

Detecting sparse signals in random fields, with an application to brain mapping

J.E. Taylor* K.J. Worsley
Stanford University McGill University

November 17, 2006

Abstract

Brain mapping data have been modeled as Gaussian random fields, and local increases in mean are detected by local maxima of a random field of test statistics derived from this data. Accurate P-values for local maxima are readily available for isotropic data based on the expected Euler characteristic of the excursion set of the test statistic random field. In this paper we give a simple method for dealing with non-isotropic data. Our approach has connections to the [Sampson and Guttorp \(1992\)](#) model for non-isotropy in which there exists an unknown mapping of the support of the data to a space in which the random fields are isotropic. Heuristic justification for our approach comes from the Nash Embedding Theorem. Formally we show that our method gives consistent unbiased estimators for the true P-values based on new results of [Taylor and Adler \(2003\)](#) for random fields on manifolds which replace the Euclidean metric by the variogram. The results are used to detect gender differences in the cortical thickness of the brain, and to detect regions of the brain involved in sentence comprehension measured by fMRI.

1 Introduction

Our motivation comes from a study by [Lerch and Evans \(2005\)](#) on cortical thickness of $n = 321$ normal adults measured by magnetic resonance imaging (MRI). The aim is to detect points s of the brain S where cortical thickness $Y_i(s)$ on subject i , $i = 1, \dots, n$, is correlated with an independent variable such as age or gender (we shall concentrate on gender). A similar study just published in *Nature* ([Shaw et al. 2006](#)) relates cortical thickness to IQ in $n = 307$ children and adolescents. To do this, cortical thickness was first smoothed

*Jonathan E. Taylor is Associate Professor, Stanford University, Stanford CA 94305 (email: jonathan.taylorstanford.edu); and Keith J. Worsley is Professor, McGill University, Montreal QC H3A2K6 Canada (email: keith.worsleymcgill.ca). This work was partially supported by NSF, NSERC and FQRNT. The authors thank the referees for helpful comments.

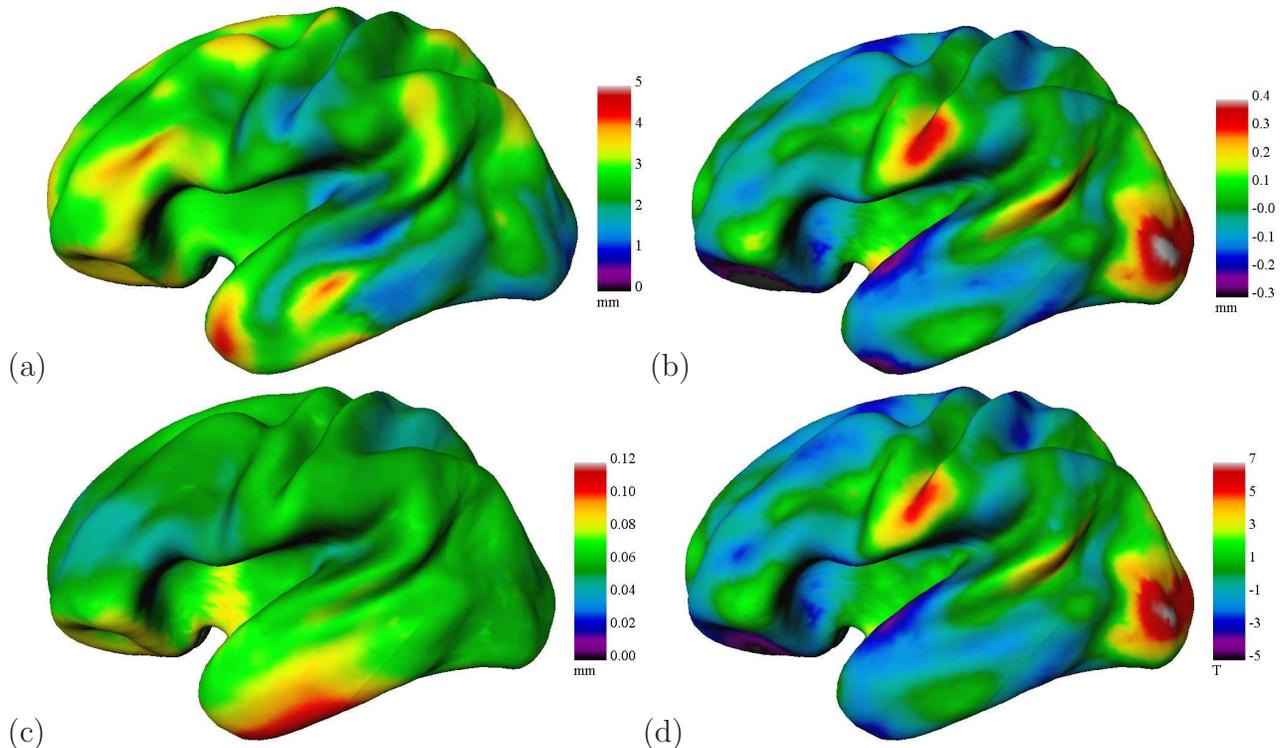


Figure 1: Cortical thickness (mm) of $n = 321$ normal adults on the average cortical surface, as defined in Section 6.1. (a) Smoothed cortical thickness of a single subject; (b) average cortical thickness of the 163 males minus the 158 females, adjusted for a linear effect of age; (c) pooled within-gender standard deviation of the difference in (b), adjusted for age, 318 df; (d) two-sample T-statistic, adjusted for age, (b) divided by (c), 318 df.

along the surface (Chung et al. 2005). After removing a linear effect of age, the usual two-sample (male-female) T-statistic was calculated at each point on the cortical surface to give a smooth T-statistic random field with 318 degrees of freedom - see Figure 1. We can detect local differences in cortical thickness by thresholding this random field at a suitably high value. The problem we shall address in this paper is where exactly to set this threshold to control the false positive rate to say 0.05.

In general we can set up a linear model for subject i , $i = 1, \dots, n$:

$$Y_i(s) = x_i\beta(s) + Z_i(s)\sigma(s), \quad (1)$$

where x_i is a p -vector of known regressors, and $\beta(s)$ is an unknown p -vector of coefficients. The errors $Z_1(s), \dots, Z_n(s)$ are independent random fields each with the same distribution as $Z(s)$, a zero mean, unit variance Gaussian random field, and $\text{Var}(Y_i(s)) = \sigma(s)^2$ is unknown. We can now detect how the regressors are related to data at point s by testing contrasts

in $\beta(s)$. Classical test statistics for these contrasts evaluated at each point s then form a random field $T(s)$, such as the two-sample T-statistic in Figure 1(d).

We suspect that non-zero contrast is confined to a small number of localized regions and the contrast in the rest of the brain is zero. It is not hard to show that if the spatial pattern of the change matches the spatial correlation function of the errors, then the spatial maximum of $T(s)$ is the likelihood ratio test for the null of zero contrast everywhere, provided $\sigma(s)$ is known (Siegmund and Worsley 1995). In essence this is just the Matched Filter Theorem from signal processing, which states that to best detect a signal added to white noise, the data should be smoothed by a filter with the same shape as the signal (Rosenfeld and Kak 1982). Thresholding $T(s)$ at that threshold which controls the P-value of its maximum should then be powerful at detecting changes, while controlling the false positive rate outside the changed region to something slightly smaller than the nominal P-value. This is because the non-activated region is slightly smaller than the brain, and the P-value of the maximum over a region increases with region size. Our main problem is therefore to find the P-value of the maximum of random fields of these test statistics.

To give the reader some idea of the rest of the paper, here is our final result for the $D = 1$ dimensional case where $S \subset \mathbb{R}$ is an interval, and a one dimensional random field of T-statistics with very high degrees of freedom, so that it is effectively a Gaussian random field with zero mean and unit variance under the null. The P-value approximation, based on upcrossings (Rice 1945), improved Bonferroni inequalities (Hunter 1976; Worsley 1982; Efron 1997), Euler characteristics (Adler 1981), or the Weyl-Hotelling tube formula (Sun 1993; Sun and Loader 1994), is

$$\mathbb{P}\left(\max_{s \in S} T(s) \geq t\right) \approx \mathbb{P}(T(s) \geq t) + \int_S \text{Var}(\dot{Z}(s))^{1/2} ds \cdot \exp(-t^2/2)/(2\pi), \quad (2)$$

where dot represents spatial derivative with respect to s .

The main problem of this paper is to estimate the coefficient

$$\mathcal{L}_1(S) = \int_S \text{Var}(\dot{Z}(s))^{1/2} ds, \quad (3)$$

known as the *Lipschitz-Killing curvature* (LKC) of S . To do this we suppose that the data $Y_i(s)$ are finely sampled at a large set of N points $P = \{s_1 < s_2 < \dots < s_N\} \subset S$. For each j let $r(s_j)$ be the n -vector of least squares residuals from the linear model (1). Specifically, if

$Y(s_j) = (Y_1(s_j), \dots, Y_n(s_j))'$ is the $n \times 1$ vector of observations at s_j and $X = (x'_1, \dots, x'_n)'$ is the $n \times p$ design matrix, then

$$r(s_j) = Y(s_j) - X(X'X)^{-1}X'Y(s_j). \quad (4)$$

Then let

$$u(s_j) = r(s_j)/\|r(s_j)\| \quad (5)$$

be the n -vector of normalized residuals. Our estimator of the LKC in this 1D case is

$$\widehat{\mathcal{L}}_1(S) = \sum_{j=1}^{N-1} \|u(s_{j+1}) - u(s_j)\|. \quad (6)$$

The rest of the paper is laid out as follows. In Section 2 we show how to generalize (2) to higher dimensions $D > 1$ and non-Gaussian test statistics $T(s)$ such as T- and F- statistics for the special case where the random field is isotropic. We generalise this to non-isotropic random fields in Section 3. Section 4 gives our main result, a consistent unbiased estimator of the LKC, with special emphasis on the 3D case in Section 5. We give two applications in Section 6, first to the 2D cortical thickness in Figure 1, then to 3D functional magnetic resonance imaging data.

2 P-value of the maximum of an isotropic random field

Generalizing (2) to higher dimensions, a very accurate approximation to the P-value of the maximum of a smooth, isotropic random field $T(s)$, $s \in S \subset \mathbb{R}^D$, at high thresholds t , is the expected Euler characteristic (EC) φ of the excursion set:

$$\mathbb{P}\left(\max_{s \in S} T(s) \geq t\right) \approx \mathbb{E}(\varphi\{s \in S : T(s) \geq t\}) = \sum_{d=0}^D \mu_d(S) \rho_d(t), \quad (7)$$

where $\mu_d(S)$ is the d -th *intrinsic volume* of S , and $\rho_d(t)$ is the d -th *EC density* of the random field above t (Adler 1981; Worsley 1995; Adler 2000; Taylor et al. 2005). Formula (7) also holds for non-isotropic fields that are functions of Gaussian fields which we shall cover in Section 3. For example, in 3D the EC of a set is given by

$$\varphi = \#\text{blob} - \#\text{handles} + \#\text{hollows}$$

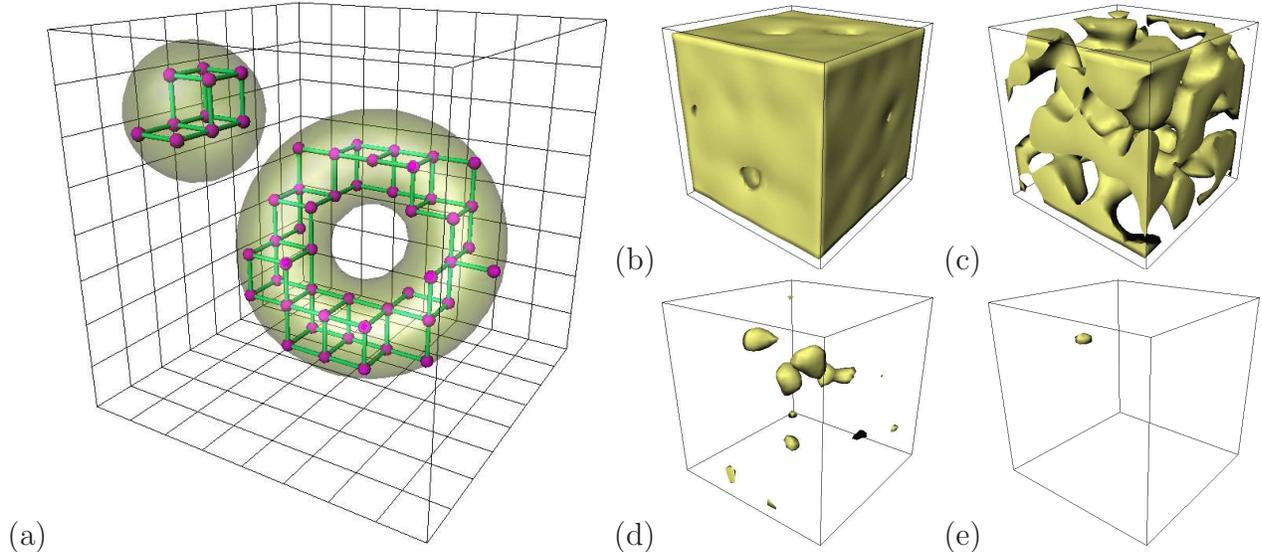


Figure 2: (a) The Euler characteristic of a 3D set, here a ball and a filled torus, is $\varphi = \# \text{blobs} - \# \text{handles} + \# \text{hollows} = 2 - 1 + 0 = 1$, or, if we fill the set with a fine rectilinear mesh, then $\varphi = \# \text{points} - \# \text{edges} + \# \text{faces} - \# \text{cubes} = 55 - 90 + 40 - 4 = 1$. (b) The excursion set of a 3D isotropic Gaussian random field with mean zero and variance one above a threshold $t = -2$; the set contains isolated hollows that each contribute $+1$ to give $\varphi = 6$ with $\mathbb{E}(\varphi) = 6.7$ from (7); (c) at $t = 0$ the handles dominate, each contributing -1 to give $\varphi = -15$ with $\mathbb{E}(\varphi) = -20.0$; (d) at $t = 2$ the handles and hollows disappear, leaving isolated blobs, each contributing $+1$ to give $\varphi = 14$ with $\mathbb{E}(\varphi) = 16.1$; (e) at $t = 3$ only one blob remains (containing the maximum value of 3.51) to give $\varphi = 1$ with $\mathbb{E}(\varphi) = 2.1$. At very high thresholds $\mathbb{E}(\varphi)$ is a good approximation to the P-value of the maximum.

