Chapter Three

Current Methods for Testing Statistic Images

Having formed a statistic image, the final difficulty in the analysis of a functional mapping experiment is assessing it. Extreme voxel statistics indicate evidence against the null hypothesis at that voxel, but how extreme is significant? In seeking to test statistic images at the voxel level we are presented with a large multiple comparisons problem, which is the subject of the remainder of this thesis.

In this chapter the current methods for assessing statistic images are reviewed. The necessary terminology and theory is introduced, and the methods expounded in sufficient detail to enable their implementation. A set of simulated Gaussian statistic images is used to illustrate the approaches, which gives an indication of the size of each test under idealised conditions. The chapter closes with analyses of a PET data set using a variety of methods. Some idea of the relative power of the tests (in 2D) can be gleaned from the simulation study presented in the Two-Stage chapter (ch.4).

3.1. Preliminaries

3.1.1. Multiple comparisons

Families of hypotheses

We have a null hypothesis, H_k, for each voxel *k*. We wish to test these hypotheses whilst controlling the probability of false rejection for any voxel hypothesis. In the language of multiple comparisons (Hochberg & Tamhane, 1987) we have a *family* of tests, a "collection of inferences for which it is meaningful to take into account some combined measure of errors". The probability of falsely rejecting any voxel hypotheses, a Type I error, is then the *familywise* error rate (FWE).

Denote by H_W, the combined voxel hypotheses over the intracerebral voxels $W=\{1,\ldots,K\}$, the *omnibus* null hypothesis. The omnibus hypothesis is the intersection of the voxel hypotheses, and is true if and only if the voxel hypotheses are all true. Rejecting any voxel hypothesis implies rejection of the omnibus hypothesis. Rejecting the omnibus hypothesis implies rejection of some (unspecified) voxel hypotheses. The FWE for a procedure is then simply the probability of falsely rejecting H_W.

Definitions: Size, level, valid, conservative, power

The *size* of a test is the probability of Type I error. The *level* of a test is the specified maximum probability of Type I error (usually denoted by α). A *valid* test has size at most the level. An *exact* test has size equal to the level. A *conservative* test has size much less than the level. Conservative tests usually have lower *power* than exact tests, where the power of the test is the probability that the test correctly rejects, given a particular departure from the null hypothesis.

For multiple comparisons problems, there are two forms of control of familywise error, weak and strong.

Weak control over FWE: "Omnibus" tests

Weak control over FWE simply requires that the test of the omnibus hypothesis is valid. That is, the probability of "rejecting" a true H_W is at most the given level α:

$$\Pr(\text{"reject } H_W\text{"} | H_W) \leq \alpha \qquad (30)$$

In the context of an activation study, with voxel hypotheses of no activation at that voxel (however expressed), the omnibus null hypothesis is of "no activation at any voxel in the volume of the brain under examination". Evidence against the null hypothesis indicates the presence of "some activation somewhere", but the test has no *localising power*, in that the Type I error for individual voxels is not controlled.[40] (If the omnibus null hypothesis has been rejected, then any set of voxels could be declared as "activated". Weak control of FWE is maintained since voxels are only declared activated if the omnibus hypothesis is rejected.)

These tests are known as *omnibus tests*, since they assess whether there is any evidence at all, anywhere, against the omnibus null hypothesis. They are useful in the present context when interest is not in the location of some effect within the brain, but whether there is any effect at all. If interest lies in reliably locating an effect, then a test procedure with strong control over FWE is required.

---

[40] "Localising power, is not an accepted statistical term, but it is widely used in the PET community.

Strong control over FWE: "Localising" tests

Strong control over FWE requires that FWE is controlled not just under H_W, when all the hypotheses are true, but under any selection of true hypotheses. That is, for any subset U of the intracerebral voxels (with indices in) W, the corresponding subfamily of voxel hypotheses are tested with (weak) control over FWE (eqn.31). Clearly strong control implies weak control.

$$\Pr(\text{"reject } H_U\text{"} | H_U) \leq \alpha \text{ for any } U \subseteq W. \tag{31}$$

Thus, for any region of the brain defined by voxels $U \subseteq W$, for which the corresponding omnibus hypothesis H_U is true (meaning the voxel hypotheses are all true in this region), the test rejects H_U (by rejecting one or more of the voxel hypotheses) with probability at most α, regardless of the truth of hypotheses for voxels elsewhere in the brain. A test procedure with strong control over FWE has localising power, and enables the results of the tests at individual voxels to be reported.

Adjusted p-values

In most cases it is possible to compute an *adjusted p-value*, $\tilde{P}_k$, for each test of H_k, such that the decision to reject H_k for a test with FWE = α is obtained merely by comparing $\tilde{P}_k$ with α. These adjusted *p*-values can then be displayed in an adjusted *p*-value image $\tilde{\boldsymbol{P}} = (\tilde{P}_1,\ldots,\tilde{P}_K)$. Formally:

$$\tilde{P}_k = \inf\{ \alpha \mid H_k \text{ is rejected at FWE} = \alpha \} \tag{32}$$

That is, for a particular multiple comparisons procedure, the adjusted *p*-value for hypothesis H_k is the smallest α such that H_k is rejected by the test at FWE = α.

Often it is expedient to think of multiple comparisons procedures in terms of *p*-value adjustment. The use of adjusted *p*-values rather than a test does not force a particular α on the consumer of the report, and allows the relative significance of different regions to be easily seen. A *p*-value is often preferable to a statistic, since the consumer is not left to account for the null distribution. Adjusted *p*-values are preferable to unadjusted *p*-values since the consumer is not left to account for the multiple comparisons problem.

Subset pivotality

The distribution of the vector $\boldsymbol{P}$ has the *subset pivotality* property if the joint distribution of the subvector $(P_k; k \in U)$ is identical under the restrictions H_U and H_W for all subsets $U \subset W$ with H_U true. (Condition 2.1 in Westfall & Young, 1993.)

For statistic images from activation experiments, this implies that the *p*-values computed for non-activated regions are unaffected by the presence or absence of activation elsewhere.

3.1.2. Random fields

Definitions

A *random field* $X(\boldsymbol{x})$, is a stochastic process whose "time" parameter, $\boldsymbol{x}$, ranges over some set more complicated than the real line. Usually this parameter space is a subset of Euclidean *D*-space, $\Re^D$. If every possible realisation $X(\boldsymbol{x})$ of the field is a continuous function of $\boldsymbol{x}$ for all points of the parameter space, then we shall refer to the field as a *continuous random field*. If the parameter space is some finite (or countably

infinite) subset of $\Re^D$, then we shall refer to the field as a *discrete random field*. We shall only consider real valued univariate random fields, where $X(\boldsymbol{x}) \in \Re$. Our parameter space will be the image space when considering continuous fields, and the set of voxel centres $\{\boldsymbol{x}_k\}_{k \in W}$ when considering images as discrete fields. For rigorous definitions, see Adler (1981).

The *finite-dimensional distributions* of a field are the joint distributions of $X(\boldsymbol{x})$ at any finite set of points $\boldsymbol{x}$. A random field is *weakly stationary* (or *weakly homogeneous*) if the marginal distributions of the field at any two points are identical. If the covariance between the values of the field at any two points is a function only of the displacement between the two points, then this function is the *covariance function* or *auto-covariance function*. A random field is *strictly stationary* (or *strictly homogeneous*) if the finite-dimensional distributions are identical for all sets of points related by a common displacement: That is, $\Pr\{X(\boldsymbol{x}_1) \le c_1, \ldots, X(\boldsymbol{x}_k) \le c_k\} = \Pr\{X(\boldsymbol{x}_1 + \boldsymbol{x}) \le c_1, \ldots, X(\boldsymbol{x}_k + \boldsymbol{x}) \le c_k\}$ for any set of points $\boldsymbol{x}, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ and any set of real numbers. If the finite-dimensional distribution of the values of the field at any two points is a function only of the distance between the two points, then the field is *isotropic*. Two fields are *identical* random fields if the finite-dimensional distributions are identical for both fields.

Gaussian Fields

A random field $X(\boldsymbol{x})$, is a *Gaussian random field* if the finite-dimensional distributions are multivariate Normal. Clearly the field is weakly stationary if the mean and variance are constant for all points $\boldsymbol{x}$. The field is strictly stationary if the covariance between the values of the field at any two points is a function of the displacement between the points, and isotropic if the covariance is only a function of the distance between the points.

χ^2, F and t fields

Adler (1981), defines χ^2, F and t fields in an analogous way to the univariate definitions. If $X(\boldsymbol{x}), X_1(\boldsymbol{x}), \ldots, X_n(\boldsymbol{x}), Y_1(\boldsymbol{x}), \ldots, Y_m(\boldsymbol{x})$ $\boldsymbol{x} \in \Psi \subset \Re^D$ are independent, identically distributed, (weakly) stationary, Gaussian random fields with zero mean and unit variance then:

$$U(\boldsymbol{x}) = \sum_{i=1}^{n} X_i(\boldsymbol{x})^2$$ is a χ^2 field with n degrees of freedom (defined only for $n > D$)

$$F(\boldsymbol{x}) = \frac{\sum_{i=1}^{n} X_i(\boldsymbol{x})^2 / n}{\sum_{i=1}^{m} Y_i(\boldsymbol{x})^2 / m}$$ is an F-field with n,m degrees of freedom, (defined only for $n + m > D$)

and $$T(\boldsymbol{x}) = \frac{X(\boldsymbol{x})}{\sqrt{\sum_{i=1}^{m} Y_i(\boldsymbol{x})^2 / m}}$$ is a t-field with m degrees of freedom.

That the fields have the implied marginal distributions is easily proved (Adler, 1981). If all the component Gaussian fields are strictly stationary or isotropic, then the χ^2, F and t fields are said to be strictly stationary or isotropic respectively.

Statistic images not discrete random fields

Recall that statistic images are said to be of a certain (univariate) distribution if the marginal distribution of the value at any voxel has that univariate distribution under the

omnibus null hypothesis. This reflects the voxel-by-voxel approach where univariate normality is assumed for each voxel rCBF value individually.

From the above definitions, it is clear that Gaussian, χ^2, F & t statistic images are not discrete random Gaussian, χ^2, F or t fields, unless the rCBF images are assumed to have multivariate normal distributions.

3.1.3. Simulated Gaussian statistic images

Various methods have been developed to address the multiple comparisons problem in this image setting. Until recently, these methods were restricted to statistics whose distribution was Gaussian under the null hypotheses, as these are more amenable to probabilistic analysis.

We shall illustrate the methods discussed in this chapter by applying them to a set of simulated standard Gaussian statistic images, the form of which we now describe.

Smoothing filter

Standard (zero mean, unit variance) Gaussian statistic images were simulated by smoothing a white noise field with a Gaussian kernel with variance-covariance matrix $\mathbf{\Sigma}$. (See appendix B:4, for a discussion of Gaussian kernels.) This implies that the Gaussian statistic image is a strictly stationary discrete Gaussian random field, with auto-correlation function $R(\boldsymbol{h}) = \exp(-\boldsymbol{h}^T(2\mathbf{\Sigma})^{-1}\boldsymbol{h}/2) / \sqrt{(2\pi)^D |2\mathbf{\Sigma}|}$ (appendix C:5), point response function equal to the filter kernel (centrally inverted) $f(\boldsymbol{x}) = \exp(-\boldsymbol{x}^T\mathbf{\Sigma}^{-1}\boldsymbol{x}/2) / \sqrt{(2\pi)^D |\mathbf{\Sigma}|}$, and variance-covariance matrix of partial derivatives $\mathbf{\Lambda} = (2\mathbf{\Sigma})^{-1}$ (appendix C:7). The Gaussian smoothing kernel chosen was spherical, with a FWHM of 10mm, so $\mathbf{\Sigma}$ is:

$$\mathbf{\Sigma} = \begin{pmatrix} 10^2 & 0 & 0 \\ 0 & 10^2 & 0 \\ 0 & 0 & 10^2 \end{pmatrix} \frac{1}{8\ln(2)}$$

Identification of intracerebral volume

A three-dimensional image space Ξ of dimensions 130mm×174mm×104mm was considered, partitioned into 65×87×26 voxels of 2mm×2mm×4mm. The Talairach co-ordinate system was adopted, graduated in millimetres. To mimic real statistic images only voxels W corresponding to the standard Talairach intracerebral volume were considered. These voxels were identified as follows.

Stereotactically normalised rCBF images from a number of subjects were obtained, resampled to the current Talairach voxellation. For each subject, the intracerebral voxels were identified as those where the rCBF was greater than a third of the maximum rCBF for that subject. A (conservative) estimate of the intracerebral voxels of the Talairach brain is then the intersection of the sets of intracerebral voxels for each subject, i.e., the voxels identified as intracerebral for all the subjects. $K = 72410$ voxels were identified as intracerebral voxels.

White noise images

White noise images were generated, assigning each intracerebral voxel an independent realisation from a standard Gaussian distribution. This white noise image was then smoothed and normalised.

Smoothing and normalisation

Since the white noise image is discrete, with voxel centres arranged in a regular lattice, the smoothing process can be implemented as a moving average filter (see

appendix B:2). The filter was computed with weights equal to the values of the filter kernel at the voxel centres. The white noise image was then smoothed with this moving average filter. To avoid edge effects, the moving average filter was truncated at the edge of the image (see appendix B:3).

