). This difference suggests the relative efficiency of the box shaped kernel is even less. Although greater fluctuations in the sample paths of the observed process when a box shaped kernel is used lead one to expect the contributions of the second term in (5.17) to be larger, even without evaluating this term it seems reasonable to conclude that use of the box shaped kernel when the signal is Gaussian leads to a substantial loss of efficiency.

There seems to be a smaller loss of efficiency if the signal is actually box shaped but one uses a Gaussian kernel. For a numerical example, suppose that $C$ is again a $100 \times 100$ square and that $f$ is box shaped with $\sigma_0 = 1$. From (7.1) and (7.2) it follows that for $b = 5.75$ we have an approximately 0.05 level test with power 0.9 at $\xi = 6.49$. If we use a Gaussian kernel, the signal to noise ratio at $t_0$ is maximized at approximately $\sigma = \sigma_0/2.8$. An approximately 0.05 level test is obtained by taking $b = 4.99$. It may be shown that (5.15) with $\eta = 0$ provides an approximation for the power of the test, so power of about 0.9 is attained at $\xi = 6.84$. This corresponds to a relative efficiency of about 90%.

## 7.3 Adaptation on Additional Parameters

The principal points of this paper are that in using a spherical Gaussian kernel, one can determine appropriate rejection thresholds when the scale parameter $\sigma$ is chosen adaptively; and if the signal is itself spherical Gaussian, adaptive choice of $\sigma$ seems a reasonable strategy with regard to power. It seems relatively straightforward to generalize most of our results to allow for different values of $\sigma$ in different coordinate directions, i.e., to allow for an elliptical Gaussian kernel oriented in an arbitrary but data independent direction. However, we do not know how much power is gained or lost by this extra degree of freedom.

A more challenging problem is to allow the Gaussian kernel also to have arbitrary covariances, so the data determine the spatial orientation of the elliptically contoured Gaussian hill.

## 7.4 Numerical Considerations and Unsolved Mathematical Problems

As suggested in (4.4), we conjecture that there exists a precise mathematical theorem, which says that the quantitities we have calculated in Sections 3 and 4 give the correct tail behavior of the maxima of these Gaussian fields as $b \to \infty$ *to the number of terms given.* Previous attempts in this direction seem either to deal with very general processes, but only first order asymptotic behavior (e.g., Adler, 1981, section 6.9, pp. 159-167), or as in Sun (1993) a special Gaussian field and second order asymptotic behavior. An interesting point of Sun's research was the numerical observation that much better agreement between theory and simulations was obtained when the second order terms are included, but in her work the formulas were too complicated to comprehend the size of these second order effects except through the numerical values they produced. In equations (3.6) and (3.7) it is clear that one often must include the top three orders of magnitude before the remaining terms become insignificant numerically.

Giving a precise mathematical relation between the quantities calculated in Section 3 and the tail probability of the maximum of the Gaussian field seems an interesting and challenging mathematical problem.

It would be interesting to derive a general lower bound on the constant $\kappa$, so that one could see generally whether the terms involving $1/\kappa$ in (3.6) and (3.7) might dominate the higher order terms in these expressions, but we have been unable to do so.

**ACKNOWLEDGMENTS.** The authors would like to thank John Cobb for asking the question that initiated their discussions of this subject. They would also like to thank Drs. A. Evans, C. Bushnell and G. Duncan of the Montreal Neurological Institute for permission to use the PET data described in Section 6.2.

# REFERENCES

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

Aldous, D. (1989). *Probability Approximations via the Poisson Clumping Heuristic*, Springer-Verlag, New York.

Belyaev, Yu.K. and Piterbarg, V.I. (1972). The asymptoptic formula for the mean number of $A$-points of excursions of Gaussian fields above high levels (in Russian). In *Bursts of Random Fields.* Moscow University Press, Moscow, 62-89.

Boothby, W. (1986). An Introduction to Differentiable Manifolds and Riemannian Geometry, Academic Press, Orlando, Florida.

Hadwiger, H. (1959). Normale Körper im euklidischen Raum und ihre topologischen und metrischen Eigenschaften. *Mathematische Zeitschrift*, **71**, 124-140.

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

James, B., James, K.L. and Siegmund, D. (1987). Tests for a change-point. *Biometrika*, **74**, 71-83.

Knowles, M. and Siegmund. D. (1989). On Hotelling's approach to testing for a nonlinear parameter in regression. *Internat. Statist. Review* **57**, 205-220.

Knowles, M., Siegmund. D. and Zhang, H.-P. (1991). Confidence regions in semilinear regression, *Biometrika* **78**, 15-31.