(see Figure 2(a)). A blob is a connected component (such as a golf ball), a handle needs one cut to open it (such as a cup handle), and a hollow is an interior hole not connected to the outside (such as the interior of a tennis ball). It can be evaluated numerically for a subset of a rectilinear mesh by

$$\varphi = \# \text{points} - \# \text{edges} + \# \text{faces} - \# \text{cubes} \quad (8)$$

where, for example, a cube is a set of 8 adjacent mesh points, differing by one mesh step along each axis, and all inside the set (Figure 2(a)). This method was used to calculate the EC of the excursion sets in Figure 2(b-e). For a full explanation of why this method works, see Adler (1977) and Adler (1981), Theorem 5.5.1.

2.1 Intrinsic volumes in 3D

Formal definitions of intrinsic volume are given in Appendix A.1, but for $D = 3$, the case most often encountered in brain imaging and the case considered here in detail in Section 5,

the intrinsic volumes are

$$\begin{aligned}
\mu_0(S) &= \varphi(S) \\
\mu_1(S) &= 2 \text{ caliper diameter}(S) \\
\mu_2(S) &= \frac{1}{2} \text{ surface area}(S) \\
\mu_3(S) &= \text{volume}(S).
\end{aligned} \tag{9}$$

For a convex set, the caliper diameter is the distance between two parallel tangent planes to the set, averaged over all rotations. For the example in Figure 1, regarded as a subset of \mathbb{R}^3 , $\mu_0(S) = 2$ (since it is a hollow surface), $\mu_1(S) = 0$ (the inside curvature cancels with the outside curvature, see Appendix A.1.3), $\mu_2(S)$ is the surface area of S (half from the inside, half from the outside), and $\mu_3(S) = 0$.

Note that for S an M -dimensional manifold embedded in a higher dimensional space $D \geq M$, the higher dimensional intrinsic volumes are all zero, so that the summation in (7) need only go as far as M , the dimensionality of the manifold, rather than the dimensionality of the embedding space.

2.2 T-statistic EC densities up to 3D

We shall assume that the error $Z(\cdot)$ is a *unit Gaussian random field* (UGRF), which we define as a Gaussian random field with zero mean, unit variance, and identity variance matrix of its vector of spatial derivatives, that is $\text{Var}(\dot{Z}(s)) = I_{D \times D}$, the $D \times D$ identity matrix (henceforth dot represents $\partial/\partial s$). Then the first four EC densities of a T-statistic random field $T(s)$ with $\nu = n - p$ degrees of freedom are

$$\begin{aligned}
\rho_0(t) &= \int_t^\infty \frac{\Gamma(\frac{\nu+1}{2})}{(\nu\pi)^{1/2}\Gamma(\frac{\nu}{2})} \left(1 + \frac{u^2}{\nu}\right)^{-(\nu+1)/2} du && \rightarrow \int_t^\infty (2\pi)^{-1/2} \exp(-u^2/2) du \\
\rho_1(t) &= (2\pi)^{-1} \left(1 + \frac{t^2}{\nu}\right)^{-(\nu-1)/2} && \rightarrow (2\pi)^{-1} \exp(-t^2/2) \\
\rho_2(t) &= (2\pi)^{-3/2} \frac{\Gamma(\frac{\nu+1}{2})}{(\frac{\nu}{2})^{1/2}\Gamma(\frac{\nu}{2})} t \left(1 + \frac{t^2}{\nu}\right)^{-(\nu-1)/2} && \rightarrow (2\pi)^{-3/2} t \exp(-t^2/2) \\
\rho_3(t) &= (2\pi)^{-2} \left(\frac{\nu-1}{\nu}t^2 - 1\right) \left(1 + \frac{t^2}{\nu}\right)^{-(\nu-1)/2} && \rightarrow (2\pi)^{-2} (t^2 - 1) \exp(-t^2/2)
\end{aligned} \tag{10}$$

as $\nu \rightarrow \infty$, to give EC densities for the UGRF itself. Note also that we do not need to specify anything about the spatial correlation function $h(s)$ of the UGRF Z other than its second derivative at the origin: $-\ddot{h}(0) = \text{Var}(\dot{Z}(s)) = I_{D \times D}$. Note that any stationary Gaussian

random field can be transformed to a UGRF by appropriate linear transformations of its domain and range. In particular, if $\text{Var}(\dot{Z}(s)) = \lambda I_{D \times D}$ for some scalar λ , then $\rho_d(t)$ is multiplied by $\lambda^{d/2}$.

Formal definitions of EC densities for higher dimensions and for other random fields built from UGRF's, such as the F-statistic random field, are given in Appendix A.2.

2.3 Accuracy of the expected EC as a P-value approximation

A heuristic explanation for why we use the expected EC is as follows. We present the heuristic in 3D, but in principle, it works in all dimensions. If the threshold t is high, the handles and hollows of the excursion set tend to disappear, leaving a set of isolated blobs, each containing one local maximum, so that the EC then counts the number of connected components (Figure 2(d)). At very high thresholds the excursion set is mostly empty with an EC of 0, or occasionally, when $\max_{s \in S} T(s) \geq t$, it will contain just one connected component with an EC of 1 (Figure 2(e)). Thus at these high thresholds the expected EC is a good approximation to the P-value of $\max_{s \in S} T(s)$. The beauty of the EC is that there is an *exact* expression for its expectation for *all* thresholds.

Moreover the approximation (7) is astonishingly accurate when $T(s)$ is Gaussian, and S is either convex or has smooth boundary, in which case

$$\mathbb{E}(\varphi\{s \in S : T(s) \geq t\}) = \mathbb{P}(T(s) > t) + (c_0 + c_1 t + \dots + c_{D-1} t^{D-1}) \exp(-t^2/2)$$

for some constants c_0, \dots, c_{D-1} (see (36) in the Appendix). Since $\mathbb{P}(T(s) > t) = O(1/t) \exp(-t^2/2)$, it might be thought that the error in the approximation (7) is simply the next term down, that is $O(1/t^2) \exp(-t^2/2)$. In fact the error is *exponentially* smaller than this: $\exp(-\alpha t^2/2)$ for some $\alpha > 1$ which is related to the curvature of the boundary of S and $\text{Var}(\ddot{T}(s))$ (Taylor et al. 2005). This means that there are in effect no further polynomial terms in the expansion, and that the expected EC captures essentially all of the P-value. The exponentially smaller error is only known for the Gaussian case and in other cases, e.g. a T field, is both unproven and probably inappropriate.

3 The non-isotropic case: Lipschitz-Killing curvature

All brain imaging data is in fact non-isotropic, though isotropy can be a good approximation for e.g. positron emission tomography data (Worsley et al. 1992). For example, the cortical thickness data is defined on a curved 2D manifold embedded in 3D, so the very notion of isotropy does not even apply, except perhaps in a local sense. What we need is a theory to cover the non-isotropic case, and, in general, the case of random fields on manifolds.

This has been accomplished in Taylor (2006) for the special case where $T(s)$ is a function of independent Gaussian random fields each with the same distribution as $Z(s)$. In a nutshell, the EC densities are first evaluated for UGRF data (assuming $\text{Var}(\dot{Z}(s)) = I_{D \times D}$), then the intrinsic volumes $\mu_d(S)$ are replaced by *Lipschitz-Killing curvatures* (LKC) $\mathcal{L}_d(S, \Lambda)$, where

$$\Lambda = \Lambda(s) = \text{Var}(\dot{Z}(s)). \quad (11)$$

Thus the LKC incorporates information from the local correlation function of the underlying Gaussian random fields as well as the search region S . In effect the spatial correlation structure of the data is transferred from the EC density to the LKC.

We now re-state the P-value approximation (7) for the non-isotropic case where $T(s)$ is a function of independent Gaussian random fields each with the same distribution as $Z(s)$:

$$\mathbb{P}\left(\max_{s \in S} T(s) \geq t\right) \approx \mathbb{E}(\varphi\{s \in S : T(s) \geq t\}) = \sum_{d=0}^D \mathcal{L}_d(S, \Lambda) \rho_d(t), \quad (12)$$

where $\rho_d(t)$ is evaluated as in Appendix A.2 but with $Z(s)$ replaced by a UGRF. Note that in this function of Gaussian case there is no rigorous result for the accuracy of the approximation.

3.1 Intuitive idea of LKC: warping to isotropy

Before giving a definition of LKC, we shall give a brief idea of how it arises. Note first that the domain of the random fields could be warped or deformed by a one-to-one smooth transform without fundamentally changing the nature of the problem. For example, we could ‘inflate’ the average cortical surface in Figure 1 to a sphere and carry out all our analysis on the sphere - some neuroscience researchers actually do this (Fischl et al. 1999; Hurdal and Stephenson 2004). Or we could use any convenient shape: the maximum of the T-statistic

would be unchanged, and so would the Euler characteristic of the excursion set. Of course the correlation structure would change, but then so would the search region, in such a way that the effects of these on the LKC, and hence the expected Euler characteristic, cancel - see Section 3.2.

This suggests that we could try to choose the warping of the parameter space S so that the random fields are isotropic, then use the isotropic theory above. In other words, we want to find a deformation $f : \tilde{S} \rightarrow S$ so that $Z(f(\tilde{s}))$ is isotropic in $\tilde{s} \in \tilde{S}$. This approach is reminiscent of that of [Sampson and Guttorp \(1992\)](#) for the analysis of geostatistical data.

First of all, is this possible? In general the answer is no, but at least one could try to achieve *local* isotropy by ensuring that neighbouring values are equally correlated at equal distances in all directions. In other words, we might try to use local multi-dimensional scaling to find new points \tilde{s} so that for any two neighbouring points \tilde{s}_1, \tilde{s}_2

$$\|\tilde{s}_1 - \tilde{s}_2\|^2 \approx \text{Var}(Z(f(\tilde{s}_1)) - Z(f(\tilde{s}_2)))$$

(see Figure 3 and Section 6.1). The fact that the Euler characteristic has a local definition (by Morse theory and (8)) suggests that local isotropy, and not global isotropy, is sufficient for our purposes. This will be confirmed in Section 3.2.

Again local isotropy can only be achieved approximately (see Figure 3(c)), but, as noted by [Sampson and Guttorp \(1992\)](#), better local isotropy might be achieved by using more dimensions (say 4 or 5) for \tilde{s} . In other words, we could embed \tilde{S} in a higher dimensional Euclidean space $\mathbb{R}^{\tilde{D}}$. As already noted in Section 2.1, this does not increase the number of terms in our P-value approximation (7) because the higher dimensional intrinsic volumes are all zero.

How many dimensions are needed? [Sampson and Guttorp \(1992\)](#) noted that if n i.i.d. random fields $Z_i(s)$ are available as in the linear model (1), then exact *sample* isotropy could be achieved by replacing the coordinates of $s \in \mathbb{R}^D$ by $\tilde{s} = (Z_1(s), \dots, Z_n(s))/\sqrt{n} \in \mathbb{R}^n$, in the sense that

$$\|\tilde{s}_1 - \tilde{s}_2\|^2 = \widehat{\text{Var}}(Z(f(\tilde{s}_1)) - Z(f(\tilde{s}_2))) = \sum_{i=1}^n (Z_i(s_1) - Z_i(s_2))^2/n$$

(see Figure 4). We shall use this idea of replacing the coordinates by the data when we construct a sample estimator of LKC in the next section.

