A Three-Dimensional Statistical Analysis for CBF Activation Studies in Human Brain

K.J. Worsley *, A.C. Evans†, S. Marrett†, and P. Neelin†

May 25, 1992

SUMMARY

Many studies of brain function with positron emission tomography (PET) involve the interpretation of a subtracted PET image, usually the difference between two images under baseline and stimulation conditions. The purpose of these studies is to see which areas of the brain are activated by the stimulation condition. In many cognitive studies, the activation is so slight that the experiment must be repeated on several subjects and the subtracted images are averaged to improve the signal to noise ratio. The averaged image is then standardized to have unit variance and then searched for local maxima (Fox et al., 1988). The main problem facing investigators is which of these local maxima are statistically significant. We describe a simple method for determining an approximate p-value for the global maximum based on the theory of Gaussian random fields as developed by Adler and Hasofer (1976) and Adler (1981). The p-value is proportional to the volume searched divided by the product of the FWHMs of the image reconstruction process, or number of resolution elements (resels). Rather than working with local maxima as in Fox et al. (1988), our method focuses on the Euler characteristic of the set of voxels with a value larger than a given threshold. The Euler characteristic depends only on the topology of the regions of high activation, irrespective of their shape. For large threshold values this is approximately the same as the number of isolated regions of activation above the threshold. We can thus determine not only if any activation has taken place, but we can also estimate how many isolated regions of activation are present.

Key Words: Activation studies – Cerebral blood flow – Postiron emission tomography – Statistical analysis.

*Department of Mathematics and Statistics, McGill University, Montreal, Quebec, Canada. † McConnell Brain Imaging Centre, Montreal Neurological Institute, Montreal, Quebec, Canada. This work was supported by a Natural Sciences and Engineering Research Council of Canada Operating Grant, the Fonds pour la Formation de Chercheurs et l'Aide a la Recherche du Gouvernement du Québec Subvention d'Équipe, and the Medical Research Council of Canada Special Program 5 Grant.

1 INTRODUCTION

An increasingly important application of positron emission tomography (PET) involves the measurement of regional cerebral blood flow (CBF) during the performance of various sensorimotor or cognitive tasks. Regional changes in CBF between two tasks reflects the mobilization of functional units specific to the new task. For subtle cognitive tasks, these changes are small ($\leq 10\%$) and are usually not apparent from a single subtraction study. The signal-to-noise ratio can be improved with an image-averaging procedure which requires the geometrical scaling and re-sampling of PET image data from each of a number of subjects, or repeated experiments in a single subject, onto a standardized stereotactic coordinate framework (Fox *et al.*, 1985, 1988; Talairach & Tournoux, 1988; Fox & Mintun, 1989). Using this transformation on each of a set of typically 8-12 change images, one can generate a 3-D composite dataset where each voxel contains a statistic for the change in CBF at that brain location when measured across the repeated experiments. This may be simply the mean relative change in CBF or a measure which includes information about the variance across subjects (repeats) such as a t-statistic. If each voxel were an independent measure such a dataset would contain 10^5-10^6 samples and, even at a threshold set to correspond to $p = 0.01$ for a significant CBF change, there would be $10^3 - 10^4$ voxels exceeding the threshold by chance alone. Hence a Bonferroni-type correction for multiple comparisons would raise the significance threshold to prohibitive levels. In reality, adjacent voxels are highly correlated by the initial image reconstruction process. This spatial coherence is related to the full-width at half-maximum (FWHM) of the reconstruction filter which is typically set at 15-20mm (this value is set higher than the intrinsic scanner resolution of 5-10mm to overcome the residual anatomical variability among subjects persisting after stereotactic transformation). Thus the appropriate sampling element for such datasets is the 'resolution element' or 'resel' which can be regarded as a block of voxels with dimensions equal to the effective FWHM of the reconstructed image in each dimension. This elemental volume is a physical entity and the number of resels in a complete dataset is fixed, unlike the number of voxels which is set arbitrarily by choice of voxel dimensions. However, since a resel volume can be centred at any voxel, the search must proceed voxel-by-voxel. Hence, the analysis of such large datasets for the occurrence of significant CBF changes require statistical methods which recognize the inherent spatial correlation among adjacent voxels and allow a means of estimating the number of truly independent samples present within the sampled volume. This paper addresses the question of identifying significant changes in spatially-correlated data using the concept of three-dimensional Gaussian random fields. The theory is presented and tested both by simulation and by analysis of real PET data taken from a study of cerebral activation in response to a painful heat stimulus and a cognitive study of single word recognition.

2 THEORY

2.1 The t-statistic image

Suppose that an activation study consisted of n subjects, each scanned under two conditions denoted by A and B. We are interested in detecting activation due to the task condition B by subtracting the image under the rest condition A.

Let $A_i(x,y,z)$ and $B_i(x,y,z)$ be the CBF of subject i, $i = 1, \cdots, n$ at voxels with coordinates (x,y,z), under the two conditions A and B. These were standardized to have the same mean and then subtracted to give normalized differences $\Delta_i(x,y,z)$ i.e. ΔCBF at (x,y,z):

$$\Delta_i(x,y,z) = (B_i(x,y,z)/\overline{B_i} - A_i(x,y,z)/\overline{A_i}) \times 100$$

where $\overline{B_i}$ and $\overline{A_i}$ are the average CBF over all intra-cerebral voxels in conditions B and A, respectively. The average subtracted image and its variance were calculated as

$$\begin{aligned} M(x,y,z) &= \sum_{i=1}^{n} \Delta_i(x,y,z)/n(x,y,z) \\ S^2(x,y,z) &= \sum_{i=1}^{n} \{\Delta_i(x,y,z) - M(x,y,z)\}^2/\{n(x,y,z)-1)\} \end{aligned}$$

where $n(x,y,z)$ is the number of subjects with a blood flow value at (x,y,z), $0 \leq n(x,y,z) \leq n$. The first summation was carried out over all voxels with $n(x,y,z) > 0$, and the second over all voxels with $n(x,y,z) > 1$. In our studies, inspection of $S(x,y,z)$ revealed no obvious spatial trends, such as higher values in the centre of the brain, nor was it markedly dependent on the average baseline CBF value or ΔCBF (see Methods section) so it was pooled over all inter-cranial voxels with $n(x,y,z) > 1$ to obtain the average variance

$$\overline{S^2} = \sum_{x,y,z} \{n(x,y,z)-1\} S^2(x,y,z) / \sum_{x,y,z} \{n(x,y,z)-1\}.$$

The t-statistic image, based on the pooled standard deviation, was then calculated as

$$T(x,y,z) = M(x,y,z)/\{\overline{S}/\sqrt{n(x,y,z)}\}.$$

Note that the standard deviation $\bar{S}$ is pooled over a very large number of voxels so the effective degrees of freedom of $T(x,y,z)$ is very large. The effective degrees of freedom will depend on the correlation structure of the voxels. If we can find N voxels sufficiently separated so that they are independent, then the effective degrees of freedom is at least $(n-1)N$. Typically $N \approx 300$ (see later) and $n \approx 10$, so the effective degrees of freedom is certainly large enough to approximate the distribution of $T(x,y,z)$ by a standard Gaussian rather than a t-distribution.

An alternative approach is to divide by the voxel standard deviation rather than the pooled standard deviation to obtain a t-statistic image with $n(x,y,z)-1$ degrees of freedom:

$$T_n(x,y,z) = M(x,y,z)/\{S(x,y,z)/\sqrt{n(x,y,z)}\}.$$

The advantage of dividing by the pooled standard deviation rather than the voxel standard deviation is that $T(x, y, z)$ is proportional to the mean $M(x, y, z)$, and so regions with high $T(x, y, z)$ correspond to regions of high mean activation. Secondly, the larger degrees of freedom of the pooled standard deviation decreases the variablity of $T(x, y, z)$, which results in increased power at detecting activation (see Methods section). A compromise between the two t-statistics is to divide by a smoothed standard deviation; in this sense $\bar{S}$ is the smoothest possible standard deviation and $S(x, y, z)$ is unsmoothed. This appears to present more theoretical difficulties than either of the two extremes and so it will not be pursued in this paper.

2.2 The maximum t-statistic $T_{\max}$ and its p-value

The statistical significance of regions of activation in the subtracted image volume was assessed by comparing the t-statistic image $T(x, y, z)$ to a level α ($\alpha = 0.05$, say) critical or threshold value t_α. If the t-statistic was larger than t_α then activation at the point (x, y, z) was declared statistically significant at level α. The critical value t_α was calculated so that if no activation was present then the maximum t-statistic $T_{\max}$, searched over a given brain region of volume V, would exceed t_α with a probability of approximately α. Thus if no activation is present, this procedure will produce a false-positive with a probability of α.

The key quantity is the probability that $T_{\max} > t$ for some fixed threshold t. The p-values based on the standard Gaussian distribution are obviously far too small since no correction is made for the large number of voxels searched to obtain the maximum. On the other hand, a Bonferroni correction is much too conservative since adjacent voxels are so highly correlated. A corrected p-value was calculated using results from the theory of Gaussian random fields (Adler, 1981; for a non-technical review of the major results see Hasofer, 1978). It was assumed that the t-statistic image $T(x, y, z)$ when no activation was present could be modeled by a smooth homogeneous Gaussian random field with zero expectation and unit variance. One of the conditions is that the distribution of $T(x, y, z)$ must be Gaussian at each voxel. The central limit theorem will ensure that the distribution of the numerator $M(x, y, z)$ is close to Gaussian since it is the average of n images. Because the standard deviation $\bar{S}$ is pooled over a very large number of voxels its effective degrees of freedom is very large. Hence the null distribution of $T(x, y, z)$ is well approximated by a standard Gaussian distribution. A condition for homogeneity is that the point response function, in particular the FWHM, should be the same at every voxel.

The main result, originally due to Adler and Hasofer (1976), is that if t is large then the probability that the maximum of such a random field exceeds t is approximately

$$\mathrm{P}(T_{\max} > t) \approx V|\Lambda|^{\frac{1}{2}}(2\pi)^{-2}(t^2 - 1)e^{-\frac{1}{2}t^2} \tag{1}$$

provided that the search volume V is large relative to the smoothness of the image. The matrix Λ is the 3×3 variance matrix of the partial derivatives of the random field in each

of the 3 variables x, y and z, and measures the roughness of the image:

$$\Lambda = \begin{pmatrix} \mathrm{Var}\left(\frac{\partial T}{\partial x}\right) & \mathrm{Cov}\left(\frac{\partial T}{\partial x}, \frac{\partial T}{\partial y}\right) & \mathrm{Cov}\left(\frac{\partial T}{\partial x}, \frac{\partial T}{\partial z}\right) \\ \mathrm{Cov}\left(\frac{\partial T}{\partial x}, \frac{\partial T}{\partial y}\right) & \mathrm{Var}\left(\frac{\partial T}{\partial y}\right) & \mathrm{Cov}\left(\frac{\partial T}{\partial y}, \frac{\partial T}{\partial z}\right) \\ \mathrm{Cov}\left(\frac{\partial T}{\partial x}, \frac{\partial T}{\partial z}\right) & \mathrm{Cov}\left(\frac{\partial T}{\partial y}, \frac{\partial T}{\partial z}\right) & \mathrm{Var}\left(\frac{\partial T}{\partial z}\right) \end{pmatrix}.$$

Thus as the search volume increases or the image becomes rougher then the p-value, and hence the critical values based on it, must increase. Note that (1) is not a probability for all values of t since it can exceed one or even take negative values. In practice (see Methods section) it appears to be a satisfactory approximation if the p-value given by (1) is less than 0.1 and the number of resels R is greater than 30 (see (5)).

Some caution should be exercised in using the result (1) since it is sensitive to the assumption of a Gaussian distribution in the extreme tails of the distribution. Even though the central limit theorem may ensure an adequate approximate Gaussian distribution over the bulk of the range, this may not be the case in the tails, unless the individual images are already close to a Gaussian distribution.