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

Leadbetter, M.R., Lindgren. G., and Rootzén. H. (1983). *Extremes and Related Properties of Random Sequences and Processes.* Springer-Verlag, New York.

Loader, C.R. (1991). Large-deviation approximations to the distribution of scan statistics, *Advances in Appl. Probab.* **23**, 751-771.

Loève, M. (1963). *Probability Theory.* Van Nostrand, Princeton, NJ.

Millman, R.S. and Parker, G.D. (1977). *Elements of Differential Geometry.* Prentice Hall, Englewood Clifts, New Jersey.

Naiman, D. Q. (1990). Volumes of tubular neighborhoods of spherical polyhedra and statisticsl inference. *Ann. Statist.* **18**, 685-716.

Parzen, E. (1961). An approach to time series analysis, *Ann. Math. Statist.* **32**, 951-989.

Rabinowitz, D. (1993). Detecting clusters in disease incidence. To appear in *Change-point Problems: Proceedings of the Mount Holyoke Conference.* IMS, Hayward California.

Santaló, L. A. (1976). *Integral Geometry and Geometric Probability. Encyclopedia of Mathematics and its Applications*, Volume 1, (Editor G-C. Rota). Addison-Wesley, Reading, Massachusetts.

Siegmund, D. (1988a). Confidence sets in change-point problems. *Internat. Statist. Rev.* **56**, 31-48.

Siegmund, D. (1988b). Tail probabilities for the maxima of some random fields. *Ann. Probab.* **16**, 487-501.

Siegmund, D. and Zhang, Heping (1994). The expected number of local maxima of a random field and the volume of tubes. *Ann. Statist.* **22**.

Sun, Jiayang (1993) Tail probabilities of the maxima of Gaussian random fields, *Ann. Probab.* **21**, 34-71.

Talbot, J.D., Marrett, S., Evans, A.C., Meyer, E., Bushnell, M.C. and Duncan, G.H. (1991). Multiple representations of pain in the human cerebral cortex, *Science* **251**, 1355-1358.

Weyl, H. (1939). On the volume of tubes. *Amer. J. Math.* **61**, 461-472.

Worsley, K.J., Evans, A.C., Marrett, S. and 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.

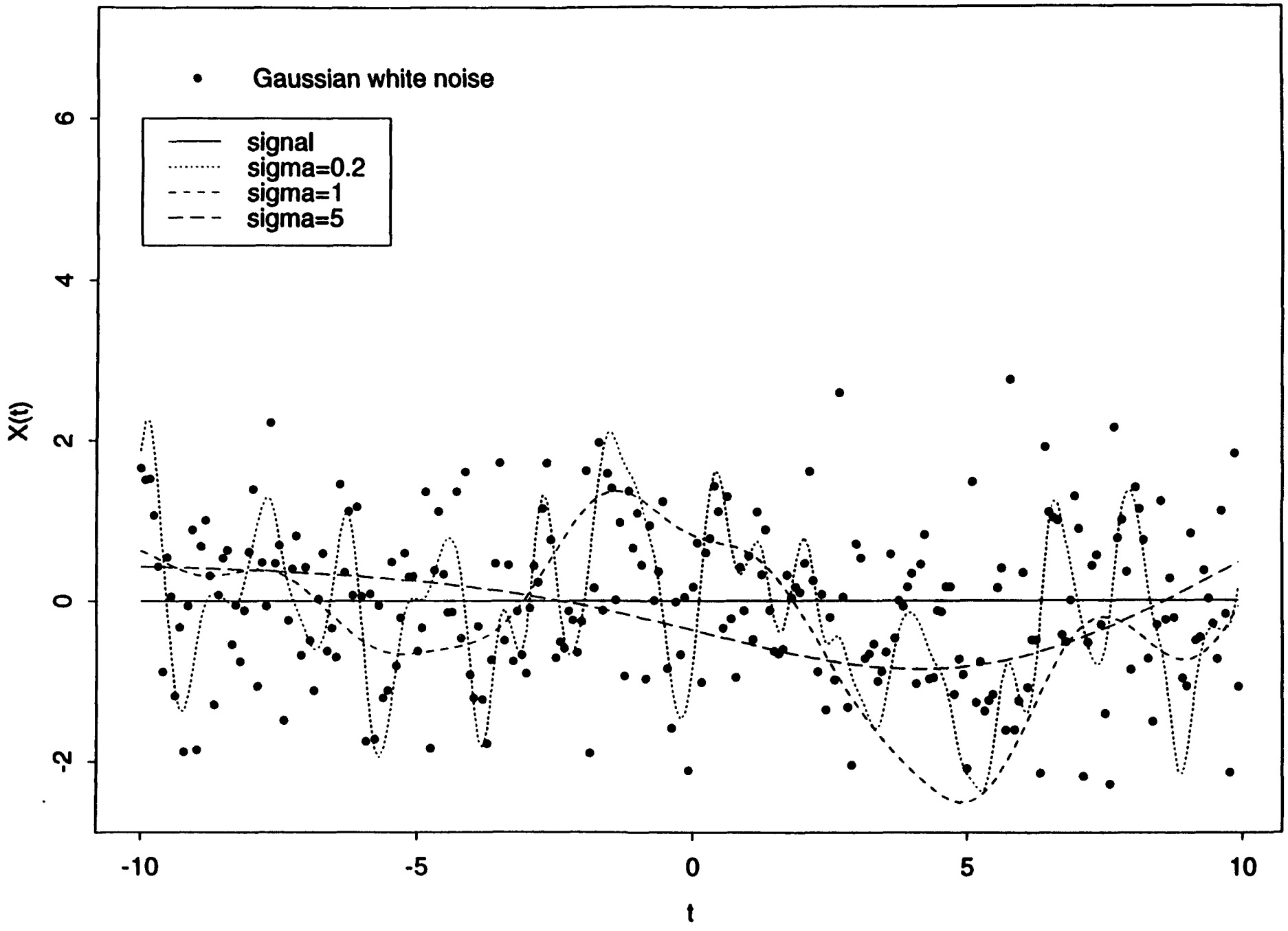
Worsley, K.J., Evans, A.C., Marrett, S. and Neelin, P. (1993). Detecting changes in random fields and applications to medical images. *Journal of the American Statistical Association*, submitted for publication.