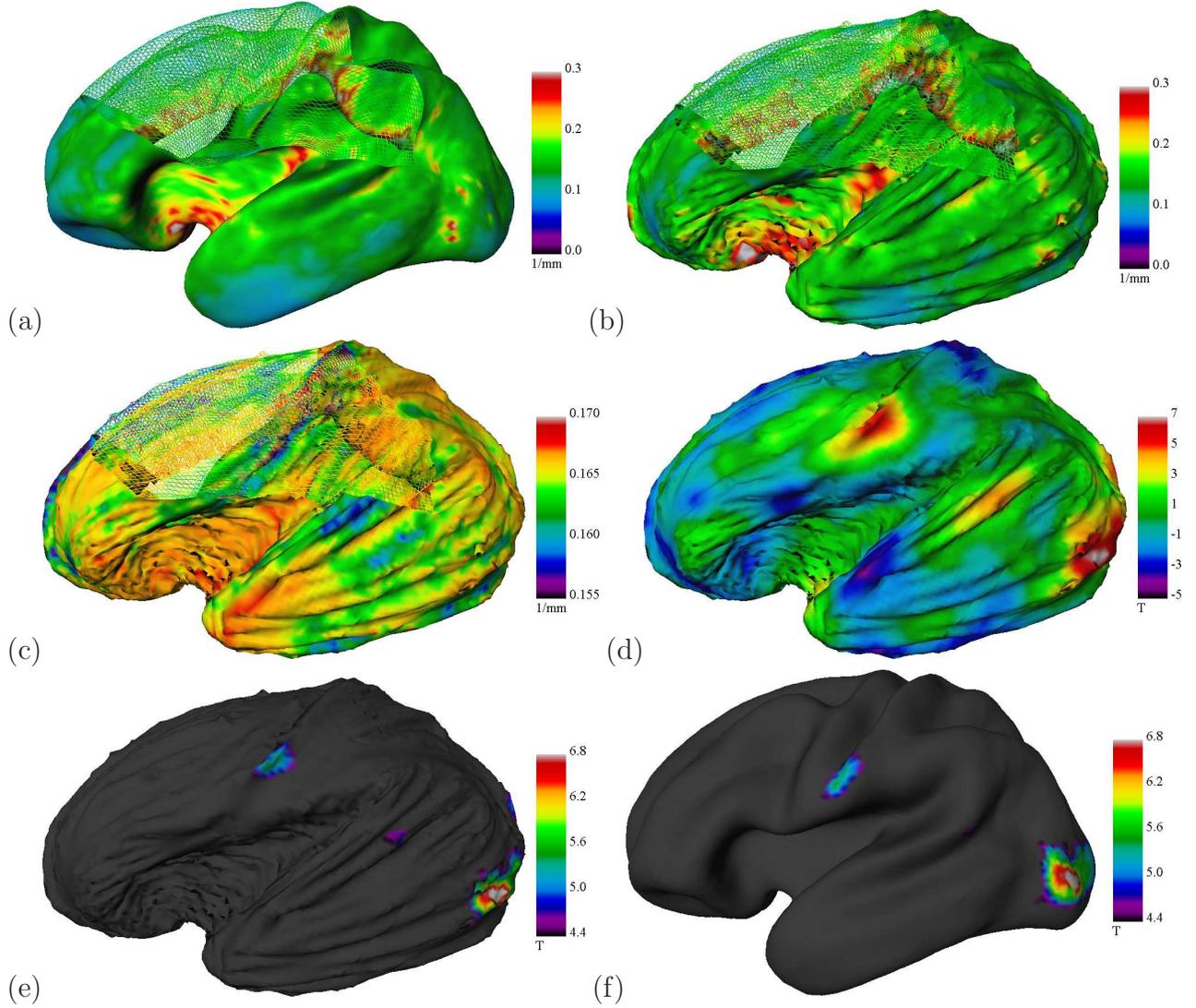


Figure 3: The same data as in Figure 1, showing part of the triangular mesh in the first three images. (a) Standard deviation of the derivative of the residuals - it varies considerably across the brain, indicating local non-isotropy; (b) warping of the data to approximate local isotropy by multiplying the edge lengths by the standard deviation of the derivative in (a) - note that regions with low values in (a) are shrunk to make them rougher, and regions with high values are expanded (by folding) to make them smoother; (c) standard deviation of the derivative recalculated using the new edge lengths (note the change in colour scale) - the values are almost constant, indicating that the data is almost locally isotropic (exact local isotropy could be achieved in 5 dimensions, by the Nash Embedding Theorem); (d) T-statistic from Figure 1 on the ‘isotropic’ mesh; (e) T thresholded at the $P = 0.05$ threshold $t = 4.43$ determined from (7) using the estimated LKC of the cortical surface from (19), well approximated by the intrinsic volumes of the ‘isotropic’ mesh; (f) the same as in (e) but back on the average cortical surface.

However a moment’s reflection will show that we need to be cautious. If we let the number of embedding dimensions \tilde{D} become infinite (e.g. as the sample size increases), then

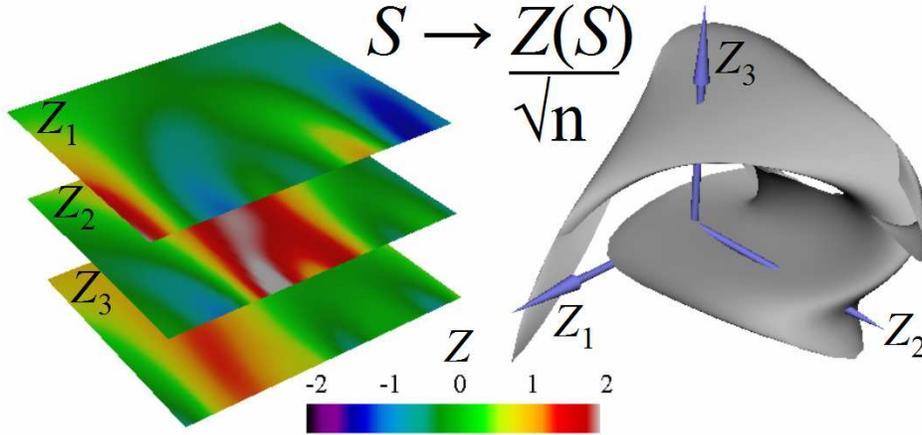


Figure 4: Example of warping S (left: square, $M = D = 2$) to sample isotropy by replacing the coordinates of $s \in S$ by the random fields $Z(S)/\sqrt{n}$, where $Z(s) = (Z_1(s), \dots, Z_n(s))$ (right: $n = \tilde{D} = 3$). Our proposed estimator of the LKC of S is the intrinsic volume of $Z(S)/\sqrt{n}$. Note that $Z(S)$ self-intersects because $n \leq 2M$; $n > 2M$ avoids this, with probability 1. These self-intersections are ignored in calculating the intrinsic volumes.

the theory behind the P-value approximation (7) might break down. Fortunately this is not a concern since the Nash Embedding Theorem says that the number of dimensions required to achieve exact local isotropy is finite, and indeed $\tilde{D} \leq M + M(M + 1)/2$ will suffice. Understanding how the Nash theorem works here is not a simple matter since we have said nothing about the Riemannian structure of the non-isotropic situation.

So in principle we can transform the parameter space S to local isotropy, then work out its intrinsic volumes, then use them in the P-value approximation (7). These new intrinsic volumes are in effect the LKC's of S .

3.2 Formal definition of LKC: variogram as metric

Taylor and Adler (2003) approach this in a different way. Instead of warping the parameter space to achieve local isotropy, they replace the Euclidean metric locally by the variogram, and define local distance between two points $s_1, s_2 \in S$ by

$$\text{Var}(Z(s_1) - Z(s_2))^{1/2}.$$

The intrinsic volumes in this new metric become the LKC's. This time the spatial correlation structure is used in the metric rather than in the warping. As suggested by the note after (10) (that the expected EC depends on the spatial correlation function only through its second

derivative at the origin), it turns out that the LKC only depends on the *local* metric or short-range spatial correlation, or equivalently the variance of the spatial derivative $\Lambda(s) = \text{Var}(\dot{Z}(s))$. Readers with some background in Riemannian geometry and some of the material in [Taylor and Adler \(2003\)](#) and [Adler and Taylor \(2006\)](#) will note that $\Lambda(s)$ is just the matrix of the afore-mentioned Riemannian metric on S in the natural basis $\partial/\partial s$. We thus write $\mathcal{L}_d(S, \Lambda)$ to denote the d -th LKC of the set S covered by the Gaussian random field $Z(s)$. In the isotropic case

$$\mathcal{L}_d(S, I_{D \times D}) = \mu_d(S).$$

It is straightforward to show that if A is a spatially constant non-singular matrix then

$$\mathcal{L}_d(S, A\Lambda A') = \mathcal{L}_d(A^{-1}S, \Lambda) = \mathcal{L}_d(S, \Lambda)/a^d$$

if $A = aI_{D \times D}$, where a is a scalar, and where $A^{-1}S = \{A^{-1}s : s \in S\}$. To simplify notation, we will sometimes drop the dependence of the LKC on the variance of the derivative of Z and write $\mathcal{L}_d(S) = \mathcal{L}_d(S, \Lambda)$. We note in passing that $\mathcal{L}_d(S)$ then satisfies the additivity property (18).

3.3 The ‘volume’ term or resels

The d -th LKC of S , and the one that makes the largest contribution to the P-value (12), is the ‘volume’ term

$$\mathcal{L}_D(S) = \int_S \det(\Lambda(s))^{1/2} ds, \tag{13}$$

a straightforward generalization of the LKC (3). In the neuroimaging literature,

$$\text{resels}_D(S) = (4 \log 2)^{-D/2} \mathcal{L}_D(S) \tag{14}$$

is known as the *resels* (*resolution elements*) ([Worsley et al. 1996](#)). The basis for this nomenclature is as follows. Suppose $Z(s)$ is modeled as Gaussian white noise convolved with an isotropic Gaussian-shaped spatial filter with Full Width at Half Maximum F . Then the D dimensional resels of S is the number of filters, or ‘resolution elements’, that fill S :

$$\text{resels}_D(S) = |S|/F^D,$$

where $|\cdot|$ is volume or Lebesgue measure. From (13) it can be shown that

$$\mathcal{L}_D(S) = (4 \log 2)^{D/2} |S|/F^D$$

which leads us to the general definition (14) in the non-isotropic case. Intuitively, the more resels in S , the more “effectively independent” observations are in S , so the higher the probability that the random field will exceed a given threshold, as seen in (12).

4 Sample estimator of LKC

Our motivation through the idea of warping to local isotropy is not wasted because we shall use it in this section to derive a sample estimator of LKC.

4.1 S a simplicial complex

In practical applications the data $Y_i(s)$ are always measured at a finite set of points $P = \{s_1, \dots, s_N\} \subset S$. What we are really after is the P-value of $\max_{s \in P} T(s)$, which we approximate by the P-value of $\max_{s \in S} T(s)$. Clearly this approximation is always conservative (since $P \subset S$), but still accurate provided P is a fine enough subsampling of the points in S .

The points in P can be joined to form a *simplicial complex* \mathcal{S} , defined as a set of subsets of P with the property that if a particular subset \mathcal{F} is in \mathcal{S} , then so are all the subsets of \mathcal{F} (Naiman and Wynn 1992, 1997) - see Figure 5(a). In the 3D case, a tetrahedral mesh is a simplicial complex if, for every tetrahedron in the mesh, all its triangles, edges and vertices are also in the mesh. With each simplex $\mathcal{F} \in \mathcal{S}$ we associate a subset F of Euclidean space consisting of the (closed) convex hull of the points in \mathcal{F} . With this convention we assume, in this section, that S is the union of all the simplices in \mathcal{S} .

In general there will be no restrictions on how \mathcal{S} is formed, but the accuracy of the P-value of $\max_{s \in S} T(s)$ as an approximation to $\max_{s \in P} T(s)$ will be best if, for fixed P , \mathcal{S} connects points that are as highly correlated as possible. This will encourage $\Lambda(s)$ to be small, so that the LKC’s of S are small, resulting in a smaller P-value (12).