2.3 The relationship of $|\Lambda|$ to the FWHMs and the number of resolution elements

.

The main computational task is to estimate Λ. This can be done by taking numerical derivatives; details are given in the Methods section. However we can derive a simple expression for Λ if we make the assumption that the image under no activation can be generated by convolving a white noise Gaussian random field with a kernel or point response function $k(\mathbf{x})$, where $\mathbf{x}$ denotes the vector $(x, y, z)^{\mathrm{T}}$, and superscript T denotes transpose. Then it can be shown (Adler, 1981, Chapter 2) that the variance matrix of the vector of partial derivatives is

$$\Lambda = \int \frac{\partial k(\mathbf{x})}{\partial \mathbf{x}} \frac{\partial k(\mathbf{x})}{\partial \mathbf{x}^{\mathrm{T}}} d\mathbf{x} \Big/ \int k^2(\mathbf{x}) d\mathbf{x}.$$

Applying this to a Gaussian kernel of the form $k(\mathbf{x}) \propto \exp\{-\mathbf{x}^{\mathrm{T}}\Sigma^{-1}\mathbf{x}/2\}$ gives $\Lambda = \Sigma^{-1}/2$. If the principal axes of Σ coincide with the x, y and z directions then the off diagonal elements of Λ are zero. If FWHM_x, FWHM_y and FWHM_z are the full widths at half maximum in the x, y and z directions then

$$\Lambda = \begin{pmatrix} 1/\mathrm{FWHM}_x^2 & 0 & 0 \\ 0 & 1/\mathrm{FWHM}_y^2 & 0 \\ 0 & 0 & 1/\mathrm{FWHM}_z^2 \end{pmatrix} (4 \log_e 2) \qquad (2)$$

and so

$$|\Lambda|^{\frac{1}{2}} = (\mathrm{FWHM}_x \mathrm{FWHM}_y \mathrm{FWHM}_z)^{-1} \quad (4 \log_e 2)^{\frac{3}{2}}. \qquad (3)$$

In practice it was found (see Methods section) that the accuracy of the relationships (2) and (3) could be greatly improved by taking into account the increased smoothness of the

image in the z-direction caused by linear interpolation and re-sampling of the 15 planes of the PET camera to produce the 80 planes of the image. The result is as follows. Suppose that the separation of the PET planes is d. The image is then re-sampled at smaller voxels of size $d_c \leq d$ by linear interpolation. If the point response function of the original image is Gaussian, then the effective FWHM of the re-sampled image increases to

$$\text{FWHM}_{zc} = d(4\log_e 2)^{\frac{1}{2}}\{1-(1-\rho)/3)\}^{\frac{1}{2}}\{2(1-\rho)-(d_c/d)(3-4\rho+\rho^4)/3\}^{-\frac{1}{2}}, \quad (4)$$

where $\rho = \exp\{-(2\log_e 2)d^2/\text{FWHM}_z^2\}$ is the correlation between planes separated by a distance d (see Figure 1 and the Appendix for a proof). In our example $\text{FWHM}_z = 6$mm, the plane separation is $d = 6.5$mm, and the size of the re-sampled voxels is $d_c = 1.5$mm, which gives an effective FWHM in the z-direction of $\text{FWHM}_{zc} = 7.6$mm. No such correction was necessary to the FWHMs in the x- and y-directions since the FWHMs were large relative to the voxel sizes.

Let $R = V/(\text{FWHM}_x \times \text{FWHM}_y \times \text{FWHM}_{zc})$, the search volume divided by the product of the full widths at half maximum, be a measure of the number of resolution elements, or resels, in the search volume. Combining (1) and (3) the approximate p-value can be simply stated as

$$\text{P}(T_{\max} > t) \approx R(4\log_e 2)^{\frac{3}{2}}(2\pi)^{-2}(t^2-1)e^{-\frac{1}{2}t^2}. \quad (5)$$

Based on (5), a table of approximate critical values t_α of $T_{\max}$, chosen so that $\text{P}(T_{\max} > t_\alpha) = \alpha$ is given in Table 1(a). For example, if the number of resels is 500 then the level $\alpha = 0.05$ critical value of the maximum t-statistic is approximately 4.47.

2.4 The number of isolated regions of activation

Our main result (1) on the p-value of $T_{\max}$ was first derived rigorously by Adler and Hasofer (1976) using the concept of the Euler characteristic χ_t of the excursion set of $T(x,y,z)$ above a threshold t. The excursion set is simply the set of voxels for which $T(x,y,z) > t$, that is the region of activation greater than t. The precise definition of the Euler characteristic involves the curvature of the boundary of the excursion set at tangent planes; see Adler (1981), page 89. Essentially it counts the number of isolated parts of the excursion set, irrespective of their shape, minus the number of 'holes'. For example the Euler characteristic of a solid ball is 1, of a doughnut is 0. If no holes are present then it counts the number of isolated regions of activation in an image above the threshold t. An illustration of the Euler characteristic of the excursion set of a two dimensional image is shown in Figure 2(a,b). As the threshold level t increases Adler shows that the holes tend to disappear and that we are left with isolated regions each of which contains just one local maximum (Figure 2(c)). Thus for large t the presence of holes is a rare occurrence and the Euler characteristic approaches the number of local maxima. For even larger t near the global maximum $T_{\max}$ the Euler characteristic takes the value 0 if $T_{\max} < t$ and 1 if $T_{\max} > t$ (Figure 2(d)). Since $\text{P}(T_{max} > t) \approx \text{P}(\chi_t \geq 1) \approx \text{E}(\chi_t)$ as $\text{P}(\chi_t > 1) \to 0$ for $t \to \infty$ (Hasofer, 1978), then the expected Euler characteristic approximates the p-value of $T_{\max}$.

The importance of the Euler characteristic, as opposed to the number of local maxima, is that it is more amenable to statistical analysis. There is no known simple result for the expected number of local maxima greater than t. On the other hand, Adler (1981), page

111, shows that for any threshold t the expected Euler characteristic of the excursion set of a zero expectation, unit variance stationary Gaussian random field is exactly

$$\mathrm{E}(\chi_t) = V|\Lambda|^{\frac{1}{2}}(2\pi)^{-2}(t^2-1)e^{-\frac{1}{2}t^2} = R(4\log_e 2)^{\frac{3}{2}}(2\pi)^{-2}(t^2-1)e^{-\frac{1}{2}t^2} \qquad (6)$$

which leads directly to the approximate p-value of $T_{\max}$ (1). Moreover, Adler (1981), page 162, shows that for large $|t|$, χ_t has a Poisson distribution and so its standard deviation is approximately $\sqrt{\mathrm{E}(\chi_t)}$. Table 1(b) is an extension of Table 1(a) which gives some values of t chosen so that the expected Euler characteristic (6) takes values 1, 2 and 5, for several values of the number of resels, R. For example, if the number of resels is 500 then we expect approximately 2 false positive isolated regions of activation above the $t = 3.38$ threshold.

Our proposed estimator of the number of isolated regions of activation is the Euler characteristic of the t-statistic image thresholded at a level α critical value, t_α, for a suitably small value of α, say 0.05. This should pick out the isolated peaks of activation in the signal, while protecting us against finding any false positive regions in the unactivated (noise-only) parts of the image. It is similar to the sequentially rejective multiple test procedure proposed by Holm (1979). A more complete picture can be obtained from a plot of the observed χ_t and $\mathrm{E}(\chi_t)$ against t - see the Results section.

2.5 Evaluation of the Euler characteristic

For a random field evaluated over a lattice of voxels, Adler (1981), page 121, shows how the Euler characteristic can be approximated using only the local properties of the excursion set over adjacent voxels. Unfortunately, if the excursion set touches the boundary of the search volume, the definition depends on the direction in which the voxels are scanned, so we have modified the definition to average over all the 8 directions in which adjacent voxels can be scanned. This does not affect its expectation (6).

Our definition is as follows. Let us represent all voxels by points at their centres. Consider a $2 \times 2 \times 2$ 'cube' of 8 adjacent voxels centred at (x, y, z), $(x + \delta x, y, z)$, $(x, y + \delta y, z)$, $(x + \delta x, y + \delta y, z), \cdots, (x + \delta x, y + \delta y, z + \delta z)$, all of which are inside the search volume (Figure 2(f)). Let P be the number of vertices of the cube inside the excursion set, that is with $T(x, y, z) > t$ ($0 \le \mathrm{P} \le 8$). Let E be the number of edges joining adjacent vertices both of which are in the excursion set ($0 \le \mathrm{E} \le 12$). Let F be the number of faces all of whose vertices are in the excursion set ($0 \le F \le 6$). Let C be the number of cubes all of whose vertices are in the excursion set, that is $C = 1$ if all vertices are in the excursion set and zero otherwise. The contribution of this cube to the Euler characteristic is $e = P/8 - E/4 + F/2 - C$. The Euler characteristic χ_t is then the sum of e over all such cubes in the search volume. On a historical note, this definition is related to the famous formula discovered by Euler in 1752: if P, E and F are the number of vertices, edges and faces of any polyhedron on a sphere then $P - E + F = 2$, which turns out to be the Euler characteristic of a hollow sphere.

It can be shown that provided the excursion set contains no holes and does not touch the boundary of the search volume then χ_t defined in this way counts the number of isolated regions. Note that for the purpose of this definition two voxels are connected if they are joined by an edge of the lattice of voxels, such as (x, y, z) and $(x + \delta x, y, z)$, but voxels separated by a diagonal such as (x, y, z) and $(x + \delta x, y + \delta y, z)$ are not connected. However

if an isolated region of the excursion set with no holes touches the boundary (Figure 2(e), $t < 2.7$) then its contribution to χ_t decreases below one. Its value depends on the shape of the boundary where the excursion set touches. If the boundary is flat then χ_t is one half; if it is convex then χ_t lies between zero and one half; if it is concave then χ_t lies between one half and one. Thus regions in the outer cortex, which lie near the boundary of the search volume, may have lower than expected χ_t if the threshold is low.

Even when no activation is present the probability that the excursion set touches the boundary of the search volume is quite high. A rough idea can be worked out using the theory of two-dimensional Gaussian fields (see next section). If the correlation structure is isotropic (all FWHM's equal) and a spherical search volume contains $R = 500$ resels, the expected number of isolated regions greater than $t = 3.62$ is 1 (from Table 1), whereas the probability that any of these regions touches the boundary is approximately 0.28. This decreases to 0.011 at the 5% critical value $t = 4.47$. To overcome this, we suggest using a generous search volume to make sure that potential regions of activation do not touch the boundary.

Finally, it should be noted that for threshold values near zero the Euler characteristic usually takes negative values because the holes outnumber the isolated regions. In fact from (6) it can be seen that the expected Euler characteristic is always negative if $|t| < 1$. The interpretation of the Euler characteristic at these thresholds seems difficult, but it is useful for verifying the agreement between the observed and expected Euler characteristic from (6).

2.6 Maxima in a slice

Adler and Hasofer (1976) and Adler (1981), page 111, give a more general result for the p-value of the maximum of a Gaussian random field in any number of dimensions. Their two-dimensional result can be used to find the p-value of the slice maximum, $T^S_{\max}$ say:

$$\mathrm{P}(T^S_{\max} > t) \approx R^S (4 \log_e 2)(2\pi)^{-\frac{3}{2}} t e^{-\frac{1}{2}t^2},$$

where R^S is the number of resels in the search area, defined as the search area divided by the product of the two FWHM's in the slice. Friston *et al.* (1991) used heuristic arguments to derive a slightly different result (formula (2)) for the p-value of a slice maximum, which in our notation is

$$P_F(T^S_{\max} > t) \approx R^S (8 \log_e 2)/\{32\pi e^{t^2} p(t)\},$$

where $p(t)$ is the probability that a standard normal variate exceeds t. Using the approximation $p(t) \approx (2\pi)^{-\frac{1}{2}} exp(-\frac{1}{2}t^2)/t$ for large t (Adler, 1981, page 135) we obtain

$$P_F(T^S_{\max} > t) \approx (\pi/4) R^S (4 \log_e 2)(2\pi)^{-\frac{3}{2}} t e^{-\frac{1}{2}t^2} \approx (\pi/4)\mathrm{P}(T^S_{\max} > t).$$

Although the p-value of Friston *et al.* (1991) formula (2) is technically incorrect, since it differs from the result of Adler and Hasofer by a factor of approximately $\pi/4 = 0.79$, it seems to be satisfactorily accurate in practice, as confirmed by the simulation results reported in Friston *et al.* (1991).