Worsley, K.J. (1993a). Detecting changes in a random field, with applications to medical images and astrophysics. *Journal of the American Statistical Association* (submitted for publication).

Worsley, K.J. (1993b). Estimating the number of "peaks" in a random field using the Hadwiger characteristic of excursion sets, with applications to medical images. *Annals of Statistics* (submitted for publication).

Worsley, K.J. (1994). Local maxima and the expected euler characteristic of excursion sets of $\chi^2$, $F$ and $t$ fields. *Advances in Applied Probability*, accepted for publication.

Figure 1(a) Smoothed noise



# Figure 1(b) Smoothed signal + noise

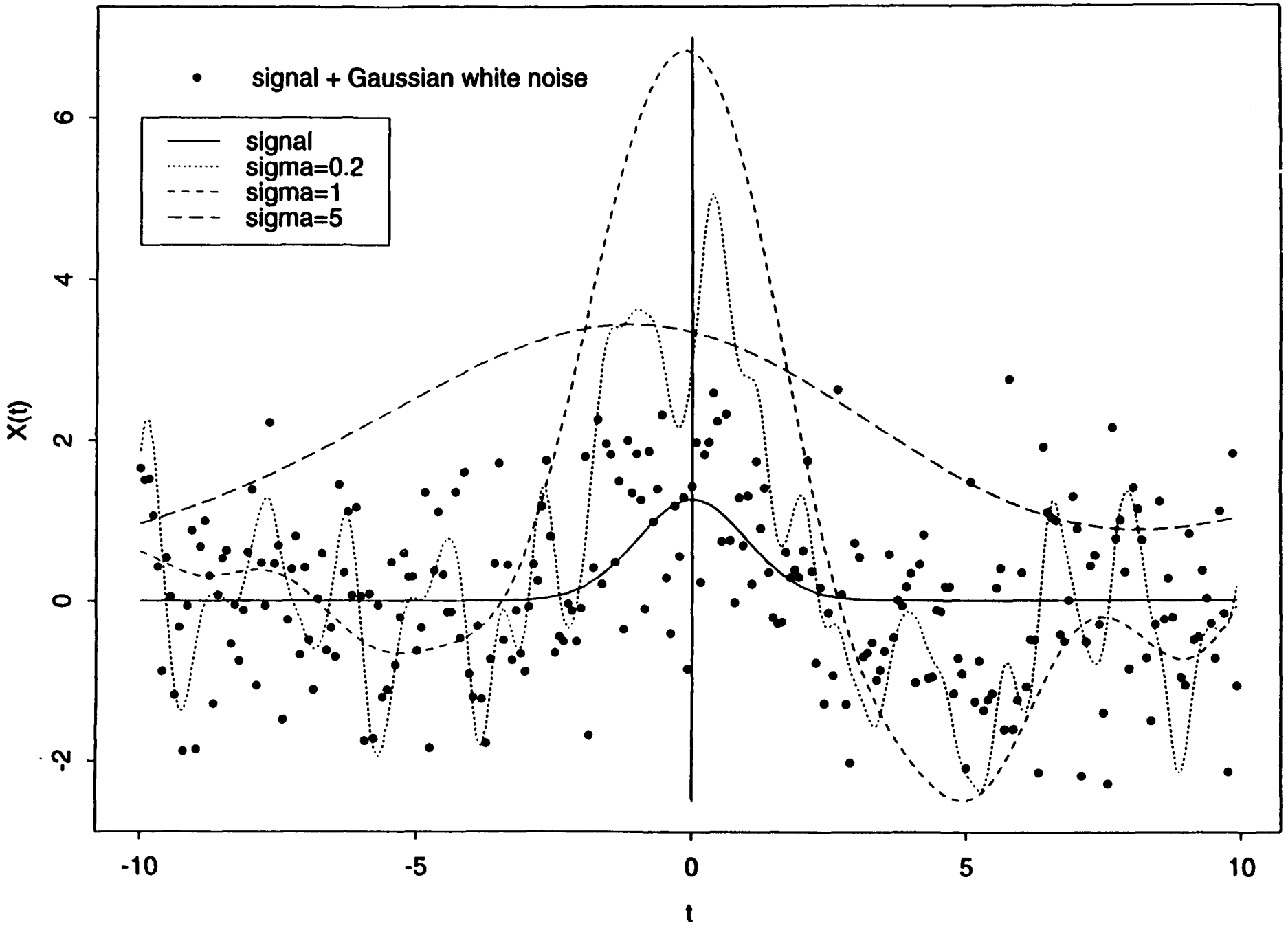
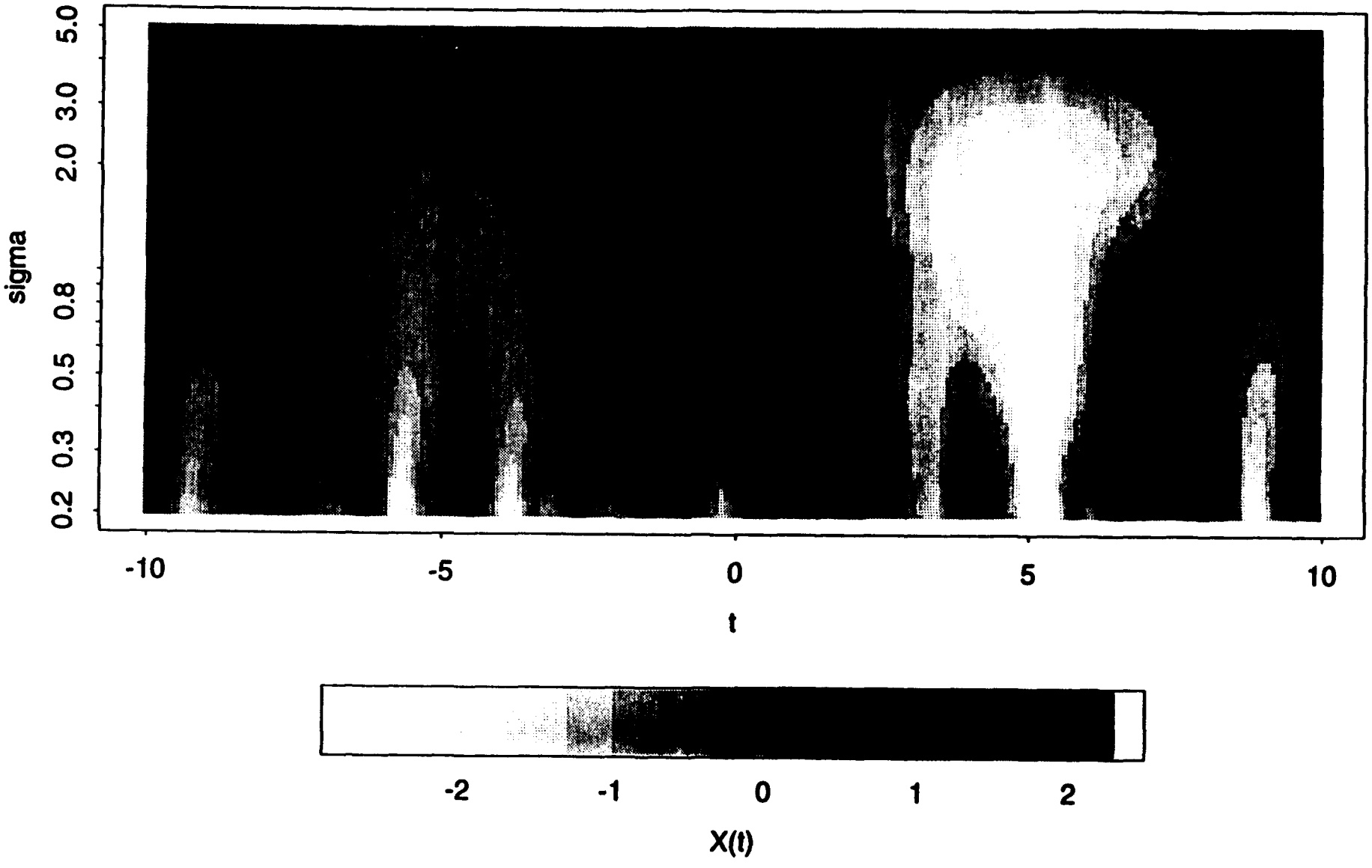
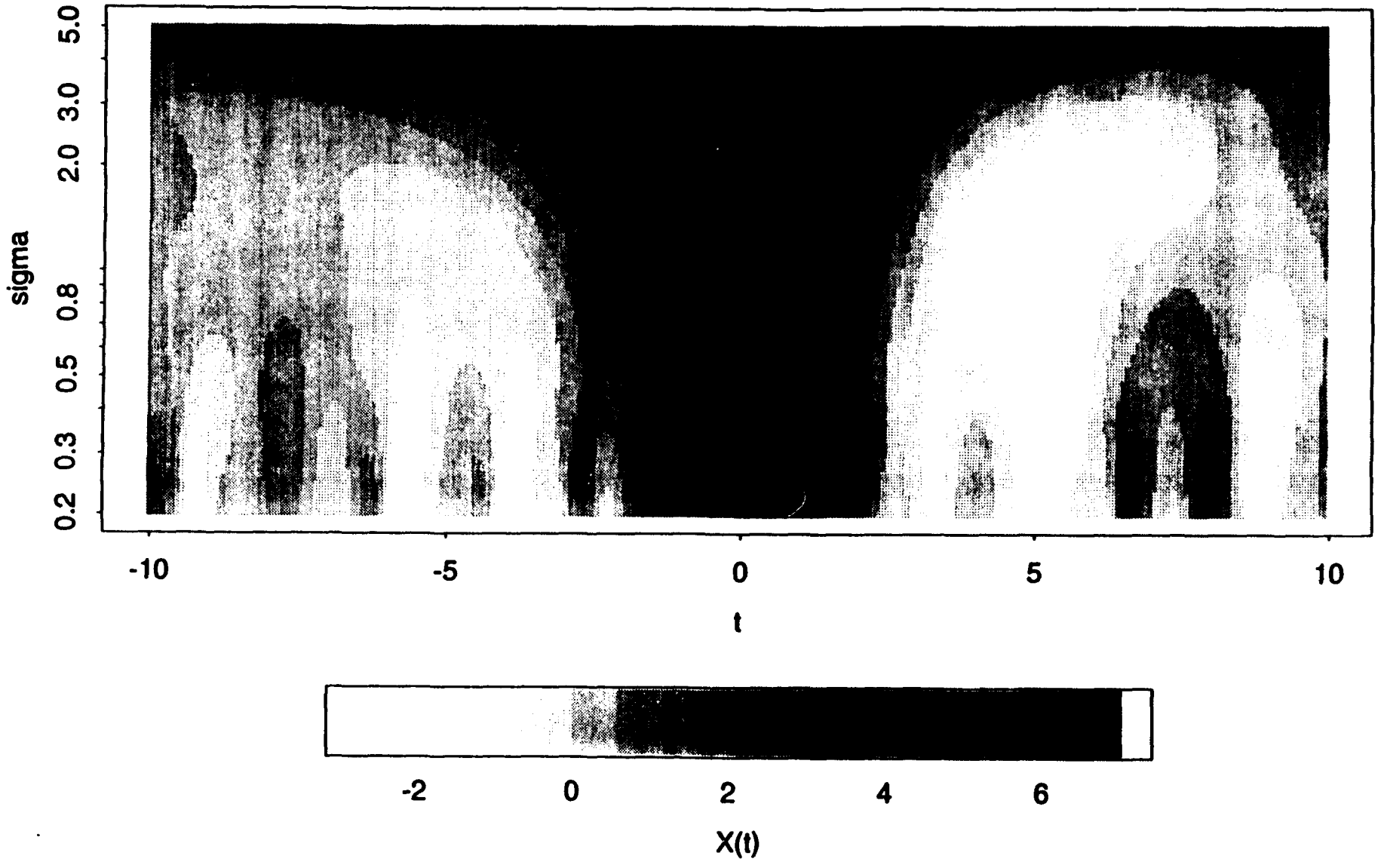