4.2 The proposed estimator

Our aim is to give a sample estimator of the LKC of S . For this purpose we have available the n -vector of least squares residuals $r(s)$, $s \in P$, from the linear model (1), as defined by (4). The components of $r(s)$ are zero-mean Gaussian random fields with the same spatial

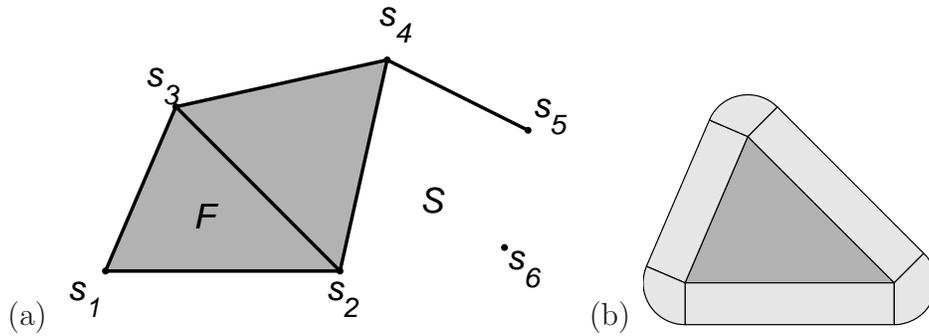


Figure 5: (a) Example of a simplicial complex $\mathcal{S} = \{\{s_1\}, \{s_2\}, \{s_3\}, \{s_4\}, \{s_5\}, \{s_6\}, \{s_1, s_2\}, \{s_1, s_3\}, \{s_2, s_3\}, \{s_2, s_4\}, \{s_3, s_4\}, \{s_4, s_5\}, \{s_1, s_2, s_3\}, \{s_2, s_3, s_4\}\}$ with $D = 2$ and $P = \{s_1, \dots, s_6\}$. F is the triangle associated with the simplex $\mathcal{F} = \{s_1, s_2, s_3\}$. The union of all the points, edges and triangles in \mathcal{S} is the set $S \subset \mathbb{R}^2$. (b) Graphical proof of the intrinsic volumes of a triangle - see Appendix A.1.1.

correlation structure as $Z(s)$. As in (5), define the normalised residuals as

$$u(s) = r(s)/\|r(s)\|. \quad (15)$$

Pursuing the idea of replacing the coordinates by the data, as suggested in the previous section and Figure 4, our estimator of the LKC of S is the intrinsic volume of the simplicial complex whose coordinates are the normalised residuals (15). Formally, let $\tilde{P} = u(P) = \{u(s) : s \in P\}$ embedded in $\tilde{D} = n$ dimensional Euclidean space, and let $\tilde{S} = u(S)$ be associated with the simplicial complex whose coordinates are \tilde{P} . Then the estimator is simply

$$\text{LKC estimator: } \hat{\mathcal{L}}_d(S) = \mu_d(u(S)), \quad (16)$$

an idea that was first proposed by Worsley et al. (1999). This equation (16) is probably the most important of the entire paper. It says that to estimate the LKC, simply replace the Euclidean coordinates by the normalised residuals, and proceed as if the data were isotropic.

The underlying idea is that $u(S)$ can be thought of as an estimator of S in isotropic space in the sense that the local geometry of $u(S)$ that is captured by the edge lengths of the simplicial complex is, on average, the same as the local geometry of S relative to $Z(s)$ or, equivalently r . By isotropic space we mean that

$$\|u(s_1) - u(s_2)\|^2 \approx \text{Var}(Z(s_1) - Z(s_2)) \quad (17)$$

so the squared Euclidean (isotropic) distances between points in $u(S)$ are approximately equal to the variogram.

4.3 A simple expression for calculating the estimator

To calculate the intrinsic volume of a simplicial complex S or \tilde{S} it would seem reasonable to work out the intrinsic volume of each component of \mathcal{S} , then use the additivity rule

$$\mu_d(A \cup B) = \mu_d(A) + \mu_d(B) - \mu_d(A \cap B) \quad (18)$$

(see Appendix A.1.4). This would lead to an inclusion-exclusion formula for $\mu_d(S)$. However a moment's reflection will show that although this simplifies the calculation, it is still far too cumbersome. The reason is that we must account not only for the intersection of two tetrahedra that share a common face, but also the intersection of two or more tetrahedra that share a common edge or vertex. In other words, there could be a huge number of intersection terms in such an inclusion-exclusion formula, many of which would simply cancel.

Instead we derive a new result using two classical combinatorial identities involving internal and external angles (31) known as *Sommerville's identities* (Grünbaum 2003). These identities, their generalizations and the intrinsic volumes of simplicial complexes are studied by physicists in quantum gravity. This area of study is usually referred to as Regge calculus in the physics literature (Budach 1989; Cheeger et al. 1984, 1986). The identities lead directly to the following proposition, which is proved in Appendix A.3:

Theorem 1

$$\mu_d(S) = \sum_{\mathcal{F} \in \mathcal{S}: \dim(\mathcal{F}) \geq d} (-1)^{\dim(\mathcal{F})-d} \mu_d(\mathcal{F}).$$

The beauty of this result is that we do not need to worry about the intersection of the simplices: Theorem 1 gives us a breakdown of the intrinsic volume of a simplicial complex purely in terms of the intrinsic volumes of its components. Using Theorem 1 with S replaced by $u(S)$, the estimator can be written as

$$\hat{\mathcal{L}}_d(S) = \mu_d(u(S)) = \sum_{\mathcal{F} \in \mathcal{S}: \dim(\mathcal{F}) \geq d} (-1)^{\dim(\mathcal{F})-d} \mu_d(u(\mathcal{F})). \quad (19)$$

All that is required are expressions for the intrinsic volumes of simplices. We shall give these expressions for a tetrahedral mesh in Section 5.2.

It will be noticed that there is still some cancelation in the breakdown in Theorem 1, and hence some unnecessary computation in (19). For example, when $d = D - 1$ all the surface area terms of the $(D - 1)$ -dimensional simplices cancel with all the surface area terms of the D -dimensional simplices whenever the former are interior to S , so that only the exterior surface areas count towards $\mu_{D-1}(u(S))$. This is because the $(D - 1)$ -dimensional intrinsic volume of a D -dimensional set is half its surface area. Since every interior $(D - 1)$ -dimensional simplex is contained in two adjacent D -dimensional simplices, then their surface area terms cancel.

It might be thought that interior contributions to other low-order intrinsic volumes also cancel, as they do for surface area. In other words, the low-order intrinsic volumes of \tilde{S} should only involve the boundary of S . But in general this is not the case. The reason is simple: if $\tilde{D} > D$, S is embedded in a higher dimensional space, so the interior of S is ‘exposed’ to become part of the boundary of \tilde{S} , as in Figure 4. Thus all simplices of \mathcal{S} are on the boundary of \tilde{S} , and potentially contribute to the intrinsic volume of \tilde{S} . In particular, if $D = 3$ then $\hat{\mathcal{L}}_1(S)$ has contributions from all simplices in \tilde{S} and these contributions do not necessarily cancel (Figure 9(a) shows an image of these contributions).

4.4 The ‘volume’ term or resels

The simplest term to work with is the ‘volume’ term $\hat{\mathcal{L}}_D(S)$ (13) or resels (14), where S is a union of D dimensional simplices in the simplicial complex \mathcal{S} . In practical applications it usually makes the largest contribution to the P-value approximation (12). From (19) the volume term is simply the sum of the volumes of the simplices in ‘isotropic’ space:

$$\hat{\mathcal{L}}_D(S) = |\tilde{S}| = \sum_{\mathcal{F} \in \mathcal{S} : \dim(\mathcal{F})=D} |\tilde{\mathcal{F}}|. \quad (20)$$

The volume of a simplex is straightforward. Let $u(s) = (u_1(s), \dots, u_n(s))'$ be the coordinates of the point s in ‘isotropic’ space, and let $u(s_0), u(s_1), \dots, u(s_D)$ be the vertices of a D -dimensional simplex \mathcal{F} in isotropic space. Let $e_j = u(s_j) - u(s_0)$ be the edge vectors, $j = 1, \dots, D$, and let $E = (e_1, \dots, e_D)$ be the $n \times D$ matrix of edge vectors in isotropic space. Then

$$|\tilde{\mathcal{F}}| = \det(E'E)^{1/2}/D!. \quad (21)$$

4.5 Consistency, bias and variance

In Appendices [A.4](#) and [A.5](#) we prove the following:

Theorem 2 *For fixed n and fixed sequence of mesh sizes Δ_m indexed by m that decrease to zero as $m \rightarrow \infty$, it is possible to construct a random sequence of meshes $\mathcal{M}_{\Delta,m}$ such that for each d*

$$\widehat{\mathcal{L}}_d(S_{\mathcal{M}_{\Delta}}) \xrightarrow{m \rightarrow \infty} \mu_d(u(S)), \quad \mathbb{P} - a.s..$$

Further, for fixed n and each d

$$\mathbb{E}(\mu_d(u(S))) = \mathcal{L}_d(S).$$

Remark: For large degrees of freedom ν and large S , the variance of the volume term of $\text{Var}(\widehat{\mathcal{L}}_D(S))$ is well approximated by

$$\frac{\text{Var}(\widehat{\mathcal{L}}_D(S))}{\mathcal{L}_D(S)^2} \simeq \frac{D(D+1)\pi^{D/2}}{4\nu\mathcal{L}_D(S)}, \quad (22)$$

in the sense that as $\nu \rightarrow \infty$,

$$\frac{\text{Var}(\widehat{\mathcal{L}}_D(S))}{\mathcal{L}_D(S)^2} = O(\nu^{-1})$$

and as some expansion factor $c \rightarrow \infty$ with ν fixed

$$\frac{\text{Var}(\widehat{\mathcal{L}}_D(cS))}{\mathcal{L}_D(cS)^2} = O(c^{-D}) = O(\mathcal{L}_D(cS)^{-1}).$$

Here $cS = \{cs : s \in S\}$. Some justification for this result is given in [Appendix A.6](#).

Since the expected EC [\(12\)](#) is a linear combination of LKC's, this implies that replacing these LKC's by our estimator $\widehat{\mathcal{L}}_d(S)$ gives a consistent and unbiased estimator of the expected EC, and hence the P-value approximation [\(12\)](#). Since the volume term is the dominant one in evaluating the P-value, [\(22\)](#) implies that the relative error in the estimated P-value decreases as the square root of the sample size times the volume term.

It is worth contrasting this with a naïve estimator formed as follows. We have at our disposal ν effectively independent copies of $Z(s)$, namely the residuals $r_i(s)$ (to within a constant). If the test statistic $T(s)$ is a T-statistic and if the degrees of freedom are large, then we can get a simple non-parametric estimator of the P-value of the maximum of $T(s)$ by the proportion of maxima of $r_i(s)$ (suitably normalized) that exceed the observed $T(s)$.

This very simple P-value estimator, which makes no assumptions at all about the correlation structure of the data, has coefficient of variation roughly $1/(P\nu)$, where P is the true (small) P-value. Contrasting this with (22) based on a random field model, we note that the ratio of the two is

$$\frac{\text{Var}(\text{random field P-value estimator})}{\text{Var}(\text{naïve P-value estimator})} \approx \frac{D(D+1)\pi^{D/2}P}{4\mathcal{L}_D(S)}.$$

In typical 3D applications in brain imaging, $\mathcal{L}_3(S) = 1000$ and $P = 0.05$, this factor is 0.0008, indicating an approximate 35-fold increase in accuracy for the random field model. However if the data is much smoother, so that $\mathcal{L}_3(S)$ is much smaller, then the increase in accuracy would not be as pronounced.

5 The 3D case

There are two reasons for considering the $D = 3$ case in detail. The first is that most data is sampled on a rectilinear mesh, and it is not easy to subdivide a rectilinear mesh into simplices in dimensions higher than 3. The second is that there are no simple workable expressions for intrinsic volumes of simplices in dimensions higher than 3. For the 3D case, we will address these two issues in the following sections.

5.1 Constructing a tetrahedral mesh to fill a set with smooth boundary

The first step, constructing a tetrahedral mesh from a rectilinear mesh, is straightforward. The ‘cubes’ of eight adjacent points on the rectilinear mesh can be divided into five tetrahedra in an alternating ‘checker board’ fashion (Figure 6(a)), or into six tetrahedra in an identical fashion (Figure 6(b)). Generalizing this to higher dimensions is straightforward (de Berg et al. 1997).

The next problem is how to fit a such a tetrahedral mesh to a smooth set. We illustrate the problem and the solution for the 2D case in Figure 7. In practice search regions S are often defined by the set of points where some smooth function $f(s)$ exceeds a threshold f_0 . For example in brain imaging, $f(s)$ might be grey matter density, and f_0 might be 0.1, so that our search region S is the set of points with at least 10% grey matter. The smooth function $f(s)$ is of course only defined at each lattice point, so we interpolate by e.g. tri-linear interpolation to all points in between.