2.7 The t-statistic image $T_n(x, y, z)$ based on the voxel standard deviation

The above theory may be used to find approximate expressions for the p-value of the maximum $T_{n\max}$ of a non-Gaussian random field such as $T_n(x, y, z)$ by transforming it to a Gaussian random variable at each point. Let $G(x, y, z) = p^{-1}(p_n(T_n(x, y, z)))$ where $p_n(t)$ is the probability that a t-random variable with $n - 1$ degrees of freedom exceeds t and $p(g)$ is the probability that a standard Gaussian random variable exceeds g. Unfortunately such a field $G(x, y, z)$ may be Gaussian at each point (x, y, z) but not necessarily multivariate Gaussian at each set of points, so that strictly speaking it is not a Gaussian random field, and the theoretical results of the previous sections do not apply. Nevertheless we can still justify p-values like (1) and (5) by appealing to the results of Vanmarcke (1983) for non-Gaussian random fields, provided that we correct for changes in the roughness Λ of the field. The appropriate correction factor λ_n is derived in the Appendix. The resulting approximate p-value of the maximum $G_{\max} = p^{-1}(p_n(T_{n\max}))$ searched over a volume V containing R resels is

$$\mathrm{P}(G_{\max} > g) \approx \lambda_n^{\frac{3}{2}} V |\Lambda|^{\frac{1}{2}} (2\pi)^{-2} (g^2 - 1) e^{-\frac{1}{2}g^2} \approx \lambda_n^{\frac{3}{2}} R (4 \log_e 2)^{\frac{3}{2}} (2\pi)^{-2} (g^2 - 1) e^{-\frac{1}{2}g^2}. \quad (7)$$

This result only holds for $n \geq 4$ subjects. Some values of λ_n for $n = 4, 5, \ldots, 20$ are 1.757, 1.434, 1.304, 1.233, 1.189, 1.159, 1.137, 1.121, 1.108, 1.097, 1.089, 1.082, 1.075, 1.070, 1.066, 1.061 and 1.058. Note that λ_n decreases to one as n increases, so that the image $G(x, y, z)$ is slightly rougher that $T(x, y, z)$. There are no theoretical results for the expected Euler characteristic of excursion sets of $T_n(x, y, z)$ or $G(x, y, z)$.

Result (7) should be used with much more caution than (1) or (5) because it is based on far more assumptions about the distribution of the images. These assumptions are that each subject's subtracted image $\Delta_i(x, y, z)$ is a stationary Gaussian random field with the same standard deviation σ. In particular, result (7) is highly sensitive to the assumption of a Gaussian distribution for each subject's subtracted image. Results (1) and (5) are more robust since the central limit theorem implies that $T(x, y, z)$ will be approximately Gaussian even if the subject's subtracted images are not, and the results (1) and (5) will remain valid (though perhaps less sensitive) even if the subject's standard deviations differ between subjects. Exact theoretical results that supercede (7) have been found for the voxel-based t-statistic (see addendum).

3 METHODS

3.1 Data Acquisition

PET scans were obtained using the Scanditronix PC-2048 system which produces 15 image slices 6.5 mm apart with a transverse image resolution of 4.6-6.4 mm and an axial resolution of 5.4-7.1 mm (5). Using the bolus $H_2^{15}O$ methodology without blood sampling (6,7,8), the relative distribution of cerebral blood flow (CBF) was measured in baseline and activation conditions. For the pain study (Talbot et al., 1991) 8 subjects underwent a procedure wherein a thermistor was applied to the forearm at both warm (42^oC) and high (48^oC) states,

each condition being studied twice in each subject. For the present work, we analyzed the difference images of the two warm conditions as a dataset which should have an expectation of zero throughout. We also analyzed 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. One subject was scanned only once in the hot condition so this subject was dropped leaving 7 subjects for this dataset. For the word-recognition study (Bub et al., in preparation), PET image data were collected from 10 normal volunteers. Visual stimuli were presented for 1 sec with an interstimulus interval of 1 sec on a monochrome monitor, suspended in front of the subject and covered by a light-tight curtain. The baseline condition was a black plus-sign on a white background and for the activation condition, single words were presented on the monitor for 1 second with an inter-stimulus interval of 1 sec. Each subject in both studies also had an MRI scan containing 64 2mm-thick T_1-weighted multi-slice spin-echo images ($T_R = 550$ msec ; $T_E = 30$ msec) for later use in the 3-D analysis.

3.2 Data Analysis

To overcome residual anatomical variability in subsequently transformed stereotactic images, a 20mm FWHM Hanning reconstruction filter was used. This has the effect of increasing signal to noise in the averaged image at the expense of resolution. In the experimental data sets studied here no axial filtering was used. Using a volumetric image registration procedure described previously (Evans *et al.*, 1989,1991), the MRI volume from each subject was aligned with the corresponding PET volume. An orthogonal coordinate frame was then established based on the anterior commissure - posterior commissural (AC-PC) line as identified in the MRI image volume (Evans *et al.*, 1992). These anatomical frame coordinates were used to apply a trilinear re-sampling of each matched pair of MRI and PET datasets into a standardized stereotactic coordinate system (Fox *et al.*, 1985; Talairach & Tournoux, 1988) with no axial filtering. PET images were then normalized by dividing each voxel by the mean value for all intra-cerebral voxels. The intra-cerebral voxels were defined as all voxels with a value greater than 150% of the mean value of the entire volume. The mean state-dependent change (ΔCBF) image volume was obtained by averaging the subtracted images across subjects. The ΔCBF volume was converted to a t-statistic volume by dividing each voxel by the mean standard deviation in normalized ΔCBF for all intra-cerebral voxels. Individual MRI images were subjected to the same averaging procedure, such that composite stereotactic image volumes, $128 \times 128 \times 80$ voxels in extent and sampled at approximately 1.4mm $\times$ 1.7mm $\times$ 1.5mm along the X, Y and Z axes respectively, were obtained for both t-statistic and MRI.

3.3 Validation of the pooled standard deviation

We shall first examine the assumption of a stationary standard deviation, which allows us to pool the standard deviation over the search volume. A formal test of equal standard deviations is complicated by the correlation of adjacent voxels; the standard Bartlett test is not valid, for example. Instead we shall assess the equality of the standard deviation by plotting it against possible explanatory variables. For the pain study, a search volume of 3×10^5 voxels or $V = 1090$ cm^3 was chosen that covered the top portion of the brain. For

the 'warm-warm' dataset the standard deviation image $S(x, y, z)$, with $n - 1 = 7$ degrees of freedom was plotted against average baseline CBF normalized to 100 (Figure 3(a)). It can be seen that the standard deviation was approximately constant. The standard deviation was also plotted against the average subtracted activation $M(x, y, z)$ (Figure 3(b)). There is some evidence of increased standard deviation in areas of high activation, perhaps attributable to misalignment of areas of activation or differing strengths of subject activation. However, over the bulk of the search volume where little activation has occurred, the standard deviation is roughly constant. Finally, the standard deviation was plotted against the distance from the slice centre, defined as the vertical line through the mid-point of the AC-PC line (Figure 3(c)). Again there is evidence of a slight decrease in standard deviation beyond 8cms, but otherwise it appears to be roughly constant over the bulk of the search volume. Accordingly, the standard deviation was pooled over the search volume to give $\bar{S} = 6.0\%$.

For the 'hot-warm' dataset, the standard deviation image $S(x, y, z)$, with $n - 1 = 6$ degrees of freedom, was plotted against average baseline CBF, average subtracted activation $M(x, y, z)$ and distance from the brain centre. Similar conclusions were obtained as above, and the standard deviation pooled over the same search volume was $\bar{S} = 4.4\%$. This reduction in standard deviation can be explained by the standard repeated measures model of equal correlations between observations taken on the same subject. Since the hot-warm dataset is an average of two differences, this model predicts a standard deviation of $6.0/\sqrt{2} = 4.2\%$, which is very close to $\bar{S} = 4.4\%$ for the hot-warm dataset.

For the word-recognition study, a search volume of 3.9×10^5 voxels or $V = 1390$ cm^3 was chosen that covered nearly all the grey matter but left out most of the white matter and ventricles. The standard deviation image $S(x, y, z)$, with $n - 1 = 9$ degrees of freedom, was also plotted against average baseline CBF, average subtracted activation $M(x, y, z)$ and distance from the brain centre. Again it appeared to be roughly constant over the bulk of the search volume and the pooled standard deviation was $\bar{S} = 5.4\%$.

3.4 Validation of the relationship between $|\Lambda|^{\frac{1}{2}}$ and FWHMs

The image of partial derivatives was evaluated on the residual images $Z_i(x, y, z)$ obtained by subtracting the ΔCBF image for subject i from the average ΔCBF image $M(x, y, z)$ and standardizing by the pooled standard deviation:

$$Z_i(x, y, z) = \{\Delta_i(x, y, z) - M(x, y, z)\}/\bar{S}.$$

These images were assumed to have the same variance structure as $T(x, y, z)$ but with the signal component removed. Derivatives in each direction were estimated by

$$\begin{aligned} Z_{xi}(x, y, z) &= \{Z_i(x + \delta_x, y, z) - Z_i(x, y, z)\}/\delta_x \\ Z_{yi}(x, y, z) &= \{Z_i(x, y + \delta_y, z) - Z_i(x, y, z)\}/\delta_y \\ Z_{zi}(x, y, z) &= \{Z_i(x, y, z + \delta_z) - Z_i(x, y, z)\}/\delta_z \end{aligned}$$

where $\delta_x = 1.4$ mm, $\delta_y = 1.7$ mm and $\delta_z = 1.5$ mm are the voxel sizes in directions x, y and z respectively. To avoid discontinuities caused by changes in $n(x, y, z)$, these calculations and all succeeding ones were only done for voxels with $n(x, y, z) = n$, that is where all subjects

contribute a value. Let N be the number of such voxels. Then the variances and covariances of these images, pooled over all voxels, were estimated by

$$
\begin{aligned}
V_{xx} &= \sum_{i,x,y,z} Z_{xi}(x,y,z)^2/N(n-1), \\
V_{yy} &= \sum_{i,x,y,z} Z_{yi}(x,y,z)^2/N(n-1), \\
V_{zz} &= \sum_{i,x,y,z} Z_{zi}(x,y,z)^2/N(n-1), \\
V_{xy} &= \sum_{i,x,y,z} \Big\{ Z_{xi}(x,y,z) + Z_{xi}(x,y+\delta_y,z) \Big\} \Big\{ Z_{yi}(x,y,z) + Z_{yi}(x+\delta_x,y,z) \Big\} \Big/ 4N(n-1) \\
V_{xz} &= \sum_{i,x,y,z} \Big\{ Z_{xi}(x,y,z) + Z_{xi}(x,y,z+\delta_z) \Big\} \Big\{ Z_{zi}(x,y,z) + Z_{zi}(x+\delta_x,y,z) \Big\} \Big/ 4N(n-1) \\
V_{yz} &= \sum_{i,x,y,z} \Big\{ Z_{yi}(x,y,z) + Z_{yi}(x,y,z+\delta_z) \Big\} \Big\{ Z_{zi}(x,y,z) + Z_{zi}(x,y+\delta_y,z) \Big\} \Big/ 4N(n-1).
\end{aligned}
$$

Note that the variances are estimated at the centres of the edges joining adjacent voxels, the covariances are estimated at the centres of the faces joining adjacent voxels. This is necessary to avoid biased covariance estimates. The matrix Λ is then estimated by

$$
\Lambda = \begin{pmatrix} V_{xx} & V_{xy} & V_{xz} \\ V_{xy} & V_{yy} & V_{yz} \\ V_{xz} & V_{yz} & V_{zz} \end{pmatrix}.
$$

Values of Λ and $|\Lambda|^{\frac{1}{2}}$ for the three datasets are shown in Table 2. They are all in good agreement, and as expected, the off-diagonal covariance terms are close to zero, showing that there has been no slippage or consistent miss-match in adjacent voxels.