The smoothed white noise image was then normalised so that marginally, each voxel was a standard univariate Gaussian variate. By the additivity property of independent Gaussian variates, the marginal distribution of any particular voxel after smoothing is Gaussian, with zero mean, and variance the sum of the squares of the smoothing weights of the moving average filter used at that voxel.[41] To obtain the simulated standard Gaussian statistic image each voxel value was divided by the square root of the variance. Thus the field is not strictly stationary near the boundaries of the intracerebral volume.

Example image

Fig.42 depicts orthogonal sections of a simulated Gaussian statistic image, generated according to the above prescription.

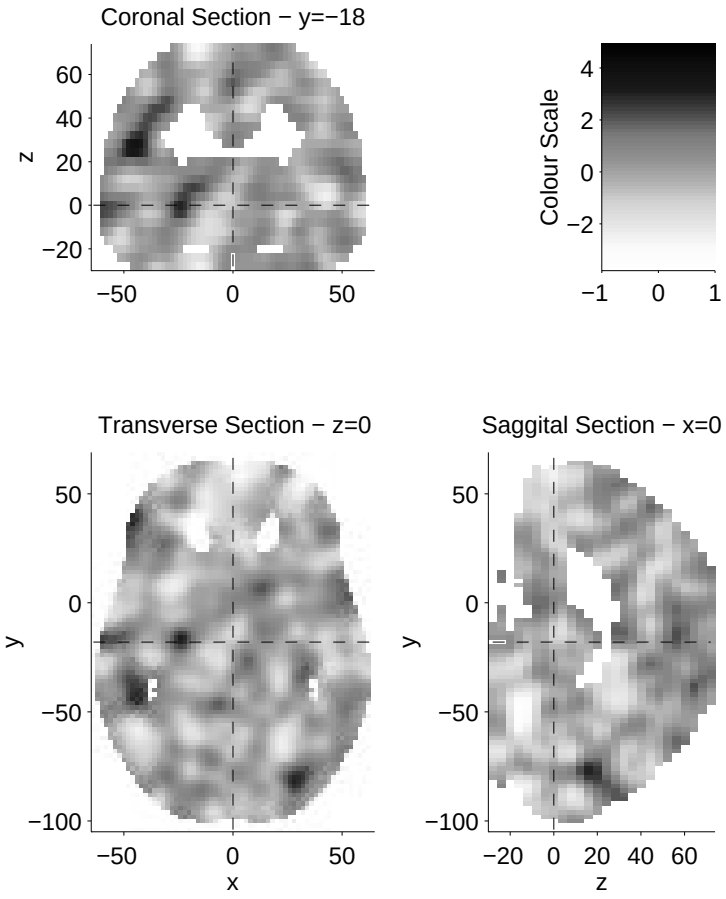

Figure 42
Orthogonal sections of a simulated standard Gaussian statistic image.

[41]Care must be taken to include only those weights used for voxels near the boundaries of the intracerebral volume where the filter was truncated. For voxels near the centre of the intracerebral volume, where little truncation takes place, the sum of squares of the weights was found to be very close to the theoretical variance of $1/(2^3\pi^{3/2}\sqrt{|\mathbf{\Sigma}|}$ derived in appendix C:5

3.1.4. Possible directions

If we assume multivariate normality for the rCBF images, with an appropriate model for the mean, then, as we saw in §2.5., multivariate procedures are precluded because the estimated variance-covariance matrix $\hat{\Sigma}$ is singular.

The rCBF images can be assumed to be strictly stationary, so that the auto-covariance function is a function only of the displacement between two voxels. From an estimate of the auto-covariance function, an inveritible estimate of Σ could be formed, and a multivariate analysis employed. Or, the rCBF images could be considered as lattice representations of continuous random fields with the same auto-covariance function, and assessed using results from continuous random field theory. The former approach has not been attempted as yet, since the latter has proved fruitful.

The simplest approaches consider the voxels individually using univariate methods, and take account of the multiplicity with a multiple comparisons procedure. These approaches only require an assumption of marginal univariate normality.

3.2. MCP Tests

3.2.1. Bonferroni

If $\{A_k\}_{k\in W}$ is any countable set of random events, then Bonferroni showed that:

$$1 - \sum_{k\in W} \Pr\left(\overline{A}_k\right) \leq \Pr\left(\bigcap_{k\in W} A_k\right) \tag{33}$$

commonly referred to as the *Bonferroni inequality*, though it is only one case of a set of inequalities. (See the *Encyclopædia of Statistical Science*, vol.1.)

Bonferroni correction for a family of tests

For a family of tests of null hypotheses H_k, let A_k be the event that the test correctly accepts H_k, by finding insufficient evidence against a true H_k at FWE level α. The right side of eqn.33 is then the probability that all tests correctly accept, given that the omnibus hypothesis H_W is true. This must be greater than $1-\alpha$ for a level α omnibus test. $\Pr(\overline{A}_k)$ is the level of the test of H_k. Setting this to α_k, then we see that weak control of familywise Type I error is maintained at level α if the α_k sum to less than α, since (from eqn.33):

$$\Pr(\text{“reject” } H_W \mid H_W) = 1 - \Pr\left(\bigcap_{k\in W} A_k\right) \leq \sum_{k\in W} \alpha_k \tag{34}$$

Strong control of FWE is also maintained by this choice of α_k, assuming subset pivotality: Subset pivotality states that the *p*-value for H_k is unaffected by the truth of $H_{k'}$ for $k\neq k'\in W$. Therefore for any set $U\subseteq W$, the probability of false rejection of H_U is at most $\sum_{k\in U} \alpha_k$ (by eqn.34), which is in turn at most $\sum_{k\in W} \alpha_k$ (since $\alpha_k \geq 0\ \forall\ k$), which is itself at most α by choice of the α_k.

Usually the α_k are chosen as α/K, where K is the number of tests (the cardinality of W), giving what's become known as the *Bonferroni correction*.[42] The level for each test is simply the chosen level of the omnibus test, divided by the number of tests. If the test statistics all have the same distribution under H_W, then the Bonferroni correction gives an identical critical value for each test. The test is then a *single threshold test*, the statistics for each test are thresholded at the critical value, and hypotheses with suprathreshold statistics are rejected. H_W is accepted only if no suprathreshold statistics are found.

Bonferroni single step adjusted p-values

If P_k is the *p*-value for H_k , then eqn.32 leads to *Bonferroni single step adjusted p-values* as:

$$\tilde{P}_k = \min\{K\times P_k, 1\} \tag{35}$$

---

[42]Prior knowledge regarding the expected location of an activation can be utilised by choosing the α_k such that they are larger in this region, resulting in a test with increased power for the region.

Conservative

For independent events the Bonferroni inequality is quite tight, even for large numbers of events, giving omnibus tests for families of independent tests with size approaching the level.

For many dependent events, the inequality can be extremely slack. Since the statistic images are smooth, tests at neighbouring voxels will be highly correlated (in terms of their *p*-values), the inequality (eqn.34) will be slack, and the omnibus test will be very conservative, with actual size much less than the specified level α.[43]

Example

For the simulated standard Gaussian statistic images, $K = 72410$. The Bonferroni correction tests each voxel hypothesis H_k: $\mu_k = 0$ against $\overline{H}_k$: $\mu_k > 0$ for $Z_k \sim N(\mu_k,1)$ at a level α/K, giving critical threshold $\Phi^{-1}(1-\alpha/K) = 4.8277$ (to 4dp) for $\alpha = 0.05$. This threshold is exceeded in only 179 of the 10^4 simulated statistic images, giving 95% confidence interval (CI) for the actual size of the omnibus test as (0.0157, 0.0201) (to 4dp, using the normal approximation to the binomial), well below the desired level of α.

3.2.2. Other MCP methods

The Bonferroni method is the simplest of the *single-step* or *simultaneous* multiple comparisons procedures, so called because all the hypotheses are tested simultaneously in a single step. Other simplistic methods ignoring the correlation between the tests, such as the Šidák method based on the Šidák-Jogdeo inequality, are also extremely conservative in the PET scenario. Improvements on the single-step methods are *multi-step* or *stepwise* methods which test the hypotheses in a certain order, usually the order of the (unadjusted) *p*-values.

For example, the simplest stepwise method is the step-down test of Holm (1979), which is based on the Bonferroni correction: Here the hypotheses are ordered according to their (unadjusted) *p*-values, from smallest to largest: $H_{(1)},\ldots,H_{(K)}$ with corresponding *p*-values $P_{(1)} \leq \ldots \leq P_{(K)}$. $H_{(1)}$ (and hence H_W) is rejected if $P_{(1)} < \alpha/K$. The test then steps down, rejecting $H_{(k)}$ if and only if all $H_{(k')}$ for $k' < k$ have already been rejected, and $P_{(k)} < \alpha/(K-k+1)$. Any untested hypotheses are accepted. The method controls FWE strongly. In essence, at each step a Bonferroni correction is used for the number of remaining hypotheses. This method is clearly more powerful for testing individual hypotheses than the Bonferroni method, but for testing the omnibus hypothesis H_W, the methods are identical.

Thus, step-down methods are as conservative as the multiple comparisons procedures on which they are based. For testing statistic images from PET, these simplistic multiple comparisons methods are very conservative, and have low power. Typically, they will only pick out large activations that are obvious to the naked eye.

---

[43]This is fairly intuitive in the current context of assessing images of statistics. At one extreme we have independent voxel values in the statistic image, and hence independent tests. In this case the Bonferroni inequality is quite tight. At the other extreme is the case where the voxel values are completely dependent. Then there is essentially only one test, and the Bonferroni correction tests this at level α/K, rather than at α as would be appropriate. The actual situation lies somewhere between these extremes.

3.3. Random Field Approaches

To obtain more sensitive tests, it is necessary to take the correlations between the voxel tests into account. That is, to account for the smoothness in the statistic images. Friston *et al*. (1991d), and Worsley *et al*. (1992), independently sought to take account of this smoothness by considering the statistic image as a lattice representation of an underlying continuous random field, and applying results from random field theory.

Random field modelling

If we assume that the rCBF images are multivariate normal, then the image regression is a multivariate regression, the fitted parameters are multivariate normal, and the univariate statistic images computed are discrete random fields. Assume also that the rCBF images are strictly stationary, so that the statistic images are also strictly stationary under H_W with covariance between values at any two points a function of the displacement between the points. Then, if the voxels are small enough, the statistic image may be approximated by a homogeneous continuous random field $Z(\boldsymbol{x})$, with the same (null) marginal distribution, covariance function and expectation, defined for $\boldsymbol{x} \in \Omega \subset \Re^3$. Here Ω is the subset of $\Re^3$ covered by the voxels in W.

3.3.1. Worsley's Euler characteristic method

Worsley *et al*. (1992), applied the results of Adler (1981) and co-workers on the theory of upcrossings in random fields, to obtain a threshold which is exceeded by a homogeneous continuous Gaussian random field with probability approximately α. This threshold is then applied to the statistic image, whose maximum value is assumed to be distributed similarly to the maximum of the continuous field. Voxels with suprathreshold statistics have their null hypotheses rejected.

Overview of the theory

Briefly, the excursion set of $Z(\boldsymbol{x})$ over a compact subset Ω of $\Re^D$, above a threshold u, $A_u(Z,\Omega) = \{\boldsymbol{x} \in \Omega : Z(\boldsymbol{x}) \geq u\}$, is characterised by the Euler (or Euler-Poincaré) characteristic $\chi(A_u(Z,\Omega))$. This essentially measures the number of isolated parts of the excursion set, less the number of holes (see Adler, 1981, §4.4, p90 [44]). For a three dimensional homogeneous Gaussian field with zero mean and unit variance, the expectation of the Euler characteristic, originally due to Adler and Hasofer (1976), is given by Adler (1981, Th.5.3.1, p111) as:

$$E[\chi(A_u(Z,\Omega))] = \lambda(\Omega)|\Lambda|^{1/2}(2\pi)^{-2}(u^2 - 1)\exp(-u^2/2) \qquad (36)$$

Here $\lambda(\Omega)$ is the Lebesgue measure of $\Omega \subset \Re^3$, the volume of the region Ω. Ω is assumed to be a compact, convex subset of $\Re^3$, whose boundary has zero Lebesgue measure. $|\Lambda|$ is the determinant of the (3×3) variance-covariance matrix of partial derivatives of the random field, with respect to the co-ordinate directions (eqn.38). This parameterises the smoothness of the field.

Adler showed that as the threshold u increases, the holes tend to disappear and the Euler characteristic tends towards the number of local maxima. For large u, near the global maxima Z_{max}, the Euler characteristic is 0 if $Z_{max} < u$ and 1 if $Z_{max} > u$. For high thresholds; $Pr(Z_{max} > u) \approx Pr(\chi(A_u(Z,\Omega)) > 1) \approx E[\chi(A_u(Z,\Omega))]$ and the expected Euler

---

[44] Adler defines the DT (*differential topology*) characteristic, which is shown to be the Euler (or Euler-Poincaré) characteristic provided that the excursion set does not touch the boundary of the region Ω.

characteristic approximates the *p*-value for Z_{max}. Setting the right side of eqn.37 to α and solving for u gives a critical threshold u_α suitable for a level α test for locating regions of Ω where the field departs from zero expectation.