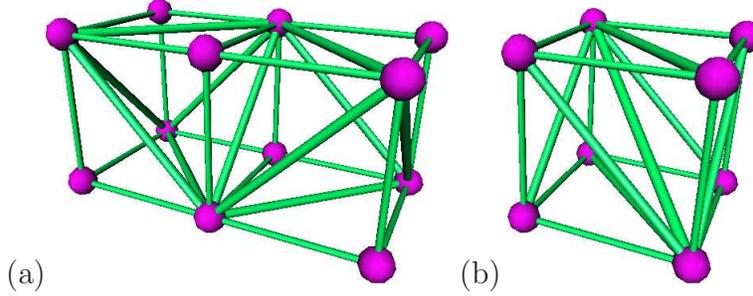


Figure 6: How to divide each ‘cube’ of a rectilinear mesh into (a) 5 tetrahedra in an alternating fashion; or (b) 6 tetrahedra in an identical fashion, each of equal volume.

There are many ways of triangulating the boundary (or outer surface) ∂S , such as the ‘marching cubes’ algorithm, but filling the interior with tetrahedra is much harder (Shephard and Georges 1991). We could identify the ‘cubes’ of 8 adjacent voxels that cut ∂S , then devise a scheme to fill the cut cubes with tetrahedra, by analogy with marching cubes. The hard part is to get the tetrahedra of one cut cube to match the tetrahedra of the next. Instead it seems far easier to first fill the lattice with tetrahedra as in Figure 6, then move the lattice points to ∂S . This is illustrated in Figure 7 for the 2D case, but the principle is the same for 3D. Specifically, if $\{s_1, s_2\}$ is an edge of the initial tetrahedral mesh such that $f(s_1) < f_0$ and $f(s_2) \geq f_0$ then the coordinates s_1 of the point outside S is moved to

$$s_1 w + s_2(1 - w), \quad \text{where } w = \frac{f(s_2) - f_0}{f(s_2) - f(s_1)}. \quad (23)$$

The same point may be moved more than once, but the mesh must be fine enough so that there is always a ‘free’ point to move. Having moved the point, the only remaining task is to interpolate the data $Y_i(s)$ to the new point. This is done by linear interpolation as in (23).

5.2 Intrinsic volumes

Let $\tilde{\cdot}$ denote replacement of Euclidean coordinates by normalised residuals, as used to define the estimator (16). Once the tetrahedral mesh is determined, the estimators of the LKC’s

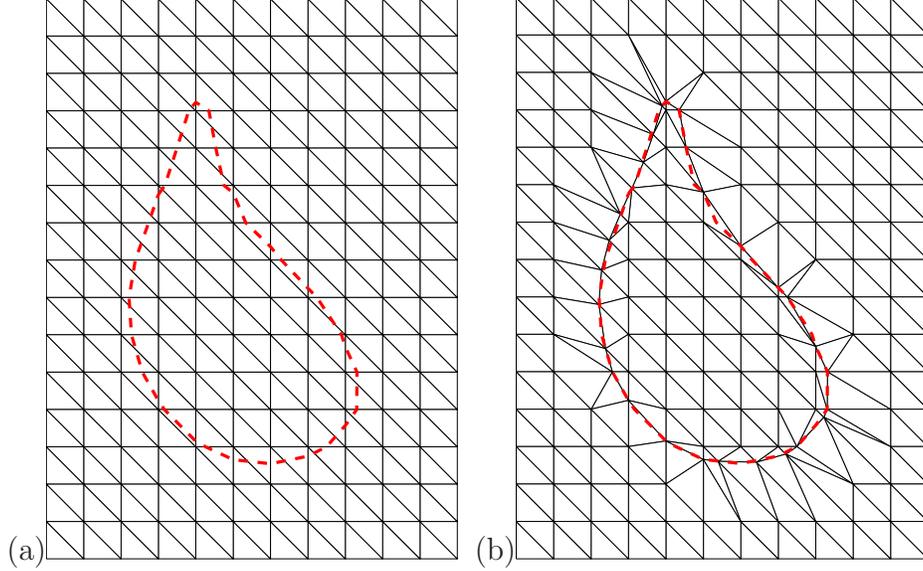


Figure 7: A smooth boundary (dashed) is filled with triangles by (a) identifying the edges that cross the boundary, then (b) moving the coordinates of the outside point to the boundary.

are, from Theorem 1:

$$\begin{aligned}
\widehat{\mathcal{L}}_0(S) &= \# \text{points} - \# \text{edges} + \# \text{triangles} - \# \text{tetrahedra}, \\
\widehat{\mathcal{L}}_1(S) &= \sum_{\text{edges}} \mu_1(\widetilde{\text{edge}}) - \sum_{\text{triangles}} \mu_1(\widetilde{\text{triangle}}) + \sum_{\text{tetrahedra}} \mu_1(\widetilde{\text{tetrahedron}}), \\
\widehat{\mathcal{L}}_2(S) &= \sum_{\text{triangles}} \mu_2(\widetilde{\text{triangle}}) - \sum_{\text{tetrahedra}} \mu_2(\widetilde{\text{tetrahedron}}), \\
\widehat{\mathcal{L}}_3(S) &= \sum_{\text{tetrahedra}} \mu_3(\widetilde{\text{tetrahedron}}),
\end{aligned} \tag{24}$$

since \mathcal{L}_0 is the Euler characteristic, which is one for all simplices. For the other intrinsic volumes,

$$\begin{aligned}
\mu_1(\text{edge}) &= \text{edge length}, \\
\mu_1(\text{triangle}) &= \frac{1}{2} \text{ perimeter length}, \\
\mu_1(\text{tetrahedron}) &= \sum_{\text{edges}} (\text{edge length}) \times (\pi - \text{interior angle}) / (2\pi), \\
\mu_2(\text{triangle}) &= \text{surface area}, \\
\mu_2(\text{tetrahedron}) &= \frac{1}{2} \text{ surface area}, \\
\mu_3(\text{tetrahedron}) &= \text{volume}.
\end{aligned} \tag{25}$$

Formula (21) allows us to calculate all the intrinsic volumes in (25) except the interior angle of the tetrahedron, which is calculated as follows. Let

$$c_{jk} = e'_j e_k - e'_j e_1 e'_1 e_k / e'_1 e_1, \quad 2 \leq j \leq k \leq 3,$$

where $e_j = u(s_j) - u(s_0)$ is the edge vector in isotropic space, as previously defined in (21). Then the interior angle along the edge $\{s_0, s_1\}$ is

$$\theta = \arccos(c_{23} / \sqrt{c_{22} c_{33}}). \quad (26)$$

As already noted after (19), there is some unnecessary computation in (24). All the surface area terms of the tetrahedra cancel with the surface area terms of the triangles whenever the triangles are inside S (i.e. the intersection of two adjacent tetrahedra). Only the triangles on the outer surface count towards $\widehat{\mathcal{L}}_2(S)$. This is because there is a $-\frac{1}{2}$ (surface area) contribution from each of the two tetrahedra on either side of the triangle that cancels with the (surface area) contribution from the triangle itself.

However in general this is not the case, and in particular it is not the case for $\widehat{\mathcal{L}}_1(S)$, where the contributions of the edges, triangles and tetrahedra do not necessarily cancel at interior edges. In fact the contributions of interior edges to $\widehat{\mathcal{L}}_1(S)$ is related to the curvature of \widetilde{S} , which we will actually estimate for the example in Section 6.2 (see Figure 9(a)).

If on the other hand \mathcal{S} is a triangular mesh, then all the interior contributions do cancel. In fact when \mathcal{S} is homeomorphic to a 2-sphere, so that all edges are interior, then $\widehat{\mathcal{L}}_1(S) = 0$, as we shall see next in Section 6.1.

Trying to devise a general way of counting the contributions of each simplex only once is a daunting task. Instead it seems easier to write computer code for (24), and in general (16), and work in double precision. This has been implemented as part of FMRISTAT, a Matlab package for the statistical analysis of fMRI data - see <http://www.math.mcgill.ca/keith/fmristat/#resels>.

6 Applications

6.1 Cortical thickness

Before proceeding, it is natural to ask how we know we are properly matching the cortical thickness at the same anatomical region, in other words, how do we address the issue of

surface-to-surface registration. This is fully discussed in [Lerch and Evans \(2005\)](#). Briefly, each brain is linearly transformed to a common stereotaxic reference space, so that overall brain sizes are equalized. This means that cortical thickness measures are relative to brain size, not absolute. The white matter, grey matter and cerebrospinal fluid are segmented. The brain is placed inside a triangulated sphere, which is then ‘shrink-wrapped’ to the grey-white boundary (the inner cortical surface), while sliding the mesh along the surface to maintain roughly equal sized triangles. This surface is then expanded out to the grey-cerebrospinal fluid boundary, to find the outer cortical surface. The cortical thickness is then measured at each node on the mid-cortical surface, which then becomes our raw data. No specific surface-to-surface registration is needed (as in other methods of cortical surface extraction) because the initial triangular mesh is identical and always positioned around the brain in exactly the same way. This means that node number 1000 (say) is always in roughly the same place on every brain. The success of this ‘registration’ can be gauged by looking at [Figure 1](#), which shows the simple average of the coordinates of each node over all 321 surfaces. The fact that [Figure 1](#) still looks like a brain, with many of the main anatomical features such as the central sulcus clearly visible, shows how well the surfaces are registered.

One final but important data preprocessing step should be mentioned. It is well known that some subjects have overall thinner cortex than others, due possibly to differing contrast between the grey and white matter that may shift the cortical boundary. This can be seen in the plot of the average cortical thickness against age shown in [Figure 8](#), which also shows that males have overall thicker cortex than females by $0.10 \pm 0.02\text{mm}$. To remove this overall effect, each subject’s average cortical thickness (averaged over all surface nodes) was subtracted from that subject’s cortical thickness, so that all thickness measurements were standardised to the same global average. This means that the male minus female cortical thickness difference ([Figure 1\(b\)](#)) must sum to zero across all surface points, or in other words, a positive difference in one part of the brain must be matched by a negative difference in another. The advantage of this is that it divorces our analysis from global effects, as in [Figure 8](#), and focuses it on localized regional effects.

Proceeding with the analysis, S is the 2D cortical surface, and since it is a closed surface, its boundary ∂S is empty. The sample standard deviation of the numerical derivative $\|e_j\|/\|s_j - s_0\|$, averaged over all neighbours s_j of s_0 , is shown in [Figure 3\(a\)](#). We can see

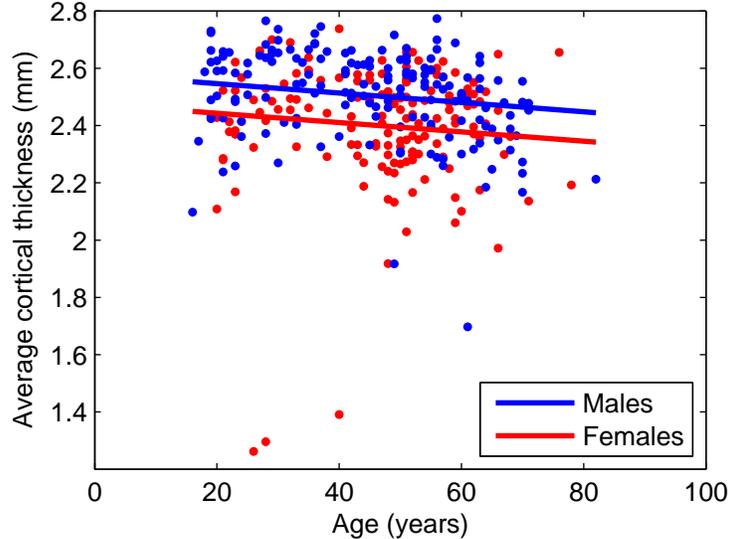


Figure 8: Average cortical thickness of the $n = 321$ subjects. Cortical thickness decreases with age, and males have overall thicker cortex than females by $0.10 \pm 0.02\text{mm}$. Average cortical thickness was subtracted from the thickness measurements at each point on the cortical surface before the analysis in Figures 1 and 3.

that it varies considerably across the brain, indicating non-isotropy. In Figure 3(b) we tried to warp the brain to approximate local isotropy by 4000 iterations of an algorithm that readjusts the coordinates \tilde{s}_0 of each point to minimise $\sum_{\text{points},j} (||\tilde{s}_j - \tilde{s}_0||^2 - ||e_j||^2)^2$ with the goal of achieving

$$||\tilde{s}_j - \tilde{s}_0||^2 \approx ||e_j||^2.$$

This tends to expand regions with high $||e_j||/||s_j - s_0||$ to make them smoother, and shrink regions with low $||e_j||/||s_j - s_0||$ to make them rougher. The success of this can be judged in Figure 3(c), where $||e_j||/||\tilde{s}_j - \tilde{s}_0||$ is almost constant. If we had an extra two dimensions, we could in principle achieve exact local isotropy in 5 dimensions by the Nash Embedding Theorem, but it is not easy to visualize a 5D object.