The nominal resolutions of the image were $\text{FWHM}_x = 20\text{mm}$, $\text{FWHM}_y = 20\text{mm}$ and $\text{FWHM}_z = 6\text{mm}$. Substituting these values into (2), assuming a Gaussian point response function, gives reasonable agreement with the empirical values of Λ except that V_{zz} is too large (Table 2). To further validate the relationship (2) we can find the effective FWHMs that would produce the same diagonal elements as the observed Λ, namely $\text{FWHM}_x^* = (4\log_e 2/V_{xx})^{\frac{1}{2}}$, $\text{FWHM}_y^* = (4\log_e 2/V_{yy})^{\frac{1}{2}}$ and $\text{FWHM}_z^* = (4\log_e 2/V_{zz})^{\frac{1}{2}}$. FWHM_x^* and FWHM_y^* are in good agreement with the nominal values of 20mm but FWHM_z^* is approximately 7.5mm, larger than the nominal value of 6mm (Table 2). However the corrected $\text{FWHM}_{zc} = 7.6\text{mm}$ from (4) is in excellent agreement with FWHM_z^* and provides a much closer estimate of $|\Lambda|^{\frac{1}{2}}$ than the uncorrected FWHMs (Table 2). More accurate results should be obtained using FWHM_{zc} in place of FWHM_z in the calculation of the number of resels, R.

3.5 Validation of the p-value for $T_{\max}$

3.5.1 Specificity

The $n = 8$ subjects in the warm-warm dataset of the pain study were divided into all possible groups of 4 subjects each. In one group of 4 subjects the subtracted ΔCBF images $\Delta_i(x, y, z)$ were replaced by their negative, $-\Delta_i(x, y, z)$. The t-statistic image $T^*(x, y, z)$ was then calculated in the usual way, using the pooled standard deviation $\bar{S} = 6.0\%$. If any activation is present in the subtracted images and if it is the same for all subjects then $T^*(x, y, z)$ should have zero expectation. Thus $T^*(x, y, z)$ can be used as a pure-noise image for validating the p-value of $T_{\max}$. The set of all 8!/(4!4!)=70 such images were created and $T_{\max}$ and its p-values using (1) were calculated over the same search volume of $V = 1090$ cm^3. The proportion of false positives at each of the $\alpha = 0.10,\ 0.05$ and 0.01 levels are shown in Table 3. The proportions are all close to the nominal levels which suggests that the critical values given by our methods are satisfactory. The number of resolution elements in the search volume was $R = 360$ resels, and almost exactly the same values and conclusions were obtained using (5).

3.5.2 Sensitivity

A phantom activation signal $\mu(x, y, z)$ was created by centring a Gaussian function at a location $\mathbf{x}_0 = (x_0, y_0, z_0)^{\mathrm{T}}$ in the anterior cingulate close to where activation was in fact detected in the hot-warm dataset (Talbot et al., 1991). The peak height was chosen to represent a percentage increase of μ_0 in CBF for the 7 subjects and the shape was chosen to be the convolution of the point response function with itself. For the case of a Gaussian point response function this is

$$\mu(x, y, z) = \mu_0 exp\{-\frac{1}{2}(\mathbf{x} - \mathbf{x}_0)^{\mathrm{T}}\Lambda(\mathbf{x} - \mathbf{x}_0)\}/\{\bar{S}/\sqrt{n}\}.$$

where $n = 7$ and $\bar{S} = 4.4\%$. This was added to each simulated image $T^*(x, y, z)$ from the warm-warm dataset and the maximum $T_{\max}$ was found over the search volume. Two different peak signal strengths of $\mu_0 = 5\%$ and $\mu_0 = 10\%$ were tried, where 100% is the mean intra-cerebral CBF. Another phantom was created with three peaks centred in the anterior cingulate, SI and SII regions, close to where activation was detected in the hot-warm dataset (Talbot et al.,1991). Finally, a third phantom was created by convolving a $20 \times 40 \times 20$ voxel (57cm^3) region of uniform height with the Gaussian point spread function to create a broader region of activation as opposed to a sharp peak. The maximum height of this region was chosen to be the same as the peak heights above. The region was located in the right hemisphere in roughly the same place where activation was detected in the hot-warm dataset. The proportion of true positives or power of the test are shown in Table 3. It appears that the test has moderate power at detecting a 5% signal but a 10% signal is almost always detected. The test is more powerful at detecting activation when three peaks are present rather than one, and more powerful at detecting a broad region than a sharp peak.

3.5.3 Sensitivity of the voxel-based t-statistic image $T_n(x, y, z)$

The t-statistic image $T_n(x, y, z)$ was calculated for the same 70 sets of simulated images that were used to validate the p-value of $T_{\max}$. Empirical critical values of the maximum $T_{n\max}$ were found by smoothing the distribution of the 70 simulated values. At levels $\alpha = 0.01$, 0.05 and 0.10 the approximate critical values were 20.2, 14.5 and 12.1 respectively. The same phantom activation signals as used previously were added to the simulated images, and the proportion of true positives, or sensitivity, was estimated. For all signals, $T_{n\max}$ was less sensitive than $T_{\max}$; $T_{n\max}$ had about the same sensitivity to a $7\% - 8\%$ signal as $T_{\max}$ had to a 5% signal, which implies that about twice as many subjects would be needed to detect the same signal with $T_{n\max}$ as with $T_{\max}$.

3.6 Validation of the expected Euler characteristic

3.6.1 Specificity

The Euler characteristic χ_t over the search volume was evaluated for excursion sets of 7 orthogonal statistically independent t-statistic images chosen from the 70 simulated images obtained from the warm-warm dataset above, for $-5 < t < 5$ at intervals of 0.1. The average of these is plotted against t in Figure 4, together with the expected Euler characteristic $\mathrm{E}(\chi_t)$ from (6) calculated using the empirical roughness measure $|\Lambda|^{\frac{1}{2}} = 1.64$ and $V = 1090$ cm^3. The agreement is very good. Approximate 95% confidence bands of ± 2 standard deviations, obtained by smoothing the sample standard deviation, have been added. It was noted that the smoothed sample standard deviation was close to the approximate theoretical standard deviation, $\sqrt{\mathrm{E}(\chi_t)}$, for $|t| > 3$.

3.6.2 Sensitivity

The Euler characteristic χ_t was also evaluated for each of the same 7 simulated t-statistic images used above, but with an added 10% peak activation in the anterior cingulate, as previously described. The average of these Euler characteristics, together with its expectation when no signal is present (6), is plotted against t in Figure 5(a). Also shown is the Euler characteristic of the signal $\mu(x, y, z)$ with no noise added; note that it takes the value 1 when t is less than the peak signal value $\mu_0/\{\bar{S}/\sqrt{n}\} = 6.01$. At the $\alpha = 0.05$ critical level $t = t_{0.05} = 4.39$ the average was 0.9 ± 0.1, giving an estimate of approximately one isolated region of activation, which agrees with the single peak of activation in the phantom. Figure 5(a) also plots the average number of local maxima greater than t. As expected, it is larger than the Euler characteristic but very close for large t. The number of local maxima greater than $t = 4.39$ was also 0.9 ± 0.1.

In addition, χ_t was evaluated for the 7 simulated t-statistic images with three added 10% peaks in the anterior cingulate, SI and SII regions (Figure 5(b)). At $t = 4.39$ the average χ_t was 3.0 ± 0.0, in agreement with the three peaks of activation in the phantom. The number of local maxima greater than $t = 4.39$ was also 3.0 ± 0.0.

Finally, χ_t was evaluated for the 7 simulated t-statistic images with the 10% broad region added in the upper right hemisphere (Figure 5(c)). At $t = 4.39$ the average χ_t was 0.9 ± 0.1, in agreement with the one broad region of activation in the phantom. However, the number

of local maxima greater than $t = 4.39$ was 4.7 ± 0.5, much greater then the singe local maximum in the added signal. The conclusion is that a broad region will appear as many local maxima; as the threshold level decreases these local maxima will at first increase the Euler characteristic above 1 ($t = 8$ down to 6.5), then as t decreases further ($t = 6.5$ down to 5) they will coalesce to form one connected region of activation and an Euler characteristic of one. In our experience holes in the excursion set are rare for these values of t and decreases in the Euler characteristic as t decreases are caused by the coalescing of isolated regions rather than the formation of holes. Thus the Euler characteristic more accurately estimates the number of regions of activation, whether they are sharp peaks or broad regions.

Note that the plot of the average Euler characteristic is relatively flat for $4 < t < 5$ so that our estimation procedure is robust against slight changes in the threshold value of t. However, if we set the threshold for the t-statistic at the uncorrected critical value taken from the t-distribution we get far more apparent regions of activation than are really present. Even if we set the threshold at $t = 3$ the number of estimated regions of activation is 4.2 when one 10% peak is present and 6.2 when three peaks are present.

3.7 Validation of the p-value for the voxel-based t-statistic image $T_n(x, y, z)$

3.7.1 Specificity

The t-statistic image $T_n(x, y, z)$ was calculated for the same 70 sets of simulated images that were used to validate the p-value of $T_{\max}$. The proportion of false positives using the theoretical p-value (7) at nominal false positive rates of $\alpha = 0.01$, 0.05 and 0.01 were 0.029, 0.143, and 0.286 respectively. Thus (7) appears to be too liberal and so we suggest using a smaller nominal false positive rate of, say $\alpha = 0.01$ instead of $\alpha = 0.05$. A possible explanation for the discrepancy between simulations and theory has already been discussed; the theoretical result (7) is highly sensitive to the assumption of a Gaussian distribution for each subject's subtracted image (see Addendum).

3.7.2 Sensitivity

Since the theoretical p-value (7) was inaccurate, empirical critical values of $T_{n\max}$ were found by smoothing the distribution of the 70 simulated values. The same phantom activation signals as used previously were added to the simulated images, and the proportion of true positives, or sensitivity, was estimated. The results were discouraging. For all signals, $T_{n\max}$ was less sensitive than $T_{\max}$; $T_{n\max}$ had about the same sensitivity to a 7-8% signal as $T_{\max}$ had to a 5% signal, which implies that about twice as many subjects would be needed to detect the same signal with $T_{n\max}$ as with $T_{\max}$.

4 RESULTS

4.1 Application to the warm-warm dataset of the pain study

The standard deviation pooled over the search volume of $V = 1090$ cm^3 was $\bar{S} = 6.0\%$ and the maximum t-statistic inside the search volume was $T_{\max} = 4.16$. The nominal resolutions of the image were FWHM$_x$ = 20mm, FWHM$_y$ = 20mm and FWHM$_z$ = 6mm. The corrected resolution in the z direction was FWHM$_{zc}$ = 7.6mm (4) which gives $R = 360$ resels. From Table 1(a) we can see that the value of $T_{\max}$ was not significant at the $\alpha = 0.1$ level; the p-value is 0.120 calculated according to (5) and 0.129 calculated according to (1) using the empirical value $|\Lambda|^{\frac{1}{2}} = 1.64$ cm^{-3}. The minimum t-statistic was -3.47 which was also not significant, indicating no evidence of de-activation. Figure 6(a) gives the observed Euler characteristic χ_t and its expected value E(χ_t) from (6) under no activation plotted against t for $-5 < t < 5$ at intervals of 0.01. The agreement seems reasonable over most values of t, confirming that no activation has taken place.