This threshold is applied to the (standard Gaussian) statistic image, giving a test with (approximate) control of FWE in the weak sense, assuming the maxima of the continuous field and the discrete statistic image are similarly distributed. Assuming subset pivotality, it is easily seen that strong control over FWE is maintained, since subsets of W correspond to regions with volume less than that of Ω.

D-dimensions

For a *D*-dimensional homogeneous continuous Gaussian random field with zero mean and variance σ^2, defined on Ω, a compact, convex subset of $\Re^D$ whose boundary has zero Lebesgue measure, Adler (1976) derived the expected Euler characteristic for excursion sets above a threshold u, given by Adler (1981, Th.5.3.1, p111) as:

$$E[\chi(A_u(Z,\Omega))] = \lambda(\Omega)|\Lambda|^{1/2}(2\pi)^{-\frac{(D+1)}{2}}\sigma^{-(2D-1)}\exp\left(-\frac{u^2}{2\sigma^2}\right)P_D(u) \qquad (37)$$

$$\text{where } P_D(u) = \sum_{j=0}^{\frac{D-1}{2}}(-1)^j\frac{(2j)!}{j!2^j}\binom{D-1}{2j}\sigma^{2j}u^{D-1-2j}$$

Thus, the approach of Worsley *et al*. (1992) can be applied to Gaussian statistic images of any dimension. Λ here is the $D\times D$ variance-covariance matrix of partial derivatives of the field with respect to the co-ordinate directions:

$$\Lambda = \underbrace{\begin{pmatrix} \text{var}\left[\partial Y/\partial x_1\right] & \text{cov}\left[\partial Y/\partial x_1, \partial Y/\partial x_2\right] & \cdots \\ \text{cov}\left[\partial Y/\partial x_1, \partial Y/\partial x_2\right] & \text{var}\left[\partial Y/\partial x_2\right] & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}}_{D\times D} \qquad (38)$$

where $\boldsymbol{x} = (x_1, x_2, \ldots, x_D)$

Results for χ^2, F and t fields

Recent work has extended these results to χ^2, *F* and *t* fields, removing the requirement for Gaussian statistic images (Worsley, 1994a). These are presented in appendix D.

Example

The simulated Gaussian statistic images described in §3.1.3. are (by construction) strictly stationary discrete three-dimensional Gaussian random fields, with variance-covariance matrix of partial derivatives $\Lambda=(2\Sigma)^{-1}$. Σ is the variance-covariance matrix of the Gaussian smoothing kernel used. The $K = 72410$ intracerebral voxels represent a volume of $1158560mm^3$. Substituting these values for Λ and $\lambda(\Omega)$ respectively in eqn.36 we find that the expected Euler characteristic is $\alpha = 0.05$ at critical threshold $u_\alpha = 4.6784$. Of the 10^4 simulated statistic images, 329 have maxima exceeding this value. A 95% CI for the FWE is (0.0300, 0.03583) to 4dp, computed using the normal

approximation to the binomial. The test is conservative, with FWE rate below the desired level α.

Estimates $\hat{\boldsymbol{\Lambda}}$ of the variance-covariance matrix of partial derivatives $\boldsymbol{\Lambda}$ can be obtained by computing numerical derivatives in the simulated statistic images and then taking their variances and covariances across the statistic image. (Estimation of smoothing is discussed fully below, in §3.3.5.) For each of the simulated images the smoothness was estimated. Converting each $\hat{\boldsymbol{\Lambda}}$ to the estimated variance-covariance matrix, $\hat{\boldsymbol{\Sigma}}$, of the point response function (assumed Gaussian in shape) gives $\hat{\boldsymbol{\Sigma}} = (2\hat{\boldsymbol{\Lambda}})^{-1}$ (appendix C:7). Expressing that in terms of FWHM in the axial directions (see appendix B:4), we find than the average estimated FWHM is 10.4mm×10.4mm×10.8mm. Increases in smoothness lower the value of the expected Euler characteristic, since $|\boldsymbol{\Lambda}|$ decreases. So, if the critical threshold is calculated for each simulated image using an estimate of $\boldsymbol{\Lambda}$ derived from that image, the test may be expected to be slightly less conservative. The critical threshold appropriate for a smoothness of 10.4mm×10.4mm×10.8mm is $u_\alpha = 4.6415$, a value exceeded in only 366 of the simulated images. A 95% CI for the FWE rate of a test on the simulated data with this threshold is (0.0335, 0.03969).

The empirical distribution function (EDF) of the maxima of the simulated Gaussian statistic images is presented in fig.43. The approximate cumulative distribution function (CDF) of the maxima by Worsley's method is the compliment of the expected Euler characteristic, $1-E[\chi(A_c(Z,\Omega))]$. This is superimposed for the theoretical and average estimated smoothness. The top 10% of the EDF is also shown, with a pointwise 95% confidence band for the true CDF, computed using the normal approximation to the binomial.

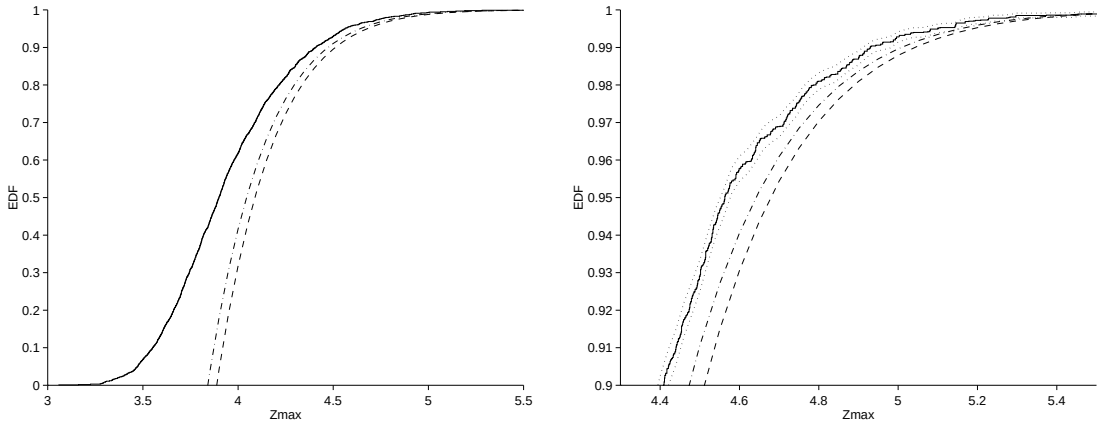

Figure 43

EDF for the maxima of the simulated Gaussian statistic images, from 10^4 simulations. Superimposed is the compliment of (one minus) the expected Euler characteristic, which approximates the CDF at high values, computed with both the theoretical smoothness of 10mm FWHM (dashed line) and the mean estimated smoothness of 10.4mm×10.4mm×10.8mm (dot-dash line). The top 10% of the EDF plot on the left is given on the right, with a pointwise 95% confidence band for the true CDF.

3.3.2. Friston's "Bonferroni" method

Brief overview of the theory

Friston *et al*. (1991d), worked with two dimensional slices of three dimensional standard Gaussian statistic images. They assumed that these planar statistic images could be approximated by the field formed by convolving a two dimensional white noise Gaussian field, with a bivariate Gaussian kernel, with variance-covariance matrix $\mathbf{\Sigma} = s^2 \mathbf{I}_2$ (for I_2 the 2×2 identity matrix). Using the one dimensional theory of level crossings in stochastic processes, they obtained the approximate probability that the centre of an elliptical region above a threshold c, occurs in the area corresponding to a single pixel, as:

$$P = \frac{1}{32\pi\,(s/h)^2 \exp(c^2)\,(1-\Phi(c))} \tag{39}$$

Here h is the length of the side of the (square) pixel.[45] The event that a pixel with H_k true is the centre of an elliptical suprathreshold area is then taken to be the false positive event, a rather heuristic arrangement. A simultaneous test for all pixel hypotheses is obtained by using a Bonferroni correction for the number of pixels in the planar image. The threshold c_α is found by equating the above expression to α/K_P. K_P is the number of intracerebral pixels in the plane under consideration. Again, the smoothness, s, is replaced with an estimate. The test was applied plane by plane without consideration for the multiplicity of testing many planes. Results were reported with a caveat like "a false positive was expected once in every 20 planes for tests at significance level $\alpha = 0.05$".

3.3.3. Transform functions

To assess the significance of χ^2, F and t statistic images, Friston *et al*. (1991) advocated transforming the statistic image to an equivalent Gaussian statistic image. This is achieved by replacing each voxel statistic with a standard Gaussian ordinate with identical probability of being exceeded: If x is drawn from a distribution with Cumulative Distribution Function $F(x)$, then the equivalent standard Gaussian variate is $z = \Phi^{-1}(F(x))$. $\Phi^{-1}(F(\bullet))$ is thus a function "transforming" a random variable from one distribution to a standard Gaussian distribution, and has become known (in PET) as a *transform function*. (See appendix E for details.) We shall refer to such transformed statistic images as *Gaussianised* χ^2, F or t statistic images.

The transform function can be modified so that two-sided hypotheses can be considered. This is necessary since we only have theory for one sided tests of (strictly stationary continuous) Gaussian random fields. A Gaussianised t-statistic for a two-sided hypothesis is the standard Gaussian ordinate z for which the probability of exceeding, is equal to the probability with which the t-statistic x is exceeded *in absolute value*, by a random variable of the specified null t-distribution. I.e. $z = \Phi^{-1}(1-2(1-F_T(|x|)))$, where $F_T(\bullet)$ is the cumulative distribution function of Student's t-distribution (with the appropriate degrees of freedom). This is equivalent to squaring the t-statistic, and "transforming" the resulting F-statistic.

Although the resulting Gaussianised statistic image is Gaussian at every voxel, it is not a discrete Gaussian field, since the condition of multivariate normal finite-

---

[45] Friston works in units of pixels, so in Friston *et al.* (1991d) eqn.39 appears with h as one. For generality, we shall choose to work in the units of the image space (which may be chosen to be pixel units).

dimensional distributions is not met. (Neither is the auto-covariance function necessarily Gaussian in shape, as is often assumed.) This is so even if the original statistic image *is* a discrete random field. The consequence of this departure from assumptions is demonstrated by Worsley (1994b) for *t*-fields simulated over a 1000cc volume, with smoothness of 20mm FWHM. For a nominal 0.05 false positive rate over a 1000cc volume, Worsley's test on the Gaussianised *t*-field gave a false positive rate of 0.069 for *t*-fields of 40 degrees of freedom, and 0.055 for *t*-fields of 120 degrees of freedom.

Considering *t*-statistics, the transform function tends (pointwise) to the identity function as the degrees of freedom tend to infinity. Thus, for *t*-statistic images that *are* discrete random *t* fields, the resulting Gaussian statistic image can be approximated by a discrete Gaussian field if the degrees of freedom are high. Worsley (1994b) suggests that the degrees of freedom should be greater than 120.

3.3.4. Comparison of Friston's and Worsley's methods

The unit variance field obtained by convolving a white noise Gaussian field with a Gaussian kernel with variance-covariance matrix $\mathbf{\Sigma}$, is a strictly stationary continuous Gaussian random field with $\mathbf{\Lambda} = (2\mathbf{\Sigma})^{-1}$ (Adler, 1981)[46]. Thus, Friston *et al*. (1991d) are considering an isotropic strictly stationary continuous Gaussian random field, with Gaussian covariance function, and $\mathbf{\Lambda} = (2s)^{-2}\,\mathbf{I}_2$. From eqn.39 Friston's *p*-value for the maxima in a plane, Z^{p}_{max}, after Bonferroni correction is:

$$\Pr(Z^{p}_{max} > c) = \frac{K_p \times pixsize^2}{32\pi\, s^2 \exp(c^2)\,(1-\Phi(c))} = \frac{\pi\,\lambda(\Omega_p)\,|\mathbf{\Lambda}|^{1/2}\,(2\pi)^{-2}}{4\exp(c^2)\,(1-\Phi(c))} \qquad (40)$$

Here Ω_p is the intracerebral area in the plane, and K_p the number of voxels in the plane. Using the approximation $\Phi(-c)=(1-\Phi(c))\approx(2\pi)^{-1/2}\exp(-c^2/2)/c$ for large c, Worsley (1992) showed that this *p*-value was a factor of $\pi/4$ smaller than the corresponding expected Euler characteristic.

The result of Worsley *et al*. (1992) is more rigorous than that of Friston *et al*. (1991d), is not restricted to two dimensions, isotropy or a Gaussian auto-correlation function. However, for 2D work, both give approximately valid tests, as we shall now see.