Figure 3(d) shows the T-statistic from Figure 1 on the ‘isotropic’ mesh. To determine the $P = 0.05$ threshold for the maximum T-statistic, the LKC’s of S were calculated from (24) and (25):

$$\begin{aligned} \hat{\mathcal{L}}_0(S) &= 2 \\ \hat{\mathcal{L}}_1(S) &= 0 \\ \hat{\mathcal{L}}_2(S) &= 2334.2. \end{aligned} \tag{27}$$

The Euler characteristic $\widehat{\mathcal{L}}_0(S) = 2$ because S is a closed surface; $\widehat{\mathcal{L}}_1(S) = 0$ because each edge is the boundary of two triangles, so all the edge contributions cancel; $\widehat{\mathcal{L}}_2(S) = 2334.2$, which would also equal the surface area of the warped brain if the warping had achieved exact local isotropy. The relative error in this last LKC, and hence the resulting P-values, is, from (22), 0.0025.

The EC densities of the T-statistic with $\nu = 318$ degrees of freedom from (10) were used in (12) to get approximate P-values. The $P = 0.05$ threshold was $t = 4.43$, and Figure 3(e) shows the excursion set of the T-statistic thresholded at this value. Figure 3(f) puts it back on the average cortical surface, allowing us to identify those regions where males have thicker cortex than females. We conclude that males have thicker cortex than females in the occipital lobe (an estimated difference of $0.44 \pm 0.06\text{mm}$), and a region near the base of the central sulcus (an estimated difference of $0.35 \pm 0.06\text{mm}$). Repeating this for negative peaks shows that females have thicker cortex in the frontal lobe. If thicker cortex correlates with increased function, then these conclusions reinforce the hypothesis that males are better at visual and motor tasks, whereas females are better at emotional and perception tasks.

6.2 fMRI

We apply our methods to a 3D data set from a functional magnetic resonance imaging (fMRI) experiment fully described in Taylor and Worsley (2006a). A subject was scanned every 2.5 seconds while listening to different sentences read by different speakers presented every 3.33 seconds, for a total of 191 scans. This was repeated 4 times on each of 14 subjects. The object of the experiment was to detect the (sparse) regions of the brain that respond to listening to different sentences, compared to listening to the same sentence. We used straightforward linear mixed effects models together with spatial smoothing to boost the degrees of freedom (Worsley et al. 2002; Worsley 2003). The result is a T-statistic image with 40 effective degrees of freedom for the contrast just described.

The search region S is the set of points inside the brain defined by the function $f(s)$ equal to the fMRI value minimized over each scan and each subject thresholded at $f_0 = 6330$. This threshold gave a reasonable brain mask that covered every scan on every subject (Figure 9). The FMRISTAT software was used to estimate LKC as follows (see in particular <http://www.math.mcgill.ca/keith/fmristat/#resels>) - note that FMRISTAT works in

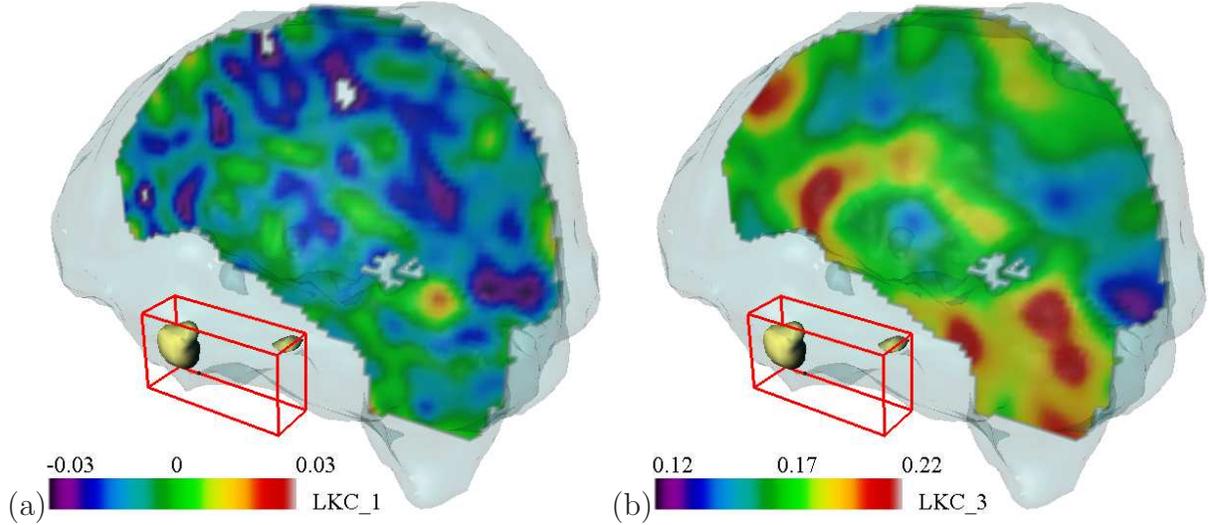


Figure 9: fMRI example. The yellow blobs inside the red box are the T-statistic (40 df) for the “different sentence” minus “same sentence” contrast thresholded at $t = 5.83$ ($P = 0.05$). There are two differences in the left mid-temporal gyrus, and a third very tiny one just visible on the lower box edge in the left inferior temporal gyrus. Inter-hemispheric slices show the cube root of the estimated LKC densities, smoothed 10mm, for (a) $d = 1$, and (b) $d = 3$, in units of mm^{-1} . The integral of these densities over the search region S (transparent) is the contribution from the interior of S to $\hat{\mathcal{L}}_d(S)$.

terms of resels, rather than LKC. The 3D rectilinear mesh was divided into five tetrahedra as in Figure 6(a), then the vertices just outside ∂S were moved to ∂S to fill S with tetrahedra as explained in Section 5.1. LKC’s of S were estimated by the method in Section 5.2 using the whitened least-squares residuals, to give

$$\begin{aligned}
 \hat{\mathcal{L}}_0(S) &= 9 \\
 \hat{\mathcal{L}}_1(S) &= 176.3 \\
 \hat{\mathcal{L}}_2(S) &= 1037.6 \\
 \hat{\mathcal{L}}_3(S) &= 9441.1.
 \end{aligned}
 \tag{28}$$

Although S is connected, it contains several small hollows, to give an EC of 9, the zeroth order LKC. The $P = 0.05$ threshold determined from (10), (12) and (28) was $t = 5.831$, and the resulting regions are shown in Figure 9. The relative error in the last LKC, and hence the resulting P-values, is, from (22), 0.0024.

It is not easy to visualize the 3D search region S warped to approximate local isotropy, as we did in the 2D example in Section 6.1. To get some idea of what this warping looks like, we can instead produce images of the LKC densities. These are defined as the contributions

from the interior of S to $\widehat{\mathcal{L}}_d(S)$ per unit volume. In other words, these densities, integrated over S , produce the contributions from the interior of S to $\widehat{\mathcal{L}}_d(S)$. The $d = 3$ density is just

$$\det(\widehat{\Lambda})^{1/2} = |\widetilde{F}|/|F|$$

for each tetrahedron F . The $d = 2$ density is zero, as explained in Section 4.3. The most interesting case is the $d = 1$ density. The contribution to $\widehat{\mathcal{L}}_1(S)$ at an interior edge \widetilde{F} is

$$|\widetilde{F}| \left(1 - \sum \theta/(2\pi)\right)$$

where θ is the interior angle (26) of a tetrahedron that contains \widetilde{F} , and the summation is over all such tetrahedra. The $d = 1$ density is positive if the sum of the interior angles is less than 2π , that is, if \widetilde{S} is (locally) positively curved. It is negative if \widetilde{S} is (locally) negatively curved, and zero if \widetilde{S} is locally flat.

Slices of these two densities are shown in Figure 9. From the $d = 1$ density we can see that the brain as a whole appears to be slightly negatively curved. The $d = 3$ density is low in grey matter regions, indicating that $|\widetilde{F}|$ is small because grey matter neighbours are more spatially correlated than neighbours in other parts of the brain. To achieve local isotropy, F should be shrunk in these regions by bringing grey matter neighbours closer together.

We should confess that all the effort of setting up the tetrahedral mesh to get the lower order LKC's makes very little difference to the final threshold in this case (see Table 1). The $d = 3$ volume term can be easily estimated, without using the tetrahedral mesh, from the variance of numerical derivatives on the rectilinear mesh. Approximating the search region \widetilde{S} by a ball with this same volume term gives exactly the same P-value and threshold, to three decimals. Using just the $d = 3$ term in (12) and ignoring the others gives a very slightly lower P-value and threshold. Finally it must be noted that, since the data is coarsely sampled, a simple Bonferroni bound for the 172,074 lattice points inside S gives a higher P-value and threshold that is almost as good. A lower bound on *lattice* maxima, found using the expected number of discrete local maxima (DLM) (Taylor et al. 2007; Worsley and Taylor 2005), gives the lowest P-value and threshold, and since it is conservative (under most conditions) it is perhaps the most accurate.

Method	P at $t = 5.831$	t at $P = 0.05$
Using the estimated LKC's	0.050	5.831
S a sphere with the same volume term	0.050	5.831
Only the volume term	0.047	5.812
Bonferroni on the 172,074 lattice points in S	0.070	5.937
Discrete Local Maxima (DLM)	0.030	5.667

Table 1: Comparison of P-values P and thresholds t for the fMRI example using different methods.

A Appendix

A.1 Intrinsic volume

The d -th intrinsic volume of a set S is a generalization of its volume to d -dimensional measures. For example the $(D - 1)$ th intrinsic volume of $S \subset \mathbb{R}^D$ is half its surface area, and the 0th intrinsic volume is the Euler characteristic of S . We give four ways of defining intrinsic volumes; the first is an implicit definition as coefficients of a polynomial, the second and third are explicit for simplices and for ‘smooth’ sets, and the fourth is a characterization through an additivity property.

A.1.1 Implicit definition

The first definition is *implicit*, identifying the intrinsic volumes as coefficients in a certain polynomial. This definition comes from the Steiner-Weyl volume of tubes formulae which states that if S is locally convex, then for r small enough

$$|\text{Tube}(S, r)| = \sum_{d=0}^D \omega_{D-d} r^{D-d} \mu_d(S) \quad (29)$$

where $|\cdot|$ denotes Lebesgue measure and $\text{Tube}(S, r) = \{x : \min_{s \in S} \|s - x\| \leq r\}$ is the tube of radius r around S and $\omega_d = \pi^{d/2} / \Gamma(d/2 + 1)$ is the Lebesgue measure of the unit ball in \mathbb{R}^d . Using (29), Figure 5(b) gives a graphical proof of the following result for a triangle ($D = 2$):

$$\begin{aligned} \mu_0(S) &= 1 = \varphi(S) \\ \mu_1(S) &= \frac{1}{2} \text{perimeter length}(S) \\ \mu_2(S) &= |S|. \end{aligned} \quad (30)$$

A.1.2 Simplex

For a closed simplex S , the intrinsic volume $\mu_d(S)$ is determined by all faces F of dimension d . The contribution of a such a face F is seen to be

$$|F| \cdot \beta(F, S) \tag{31}$$

where $|F|$ is the d dimensional surface area of F and $\beta(F, S)$ is the *external angle* of F in S defined to be the proportion of unit normal vectors in $D - d$ dimensions perpendicular to F that project from the tube to F . Therefore,

$$\mu_d(S) = \sum_{F \subseteq S : \dim(F)=d} |F| \beta(F, S). \tag{32}$$

In Figure 5(b) it is easy to see that the external angle of a vertex F in the triangle S is just

$$(\pi - \text{internal angle at } F)/(2\pi).$$

A.1.3 Smooth sets

If, on the other hand S is a compact domain in \mathbb{R}^2 with smooth boundary, then it is easy to see that (30) still holds. In higher dimensions, using the implicit definition as coefficients in the volume of tube formula and using some (straightforward, but detailed) calculus, it is possible to derive the intrinsic volumes of a smooth domain in \mathbb{R}^D , which we now describe. Let $C(s)$ be the $(D - 1) \times (D - 1)$ inside curvature matrix at $s \in \partial S$, the boundary of S . To compute the intrinsic volumes, we need the *det-traces* of a square matrix: for a $d \times d$ symmetric matrix A , let $\text{detr}_j(A)$ denote the sum of the determinants of all $j \times j$ principal minors of A , so that $\text{detr}_d(A) = \det(A)$, $\text{detr}_1(A) = \text{tr}(A)$, and we define $\text{detr}_0(A) = 1$. Let $a_d = 2\pi^{d/2}/\Gamma(d/2)$ be the $(d - 1)$ -dimensional Hausdorff (surface) measure of a unit $(d - 1)$ -sphere in \mathbb{R}^d . For $d = 0, \dots, D - 1$ the d -th intrinsic volume of S is

$$\mu_d(S) = \frac{1}{a_{D-d}} \int_{\partial S} \text{detr}_{D-1-d}\{C(s)\} ds, \tag{33}$$

and $\mu_D(S) = |S|$, the Lebesgue measure of S . Note that $\mu_0(S) = \varphi(S)$ by the Gauss-Bonnet Theorem, and $\mu_{D-1}(S)$ is half the surface area of S . This leads immediately to (9).