4.2 Application to the hot-warm dataset of the pain study

The standard deviation pooled over the search volume of $V = 1090$ cm^3 was $\bar{S} = 4.4\%$ and the maximum t-statistic inside the search volume was $T_{\max} = 4.99$. From Table 1(a) we can see that the value of $T_{\max}$ was significant at the $\alpha = 0.01$ level; the p-value is 0.00393 calculated according to (5) and 0.00392 calculated according to (1) using the empirical value $|\Lambda|^{\frac{1}{2}} = 1.52$ cm^{-3}.

Figure 6(b) gives the observed Euler characteristic χ_t and its expected value E(χ_t) from (6) under no activation plotted against t for $-5 < t < 5$ at intervals of 0.01. An enlargement of the upper tail also shows the number of local maxima above t for comparison with the Euler characteristic (Figure 6(d)). The agreement is close for $t > 3$, which suggests that the regions of activation are sharp peaks as in Figure 5(b).

To determine the regions of activation, the level $\alpha = 0.05$ critical value $t_{0.05}$ was found by solving (5), which for $R = 360$, gives $t_{0.05} = 4.40$. Using (1) and $|\Lambda|^{\frac{1}{2}} = 1.52$ cm^{-3} gives $t_{0.05} = 4.39$. The Euler characteristic at either threshold level is one, indicating one isolated region of activation. At the $\alpha = 0.1$ level, $t_{0.1} = 4.21$, the Euler characteristic increases to three, indicating three isolated regions of activation. These are in the anterior cingulate and SI regions (Talbot et al., 1991).

The Euler characteristic starts to take on non-integer values for $t < 4$ as the excursion set starts to intersect the boundary of the search volume. However, it is still much larger than its expected value for $t > 3.5$, indicating substantial evidence of activation. The Euler characteristic for $-2 < t < 2$ is hard to interpret, but the fact that Figure 6(b) shows such a large discrepancy between χ_t and E(χ_t) in this range indicates that some low-level activation may also be present. At the other extreme, The minimum t-statistic was -4.32 which was not quite significant at the $\alpha = 0.05$ level, indicating no strong evidence of de-activation.

4.3 Application to the word-recognition study

The standard deviation pooled over the search volume of $V = 1390$ cm^3 ($R = 457$ resels) was $\bar{S} = 5.4\%$, and the maximum t-statistic inside the search volume was $T_{\max} = 5.58$. From Table 1(a) we can see that the value of $T_{\max}$ was significant at the $\alpha = 0.01$ level. Its approximate p-value is 0.00028 calculated according to (5) and 0.00030 calculated according to (1) using the empirical value $|\Lambda|^{\frac{1}{2}} = 1.61$ cm^{-3}. Figure 6(c) gives the observed Euler characteristic χ_t and its expected value $\mathrm{E}(\chi_t)$ from (6) under no activation plotted against t for $-6 < t < 6$ at intervals of 0.01. An enlargement of the upper tail also shows the number of local maxima above t for comparison with the Euler characteristic (Figure 6(e)). The agreement is close for $t > 5$, but for $t < 5$ the number of local maxima is much larger.

To determine the regions of activation, the level $\alpha = 0.05$ critical value $t_{0.05}$ was found by solving (5), which for $R = 457$, gives $t_{0.05} = 4.45$. Using (1) and $|\Lambda|^{\frac{1}{2}} = 1.61$ cm^{-3} we get $t_{0.05} = 4.47$. The Euler characteristic at either threshold level is three with no holes in the excursion set, indicating three isolated regions of activation. These were identified in the occipital region, the left inferior temporal region and the left frontal region. Note that the number of local maxima greater than $t = 4.47$ is six, indicating six centres of activation. Four of these local maxima coalesce in the occipital region to form only one isolated region of activation, which explains why the Euler characteristic at $t = 4.47$ is only three. This behaviour is reminiscent of the broad phantom signal in Figure 5(c). For values of t closer to zero the Euler characteristic looks very similar to that of the hot-warm dataset of the pain study (compare Figures 5(b) and 5(d)). At the other extreme, the minimum value of $T(x, y, z)$ was -4.98 at a point near the right anterior temporal lobe. Its p-value was 0.005 and its Euler characteristic at $t = -t_{0.05} = -4.47$ was one, indicating one possible region of de-activation.

5 CONCLUSIONS

This paper has described a method which provides for the direct determination in three dimensions of the significance of focal physiological response in a PET activation study. From a practical point of view, there are two principal results. The first, and most directly useful result, is Table 1(a) which gives some critical values for the global maximum $T_{\max}$ of the t-statistic image based only on the number of resolution elements or resels R in the search volume (5). This number is defined simply as the search volume divided by the product of the three FWHMs of the image reconstruction process. More accurate results are obtainable if the FWHM in the z-direction is corrected for interpolation and re-sampling (4). These critical values increase as the search volume increases, so that the search volume should be chosen *a priori* to be as small as possible. If activation is suspected in the left hemisphere, for example, the search volume should be restricted to the left hemisphere. The search volume could also be confined to grey matter regions identified by MRI. An alternative but more time-consuming method of obtaining p-values of $T_{\max}$ is to evaluate the measure of roughness $|\Lambda|^{\frac{1}{2}}$ from the data directly using empirical numerical derivatives, as outlined in the Methods section, and use (1). In practice this seems to give very similar results to

the p-values obtained by using the number of resels R and (5), which requires very little computation.

We have suggested using a t-statistic image based on a pooled standard deviation rather than a voxel based standard deviation, provided that the standard deviation image appears to be flat. Our simulations show that the sensitivity of the voxel based standard deviation t-statistic appears to be less than the pooled standard deviation based t-statistic. The arguments that have been advanced in this paper have assumed a constant underlying standard deviation. Figure 3 indicates no significant change in standard deviation against three different variables. Therefore pooling the standard deviation gives greater sensitivity to activation. In the event that there is some evidence of non-constant standard deviation then it might be appropriate to use a local standard deviation in the t-statistic image, but this would have to be demonstrated on the data to justify the loss in sensitivity.

Recent work by Friston *et al.* (1991) for the two dimensional situation is very close in spirit to our three dimensional approach. Working independently, these authors have derived an expression for the p-value of the maximum t-statistic searched over a two-dimensional slice of the data. The form of their p-value is similar to our (1) and (5), and it depends on a measure of the smoothness of the image in an analogous way. It is interesting to note that the critical values that they obtain for the maximum over a slice are in the range of 3 to 4, whereas the critical values we have obtained for the maximum over a volume are higher, in the range of 4 to 5. The implication for this difference is that setting the threshold at a typical value of say 4.5 for $\alpha = 0.05$ would provide an expectation of one false positive per 20 volumes; setting it at say 3.5 might provide an expectation of one false positive per 20 slices, but one per volume in a 20 slice volume.

The final choice of threshold will depend upon how conservative one wishes to be when interpreting physiological activation datasets. It may be acceptable if one observes 10 peaks above $t = 3.5$ to know that one peak is on average false. To choose a threshold of $t = 4.5$ would virtually guarantee that any peak quoted was real at the expense of rejecting data which was consistent with physiological evidence from other studies. For either choice, Table 1(b) and Figure 7 show the statistical price to be paid.

The second principal result is that the number of isolated regions of activation appears to be well estimated by the Euler characteristic of the region above the critical value of $T_{\max}$, whereas the number of local maxima appears to be an overestimate, particularly when the region is broad. Although change distribution analysis (Fox, 1988) has proven to be extremely useful in the statistical analysis of subtractive PET activation studies, it does have some limitations in both theory and practice. The g_2 omnibus statistic provides a sensitive and specific test for determining if a significant change in state exists, but it does not identify which peaks are significant. Post-hoc z-scores do not explicitly deal with the issue of multiple comparisons. Instead, we have suggested determining the number of regions of activation by thresholding the t-statistic image at the level α critical value for $T_{\max}$, which for $\alpha = 0.05$ corresponds to a z-score of approximately 3.0 on typical datasets in our laboratory (Figure 7).

This procedure protects us against finding some regions of activation when in fact none are present. In this respect our approach is no different from using an omnibus test such as g_2 followed by thresholding the t-statistic image at the uncorrected critical value of the t-distribution, or indeed any critical value; both methods still protect against finding some

activation when none is present. However if say two large peaks are present, large enough to push an omnibus test statistic such as g_2 over its critical value, then activation will be detected. Following this by thresholding the t-statistic at an uncorrected value will yield not only the two large peaks, but also a large number of spurious peaks from the unactivated part of the image. In other words, such a procedure does not protect us against declaring that there are say 20 regions of activation when in fact there are only two. In the same situation, our approach of thresholding at the critical values of $T_{\max}$ should still protect us against finding too many regions of activation, even when large peaks are present, because we are still protected in the unactivated part of the image. Although this is outside the scope of current theoretical results, it is supported by our simulations, which show that the Euler characteristic, thresholded at the critical values of $T_{\max}$, does appear to correctly predict the number of isolated regions of activation, whether they are broad regions or sharp peaks.

Acknowledgements: We would like to thank the staff of the Positron Imaging Laboratories and the Medical Cyclotron Unit for their help. Special thanks to Sylvain Milot for technical assistance. We are grateful to Drs. Cathy Bushnell and Gary Duncan, as well as to Drs. Daniel Bub and Howard Chertkow for allowing us to use data collected in their experiments. Funding for this work was provided by the McDonnell-Pew Program in Cognitive Neurosciences, the Medical Research Council of Canada, the Natural Sciences and Engineering Research Council of Canada and the Fonds pour la Formation de Chercheurs et l'Aide á la Recherche de Québec.

APPENDIX

Proof of (4)

The result is derived by finding the mean variance of the derivative of the interpolated process, dividing by its mean variance, and equating this to the corrected FWHM using (2). Consider a voxel of the re-sampled image at distances ad and $(1-a)d$, $0 \leq a \leq 1$, from two voxels in two adjacent planes of the PET camera with values X_1 and X_2 respectively. Then by linear interpolation the value of the re-sampled voxel is $T = (1-a)X_1 + aX_2$. Without loss of generality we can assume the X_1 and X_2 have zero mean, unit variance, and correlation coefficient ρ. The variance of T is $\mathrm{Var}(T|a) = 1 - 2a(1-a)(1-\rho)$. Averaged over all values of a, $0 \leq a \leq 1$, the mean variance is $\mathrm{Var}(T) = 1 - (1-\rho)/3$.

To determine the numerical derivative D of the re-sampled image two situations can arise: the pair of adjacent re-sampled voxels either lie between the same pair of adjacent planes of the PET camera (event E_1), or are separated by a plane of the PET camera (event E_2). In the former situation D equals the derivative of the voxels in the PET planes, and so $\mathrm{Var}(D|E_1) = \mathrm{Var}\{(X_1 - X_2)/d\} = 2(1-\rho)/d^2$. In the latter situation, suppose that X_1, X_2 and X_3 are the values of voxels in three adjacent PET planes, and that the middle plane is at distances bd_c and $(1-b)d_c$, $0 \leq b \leq 1$, from the two adjacent voxels of the re-sampled image. Then it can be shown that $D = \{bX_1 + (1-2b)X_2 + (b-1)X_3\}/d$. Since the correlations between X_1 and X_2, X_2 and X_3, and X_1 and X_3 are ρ, ρ and ρ^4 respectively then $\mathrm{Var}(D|E_2, b) = \{2(1-\rho) - 2b(1-b)(3 - 4\rho + \rho^4)\}/d^2$. Averaged over all values of b, $0 \leq b \leq 1$, the mean variance is $\mathrm{Var}(D|E_2) = \{2(1-\rho) - (3 - 4\rho + \rho^4)/3\}/d^2$.