Example

To illustrate these two-dimensional methods, consider only the AC-PC plane of the simulated Gaussian statistic images. The $K_{p8} = 4079$ intracerebral voxels in this plane constitute a 4mm thick slice of the image space, with faces parallel to the X-Y plane of area 16316mm^2. Ignoring the third dimension, the expected Euler characteristic for a two-dimensional strictly stationary continuous standard Gaussian random field (eqn.37) with this area and FWHM of 10mm is equal to $\alpha = 0.05$ at critical threshold $u_\alpha = 3.9299$ (to 4dp). This value is exceeded by the maximum statistic in the AC-PC plane of 436 of the 10^4 statistic images, giving a 95% CI for the FWE of (0.0402,0.0470) to 4dp, computed using the normal approximation to the binomial. The critical threshold for a level $\alpha = 0.05$ test by Friston's method (eqn.40) is slightly lower, at $c_\alpha = 3.8796$ (to 4dp). This level is exceeded by the maximum statistic in the AC-PC plane of 526 of the 10^4 statistic images, giving (0.0489,0.0563) as a 95% CI for the FWE.

If the mean estimated in-plane smoothness of 10.4mm FWHM is used, then the critical value for Worsley's test drops to $u_\alpha = 3.9085$ (4dp), a value exceeded in 501 of

[46]See appendix C for a summary of results.

the 10^4 simulations, giving (0.0465, 0.0537) as a 95% CI for the FWE. Friston's method gives a critical threshold of $c_\alpha = 3.8580$ (4dp), exceeded in 560 simulated images, giving 95% CI for the FWE as (0.0522, 0.0598).

The top 10% of the EDF of the maximum voxel value in this plane for the 10^4 simulations is presented in fig.44, together with a pointwise 95% confidence band for the true CDF, computed using the normal approximation to the binomial. Superimposed are the theoretical CDFs of Friston *et al.* (1991d) and Worsley *et al.* (1992). The agreement between the EDF and the theoretical CDFs is remarkable. For critical thresholds c in the range depicted, [3.5,5], the ratio of Friston's p-value (eqn.40) to the expected Euler characteristic (eqn.37) is close to the value of $\pi/4$ computed by Worsley *et al.* (1992) .

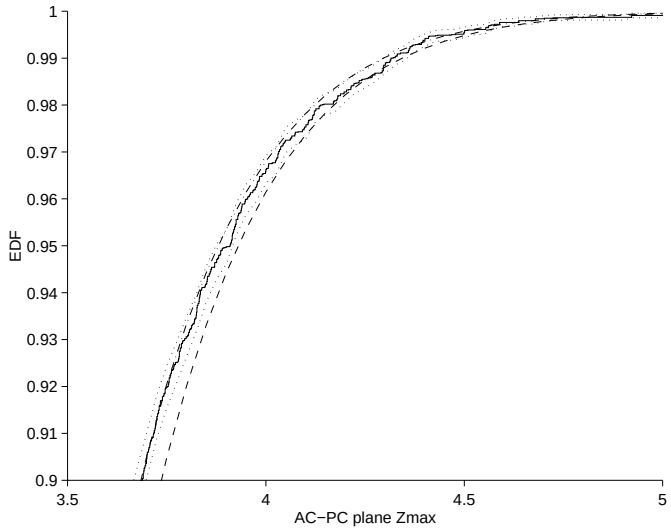

Figure 44

Top 10% of the EDF of the maxima in the AC-PC plane of the simulated Gaussian statistic images, from 10^4 simulations, with pointwise 95% confidence band (dotted lines) for the true CDF, computed using the normal approximation to the binomial. Superimposed is the compliment of (one minus) the expected Euler characteristic (dot-dash line) for a 2D field of this area (eqn.37). Also superimposed (dashed line) is the CDF as predicted by Friston *et al.* (1991d) (1-eqn.40). Both theoretical CDFs were computed with the theoretical smoothness corresponding to a Gaussian point response function of 10mm FWHM.

3.3.5. Estimation and specification of smoothness

To apply these random field methods, an estimate of $|\Lambda|^{1/2}$ is required, necessitating the estimation of Λ, the variance-covariance matrix of the partial derivatives of the field, or of the component fields in the case of χ^2, F or t fields.

Within image estimation (Friston)

The most common approach, utilised by Friston *et al.* (1991d), is to estimate smoothness within the statistic image. Numerical partial derivatives are computed for each voxel location, and their variances and covariances taken across all voxels.

For strictly stationary χ^2, F or t fields, an estimate of the variance-covariance matrix of partial derivatives of the component fields must be derived from the (estimated) variance-covariance matrix of partial derivatives of the statistic image. Adler (1981, p.169) notes that for a chi-squared field, the variance of the partial

derivatives relative to the variance of the field is twice that of its component fields. Worsley *et al*. (1992) derive a relationship between the variance-covariance matrix of partial derivatives of a *t*-field and its Gaussianised counterpart. A simplification of their argument yields the relationship between the variance-covariance matrix of partial derivatives of a *t*-field and its component field, and is given in appendix G. Further relationships can be derived from Worsley (1994a).

Strong control?

Departures from the null hypothesis due to activation (or de-activation) will appear as increases (decreases) in the statistic image, so the (numerical) partial derivatives of the statistic image will be increased in the locality of the activation. The variances and covariances of the partial derivatives will therefore be greater than were there no activation present, and smoothness will be underestimated. Subset pivotality is not maintained, even if the distribution of a non-activated part of the statistic image is identical under the omnibus hypothesis and in the presence of activation elsewhere: The *p*-values for voxels k with H_k true will be greater in the presence of an activation than under the omnibus hypothesis H_W. Thus the test still has strong control over FWE, but becomes more conservative, and less powerful. Usually activations are very localised, or perhaps diffuse but of low magnitude, so the effect on the estimated smoothness is negligible.

Across image estimation from component fields (Worsley)

A more rigorous approach was advocated by Worsley *et al*., (1992). Λ is estimated for each voxel across the component images from which the statistic image is formed, accounting for any change in mean due to activation (de-activation). The variances and covariances computed for each voxel are then pooled over all voxels to obtain an estimate of the variance-covariance matrix of partial derivatives of the component fields, assumed strictly stationary. See Worsley *et al*. (1992), pp907–909 for details, where Λ is estimated for a Gaussian field of paired *t*-statistics computed with pooled variance estimate.

Resels

Worsley *et al*. (1992) introduced the concept of *resels*, or *resolution elements*, a measure of the number of independent measurements in a strictly stationary standard Gaussian random field with Gaussian point response function (PRF). (The PRF for pre-processed PET rCBF (rA) images is reasonably approximated by a Gaussian density function, particularly if primary smoothing with a Gaussian kernel is performed!) The size of the PRF is usually specified in terms of FWHM in the axial dimensions, FWHM_X, FWHM_Y & FWHM_Z respectively, and the number of resels R in a volume Ω was defined as the volume divided by the product of the FWHM:

$$R = \lambda(\Omega) \;/\; (\text{FWHM}_X \times \text{FWHM}_Y \times \text{FWHM}_Z)$$

This is related to the variance-covariance matrix of partial derivatives of the field as follows: If the PRF of the field is Gaussian, with variance-covariance matrix Σ. Then (appendix B:4) Σ is given in terms of the FWHM as:

$$\Sigma = \begin{pmatrix} \text{FWHM}_X^2 & 0 & 0 \\ 0 & \text{FWHM}_Y^2 & 0 \\ 0 & 0 & \text{FWHM}_Z^2 \end{pmatrix} \frac{1}{8\ln(2)}$$

The field can be generated by convolving a weakly stationary continuous white noise Gaussian random field (of appropriate variance) with a Gaussian kernel with

variance-covariance matrix $\boldsymbol{\Sigma}$.[47] (This in turn implies, and is implied by, the fact that the auto-correlation function of the strictly stationary standard Gaussian random field is a Gaussian kernel with variance-covariance $2\boldsymbol{\Sigma}$ (appendix C:5). Thus the assumption of Gaussian PRF is equivalent to one of Gaussian auto-correlation.)

The variance-covariance matrix of partial derivatives of a strictly stationary continuous standard Gaussian random field, formed by convolving a weakly stationary continuous white noise Gaussian random field with a Gaussian kernel with variance-covariance matrix $\boldsymbol{\Sigma}$, is $\boldsymbol{\Lambda} = (2\boldsymbol{\Sigma})^{-1}$ (see appendix C:7). Therefore:

$$\boldsymbol{\Lambda} = \begin{pmatrix} 1/\text{FWHM}_X{}^2 & 0 & 0 \\ 0 & 1/\text{FWHM}_Y{}^2 & 0 \\ 0 & 0 & 1/\text{FWHM}_Z{}^2 \end{pmatrix} 4\ln(2)$$

and

$$|\boldsymbol{\Lambda}|^{1/2} = (4\ln(2))^{D/2} \ / \ (\text{FWHM}_X \times \text{FWHM}_Y \times \text{FWHM}_Z)$$

for $D = 3$ dimensions, with the obvious modification for $D = 2$ dimensions. Thus $(4\ln(2))^{D/2} R = \lambda(\Omega) |\boldsymbol{\Lambda}|^{1/2}$.

Smoothness in terms of FWHM

As just discussed, if the point response function of a strictly stationary continuous standard Gaussian random field is a Gaussian kernel, with variance-covariance matrix $\boldsymbol{\Sigma}$, then $\boldsymbol{\Lambda}$, the variance-covariance matrix of partial derivatives of the field, is given by $\boldsymbol{\Lambda} = (2\boldsymbol{\Sigma})^{-1}$. This result is often used in reverse to specify smoothness in terms of the FWHM of an assumed Gaussian point response function, as follows:

Suppose a strictly stationary continuous standard Gaussian random field has estimated variance-covariance matrix of partial derivatives $\hat{\boldsymbol{\Lambda}}$. Assuming the field has a Gaussian PRF with variance-covariance matrix $\boldsymbol{\Sigma}$, this is estimated by $\hat{\boldsymbol{\Sigma}} = (2\hat{\boldsymbol{\Lambda}})^{-1}$. (Recall this is equivalent to assuming that the field has Gaussian auto-correlation function with variance-covariance matrix $2\boldsymbol{\Sigma}$.) It is common to consider only the variances of the numerical partial derivatives and assume that the covariances are zero. (Equivalent to assuming that the PRF is ellipsoidal, with axes parallel to the co-ordinate axes.) Thus $\hat{\boldsymbol{\Sigma}}$ will be a diagonal $D \times D$ matrix of variances in the axial directions. The smoothness is then specified as the product of the FWHM in the D axial directions, $\text{FWHM}_X \times \text{FWHM}_Y \times \text{FWHM}_Z$. FWHM is derived from a variance σ^2 by $\text{FWHM} = \sigma\sqrt{8\ln(2)}$. (See appendix B:4 for a review of Gaussian kernels and FWHM.) If the FWHM is the same in each of the axial directions under consideration, then a single FWHM is stated.

---

[47] Convolving a delta function with the filter kernel gives the point response function. The convolution also results in a central inversion of the filter kernel. Thus the point response function is equal to the filter kernel, centrally inverted. In this case the filter kernel is symmetric.

3.3.6. Discussion of random field approaches

There are many assumptions in the use of the random field method of Worsley *et al*. (1992), both implicit and explicit. Difficulties with this approach stem from the assumptions, so we shall begin by stating the assumptions, and then discuss some of their infringements and implications.

3.3.6.1. Assumptions

The key assumption is:

- Under the null hypothesis, the maximum voxel statistic in the standard Gaussian (χ^2, *F* or *t*) statistic image, is distributed approximately the same as the maxima of a strictly stationary continuous standard Gaussian (χ^2, *F* or *t*) random field, with (component fields with) variance-covariance matrix of partial derivatives Λ, estimated from the (component fields of the) statistic image.

This implies:

- † Under the null hypothesis, the standard Gaussian (χ^2, *F* or *t*) statistic image is a realisation of a strictly stationary, discrete, standard Gaussian (χ^2, *F* or *t*) random field.
- The voxel dimensions are small relative to the smoothness of the field (measured as the FWHM of the PRF). This is so that the statistic image can be considered as a good lattice representation of a continuous random field, where the lattice is the set of voxel centres.
- The dimensions of the image are much greater than the smoothness of the field (measured as the FWHM of the PRF). This is so that the variance-covariance matrix of partial derivatives of the statistic image can be well estimated by numerical derivatives.

Considering the formation of the statistic image, the assumption that the statistic image is a strictly stationary discrete random field (†), implies:

- The rCBF (rA) images themselves are discrete Gaussian random fields, with mean given by the assumed model (see ch.2), and variance image $\boldsymbol{v} = (\sigma_1^2,\ldots,\sigma_K^2)$.
- The error images $\boldsymbol{\varepsilon}_{ijq} = (\varepsilon_{ijq1},\ldots,\varepsilon_{ijqK})$ (formed from the rCBF (rA) images by subtracting the mean from the model, see ch.2) on normalisation to unit variance by division of each voxel value by the (unknown) true standard deviation σ_k, are strictly stationary discrete standard (zero mean and unit variance) Gaussian random fields.

3.3.6.2. Normality

Univariate normality

That rA images have normally distributed voxel values under repeat scans on the same subject under the identical conditions would appear to be reasonable. Filtered backprojection gives reconstructed rA images formed by multiple convolutions of the Poisson VOR counts data, which themselves have high values. Consider the reconstruction of rA value in a single voxel over repeat scans on the same individual under *identical* conditions (identical blood tracer activity as a function of time, identical scan protocol and rCBF). An appeal to the Central Limit Theorem would suggest normality for the distribution of the particular voxel value, since the number of VOR contributing to the reconstruction of a rA over a single voxel is great. Reconstruction and primary smoothing would reinforce this assertion. Since calibration of rA to rCBF is almost linear, (see ch.1.) rCBF at a voxel (over repeat scans under identical conditions) may also be assumed normal.