A.1.4 Characterization through additivity

Finally, intrinsic volumes satisfy the additivity rule (18)

$$\mu_d(A \cup B) = \mu_d(A) + \mu_d(B) - \mu_d(A \cap B),$$

for any two sets A, B , and are invariant under rigid motions, that is, for any rigid motion g

$$\mu_d(gA) = \mu_d(A)$$

and any scaling $c > 0$

$$\mu_d(cA) = \mu(\{cs : s \in A\}) = c^d \mu_d(A).$$

These three relations actually characterize intrinsic volumes up to a constant (Hadwiger 1957), at least on the convex ring formed by finite unions of convex sets.

A.2 EC density

For $d > 0$ the EC density of a smooth stationary random field is, writing $T = T(s)$,

$$\rho_d(t) = \mathbb{E} \left((T \geq t) \det(-\ddot{T}_d) \mid \dot{T}_d = 0 \right) \mathbb{P}(\dot{T}_d = 0), \quad (34)$$

where dot notation with subscript d denotes differentiation with respect to the first d components of s , first as a row vector, then as a column vector. For $d = 0$, $\rho_0(t) = \mathbb{P}(T \geq t)$.

There is a much simpler way of getting EC densities when T is built from i.i.d. Gaussian random fields, such as T- and F-statistic random fields for contrasts from the linear model (1). This very powerful result is based on the *Gaussian Kinematic Formula* discovered by Taylor (2006). The idea is to take the Steiner-Weyl volume of tubes formula (29) and replace the search region by the rejection region, and volume by probability. The coefficients of powers of the tube radius are (to within a constant) the EC densities we seek.

The details are as follows. Suppose $T = f(Z_1, \dots, Z_n)$ is a function of i.i.d. UGRF's Z_1, \dots, Z_n . Put a tube of radius r about the rejection region $R = \{z : f(z) \geq t\} \subset \mathbb{R}^n$, evaluate the probability content of the tube (using the $N_n(0, I_{n \times n})$ distribution of Z_1, \dots, Z_n), and expand as a power series in r . Then

$$\mathbb{P}((Z_1, \dots, Z_n) \in R \oplus B(0, r)) = \sum_{d=0}^{\infty} \frac{r^d}{d!} (2\pi)^{d/2} \rho_d(t). \quad (35)$$

For T a UGRF, $n = 1$, $f(z) = z$, and this leads directly to

$$\rho_d(t) = \left(\frac{-1}{\sqrt{2\pi}} \frac{\partial}{\partial t} \right)^d \mathbb{P}(T \geq t). \quad (36)$$

For T a T-statistic random field with $\nu = n - 1$ degrees of freedom,

$$f(z_1, \dots, z_n) = z_1 \left(\sum_{j=2}^n z_j^2 / \nu \right)^{-1/2}.$$

The rejection region R is a cone, and after a little elementary geometry,

$$\mathbb{P}((Z_1, \dots, Z_n) \in R \oplus B(0, r)) = \mathbb{P}\left(Z > t\sqrt{V/\nu} - r\sqrt{1 + t^2/\nu}\right) + O(r^{\nu+1}), \quad (37)$$

where $V \sim \chi_\nu^2$ independently of $Z \sim N(0, 1)$. Expanding the right hand side of (37) in powers of r and equating it to that of (35) gives the EC densities (10) for $d \leq \nu$. EC densities for other non-Gaussian random fields built from Gaussians can be found in Worsley (1994), Cao and Worsley (1999a), Cao and Worsley (1999b) and Taylor and Worsley (2006b).

A.3 Proof of Theorem 1

Sommerville's identities, which can be proved by a fairly straightforward inclusion-exclusion argument, are the following: let S be a simplex and $F \subseteq S$ a face. Then

$$\begin{aligned} \beta(F, S) &= \sum_{F \subseteq G \subseteq S} (-1)^{\dim(G) - \dim(F)} \beta(F, G) \\ \alpha(F, S) &= \sum_{F \subseteq G \subseteq S} (-1)^{\dim(S) - \dim(G)} \alpha(G, S) \end{aligned}$$

where $\dim(F)$ is the dimensionality of F , equal to one less than the number of points in \mathcal{F} and $\alpha(F, S)$ is the internal angle of F in S which is the proportion of unit normals leaving F but staying in S . It actually holds for arbitrary polytopes (the convex hull of a finite pointset) in Euclidean space. For completeness, we first sketch a proof of Somerville's identity for internal angles, from which the corresponding one for external angles follows by a duality argument. We then prove the result true for simplices, then appeal to additivity (18).

The internal angle of a face F of dimension d in a D -dimensional simplex S is the normalized volume of the spherical cross-section of a convex cone determined by the intersection of $k = D - d$ of half-spaces $\{H_1, \dots, H_k\}$ each perpendicular to $\text{sp}(F)$, the linear space spanned by F . Further, each subset of $\{H_1, \dots, H_k\}$ corresponds to some face $G \supseteq F$ contained in S .

Let V_{D-d-1} denote the surface measure on $S^\perp(F)$, the sphere of unit normals perpendicular to $\text{sp}(F)$, then

$$\alpha(F, S) = V_{D-d-1} \left(\prod_{j=1}^k 1_{H_j} \right)$$

where we have committed the slight abuse of notation of identifying the half-spaces with the half-spheres in $S^\perp(F)$ they determine. By inclusion-exclusion

$$\prod_{j=1}^k 1_{H_j} = 1 - \sum_j 1_{H_j^c} + \cdots + (-1)^k \sum_{j_1 < \cdots < j_k} \prod_{l=1}^k 1_{H_{j_l}^c}$$

and by symmetry

$$\begin{aligned} \alpha(F, S) &= 1 - \sum_j V_{D-d-1} (1_{H_j^c}) + \cdots + (-1)^k \sum_{j_1 < \cdots < j_k} V_{D-d-1} \left(\prod_{l=1}^k 1_{H_{j_l}^c} \right) \\ &= 1 - \sum_j V_{D-d-1} (1_{H_j}) + \cdots + (-1)^k \sum_{j_1 < \cdots < j_k} V_{D-d-1} \left(\prod_{l=1}^k 1_{H_{j_l}} \right) \\ &= \sum_{F \subseteq G \subseteq S} (-1)^{\dim(S) - \dim(G)} \alpha(G, S). \end{aligned}$$

Turning to the main result of the Theorem, by (32), if S is a simplex

$$\begin{aligned} \mu_d(S) &= \sum_{G \subseteq S: \dim(G)=d} |G| \cdot \sum_{G \subseteq F \subseteq S} (-1)^{\dim(F) - \dim(G)} \beta(G, F) \\ &= \sum_{F \subseteq S: \dim(F) \geq d} (-1)^{\dim(F) - d} \sum_{G \subseteq F: \dim(G)=d} |G| \cdot \beta(G, F) \\ &= \sum_{F \subseteq S: \dim(F) \geq d} (-1)^{\dim(F) - d} \mu_d(F). \end{aligned}$$

A.4 Proof of Theorem 2: consistency

Turning to consistency, we must impose some conditions on the UGRFs $Z_i, 1 \leq i \leq n$ involved in defining $\tilde{S} = u(S)$. We must assume that the collection of random variables $Z(s), s \in S$ considered as a subset of L^2 is a smooth manifold with boundary and that the triangulations of \tilde{S} “respect” this structure, at least asymptotically. This will generally be the case if the map from M to $L^2, s \mapsto Z(s)$ (i.e. the random field) is smooth and 1:1, so that the random field can be thought of as an embedding of M in L^2 . In this case, the variogram is a true metric, and most reasonable triangulation schemes of S will respect the physical shape of S .

Assuming such sufficient regularity, without loss of generality, we can simplify the problem to studying the trivial linear model

$$Y_i(s) = Z_i(s), \quad 1 \leq i \leq n \quad (38)$$

so that (17) becomes exact as $n \rightarrow \infty$, i.e. so that

$$\lim_{n \rightarrow \infty} \|u(s_1) - u(s_2)\|^2 = \text{Var}(Z(s_1) - Z(s_2)). \quad (39)$$

Working under (38), let us fix a triangulation \mathcal{M} of S and set $\tilde{S}_{n,\mathcal{M}}$ to be the simplicial complex with edge lengths $\|u(s_1) - u(s_2)\|$. It is not hard to see that

$$\hat{\mathcal{L}}_d(\tilde{S}_{n,\mathcal{M}}) \xrightarrow{n \rightarrow \infty} \mathcal{L}_d(\tilde{S}_{\mathcal{M}})$$

where $\tilde{S}_{\mathcal{M}}$ is a triangulation of S with edge lengths $\text{Var}(Z(s_1) - Z(s_2))^{1/2}$. Having sent $n \rightarrow \infty$ we now can take a fixed sequence \mathcal{M}_k of finer and finer triangulations and the results of Cheeger et al. (1984) imply

$$\mathcal{L}_d(\tilde{S}_{\mathcal{M}_k}) \xrightarrow{k \rightarrow \infty} \mathcal{L}_d(S).$$

This is different than saying that the estimates recover the local curvature matrix. In fact the results in Cheeger et al. (1984), which seem to be the strongest available for convergence of intrinsic volumes, are not even strong enough to imply that the (signed) curvature measures, of which the intrinsic volumes are the (signed) total mass, converge vaguely as finite (signed) measures. However, if one integrates the curvature measures over small balls, and other sets with smooth boundaries, these quantities do converge (Cheeger et al. 1984).

A.5 Proof of Theorem 2: unbiasedness

In order to apply the results in Cheeger et al. (1984), the triangulations \mathcal{M}_k must satisfy some uniform non-degeneracy conditions. This makes claims about unbiasedness rather tricky: for instance, it seems difficult to prove that for fixed n and a *fixed* sequence of triangulations \mathcal{M}_k that the estimates

$$\hat{\mathcal{L}}_d(\tilde{S}_{n,\mathcal{M}_k}) \quad (40)$$

are consistent. However, if we allow the triangulations to be random then it is possible to define a sequence of triangulations \mathcal{M}_k such that the estimates (40) converge to $\mu_d(u(S))$.

It seems then that the best result in terms of unbiasedness we can hope for is the following

$$\mathbb{E}(\mu_d(u(S))) = \mathcal{L}_d(S), \quad 0 \leq d \leq D. \quad (41)$$

Without loss of generality, we work now with the trivial model (38) in which case

$$u(s) = Z(s)/\|Z(s)\| \quad (42)$$

where, with some abuse of notation, $Z(s) = (Z_1(s), \dots, Z_n(s))'$ is the n -vector of independent Gaussian random fields with zero mean and unit variance each with the same distribution as $Z(s)$.

The simplest term to work with is the volume term $\mu_D(u(S))$. In practical applications this term usually makes the largest contribution to the P-value approximation (12). The estimators are indeed unbiased for all d , but we omit these details which can be found in the upcoming monograph (Adler et al. 2006).

Unbiasedness of the volume term has already been shown by Kiebel et al. (1999) and Hayasaka et al. (2004) but we reproduce the result here. Standard formulas from multivariate calculus for computing surface area show that

$$\mu_D(u(S)) = \int_S \det(\dot{u}(s)' \dot{u}(s))^{1/2} ds$$

while

$$\mathcal{L}_D(S) = \int_S \det(\Lambda(s))^{1/2} ds.$$

Therefore, it is sufficient to show that

$$\mathbb{E}(\det(\dot{u}(s)' \dot{u}(s))^{1/2}) = \det(\Lambda(s))^{1/2}$$

for each $s \in S$.