Since the proportions of resampled voxels from the first situation is $(1-d_c)/d$ and d_c/d from the second, the overall mean variance of the derivative, found by the weighted average of $\mathrm{Var}(D|E_1)$ and $\mathrm{Var}(D|E_2)$ is $\mathrm{Var}(D) = \{2(1-\rho) - (d_c/d)(3-4\rho+\rho^4)/3\}/d^2$. Equating the mean measure of roughness $\mathrm{Var}(D)/\mathrm{Var}(T)$ to $(4log_e2)/(\mathrm{FWHM}_{zc})^2$ from (2) gives the corrected FWHM of the re-sampled image, FWHM_{zc}.

Proof of (7)

Let $\mathbf{x} = (x, y, z)'$. Assume that $\Delta_i = \Delta_i(\mathbf{x})$ is a Gaussian random field with expectation zero and standard deviation σ, and that $Var(\partial\Delta_i/\partial\mathbf{x}) = \Lambda\sigma^2$. Now since $\Delta_1, ..., \Delta_n$ and their partial derivatives are all independent Gaussian random random variables with zero expectations (Adler, 1981, page 31) then by conditioning on $\Delta_1, ..., \Delta_n$ we have

$$\begin{aligned}
\mathrm{Var}\left(\frac{\partial G}{\partial \mathbf{x}}\right) &= \mathrm{Var}\left(\frac{\partial G}{\partial T_n}\sum_{i=1}^{n}\frac{\partial T_n}{\partial \Delta_i}\frac{\partial \Delta_i}{\partial \mathbf{x}}\right) \\
&= \mathrm{E}\left\{\left(\frac{\partial G}{\partial T_n}\right)^2\sum_{i=1}^{n}\left(\frac{\partial T_n}{\partial \Delta_i}\right)^2 Var\left(\frac{\partial \Delta_i}{\partial \mathbf{x}}\right)\right\} \\
&= \mathrm{E}\left\{\left(\frac{\partial G}{\partial T_n}\right)^2\sum_{i=1}^{n}\left(\frac{\partial T_n}{\partial \Delta_i}\right)^2\right\}\Lambda\sigma^2.
\end{aligned}$$

From the definition of T_n we have

$$\frac{\partial T_n}{\partial \Delta_i} = \frac{\partial}{\partial \Delta_i}\frac{M}{S/\sqrt{n}} = \frac{1}{\sqrt{n}S} - \frac{\sqrt{n}M(\Delta_i - M)}{(n-1)S^3},$$

and so

$$\sum_{i=1}^{n}\left(\frac{\partial T_n}{\partial \Delta_i}\right)^2 = \frac{1}{S^2} + \frac{nM^2}{(n-1)S^4}.$$

Hence the correction factor to the variance matrix of the partial derivatives of G is

$$\lambda_n = \mathrm{E}\left\{\left(\frac{p_n'(T_n)}{p'(G)}\right)^2\left(\frac{1}{S^2} + \frac{nM^2}{(n-1)S^4}\right)\right\}\sigma^2,$$

where $p_n'(t)$ is the probability density function of a t-random variable with $n-1$ degrees of freedom and $p'(z)$ is the probability density of a standard normal random variable. This can be simplified further using the fact that $U = (nM^2 + (n-1)S^2)/\sigma^2$ has a χ^2-distribution with n degrees of freedom, independent of T_n, and that $E(1/U^2) = 1/(n-2)$, to give

$$\lambda_n = \mathrm{E}\left\{\left(\frac{p_n'(T_n)}{p'(G)}\right)^2\frac{(T_n^2+n-1)^2}{(n-1)(n-2)}\right\}.$$

Integrating over the density of T_n gives

$$\lambda_n = \int_{-\infty}^{\infty}\frac{(t^2+n-1)^2}{(n-1)(n-2)}\frac{p_n'(t)^3}{p'(g)^2}dt,$$

where $g = p^{-1}(p_n(t))$. The integral is finite only if $n \geq 4$.

ADDENDUM

This article describes exact results for a Gaussian random field and an approximation to the non-Gaussian case for the voxel-based t statistic image $T_n(x, y, z)$. Since the submission of this article, Worsley (1992) has extended the results of Adler (1981) to find an approximate p-value for $T_{n\max}$ and an exact expression for the expected Euler characteristic for the t field $T_n(x, y, z)$, as well as the χ^2 field $S^2(x, y, z)$. These results supercede (7) and will be the subject of a forthcoming article.

REFERENCES

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

Adler RJ, Hasofer AM (1976) Level crossings for random fields *Annals of Probability, 4*: 1-12.

Evans AC, Marrett S, Neelin P, Gum T, Dai W, Milot S, Meyer E, Bub D (1992) Anatomical mapping of functional activation in stereotactic coordinate space *NeuroImage*, 1:43-53.

Evans AC, Marrett S, Peters TM (1989) Anatomical-functional correlative analysis of the human brain using three-dimensional imaging systems *Proceedings of the International Society of Optical Engineering (SPIE): Medical Imaging III* 264-274.

Evans AC, Marrett S, Torrescorzo J, Ku S, Collins L (1991) MRI-PET correlative analysis using a volume of interest (VOI) atlas *J Cereb Blood Flow Metab 11(2)*: A69-A78.

Evans AC, Thompson CJ, Marrett S, Meyer E, Mazza M (1991) Performance characteristics ofthe PC-2048: A new 15-slice encoded-crystal PET scanner for neurological studies *IEEE Transactions on Medical Imaging 10(1)*: 90-98.

Fox PT, Mintun ME (1989) Non-invasive functional brain mapping by change-distribution analysis of averaged PET images of $H_2^{15}O$ tissue activity *J Nucl Med*: 141-149.

Fox PT, Mintun MA, Reiman EM, Raichle ME (1988) Enhanced detection of focal brain responses using intersubject averaging and distribution analysis of subtracted PET images *J Cereb Blood Flow Metab 8*: 642-653.

Fox PT, Perlmutter JS, Raichle ME (1985) A stereotactic method of anatomical localization for positron emission tomography *J Comput Assist Tomogr 9(1)*: 141-153.

Fox PT, Raichle ME (1984) Stimulus rate dependence of regional cerebral blood flow in human striate cortex, demonstrated with positron emission tomography *J Neurophysiology 51*: 1109-1121.

Friston KJ, Frith CD, Liddle PF, Frackowiak RSJ (1991). Comparing functional (PET) images: the assessment of significant change *J Cereb Blood Flow Metab 11*: 690-699.

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

Herscovitch P, Markham J, Raichle ME (1983) Brain blood flow measured with intravenous $H_2^{15}O$ I. Theory and error analysis. *J Nucl Med 24*: 782-789.

Holm S (1979) A simple sequentially rejective multiple test procedure *Scandinavian Journal of Statistics, 6*: 65-70.

Raichle ME, Martin WRW, Herscovitch P, Mintun MA, Markham J (1983) Brain blood flow measured with intravenous $H_2^{15}O$ II. Implementation and validation. *J Nucl Med 24*: 790-798.

Talairach J, Tournoux P (1988) *Co-planar stereotactic atlas of the human brain: 3-Dimensional proportional system: an approach to cerebral imaging* Georg Thieme Verlag, Stuttgart, New York

Talbot JD, Marrett S, Evans AC, Meyer E, Bushnell MC, Duncan GH (1991) Multiple representations of pain in human cerebral cortex *Science 251*: 1355-1358.

Vanmarcke E (1983) *Random Fields. Analysis ans Synthesis.* Cambridge, MIT Press.

Worsley KJ (1992) Local maxima and the expected Euler characteristic of excursion sets of χ^2, t and F fields. *Journal of Applied Probability* (in press).

TABLE 1. *Values of the threshold t for controlling (a) the p-value of $T_{\max}$, $\alpha = \mathrm{P}(T_{\max} > t)$ from (5), and (b) the expected number of false positive isolated regions of activation, $\mathrm{E}(\chi_t)$ from (6), for values of R, the number of resels or resolution elements in the search volume.*

	(a) p-value of $T_{\max}$			(b) regions of activation		
R	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\mathrm{E}(\chi_t) = 1$	$\mathrm{E}(\chi_t) = 2$	$\mathrm{E}(\chi_t) = 5$
100	4.47	4.05	3.84	3.02	2.68	1.91
200	4.64	4.24	4.05	3.30	3.02	2.55
300	4.74	4.34	4.16	3.45	3.19	2.78
400	4.81	4.42	4.24	3.55	3.30	2.92
500	4.86	4.47	4.30	3.62	3.38	3.02
1000	5.01	4.64	4.47	3.84	3.62	3.30
2000	5.16	4.81	4.64	4.05	3.84	3.55

TABLE 2. *Validation of the relationships (2) and (3) between the varaince matrix Λ of the derivatives of the image and the FWHMs of the image. Values of Λ from three datasets are given, together with their effective FWHMs determined by inverting the relationship (2). Also given are values of Λ determined from (2) using the nominal FWHMs of the image, both uncorrected and corrected for interpolation and re-sampling in the z-direction. Note that the agreement between values of $|\Lambda|^{\frac{1}{2}}$ from the three datasets is in good agreement with that determined by (3) provided that the corrected FWHM_{zc} is used.*

Dataset	Direction	$\Lambda(\mathrm{cm}^{-2})$			$\lvert\Lambda\rvert^{\frac{1}{2}}(\mathrm{cm}^{-3})$	FWHM(mm)
Pain study,	x	0.754	-0.014	-0.008		19.2
warm-warm.	y	-0.014	0.711	-0.088	1.64	19.8
	z	-0.008	-0.088	5.033		7.4
Pain study,	x	0.753	-0.011	-0.007		19.2
hot-warm.	y	-0.011	0.636	0.005	1.52	20.9
	z	-0.007	0.005	4.840		7.6
Word-recognition	x	0.766	0.007	0.003		19.1
study.	y	0.007	0.785	0.280	1.61	18.8
	z	0.003	0.280	4.400		8.0
From (2), using	x	0.693	0	0		20
uncorrected	y	0	0.693	0	1.92	20
FWHM_z	z	0	0	7.702		6
From (2), using	x	0.693	0	0		20
corrected	y	0	0.693	0	1.51	20
FWHM_{zc}	z	0	0	4.762		7.6

TABLE 3. *Specificity and sensitivity of* $T_{\max}$. *Proportion of simulated data sets significant at level* α, *based on the theoretical p-values (1) and (5). The number of resels in the search volume of* $V = 1090cm^3$ *was* $R = 360$ *and the number of simulations was* 70.