Univariate normality: assessment using Q-Q correlation images

Various authors have illustrated the veracity of the univariate normality assumption for each voxel, by imaging the correlation coefficient of a Q-Q plot for each voxel (as we did for the "V5" data in §2.6.1.). As usual, there is the problem of multiplicity, and the low power of tests due to the small numbers of replications. Departures from normality in the extreme tails of the distribution are virtually undetectable with small sample sizes. A general rule of thumb is that more than 50 observations are required to test for normality.

Thus, the results of such an assessment should be interpreted with caution.

Multivariate normality

An assumption of multivariate normality for reconstructed rA images from scans on the same individual under identical conditions is not so convincing. Although the number of VOR contributing to the reconstruction of a rA over a single voxel is great, so is the number of voxels. When considering the reconstruction of an entire image, the reduction in dimensionality from the number of VOR to the number of voxels may be insufficient for the asymptotic result of the (multivariate) central limit theorem to apply.

Again, primary smoothing would make the assumption less contentious.

Departures from Normality

It should be noted that it is in the extreme tails of the distribution that the assumption of normality is critical. It is here that departures from normality will greatly affect the level of the tests.

3.3.6.3. Model assumptions

Adequacy of model fit was discussed in chapter 2. Henceforth we shall assume that the model fits, and thus that the error images $\boldsymbol{\varepsilon}_{ijq}$ have zero mean.

3.3.6.4. Strict stationarity of statistic images

That the error images ($\boldsymbol{\varepsilon}_{ijq}$) are drawn from identical multivariate normal distributions seems unlikely, since the pattern of inter-regional correlations of rCBF in different individuals is likely to be different.

Biological considerations aside, it still seems unlikely that the (normalised) error images are strictly stationary. Current work by JB Poline indicates that reconstructed rA

images are smoother towards the centre of the tomograph image space, in that the variances of the partial derivatives are smaller. Primary smoothing of reconstructed images evens out small regional differences in the auto-correlation function, and results in error images whose local auto-correlation structure is similar to the filter kernel used. The effect of departures from strict stationarity on the proposed tests remains to be investigated.

3.3.6.5. Low df, rough random fields, noisy statistic images

F and *t* statistic images are calculated using an estimate of the variance of the observations about the assumed model. Usually this variance is not assumed to be constant throughout the intracerebral volume, and is estimated separately for each voxel. The low numbers of degrees of freedom available for variance estimation, after fitting model parameters, lead to estimates of the voxel variances that have large variance. This variance manifests itself as high (spatial) frequency noise in images of the estimated standard deviation, even though the actual (population) standard deviations would most likely give a very smooth image. This noise is then propagated through to the statistic images, which are formed with the sample standard deviation in the denominator, giving statistic images with low smoothness.

These noisy statistic images are not well approximated by continuous random fields, since continuous fields with low smoothness will have features (such as peaks) smaller in spatial extent than the voxel dimensions. Worsley *et al.* (1993a) argue that a continuous three dimensional χ^2-field with three (or less) degrees of freedom almost certainly has a zero. This is addressed rigorously in Worsley (1994a), where it is demonstrated that a *D*-dimensional χ^2 field with degrees of freedom less than or equal to *D* almost surely has a zero. So, three dimensional *t*-fields on three or less degrees of freedom almost certainly have a singularity, as do *F*-fields with denominator degrees of freedom ≤ 3! For small degrees of freedom (greater than three), Worsley (1993b) obtained critical thresholds that were well in excess of those from a highly conservative Bonferroni correction. Current thinking is that χ^2, *F*, and *t* fields should only be considered as good lattice representations of random fields if the degrees of freedom are at least 24. For lower degrees of freedom the high thresholds derived lead to conservative tests.

These difficulties are overcome if the variance image is assumed constant, when the estimate of the common variance is regarded as exact. We have already discussed the dangers of assuming homoscedascity in §2.7. A more common approach is to smooth the statistic images.

3.3.6.6. Smoothing Statistic images

Theory

Consider a strictly stationary continuous Gaussian random field with zero mean, unit variance, variance-covariance matrix of partial derivatives $\mathbf{\Lambda}_Y$, and Gaussian auto-correlation function (equivalently Gaussian PRF). If this field is smoothed by convolution with a Gaussian kernel with variance-covariance matrix $\mathbf{\Sigma}$, then the resulting field is a strictly stationary continuous Gaussian random field with zero mean and variance $c = 1/\sqrt{\left|2\mathbf{\Lambda}_Y\mathbf{\Sigma} + \mathbf{I}_D\right|}$. On normalisation to unit variance by division by $\sqrt{c}$, the resulting

field has variance-covariance matrix of partial derivatives $\mathbf{\Lambda} = (2\mathbf{\Sigma} + \mathbf{\Lambda}_Y^{-1})^{-1}$. (A proof appears in appendix C:8.)

Practice

This result forms the basis for *statistic image smoothing*, sometimes called *secondary smoothing* in the PET functional mapping literature. Statistic images with non-Gaussian (null) distributions are first transformed to have univariate standard Gaussian marginal distributions. The resulting standard Gaussian statistic image is then assumed to be a strictly stationary Gaussian random field with zero mean, unit variance, and Gaussian auto-correlation function. The variance-covariance matrix of partial derivatives is estimated as $\hat{\mathbf{\Lambda}}_Y$, leading to $\hat{c}$ as an estimate of c. The image is smoothed with a moving average filter with weights given by a Gaussian kernel with variance-covariance matrix $\mathbf{\Sigma}$ (appendix B:2), and the resulting image scaled by a factor of $1/\sqrt{\hat{c}}$ to obtain the smoothed statistic image.

Pros

The smoothed Gaussian statistic image will have local auto-correlation structure dominated by the filter kernel. Thus, secondary smoothing leads to images which are closer to being strictly stationarity Gaussian random fields. In addition, the "smoothness" of the smoothed statistic image is much larger (in terms of FWHM) than the voxel size, so the image is more amenable to consideration as an approximation of a smooth continuous random field.

Smoothing trades spatial resolution for noise reduction, increasing the signal to noise ratio for signals greater in extent than the filter kernel. However, the effects are considerable, and different kernels can give radically different results. Thus smoothing is undesirable when an activation is expected to be fairly localised, or strong enough to be detectable without secondary smoothing.

Cons

The transforming of non-Gaussian statistic images is necessary because the normal family of distributions is the only family, of those under consideration, that is closed under addition. However, as we have already discussed, these Gaussianised statistic images are not Gaussian random fields, and therefore the theory is not directly applicable. (See §3.3.3. on "Transform functions", p.107.) Also, estimation of $\mathbf{\Lambda}_Y$ is likely to be poor since the statistic image is rough, leading to inappropriate estimation of the scaling factor c, giving smoothed statistic images with non-unit variance which may in turn lead to an increased risk of false-positives when analysed assuming unit variance.

Secondary smoothing is not (in general) equivalent to smoothing the scan data, and as such is difficult to interpret.

3.4. Omnibus Tests

Consider now omnibus tests, with only weak control over familywise Type I error.

3.4.1. Friston's exceedence proportion test

Exceedence proportion

Friston *et al*. (1990) proposed using as a test statistic summarising evidence against H_W, the proportion P^e of voxels exceeding a given threshold:

$$P^e = \sum_{k \in W} \{Z_k > F^{-1}(1-\eta)\} / K \qquad (41)$$

The logical expression in braces ("{•}") takes the value one if the expression is true and zero otherwise, following Knuth (1992). Here $F(\bullet)$ is the cumulative distribution function of the hypothesised null distribution at each voxel. Thus, η specifies the threshold as an *upper tail probability threshold*. Recall $W=\{1,\ldots,K\}$.

The null distribution given by Friston *et al*. (1990) is incorrect. They assumed that under H_W, P^eK had a Binomial $Bin(K, \eta)$ distribution. Whilst individually the events $Z_k > F^{-1}(1-\eta)$ under H_k are Bernoulli trials with "success" probability η, they are clearly not independent. The strong positive correlations between the trials at neighbouring voxels (due to the smoothness of the statistic images) results in the exceedence proportion having much greater variability than the simple binomial model would suggest. Thus, the binomial test is far too sensitive, with actual FWE far above the specified level α.

Example: Exceedence proportions for simulated images

For each of the 10^4 simulated Gaussian statistic images, the exceedence proportion P^e was computed for upper tail probability thresholds of $\eta = 0.05$, $\eta = 0.01$, and $\eta = 0.001$. The empirical distribution functions (EDFs) are displayed in fig.45. The cumulative distribution function of a $N(\eta, \eta(1-\eta)/K)$ distribution is superimposed, as an approximation to the distribution of P^e under the hypothesis $P^eK \sim Bin(K, \eta)$.[48] These plots clearly show the magnitude of the error in the binomial assumption of Friston *et al*. (1990)

The critical value for Friston's exceedence proportion test at level α, for a probability threshold η, is $\eta + \Phi^{-1}(1-\alpha)\times\sqrt{(\eta(1-\eta)/K)}$, for K and η such that the normal approximation is reasonable. For $\alpha = 0.05$; $\eta = 0.05$, 0.01, & 0.001 this gives critical exceedence proportions of 0.051332, 0.010608, & 0.0011193 respectively (6dp). From the simulated data, 95% confidence intervals for the actual size of the tests with these critical exceedence proportions are (0.3951,0.4113), (0.3492,0.3650), & (0.3032,0.3184) respectively (4dp), computed via the normal approximation to Binomial.

---

[48] The usual criteria for a $Bin(K, \eta)$ RV being well approximated by a $N(K\eta, K\eta(1-\eta))$ distribution, is that the expected value $K\eta$ is greater than 3 standard deviations from zero, i.e. $K > 9(1-\eta) / \eta \Leftrightarrow \eta > 9/(K+9)$. Here $K = 72410$, so reasonable approximations are afforded for $\eta > 0.00012$ (5dp). For η too small for a reasonable normal approximation, the PDF of $Bin(K, \eta)$ has most of its weight near zero, so direct computation of the CDF (and therefore of most percentiles) from the binomial probabilities in this limited range involves few terms, and is not prohibitive.

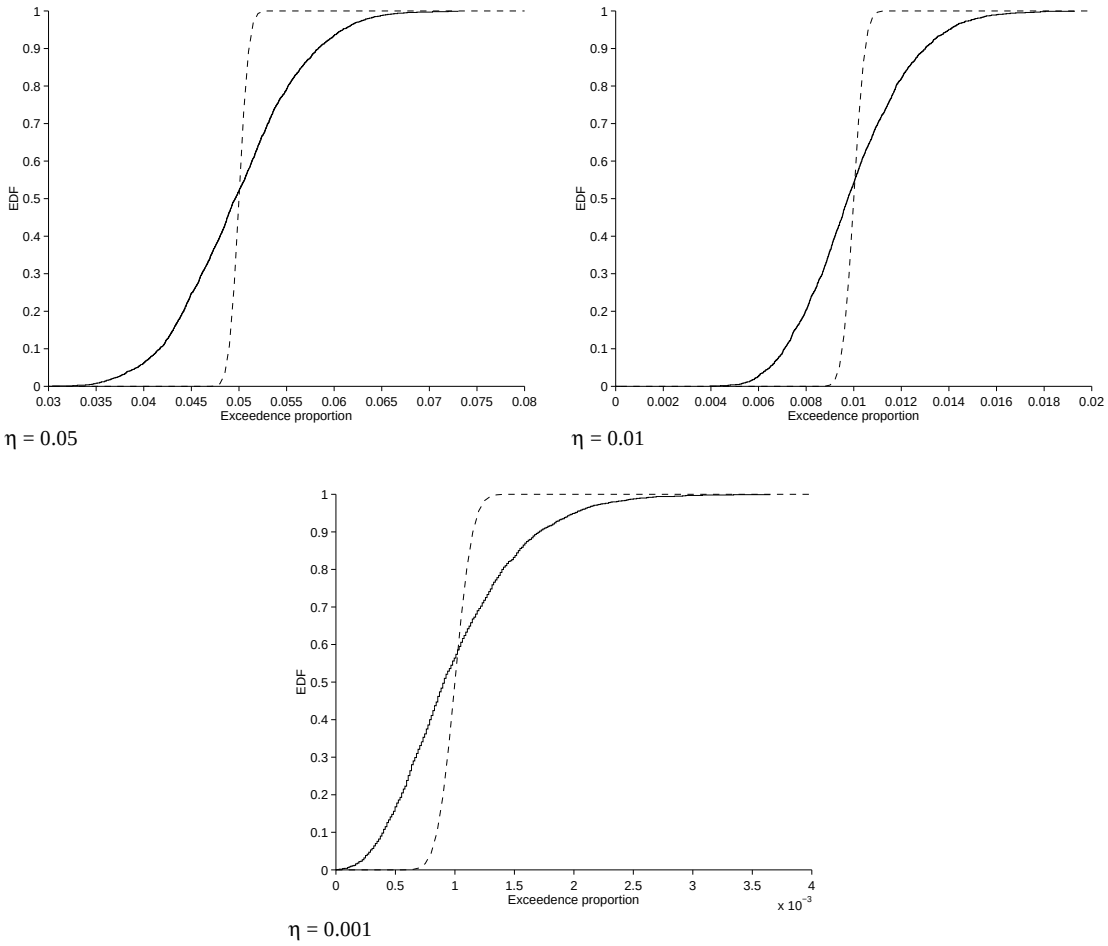

Figure 45
EDFs of the proportion of voxels with values exceeding the 95th, 99th & 99.9th percentiles of the Gaussian distribution, from 10^4 simulated Gaussian statistic images. Superimposed (dashed lines) are the CDFs predicted under the binomial hypothesis of Friston *et al.* (1990).