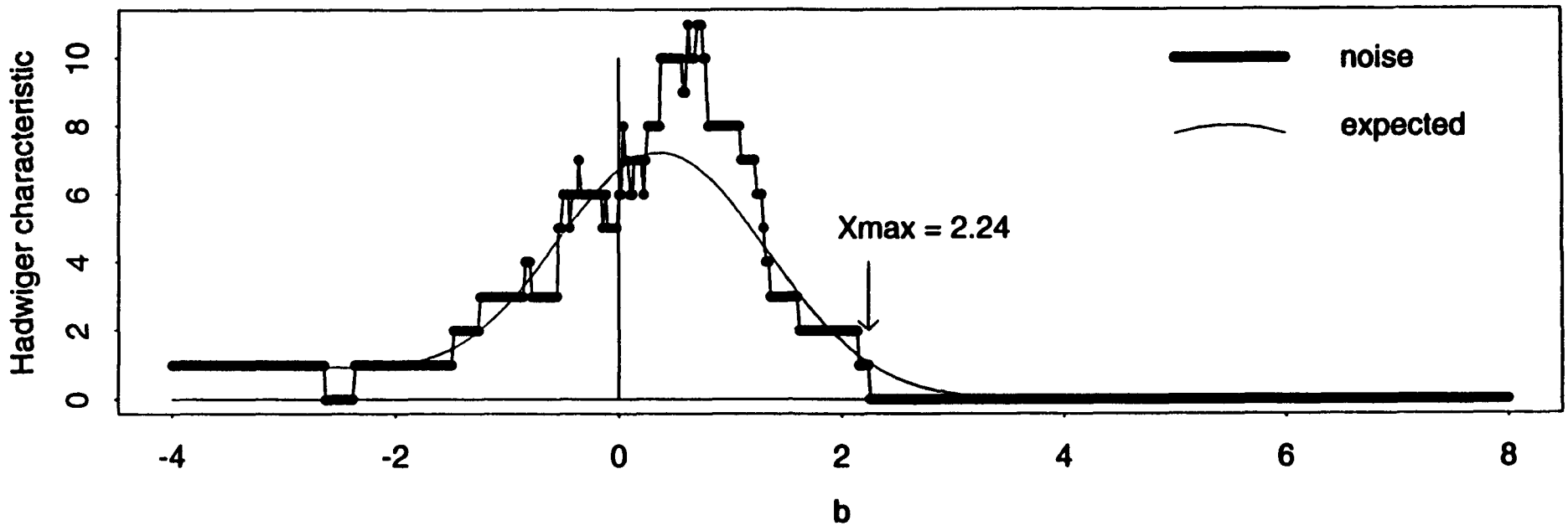
Figure 2(a) Smoothed noise

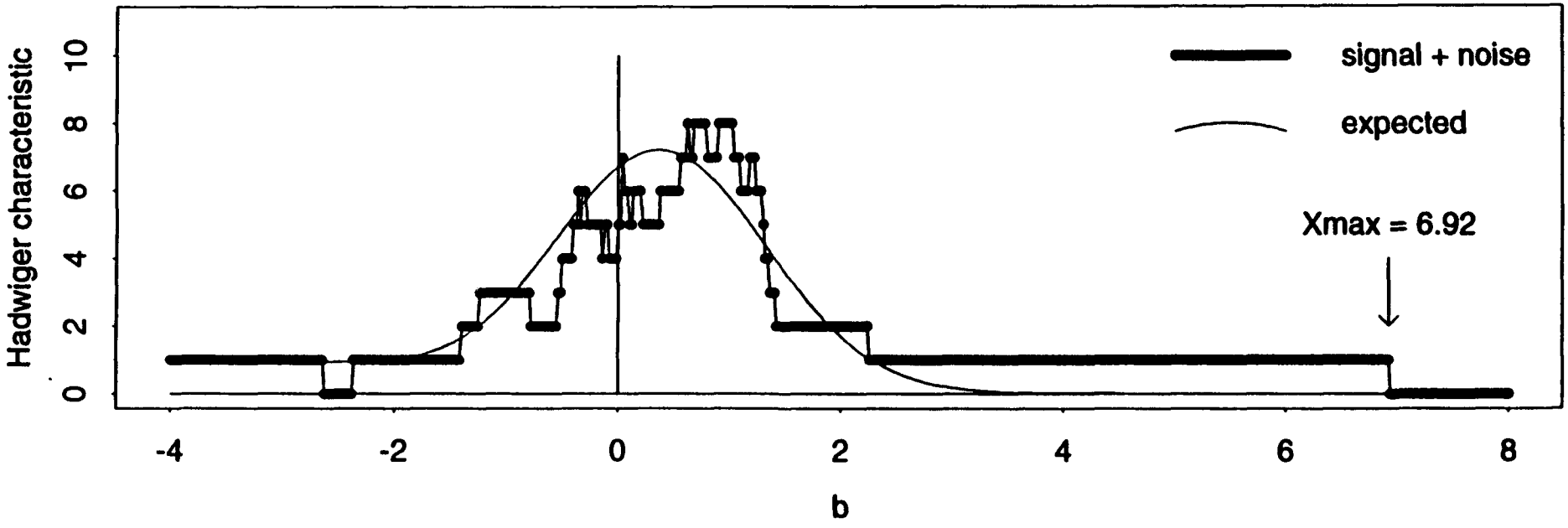


# Figure 2(b) Smoothed signal + noise

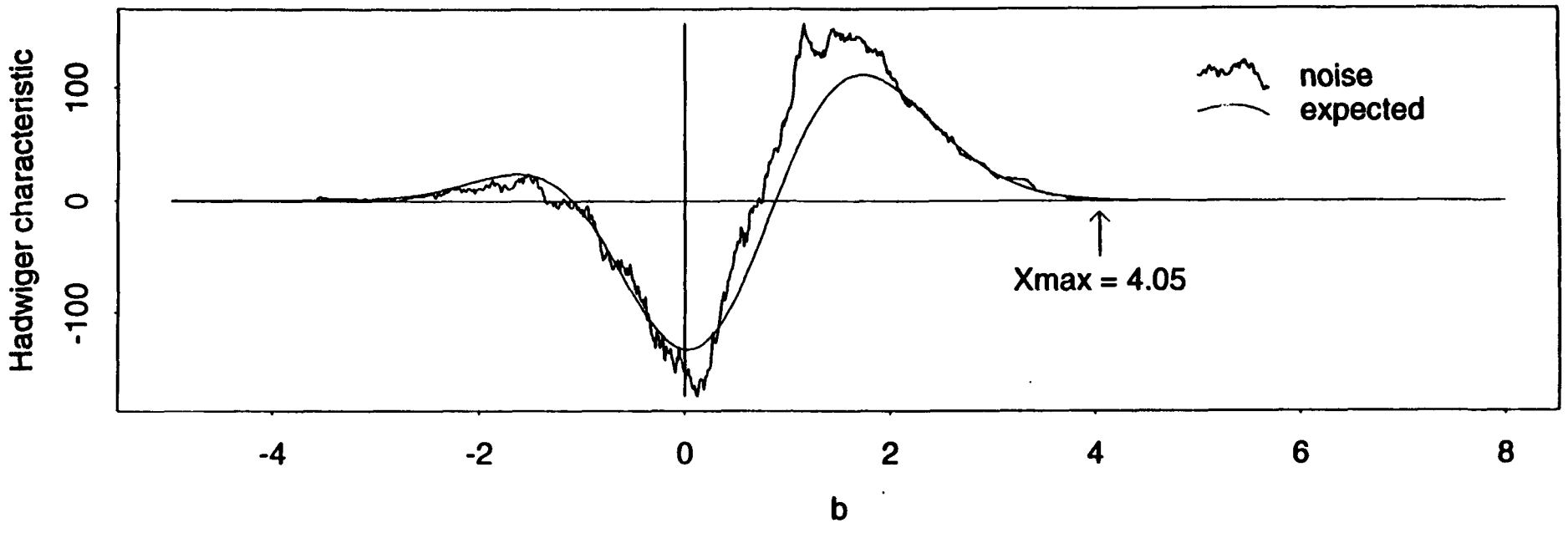


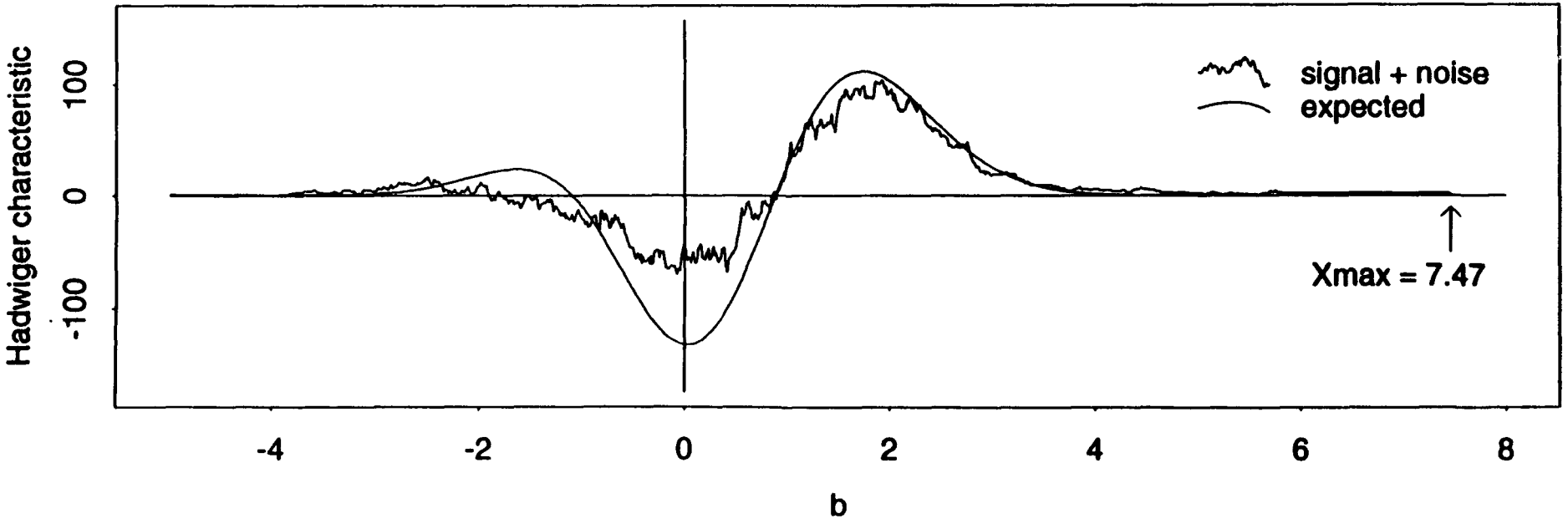
# Figure 3 Hadwiger characteristic



# Figure 4 Hadwiger characteristic, PET data



UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

| REPORT DOCUMENTATION PAGE | | READ INSTRUCTIONS BEFORE COMPLETING FORM |
|---|---|---|
| 1. REPORT NUMBER<br>477 | 2. GOVT ACCESSION NO. | 3. RECIPIENT'S CATALOG NUMBER |
| 4. TITLE (and Subtitle)<br>Testing for a signal with unknown location and sacle in a stationary Gaussian random field | | 5. TYPE OF REPORT & PERIOD COVERED<br>Technical<br>6. PERFORMING ORG. REPORT NUMBER |
| 7. AUTHOR(s)<br>David O. Siegmund & Keith J. Worsley | | 8. CONTRACT OR GRANT NUMBER(s)<br>N0025-92-J-1264 |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS<br>Department of Statistics<br>Stanford University<br>Stanford, CA 94305-4065 | | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS<br>NR-042-267 |