Signal added	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
None	0.014	0.057	0.100
One 5% peak	0.271	0.343	0.414
Three 5% peaks	0.186	0.414	0.557
One 10% peak	0.829	0.857	0.900
Three 10% peaks	1.000	1.000	1.000
Broad 5% region	0.543	0.800	0.886
Broad 10% region	1.000	1.000	1.000

Figure 1. Corrected FWHM

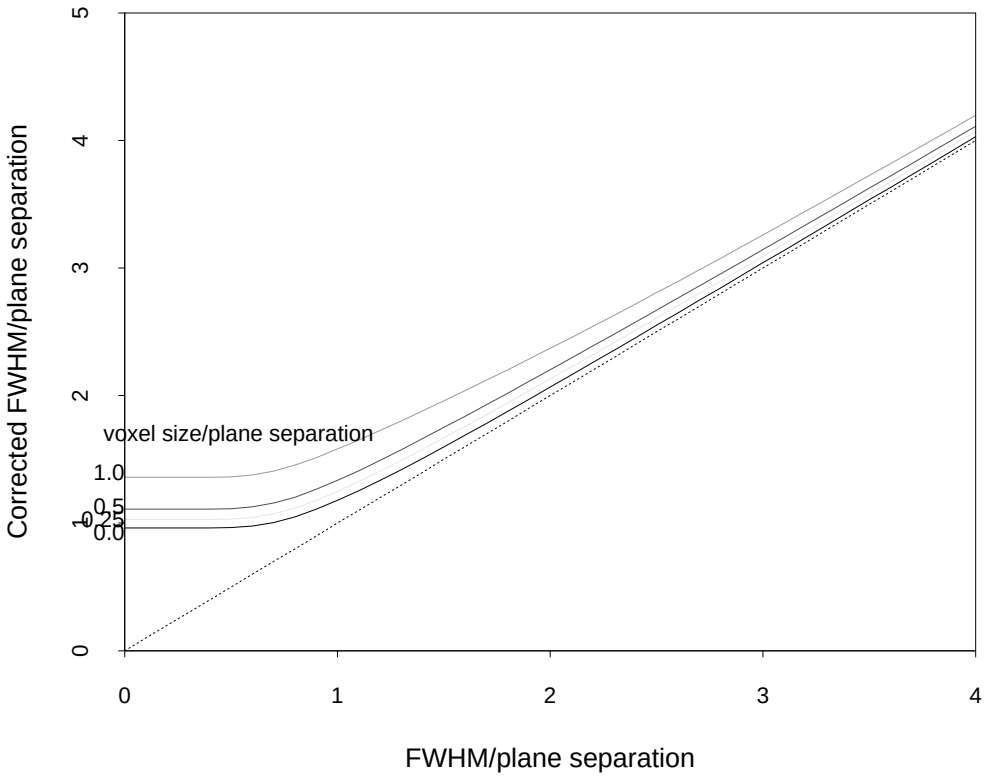

Figure 1: The corrected FWHM in the z-direction, FWHM_{zc}, due to linear interpolation and re-sampling between planes of the PET camera. The graph plots FWHM_{zc}/d against FWHM_z/d, where FWHM_z is the nominal FWHM of the PET camera in the z direction and d is the separation between planes of the PET camera, for fixed values of d_c/d , where d_c is the voxel size of the re-sampled image. The values of d_c/d are 1.0, 0.5, 0.25 and 0.0 (solid lines, from top to bottom, respectively). For comparison a diagonal line has been added (dashed line). Note that $\mathrm{FWHM}_{zc} > \mathrm{FWHM}_z$, so that the interpolated image is smoother than the original image, particularly if FWHM_z/d is small and d_c/d is close to one.

Figure 2. The Euler characteristic for an artificial image in two dimensions

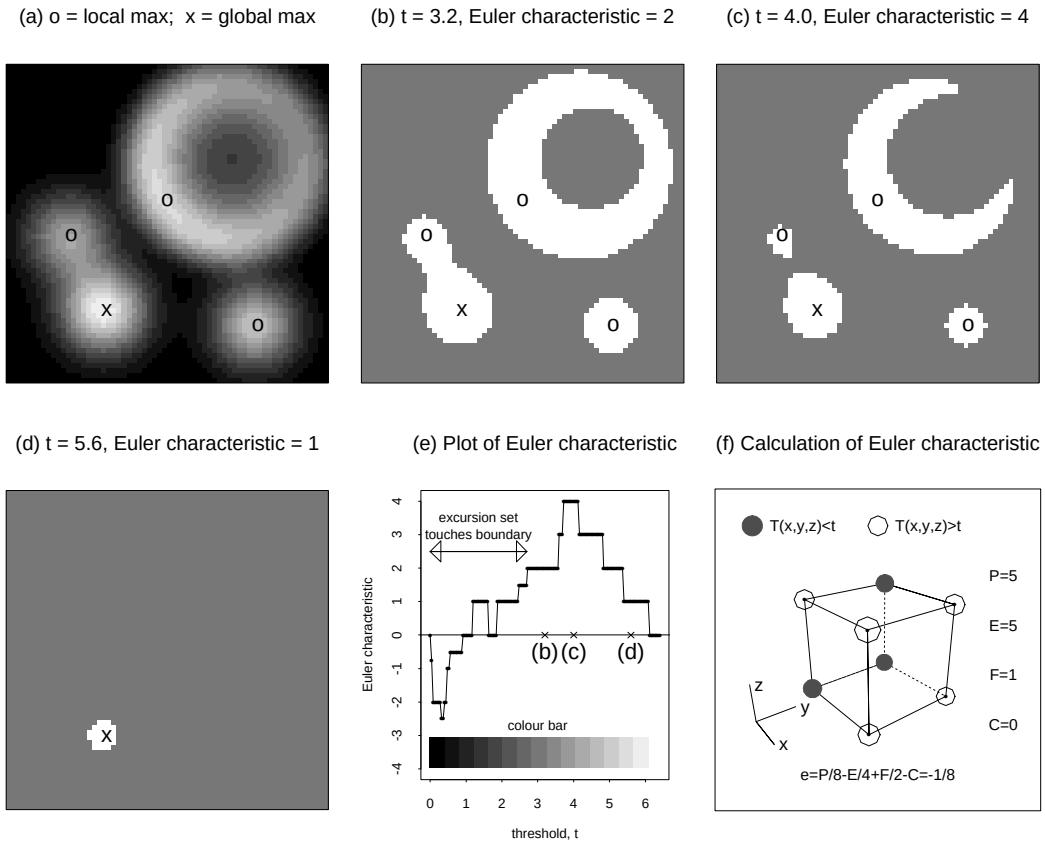

Figure 2: The Euler characteristic of an artificial image in two dimensions. (a) The image, with colour bar in (e). Local maxima are indicated by o and the global maximum is indicated by x. (b) Excursion set (white areas) above a threshold $t = 3.2$. Since the excursion set does not touch the boundary the Euler characteristic counts the number of isolated regions minus the number of 'holes', giving $\chi_t = 2$. (b) As the threshold is increased to $t = 4.0$ the holes disappear and the Euler characteristic counts the number of local maxima, giving $\chi_t = 4$. (c) At even higher levels $t = 5.6$, the Euler characteristic takes the value one if the global maximum exceeds t and zero otherwise, giving $\chi_t = 1$. (e) A plot of Euler characteristic against t. Note that for $t < 2.7$ the excursion set touches the boundary and the Euler characteristic can take non-integer values. (f) A $2 \times 2 \times 2$ 'cube' of 8 adjacent voxels in the search volume of a three-dimensional image, to illustrate the calculation of the Euler characteristic. Voxels below the threshold t are shaded and those above are white. The contribution of this cube to the Euler characteristic is $e = -1/8$.

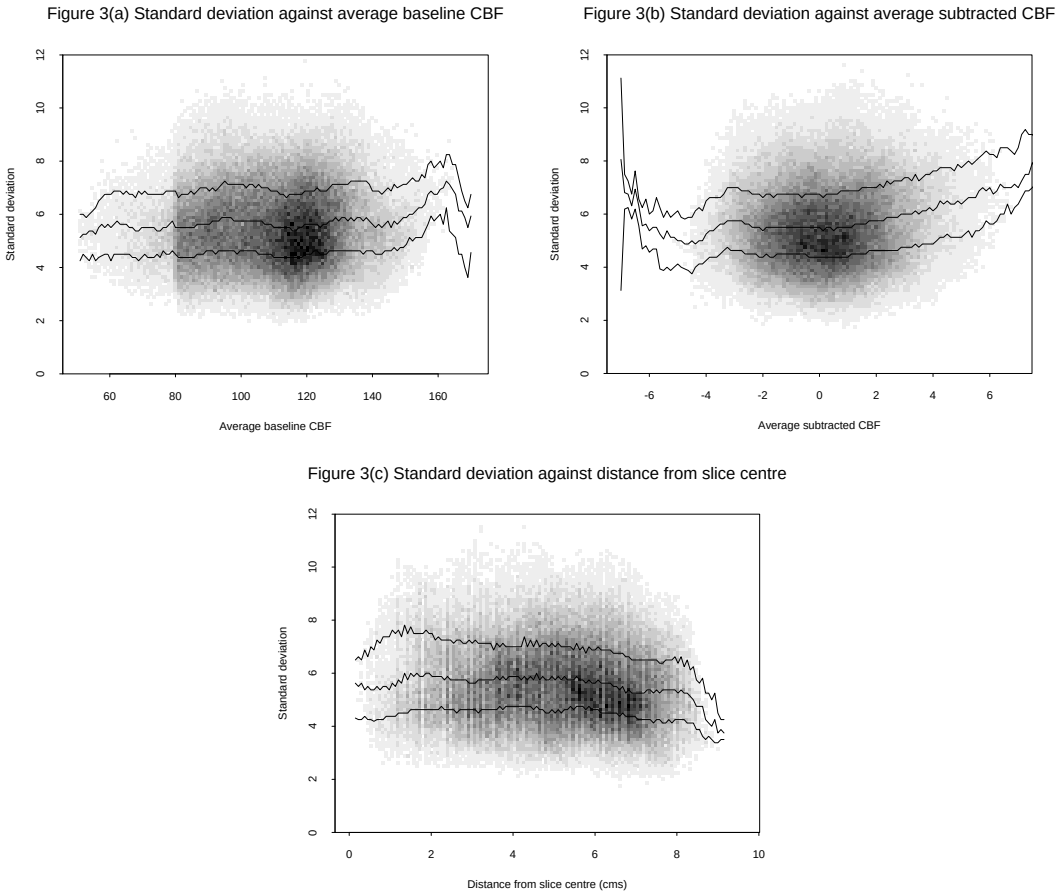

Figure 3: Validation of the standard deviation. Scatter plots of the standard deviation $S(x, y, z)$ of the warm-warm dataset from the pain study data for all voxels in the search volume. $S(x, y, z)$ is plotted against (a) average baseline CBF, (b) average ΔCBF $M(x, y, z)$ and (c) distance from the slice centre (cms). Darker areas indicate greater voxel density. Medians (middle line) and upper and lower quartiles (upper and lower lines) have been added. $S(x, y, z)$ appears to be roughly constant, allowing us to pool the standard deviation over the search volume.

Figure 4. Validation of the Euler characteristic

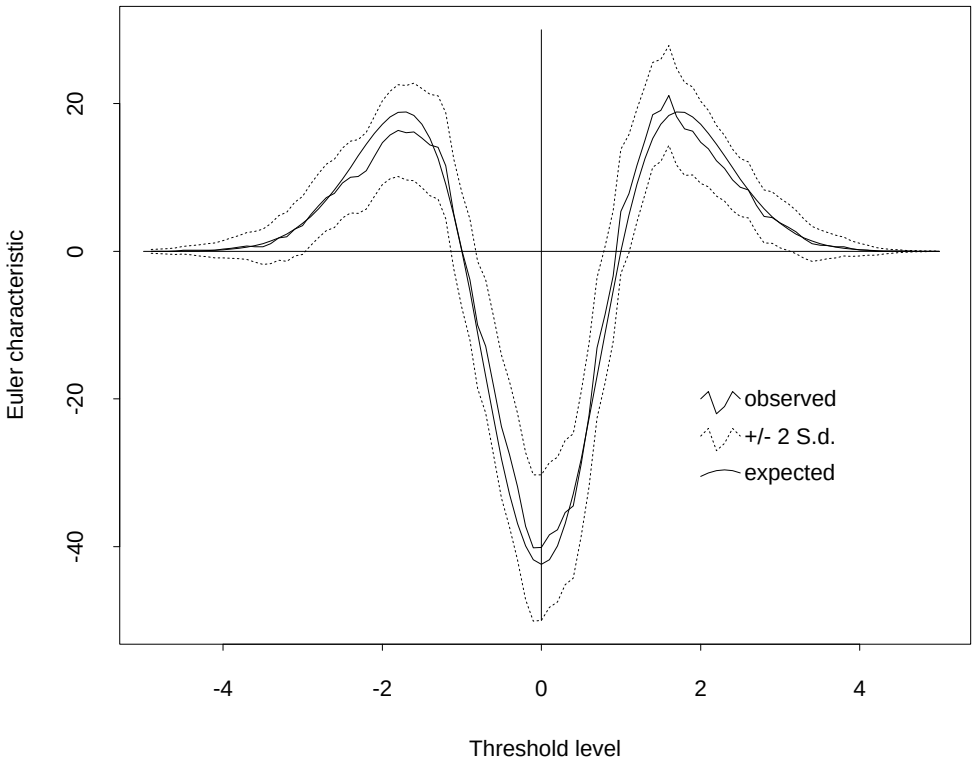

Figure 4: Validation of the expected Euler characteristic (6) using simulated data from the warm-warm dataset of the pain study. The average Euler characteristic χ_t (jagged line) and the expected Euler characteristic $\mathrm{E}(\chi_t)$ from (6) (smooth line) plotted against the threshold level t at intervals of 0.1, for 7 orthogonal statistically independent simulated images $T^*(x, y, z)$ formed by changing the sign of 4 of the 8 subjects' subtracted images. An approximate 95% confidence band of width ± 2 smoothed standard deviations has been added (dotted lines).