SPM implementation

Although flawed, this test is still widely used due to its inclusion in versions of the SPM package prior to "SPM94". The actual implementation deserves comment.

Friston *et al.* (1990) proposed using the Poisson "rare events" approximation to the Binomial, comparing $P^e K$ with $Po(\eta K)$. The *p*-values given by the SPM package for an exceedence proportion P^e over the K intracerebral voxels thresholded at probability level η is the compliment of the appropriate Poisson CDF (eqn.42). This gives a good approximation to the binomial for the number of voxels and thresholds usually considered ($\eta \leq 0.05$), even if only one plane is being considered ($K_p \approx 4000$). There is no computational advantage in making this approximation.

$$p = 1 - \sum_{u=0}^{P^e K - 1} e^{-\eta K} (\eta K)^u / u! \qquad (42)$$

In addition to this (erroneous) omnibus *p*-value, many authors quote a "Chi-Squared" value, computed by the SPM package as:[49]

$$\text{Chi-Squared} = \frac{2K(P^e-\eta)^2}{(P^e+\eta)(2-P^e-\eta)} \tag{43}$$

This is the Chi-squared statistic for a test of homogeneity between the rows of the following table. Clearly this is an inappropriate situation for this Chi-Squared statistic since the second row contains expected values rather than observations from a second multinomial distribution.

P^eK	$K-P^eK$
ηK	$K-\eta K$

3.4.2. Worsley's exceedence proportion test

Recent work by Worsley (1994c) considers the exceedence proportion of a strictly stationary continuous Gaussian random field $Z(\boldsymbol{x})$, $\boldsymbol{x} \in \Psi \subseteq \Re^D$, with unit variance, and hypothesised zero mean.

Theory

The exceedence proportion P^c, for a continuous standard Gaussian random field, above the (upper tail probability) threshold η, is defined analogously to the discrete case (eqn.41) as:

$$P^c = \int_{\boldsymbol{x}\in\Psi} \{Z(\boldsymbol{x}) > z\} \,/\, \lambda(\Psi) \tag{44}$$

for $z = \Phi^{-1}(1-\eta) = -\Phi^{-1}(\eta)$
and $\lambda(\Psi)$ the Lebesgue measure of Ψ, its volume

Let $R(\boldsymbol{h}) = \text{Cov}[Z(\boldsymbol{x}), Z(\boldsymbol{x}+\boldsymbol{h})]$ be the covariance function of the field, independent of $\boldsymbol{x}$ since the field is strictly stationary. Worsley (1994c, §3, eqn.3.1) shows that:

$$\text{Var}[P^c] \le g(z)/\lambda(\Psi), \text{ where } g(z) = \int_{\Re^D} [\Pr(Z(\mathbf{0}) \ge z, Z(\boldsymbol{h}) \ge z) - \Phi(-z)^2]\, d\boldsymbol{h}$$

Under mild conditions the above inequality tends to equality as Ψ becomes large, and the null distribution of P^c tends to normality, with mean η and the given variance.

If the correlation function is Gaussian, $R(\boldsymbol{h}) = \exp(-\boldsymbol{h}^T(2\boldsymbol{\Sigma})^{-1}\boldsymbol{h}/2)$ then Worsley (1994c, §6, eqn.6.4) shows that the limiting value of $\text{Var}[P^c]$ is then

$$\text{Var}[P^c] \approx g_I(z)\, |2\Sigma|^{1/2} / \lambda(\Psi) \tag{45}$$

where

$$g_I(z) = \int_0^\infty \frac{\pi^{D/2-1} y^{D+1}}{D\Gamma(D/2)\sqrt{1-\exp(-y^2)}} \exp\left(\frac{-z^2}{1+\exp(-y^2/2)} - \frac{y^2}{2}\right) dy \tag{46}$$

---

[49] The *p*-value given by the SPM package is obtained as stated, and not by comparing the Chi-Squared statistic with a Chi-Squared distribution with one degree of freedom.

Since a strictly stationary continuous standard Gaussian random field with Gaussian covariance function $R(\boldsymbol{h}) = \exp(-\boldsymbol{h}^T(2\boldsymbol{\Sigma})^{-1}\boldsymbol{h}/2)$ is equivalent to the field formed by convolving a white noise field (of appropriate variance) with a Gaussian kernel with variance-covariance matrix $\boldsymbol{\Sigma}$, the variance-covariance matrix of partial derivatives of the field $\boldsymbol{\Lambda}$ is related to $\boldsymbol{\Sigma}$ by $\boldsymbol{\Lambda} = (2\boldsymbol{\Sigma})^{-1}$. (See appendix C for results.) The limiting value of $\mathrm{Var}[P^c]$ is then:

$$\mathrm{Var}[P^c] \approx g_I(z) / (|\boldsymbol{\Lambda}|^{1/2} \lambda(\Psi)) \tag{47}$$

Application

This result allows us to apply an omnibus exceedence proportion test to Gaussian statistic images. It must be assumed that, under H_W, the exceedence proportion of the statistic image P^e, over set of voxels W, is distributed approximately as the exceedence proportion P^c, over $\Omega \subset \Re^D$ of a strictly stationary continuous Gaussian random field with the same smoothness $\boldsymbol{\Lambda}$ and Gaussian covariance function. Assuming additionally that Ω is large enough to use the asymptotic result, and that K and η are such that a normal approximation to P^e is reasonable, the null distribution of P^e is approximately normal with mean η and variance:

$$\begin{aligned}\mathrm{Var}[P^e] &\approx \mathrm{Var}[P^c] \\ &\approx g_I(-\Phi^{-1}(\eta)) / (|\boldsymbol{\Lambda}|^{1/2} \lambda(\Omega)) \\ &\approx g_I(-\Phi^{-1}(\eta)) / (|\hat{\boldsymbol{\Lambda}}|^{1/2} h_x \times h_y \times h_z \times K)\end{aligned} \tag{48}$$

where the K voxels are of dimension $h_x \times h_y \times h_z$ and $\hat{\boldsymbol{\Lambda}}$ is an estimate of $\boldsymbol{\Lambda}$. This gives critical exceedence proportion c for a test at (approximate) level α of:

$$c = \eta + \Phi^{-1}(1-\alpha) \times \sqrt{g_I(-\Phi^{-1}(\eta)) / (|\hat{\boldsymbol{\Lambda}}|^{1/2} h_x \times h_y \times h_z \times K)} \tag{49}$$

Alternatively, an approximate p-value for an exceedence proportion P^e is given by

$$p \approx 1 - \Phi\left(\frac{(P^e - \eta)\sqrt{|\hat{\boldsymbol{\Lambda}}|^{1/2} h_x \times h_y \times h_z \times K}}{\sqrt{g_I(-\Phi^{-1}(\eta))}} \right) \tag{50}$$

Some values of $g_I(-\Phi^{-1}(\eta))$ for $\eta \in \{0.05, 0.01, 0.005, 0.001, 0.0001\}$ and $D = 2,3$ computed 9dp using an adaptive 8 panel Newton-Cotes rule are as follows:

$g_I(-\Phi^{-1}(\eta))$	$D = 2$	$D = 3$
$\eta = 0.05$	0.132138364	0.271976144
$\eta = 0.01$	0.018522406	0.032154139
$\eta = 0.005$	0.007965094	0.012816709
$\eta = 0.001$	0.001168624	0.001593527

Reservations

This method is a “random field” method, and many of the reservations expressed during the discussion of random field approaches of §3.3.6. are relevant here.

Example: Simulated images

The simulated images were formed by smoothing a white noise field with a Gaussian moving average filter of FWHM 10mm. So, the variance-covariance matrix of the kernel is $\mathbf{\Sigma} = 10^2/8\ln(2) \times \mathbf{I}_3$, so $|\mathbf{\Lambda}| = |(2\mathbf{\Sigma})^{-1}| = (8\ln(2)/(2\times10^2))^3$

Critical exceedence proportions for an approximate level $\alpha = 0.05$ test with upper tail probability thresholds of $\eta = 0.05$, 0.01, 0.001 & 0.0001; are 0.061729, 0.014033, 0.001898 & 0.000317 respectively (6dp). From the simulated data, 95% CIs for the actual sizes of the tests with these critical exceedence proportions are (0.0356, 0.0420), (0.0458, 0.0530), (0.0623, 0.0705), & (0.0745, 0.0833) respectively, computed using the normal approximation to the binomial. Thus, for a full brain volume at this smoothness, the test appears to be slightly conservative for a 95% threshold ($\eta = 0.05$), exact for a 99% threshold ($\eta = 0.01$), slightly lax for a 99.9% threshold ($\eta = 0.001$), and very lax for a 99.99% threshold ($\eta = 0.0001$). A full simulation for other smoothness values ($\sqrt{|\Lambda|}$) is required to assess the test fully.

As noted earlier, the average estimated smoothness of the simulated data in terms of FWHM is 10.4mm×10.4mm×10.8mm. This gives $|\hat{\mathbf{\Lambda}}|$ smaller than that computed with the theoretical smoothness of 10mm FWHM, and therefore slightly larger *p*-values. The critical exceedence proportions with the estimated smoothness are 0.0627, 0.014359, 0.001970, 0.000335 (6dp) for $\eta = 0.05$, 0.01, 0.001 & 0.0001 respectively. 95% CIs for the actual sizes of the tests are (0.0262, 0.0318), (0.0336, 0.0398), (0.0513, 0.0587) & (0.0627, 0.0709) respectively. The test becomes more conservative for $\eta = 0.05$ & $\eta = 0.01$, and less lax for $\eta = 0.001$ & $\eta = 0.0001$.

The top 10% of the EDFs of the exceedence proportion above upper tail probability thresholds of $\eta = 0.05$, 0.01, 0.001 & 0.0001 are depicted in fig.46, with pointwise 95% confidence bands computed using the normal approximation to the binomial. Superimposed is the theoretical CDF given by Worsley (1994c), the compliment of the *p*-value given in eqn.50, computed for both the theoretical and estimated smoothness. The discreteness of the exceedence proportion (a multiple of the voxel volume) is evident in the EDF for $\eta = 0.0001$, which has large steps but a narrow confidence band.

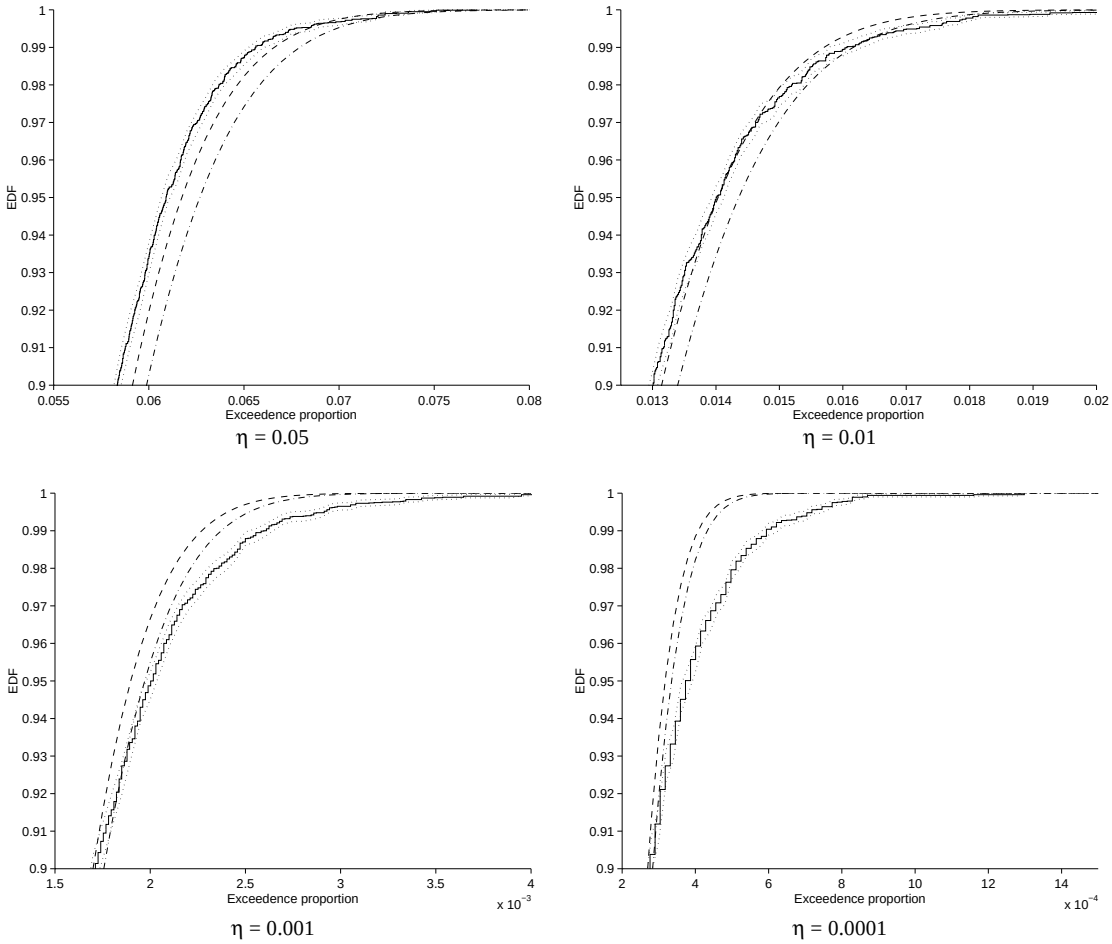

Figure 46

Top 10% of the EDFs of the proportion of voxels exceeding the 95th, 99th, 99.9th &99.99th percentiles of the Gaussian distribution, from 10^4 simulated Gaussian statistic images, with pointwise 95% confidence band for the true CDF, computed using the normal approximation to the binomial. Superimposed are the EDFs predicted by Worsley (1994c), for the theoretical smoothness of 10mm FWHM (dashed line) and the estimated smoothness of 10.4mm×10.4mm×10.8mm (dot-dash line).