In the trivial model (38)

$$\dot{u}(s) = \left(I_{n \times n} - \frac{Z(s)Z(s)'}{Z(s)'Z(s)} \right) \frac{\dot{Z}(s)}{\|Z(s)\|}. \quad (43)$$

Now since $\text{Var}(Z(s)) = 1$ does not depend on s then $Z(s)$ and $\dot{Z}(s)$ are uncorrelated and hence independent. To see this, let $h(s_1, s_2) = \text{Cov}(Z(s_1), Z(s_2))$ and let $h_j(s_1, s_2)$ denote its derivative with respect to s_j , $j = 1, 2$. Then $\text{Cov}(\dot{Z}(s), Z(s)) = h_1(s, s) = h_2(s, s)$ by

symmetry. Since the variance is constant then $\partial \text{Var}(Z(s))/\partial s = h_1(s, s) + h_2(s, s) = 0$, so that $h_j(s, s) = 0$. Conditional on $Z(s)$,

$$\dot{u}(s)' \dot{u}(s) \sim \text{Wishart}_D(\Lambda(s)/\|Z(s)\|^2, n-1).$$

It therefore follows that unconditionally

$$\det(\dot{u}(s)' \dot{u}(s))^{1/2} \stackrel{D}{=} \det(\Lambda(s))^{1/2} \det(W)^{1/2} V^{-D/2} \quad (44)$$

in distribution, where $W \sim \text{Wishart}_D(I_{D \times D}, n-1)$ and $V \sim \chi_n^2$, independently, provided $n > D$. Note that if $n \leq 2D$ then \tilde{S} can self-intersect with non-zero probability, as in Figure 4, in which case the intersections are ignored when calculating the intrinsic volumes. In other words, the LKC is computed based on the local geometry as determined by $\dot{u}(s)' \dot{u}(s)$, and not on the global geometry of the actual image \tilde{S} of the search region in \mathbb{R}^n . Formally, $\dot{u}(s)' \dot{u}(s)$ defines a Riemannian metric (see Adler and Taylor (2006) for a full discussion of the relevant Riemannian geometry as it relates to random fields), which agrees with the Riemannian metric \tilde{S} inherits as a subset of \mathbb{R}^n only when \tilde{S} has no intersections.

It is well known in multivariate analysis that

$$\det(W) \stackrel{D}{=} \prod_{j=1}^D V_j \quad (45)$$

where $V_j \sim \chi_{n-j}^2$, independently, $j = 1, \dots, D$. It can be checked that after taking expectations,

$$\mathbb{E}(\det(W)^{1/2} V^{-D/2}) = 1, \quad (46)$$

so that combining (44-46) with (20), the volume term is unbiased for any D , i.e.

$$\mathbb{E}(\mu_D(u(S))) = \mathcal{L}_D(S).$$

A.6 Variance of the LKC estimator

The next question is its variance. The variance of $\det(\hat{\Lambda}(s))^{1/2}$ relative to the square of its mean (the *coefficient of variation*) is

$$\text{Var}(\det(W)^{1/2} V^{-D/2}) = \frac{\Gamma(\nu) \Gamma(\frac{\nu}{2} - D) 2^D}{\Gamma(\nu - D) \Gamma(\frac{\nu}{2})} - 1 \approx \frac{D(D+1)}{2\nu}$$

for large ν . To get a rough idea of the variance of $\widehat{\mathcal{L}}_D(S)$ we need the spatial correlation of $\det(\widehat{\Lambda}(s))^{1/2}$.

Let $h = h(s_1, s_2) = \text{Cor}(Z(s_1), Z(s_2))$ be the spatial correlation function of Z , and let subscript j denote differentiation with respect to s_j , $j = 1, 2$. Then the spatial covariance function of \dot{Z} is h_{12} and, from standard multivariate analysis, the spatial covariance function of $\dot{Z}'\dot{Z}$ is $2\text{tr}(h_{12}^2)$. To take this a bit further we need to assume that Z is stationary, in which case the spatial correlation function of $\dot{Z}'\dot{Z}$ is

$$H(s_1, s_2) = \text{tr}(h_{12}(s_1, s_2)^2) / \text{tr}(h_{12}(0, 0)^2).$$

We shall now argue that H is also roughly the spatial correlation function of $\det(\widehat{\Lambda}(s))^{1/2}$ for large ν . First of all for large ν we can replace \dot{U} by $\dot{r}/\sqrt{\nu}$ whose rows have the same spatial correlation structure as \dot{Z} . The spatial correlation function of $\dot{r}'\dot{r}/\nu$ is H . Consider the case of $D = 1$, so that $\det(\widehat{\Lambda}(s))^{1/2} \approx (\dot{r}'\dot{r}/\nu)^{1/2}$. Taking a smooth function of a random field with small variance does not appreciably alter its spatial correlation function, by using a simple linear approximation argument. Thus in the case $D = 1$, H is approximately the spatial correlation function of $\det(\widehat{\Lambda}(s))^{1/2}$. We can apply the same argument to higher dimensions by breaking down the determinant of the Wishart matrix $\dot{r}'\dot{r}$ into component χ^2 random variables by a sequence of Gram-Schmidt orthogonalizations in exactly the same way as we prove (45). Each of these components has the same spatial correlation H , and, by a linear approximation, so does $\det(\widehat{\Lambda}(s))^{1/2}$. We conclude finally that for large search regions S , large ν and stationary Z ,

$$\frac{\text{Var}(\widehat{\mathcal{L}}_D(S))}{\mathcal{L}_D^2(S)} \approx \frac{D(D+1)}{2\nu|S|} \int_S H(0, s) ds. \quad (47)$$

The coefficient of variation of the estimated P-value (12) is also approximated by the right hand side of (47). A common model for the spatial correlation function of Z is the Gaussian function

$$h(s_1, s_2) = \exp(-(s_1 - s_2)' \Lambda (s_1 - s_2) / 2), \quad (48)$$

which gives (22).

References

Adler, R. J. (1977), "A spectral moment estimation problem in two dimensions," *Biometrika*, 64, 367–373.

- (1981), *The Geometry of Random Fields*. Chichester: John Wiley & Sons.
- (2000), “On excursion sets, tube formulae, and maxima of random fields,” *Annals of Applied Probability*, 10, 1–74.
- Adler, R. J. and Taylor, J. E. (2006), *Random Fields and Geometry*. Boston: Birkhäuser. To appear.
- Adler, R. J., Taylor, J. E., and Worsley, K. J. (2006), *Random Fields and Geometry: Applications*. In preparation.
- Budach, L. (1989), “On the combinatorial foundations of Regge-calculus,” *Annalen der Physik*. 7. Folge, 46, 1–14.
- Cao, J. and Worsley, K. J. (1999a), “The detection of local shape changes via the geometry of Hotelling’s T^2 fields,” *Annals of Statistics*, 27, 925–942.
- (1999b), “The geometry of correlation fields with an application to functional connectivity of the brain,” *Annals of Applied Probability*, 9, 1021–1057.
- Cheeger, J., Müller, W., and Schrader, R. (1984), “On the curvature of piecewise flat spaces,” *Communications in Mathematical Physics*, 92, 405–454.
- (1986), “Kinematic and tube formulas for piecewise linear spaces,” *Indiana University Mathematics Journal*, 35, 737–754.
- Chung, K. K., Robbins, S. M., Dalton, K. M., Davidson, R. J., Alexander, A. L., and Evans, A. C. (2005), “Cortical thickness analysis in autism with heat kernel smoothing,” *NeuroImage*, 25, 1256–1265.
- de Berg, M., van Kreveld, M., Overmars, M., and Schwarzkopf, O. (1997), *Computational geometry: algorithms and applications*. Berlin: Springer.
- Efron, B. (1997), “The length heuristic for simultaneous hypothesis tests,” *Biometrika*, 84, 143–157.
- Fischl, B., Sereno, M. I., and Dale, A. M. (1999), “Cortical Surface-Based Analysis: II: Inflation, Flattening, and a Surface-Based Coordinate System,” *NeuroImage*, 9, 195–207.
- Grünbaum, B. (2003), *Convex Polytopes*, volume 221 of *Graduate Texts in Mathematics*. New York: Springer-Verlag, second edition.
- Hadwiger, H. (1957), *Vorlesungen Über Inhalt, Oberfläche und Isoperimetrie*. Berlin: Springer-Verlag.
- Hayasaka, S., Luan-Phan, K., Liberzon, I., Worsley, K. J., and Nichols, T. E. (2004), “Non-stationary cluster-size inference with random field and permutation methods,” *NeuroImage*, 22, 676–687.
- Hunter, D. (1976), “An upper bound for the probability of a union,” *Journal of Applied Probability*, 13, 597–603.
- Hurdal, M. K. and Stephenson, K. (2004), “Cortical cartography using the discrete conformal approach of circle packings,” *NeuroImage*, 23, S119–S128.

- Kiebel, S. J., Poline, J. B., Friston, K. J., Holmes, A. P., and Worsley, K. J. (1999), “Robust smoothness estimation in statistical parametric maps using normalised residuals from the general linear model,” *NeuroImage*, 10, 756–766.
- Lerch, J. and Evans, A. C. (2005), “Cortical thickness analysis examined through power analysis and a population simulation,” *NeuroImage*, 24, 163–173.
- Naiman, D. Q. and Wynn, H. P. (1992), “Inclusion-exclusion-Bonferroni identities and inequalities for discrete tube-like problems via Euler characteristics,” *Annals of Statistics*, 20, 43–76.
- (1997), “Abstract tubes, improved inclusion-exclusion identities and inequalities and importance sampling,” *Annals of Statistics*, 25, 1954–1983.
- Rice, S. O. (1945), “Mathematical analysis of random noise,” *Bell System Technical Journal*, 24, 46–156. Also in Wax, N. (Ed.) (1954), *Selected Papers on Noise and Stochastic Processes*, Dover, New York.
- Rosenfeld, A. and Kak, A. C. (1982), *Digital Picture Processing, Volume 2*. Orlando: Academic Press.
- Sampson, P. D. and Guttorp, P. (1992), “Nonparametric estimation of nonstationary spatial covariance structure,” *Journal of the American Statistical Association*, 87, 108–119.
- Shaw, P., Greenstein, D., Lerch, J., Clasen, L., Lenroot, R., Gogtay, N., Evans, A., Rapoport, J., and Giedd, J. (2006), “Intellectual ability and cortical development in children and adolescents,” *Nature*, 440, 676–679.
- Shephard, M. S. and Georges, M. K. (1991), “Automatic Three-Dimensional Mesh Generation by the Finite Octree Technique,” *International Journal for Numerical Methods in Engineering*, 32, 709–749.
- Siegmund, D. O. and Worsley, K. J. (1995), “Testing for a signal with unknown location and scale in a stationary Gaussian random field,” *Annals of Statistics*, 23, 608–639.
- Sun, J. (1993), “Tail probabilities of the maxima of Gaussian random fields,” *Ann. Probab.*, 21, 34–71.
- Sun, J. and Loader, C. R. (1994), “Simultaneous confidence bands for linear regression and smoothing,” *Annals of Statistics*, 22, 1328–1345.
- Taylor, J. E. (2006), “A Gaussian kinematic formula,” *Annals of Probability*, 34, 122–158.
- Taylor, J. E. and Adler, R. J. (2003), “Euler characteristics for Gaussian fields on manifolds,” *Annals of Probability*, 31, 533–563.
- Taylor, J. E., Takemura, A., and Adler, R. J. (2005), “Validity of the expected Euler characteristic heuristic,” *Annals of Probability*, 33, 1362–1396.
- Taylor, J. E. and Worsley, K. J. (2006a), “Inference for magnitudes and delays of responses in the FIAC data using BRAINSTAT/FMRISTAT,” *Human Brain Mapping*, 27, 434–441.

- (2006b), “Random fields of multivariate test statistics, with applications to shape analysis and fMRI,” *Annals of Statistics*. Under revision.
URL http://www.math.mcgill.ca/keith/roy/roy_abstract.htm
- Taylor, J. E., Worsley, K. J., and Gosselin, F. (2007), “Maxima of discretely sampled random fields, with an application to ‘bubbles’,” *Biometrika*. In press.
URL <http://www.math.mcgill.ca/keith/bubbles/dlmstat3.pdf>
- Worsley, K. (1982), “An improved Bonferroni inequality and applications,” *Biometrika*, 69, 297–302.
- Worsley, K. J. (1994), “Local maxima and the expected Euler characteristic of excursion sets of χ^2 , F and t fields,” *Advances in Applied Probability*, 26, 13–42.
- (1995), “Boundary corrections for the expected Euler characteristic of excursion sets of random fields, with an application to astrophysics,” *Advances in Applied Probability*, 27, 943–959.
- (2003), “Detecting activation in fMRI data,” *Statistical Methods in Medical Research*, 12, 401–418.
- Worsley, K. J., Andermann, M., Koulis, T., MacDonald, D., and Evans, A. C. (1999), “Detecting changes in nonisotropic images,” *Human Brain Mapping*, 8, 98–101.
- 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., Liao, C., Aston, J., Petre, V., Duncan, G. H., Morales, F., and Evans, A. C. (2002), “A general statistical analysis for fMRI data,” *NeuroImage*, 15, 1–15.
- Worsley, K. J., Marrett, S., Neelin, P., Vandal, A. C., Friston, K. J., and Evans, A. C. (1996), “A unified statistical approach for determining significant signals in images of cerebral activation,” *Human Brain Mapping*, 4, 58–73.
- Worsley, K. J. and Taylor, J. E. (2005), “An improved theoretical P-value for SPMs based on discrete local maxima,” *NeuroImage*, 28, 1056–1062.