Figure 5(a). Sensitivity of the Euler characteristic to one peak

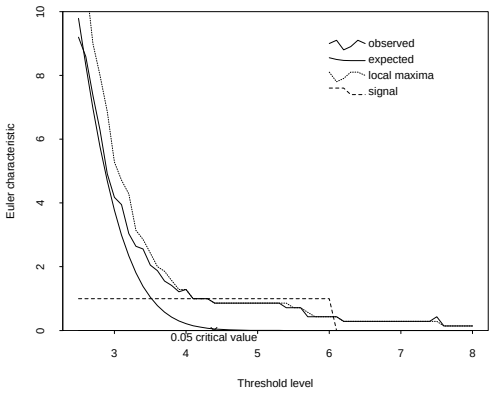

Figure 5(b). Sensitivity of the Euler characteristic to three peaks

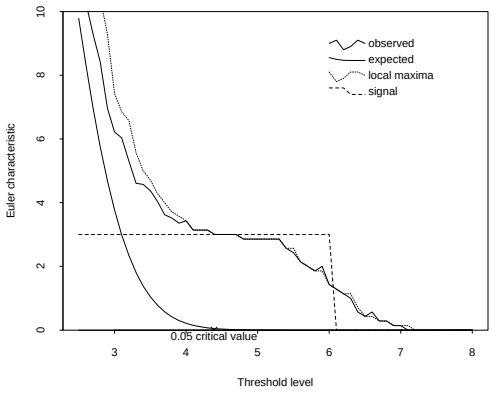

Figure 5(c). Sensitivity of the Euler characteristic to a broad region

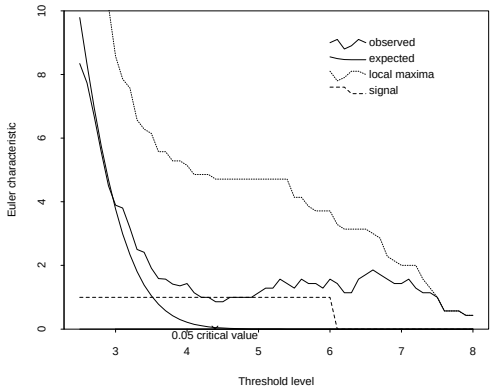

Figure 5: Specificity of the Euler characteristic, assessed by adding phantom signals to the simulated data from the warm-warm dataset used in Figure 4. The average Euler characteristic χ_t (jagged line), the expected Euler characteristic $\mathrm{E}(\chi_t)$ from (6) (smooth line) plotted against the threshold level t, $t > 2.5$ at intervals of 0.1, for the 7 orthogonal simulated images $T^*(x, y, z)$ formed by changing the sign of 4 of the 8 subjects' subtracted images. Also shown is the average number of local maxima greater then t (dotted line) and the Euler characteristic of the signal alone (dash-dot line). In (a) a phantom signal of peak ΔCBF $= 10\%$ in the anterior cingulate was added to each simulated image; in (b) a phantom signal of three peaks of ΔCBF $= 10\%$ in the anterior cingulate, SI and SII regions were added to each simulated image, and in (c) a broad region of uniform activation of ΔCBF $= 10\%$ in the upper right hemisphere was added. The level $\alpha = 0.05$ critical value $t_{0.05} = 4.39$ is marked on the horizontal axis.

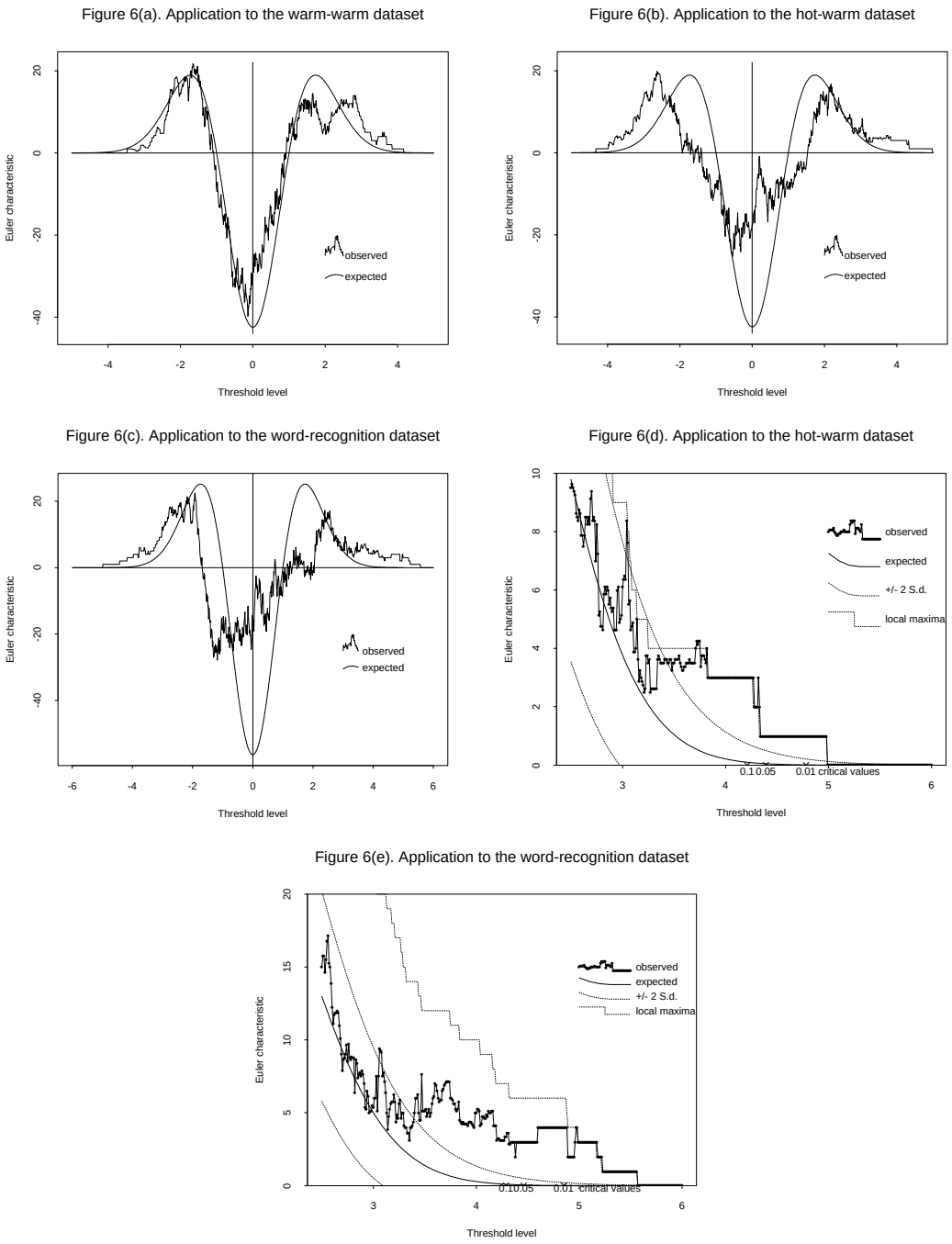

Figure 6: Application of the Euler characteristic to the three datasets. The observed Euler characteristic χ_t for the t-statistic image $T(x, y, z)$ (jagged line) plotted against the threshold level t at intervals of 0.01 for (a) the warm-warm dataset, (b) the hot-warm dataset, and (c) the word recognition study. The smooth line is the expected Euler characteristic $\mathrm{E}(\chi_t)$ from (6). (d) and (e) Enlargements of the upper tails ($c > 2.5$) of Figures 5(b) and (c) respectively. Also shown is the cumulative number of local maxima above t (stepped dotted line). Approximate 95% confidence bands of $\pm 2\sqrt{\mathrm{E}(\chi_t)}$ have been added about $\mathrm{E}(\chi_t)$ (smooth dotted lines). The level $\alpha = 0.1, 0.05$ and 0.01 critical values are marked on the horizontal axes.

Figure 7. Expected number of false positive isolated regions

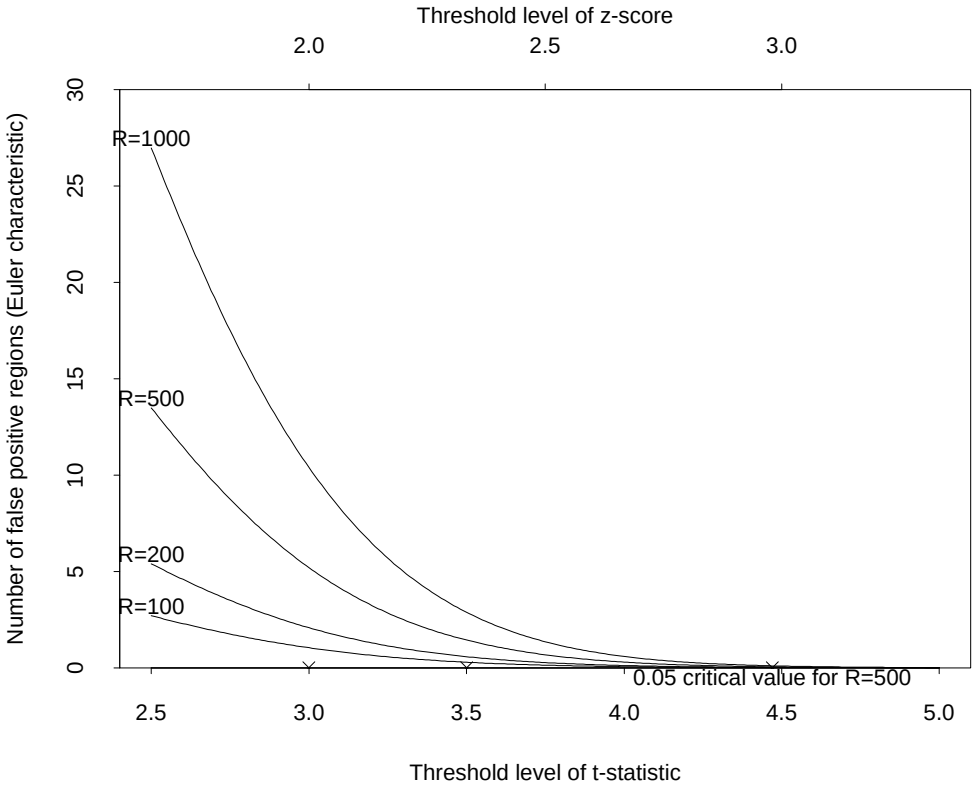

Figure 7: The expected number of false positive isolated regions of activation, as measured by the Euler characteristic (6), plotted against the threshold value of the t-statisitic, t. Four values of the number of resolution elements, or resels, are plotted: R=100,200,500 and 1000 (from bottom to top, respectively). A second horizontal axis has been added at the top of the plot which corresponds to typical z-scores for local maxima from change distribution analysis (Fox et al.,1988) on datasets in our laboratory. Thus thresholding an image at a z-score value of 2.0 will yield on average approximately 5.2 false positive regions of activation in a volume of 500 resels; thresholding at a t-statistic value of 3.5, the approximate critical value for local maxima in a slice (Friston et al.,1991), will yield on average approximately 1.4 false positives per volume. Note that thresholding at a t-statistic value of 4.47 from Table 2 will yield on average approximately 0.05 false positives per volume.