3.4.3. Change distribution analysis

The original PET omnibus test, now seldom used, is *Change distribution analysis*, proposed by Fox *et al*. (1988).

Overview

The statistic image is collapsed into its local maxima (or minima), the values of voxels with value greater (smaller) than those of the 26 neighbouring voxels. These local maxima (minima) were assumed independent. The distribution of these local maxima (minima) was then examined for outliers, indicating departures from H_W. The test statistic adopted was the sample kurtosis g_2, a statistic usually used to assess departures from normality (Snedecor & Cochran, 1967, §3.14; D'Agostino, 1971), and advocated by Grubbs (1969, §4.10) to test for multiple outliers in a normal distribution, which is mesokurtotic (kurtosis $\gamma_2=0$).[50] Critical values for the normal distribution were used, using the tables of D'Agostino & Tietjen (1971) for small samples, and the normal approximation of Zar (1984, p119, & definitions, p83), the latter being a slight complication of that of Snedecor & Cochran (1967). Although the test only has weak control, if significant, the biggest outlier in the distribution of local maxima indicates significant evidence against the respective voxel hypothesis.

$$\text{The kurtosis } g_2 \text{ for a sample } \{X_1,\ldots,X_N\} \text{ is } g_2 = \frac{\Sigma(X - \overline{X})^4/N}{(\Sigma(X - \overline{X})^2/N)^2} - 3$$

For null study mean difference images $\overline{\Delta}_\bullet$, Fox *et al*. found that the distribution of local maxima (minima) was slightly platykurtotic ($g_2<0$, flatter than a normal distribution, with heavier tails), so a test of H_W via a test of $H:\gamma_2=0$ against $\overline{H}:\gamma_2>0$ (leptokurtosis, more peaked than a normal) might be expected to be slightly conservative. Simulation and routine use confirmed this to be the case, and the test was abandoned in favour of more powerful approaches with strong control over FWE.

Implementation

The method was actually implemented by considering the local maxima and minima together, leading to a bimodal distribution which is distinctly platykurtotic, leading to a very conservative test for outliers based on detecting leptokurtosis. To overcome this, "one-sided" g_2 statistics were computed for the positive and negative values separately. These two statistics were assessed separately, again by reference to the distribution of g_2 for normal samples, to give separate one sided omnibus tests of increases and decreases. Splitting the distribution of local extrema about zero is not equivalent to forming the distributions of local maxima and minima separately.

---

[50] This was used in preference to a one sided outlier test based on the skewness of the distribution, presumably because the distribution of the local maxima is likely to be skew under the null hypothesis.

3.5. Suprathreshold Cluster Tests

Voxel-by-voxel approaches with strong control over FWE test the statistic image at each voxel with the omnibus null hypothesis, H_W, of zero mean, by comparing each voxel value with a high threshold. Only the magnitude of a departure from H_W at each voxel is assessed. The spatial structure of departures from H_W is not considered. Regarding departures from H_W as a signal added to a null statistic image, signals of low magnitude will (probably) not be detected, even if they are distinguished by having spatial extent (measured in FWHM) greater than the FWHM of the statistic image. Including spatial considerations could therefore increase the sensitivity of tests.

Suprathreshold cluster tests

Suprathreshold cluster tests, a recent development in PET statistics, attempt to include the structure of signals into the test. The statistic image is thresholded at a fairly low threshold, yielding clusters of voxels with suprathreshold values. These clusters are then assessed individually for significance. The approach is half way between ROI and voxel-by-voxel methods, testing the evidence against the null hypothesis for regions of interest defined voxel-by-voxel from the statistic image. Suprathreshold cluster tests do not have strong control over FWE at the voxel level, but may be said to have strong control at the cluster level if, for every identified cluster of voxels $U \subseteq W$, the probability of false rejection of H_U is at most α, the level of the test, regardless of the truth of H_k at voxels k outside U (eqn.51).

$$\Pr(\text{“reject } H_U\text{”} | H_{U'}) \leq \alpha \text{ for any } U \subseteq U' \subseteq W. \tag{51}$$

Suprathreshold cluster size tests

Suprathreshold cluster size tests, assess the suprathreshold clusters by their size, measured in voxels, or as the volume (area) covered by the voxels (pixels) in the cluster. The statistic image is thresholded at a fairly low threshold, yielding clusters of suprathreshold voxels. Clusters of size greater than or equal to a critical size are declared significant. For a given test level α, and upper tail probability threshold η, the problem is to obtain the critical cluster size s_α that is attained with probability at most α under H_W. Strong control at the cluster level then follows, assuming subset pivotality. Single threshold methods can be viewed within this framework as seeking the threshold which yields a cluster of size at most 1 voxel with probability at most α, under H_W.

This problem has been addressed by three groups; Poline & Mazoyer (1993), Roland *et al*. (1993), and Friston *et al*. (1993). The first two attempt to estimate critical cluster size thresholds using simulation, while Friston and co-workers use approximate theory for continuous random fields.

3.5.1. Simulation approaches

Conceptually, the simulation approach is simple: Simulate a large number of null statistic images, threshold them at the specified level, identify the suprathreshold clusters, compute their sizes and note the size of the largest cluster for each image. The empirical distribution function of suprathreshold cluster size thus obtained estimates the true distribution, and can then be used for (approximate) inference. For example, the critical cluster size s_α for a test with level α is the $1-\alpha$ point of the distribution of maximum cluster size, with estimate $\hat{s}_\alpha$ as the $100(1-\alpha)^{th}$ percentile of the empirical distribution. An approximate *p*-value for a cluster of size *s* could be similarly obtained as the

proportion of the empirical distribution greater than *s*. The *p*-value of the largest suprathreshold cluster of a statistic image is in this scenario the *p*-value for H_W.

This straightforward approach is not taken by either Poline & Mazoyer (1993), or Roland *et al*. (1993).

Poline's approach: Poisson assumptions

Poline & Mazoyer (1993), working in 2D, assume that for a null statistic image suprathreshold clusters of size at least *s* occur according to a planar Poisson process with intensity $\theta^{\geq}_s$ per unit area. Then, over a plane of area $\lambda(\Omega_P)$ the number $X^{\geq}_s$ of clusters with size at least *s* is distributed $X^{\geq}_s \dot{\sim} Po(\lambda(\Omega_P)\times\theta^{\geq}_s)$. (The Poisson approximation enables application for the different sized areas of interest on different planes.) The rates $\theta^{\geq}_s$, for each possible cluster size *s*,[51] are estimated from simulated null statistic images as $\hat{\theta}^{\geq}_s$. For large *s*, insufficient evidence against the Poisson assumption was found by a Chi-squared goodness of fit test applied to simulated null statistic images.

The critical cluster size s_α for a level α test is estimated by $\hat{s}_\alpha$, the smallest *s* such that $\Pr(X^{\geq}_s \geq 1) \leq \alpha$, with probabilities given by the Poisson assumption with the estimated rates. (This is uniquely defined: The rates are estimated from the same simulated data, and therefore must form a sequence monotonically decreasing as *s* increases.) The *p*-value for an individual cluster of size *s* is then computed as $\Pr(X^{\geq}_s \geq 1)$, using the Poisson assumption, with the estimated rates. The step-down procedure proposed by Poline *et al*. (1993), using these *p*-values, distinguishes itself from all other step-down procedures by getting harsher as the test steps down, leading to a less powerful test than the single-step test rejecting the null hypothesis for clusters with *p*-value less than α.

Roland's approach: Poisson assumptions

Roland *et al*. (1993) adopt a similar approach. Working in 3D, they assume that the number $X^{=}_s$ of suprathreshold clusters of size *s* has a Poisson distribution with rate $\theta^{=}_s$, *independently* for each size *s*. With these assumptions, the number of clusters of size at least *s*, $X^{\geq}_s$, is, by the closure property of independent Poisson variates, distributed $X^{\geq}_s \sim \mathrm{Po}(\sum_{s'\geq s}\theta^{=}_s)$. The rates $\theta^{=}_s$ are estimated from simulations, from which the critical cluster size s_α can be estimated, or approximate *p*-values for individual clusters (and hence H_W) computed.

Reservations: Simulating null statistic images

The unnecessary Poisson assumptions cast doubt on the validity of the suprathreshold cluster size tests of Poline & Mazoyer (1993) and Roland *et al*. (1993).

That aside, a more fundamental reservation with the approach concerns the matching of simulated and real null statistic images. In particular, whether the suprathreshold cluster size achieved with probability at most α in a null statistic image from a real experiment is close to the critical value for the simulated statistic images. This is difficult to assess, since very few true "null" experiments are carried out. Even considering scans acquired under the same experimental conditions, there is still very little data for any particular combination of scanning protocol, reconstruction method and pre-processing.

Most experimenters attempt to mimic real null statistic images by simulating rCBF images whose marginal distribution and auto-correlation function match those of the real

---

[51] Since the clusters are identified voxel-by-voxel, their sizes are multiples of the pixel sizes, assuming a constant pixel size. Thus *s* is discrete.

data in question. Both Poline & Mazoyer and Roland *et al*. simulate paired *t*-statistic images from simulated (mean) difference images generated by smoothing a white noise Gaussian field with a Gaussian kernel, the variance of the noise, and the FWHM of the kernel, chosen to match the (estimated) variance and FWHM of the differences images in question.[52] Whether this is adequate remains to be seen. In particular, the sensitivity of estimated critical values to changes in the simulation parameters is seldom investigated, though Poline found a 20% change in pixel variance had little effect on size of the test for the simulated data.

Since the data from each study is different, and as the robustness of the critical values to changes in the simulation parameters is not known, these methods require a simulation to be carried out for every new data set. This requires substantial computer time.

Due to these problems, the simulation approaches of Poline & Mazoyer (1993), and Roland *et al*. (1993) have been little used.

3.5.2. Friston's Theoretical treatment

Friston *et al*. (1994d) have produced an approximate expression for the CDF of the size of the largest component of the excursion set of a strictly stationary continuous standard Gaussian field, the continuous analogue of suprathreshold cluster size.

Overview of theory

Let M be the number of distinct components[53] of the excursion set $A_u(Z,\Psi) = \{\boldsymbol{x} \in \Psi : Z(\boldsymbol{x}) \geq u\}$, of a strictly stationary continuous Gaussian random field $Z(\boldsymbol{x})$, with zero mean and unit variance, defined over compact subset Ψ of $\Re^D$, and thresholded at a level $u = \Phi^{-1}(1-\eta)$. Let S be the size (the Lebesgue measure) of a component of the excursion set, given that it exists, and let $S_{\max}$ be the size of the largest component of the excursion set.

Adler (1981, Th.6.9.3, p.161) shows that the number of points which contribute to make up the Euler characteristic has a Poisson distribution in the limit as the threshold tends to infinity. For high thresholds u, the number of components of the excursion set A_u, and the number of local maxima above u, are essentially the same as the Euler characteristic, and should therefore also have the same limiting distribution, although this has never been rigorously proven. So, Friston *et al*. (1994d) assume a Poisson form for M, with mean θ, the expected number of local maxima above u, given by Hasofer (1976) as:

$$\theta = E[M] = \lambda(\Omega)|\Lambda|^{1/2}(2\pi)^{-(D+1)/2} u^{D-1} \exp(-u^2/2) \qquad (52)$$

The right hand side of eqn.52 is the expected Euler characteristic (eqn.37), with the polynomial $P_D(u)$ approximated by its leading term.

---

[52]The reader is referred to Poline & Mazoyer (1993) and Roland *et al*. (1993) for full details of the method of simulation. Both authors use more complicated methods than indicated here, each with its own quirks, advantages and disadvantages. Poline & Mazoyer model the covariance between scans on the same individual; Roland *et al*. do not. Roland *et al*. estimate the auto-correlation of the difference images from "noise" images with physiological correlations "removed". This underestimates the smoothness of the difference images, resulting in simulated statistic images with lower FWHM than appropriate, possibly leading to an underestimation of the critical suprathreshold cluster size.