| 11. CONTROLLING OFFICE NAME AND ADDRESS<br>Office of Naval Research<br>Statistics & Probability Program<br>Code 111 | | 12. REPORT DATE<br>January 7, 1994<br>13. NUMBER OF PAGES<br>37 |
| 14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) | | 15. SECURITY CLASS. (of this report)<br>Unclassified<br>15a. DECLASSIFICATION/DOWNGRADING SCHEDULE |

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Prepared also under the auspices of the National Science Foundation, Grant No. DMS 92-04432

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

*Key words and phrases.* Euler characteristic. integral geometry. image analysis. Gaussian fields. volume of tubes.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

See Reverse Side

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

## SUMMARY

We suppose that our observations can be decomposed into a fixed signal plus random noise, where the noise is modelled as a particular stationary Gaussian random field in $N$-dimensional Euclidean space. The signal has the form of a known function centered at an unknown location and multiplied by an unknown amplitude, and we are primarily interested in a test to detect such a signal. There are many examples where the signal scale or width is assumed known, and the test is based on maximising a Gaussian random field over all locations in a subset of $N$-dimensional Euclidean space. The novel feature of this work is that the width of the signal is also unknown and the test is based on maximising a Gaussian random field in $N+1$-dimensions, $N$ dimensions for the location plus one dimension for the width. Two convergent approaches are used to approximate the null distribution: one based on the method of Knowles and Siegmund (1989), which uses a version of Weyl's (1939) tube formula for manifolds with boundaries, and the other based on some recent work by Worsley (1993b), which uses the Hadwiger characteristic of excursion sets as introduced by Adler (1981). Finally we compare the power of our method with one based on a fixed but perhaps incorrect signal width.