[53]For a continuous field the excursion set over a compact domain is the union of compact subsets of the domain, the *components* of the excursion set.

Assuming independence of the sizes of components of the excursion set, the CDF, $F_{S_{max}}(s)$, of S_{max} is given by:

$$\begin{aligned} F_{S_{max}}(s) &= \Pr(S_{max} \le s) \\ &= \sum_{m=0}^{\infty} \Pr(M = m)\,(\Pr(S \le s \mid S > 0))^m \\ &= \sum_{m=0}^{\infty} \frac{1}{m!}\theta^m e^{-\theta}\,(\Pr(S \le s \mid S > 0))^m \\ &= e^{-\theta \Pr(S > s \mid S > 0)} \end{aligned} \tag{53}$$

It remains to find an expression for $\Pr(S > s \mid S > 0)$. The expectation of S, given $S > 0$, is easily obtained since the sum of the sizes of the components of the excursion set is the size of the excursion set, the exceedence proportion multiplied by the size of the domain of the field, with expected value $\lambda(\Psi)\,\Phi(-u)$. Since the sizes of components are assumed independent under the null hypothesis:

$$E\left[\sum_{m=1}^{M} S_m\right] = E\left[E\left[\sum_{m=1}^{M} S_m \,\middle|\, M\right]\right] = E[M\ \ E[S \mid S > 0]\ ]$$

$$= E[M]\,E[S \mid S > 0] = \theta\,E[S \mid S > 0] = \lambda(\Psi)\,\Phi(-u) \tag{54}$$

Adler (1981, p.158) reports results given by Nosko, stating that the size of a component of the excursion set of a continuous strictly stationary two-dimensional standard Gaussian random field, has asymptotically an exponential distribution as u tends to infinity, with mean $2\pi/(u^2\,|\Lambda|^{1/2})$. This result can be extended to D dimensions since the shape of an excursion of the field above a threshold u is (asymptotically) parabolic with curvature matrix $-u\Lambda$ (Adler, 1981, p157), and since (asymptotically) the height of the excursion above the threshold plane (the "excess height"), given that the local maxima exceeds u, is exponential with mean $1/u$ (Adler, 1981, Th.6.8.2). Worsley (in Friston *et al*., 1994) put these two facts together to obtain that the size of a component of the excursion set, to the power $2/D$, has asymptotically an exponential distribution as the threshold u tends to infinity, with mean:

$$E[S^{2/D}] = \frac{2\pi}{u^2\,\Gamma(D/2+1)^{2/D}\,|\Lambda|^{1/D}} \tag{55}$$

Friston *et al*. (1994d) found that $E[S^{2/D}]$ is substantially over-estimated by eqn.55 for low thresholds u, and proposed the following approximate PDF for S, for use at low thresholds:

$$f_S(s) = \frac{2\beta}{D}\,s^{2/D-1}\,e^{-\beta\,s^{2/D}} \tag{56}$$

giving CDF for S as:

$$F_S(s) = 1 - e^{-\beta\,s^{2/D}} \quad \Rightarrow \quad \Pr(S > s \mid S > 0) = e^{-\beta\,s^{2/D}} \tag{57}$$

The mean of the distribution of eqn.56 is $\Gamma(D/2 -1)\ \beta^{-D/2}$, and β is chosen so that this matches E[S] as given by eqn.54:

$$\beta = \left(\frac{\Gamma(D/2 -1)\ \theta}{\lambda(\Psi)\ \Phi(-u)} \right)^{2/D} \tag{58}$$

Clearly $S^{2/D}$ has an exponential distribution with mean $1/\beta$. Asymptotically, the mean assumed by Friston is equal to that given by Nosko (eqn.55), as is seen by substituting the high value approximation to the normal CDF ($\Phi(-u) \approx (2\pi)^{-1/2} \exp(-u^2/2) / c$), in the above expression for β (eqn.58).

With this assumed form for $\Pr(S > s \mid S > 0)$ (eqn.57), eqn.53 gives the CDF of the size S_{max} of the largest component of the excursion set:

$$F_{S_{max}}(s) = \exp(-\theta \exp(-\beta\ s^{2/D})) \tag{59}$$

The p-value for a component of size s is then $1-F_{S_{max}}(s)$. Setting eqn.59 to $1-\alpha$ and solving for s gives the critical size s_α for a test with approximate level α, eqn 60.

$$s_\alpha = \left[\beta^{-1} \ln\left(\frac{-\theta}{\ln(1-\alpha)} \right) \right]^{D/2} \tag{60}$$

The theory is applied to standard Gaussian statistic images under the usual assumption that the statistic image is a good lattice representation of a strictly stationary continuous standard Gaussian field with the same variance-covariance matrix of partial derivatives, and in particular that the distribution of the maximum suprathreshold cluster size for the discrete field is similar to that of the largest component of the excursion set of the continuous field.

Example: Simulated images

For each of the simulated Gaussian statistic images, the maximum suprathreshold cluster size was computed for upper tail probability thresholds of $\eta = 0.01$, 0.001 & 0.0001. With the theoretical smoothness of $|\Lambda| = (8\ln(2)/(2\times10^2))^3$, $D = 3$ and $u = -\Phi^{-1}(\eta)$, critical suprathreshold cluster sizes for the three thresholds are obtained from eqn.60 as 3197.9mm^3, 990.6mm^3 & 318.9mm^3 respectively (1dp). From the simulated data, 95% CIs for the actual sizes of tests with these critical cluster sizes are (0.0366, 0.0430), (0.0329, 0.0391) & (0.0470, 0.0542) (4dp), for upper tail probability thresholds $\eta = 0.01$, 0.001 & 0.0001 respectively.

The top 10% of the EDFs of the maximum suprathreshold cluster size, S_{max}, for the three thresholds considered, are depicted in fig.47. Superimposed are the theoretical CDFs of eqn.59, computed with the theoretical smoothness of 10mm FWHM, and the estimated smoothness of 10.4mm×10.4mm×10.8mm. As smoothness (in terms of FWHM) increases, larger suprathreshold regions become increasingly likely, so overestimating smoothness results in more conservative (and therefore less powerful) tests, as is apparent from the two theoretical CDFs. The discreteness of the suprathreshold cluster size (a multiple of the voxel volume) is evident in the EDF for $\eta = 0.0001$, which has large steps but a narrow confidence band.

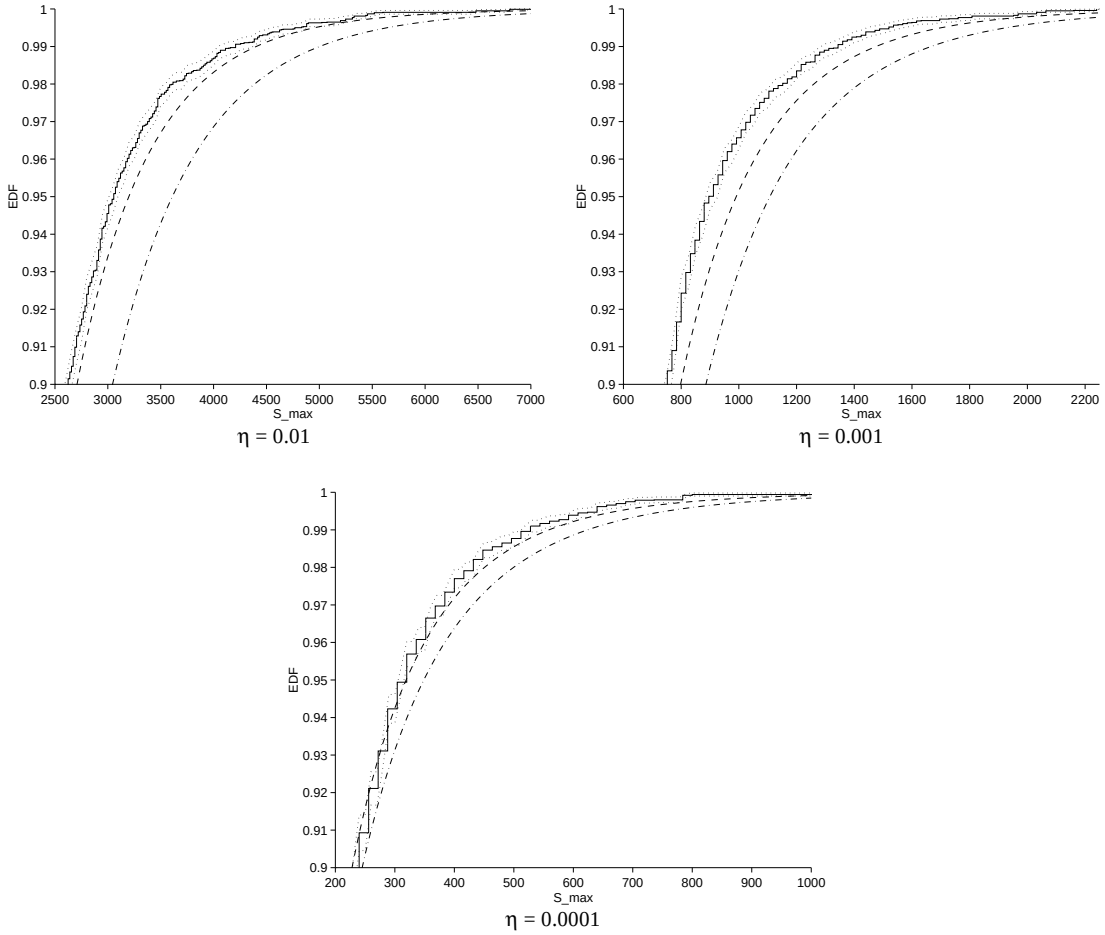

Figure 47

Top 10% of the EDFs of the maximum suprathreshold cluster size above a threshold $-\Phi^{-1}(\eta)$, from 10^4 simulated Gaussian statistic images, with pointwise 95% confidence band for the true CDFs, computed using the normal approximation to the binomial. Superimposed are the CDFs predicted by Friston *et al*. (1994b) (eqn.59), for the theoretical smoothness of 10mm FWHM (dashed lines), and the estimated smoothness of 10.4mm×10.4mm×10.8mm (dot-dash line).

Reservations

We have already noted many problems with random field methods, when considering single threshold methods. Many of these points are relevant here. Of additional concern is the fact that the asymptotic results are being applied at low thresholds.

3.5.3. Comments on suprathreshold cluster size methods

More powerful

Friston *et al*. (1993) demonstrated the improved power of their suprathreshold cluster size tests over Worsley's Z_{max} test for Gaussian fields. Further insight into the relative power of Friston's S_{max} approach and other voxel-by-voxel approaches may be gleaned from the (two-dimensional) simulation study of ch.4. In general, suprathreshold cluster methods will outperform single threshold methods. This is not surprising, since

the latter offer strong control at the voxel level, rather than at the region level as with the suprathreshold cluster tests.

Resolving power

Since a suprathreshold cluster size test only has strong control at the region level, evidence of activations can only be reported as "within this region". For low thresholds when the critical cluster size is large this leaves the test with low resolving power. This makes the use of suprathreshold cluster methods undesirable when an activation is strong enough to be detected by a method offering strong control at the voxel level.

Threshold choice

Activations vary in shape and intensity from extremely intense focal activations (usually associated with primary functions such as motor control), to widespread activation at a low level (usually associated with more subtle cognitive tasks such as word recognition). Perhaps the main criticism of the suprathreshold cluster size approach is that the choice of threshold determines the type of activations the method is most sensitive to.

If a low threshold is chosen ("large" η), then the critical cluster size is large. The test will be sensitive to disperse activations which exceed the threshold over large clusters, but will miss focal activations. Vice-versa, if a high threshold is chosen (small η), then the critical cluster size is small. The test will be sensitive to intense focal activations, but will miss disperse low level activations which fail to exceed the threshold. (At the extreme, if a very high threshold is chosen such that the critical cluster size is less than one voxel, then we are left with a single threshold approach, the criticisms of which motivated suprathreshold cluster tests in the first place.)

Suprathreshold cluster approaches including the magnitude of the activation

To overcome the problem of intense focal activations being missed by suprathreshold cluster size tests with low thresholds, the magnitude of the activation can be considered. Poline *et al*. (1994a) consider such an approach, using the bivariate distribution of suprathreshold cluster size and mean amplitude, which is described in appendix H. Clearly for simulation approaches, any statistic characterising a suprathreshold region of a statistic image can be considered.

Suprathreshold cluster excess mass tests

One statistic that springs to mind is the "excess mass" of a suprathreshold portion of the statistic image, the size of the region enclosed between the statistic image and the threshold plane.[54] This is simply computed as the sum of the voxel values less the threshold level, over the voxels of a suprathreshold cluster. The null distribution of the maximum cluster weight could be obtained easily from simulated statistic images.

However, an approximate theoretical distribution could be derived for Gaussian statistic images along the lines of that given by Friston *et al*. (1994) for the suprathreshold cluster size (§3.5.2.). This is because it is possible to compute the (asymptotic) distribution of the size of the region enclosed by an excursion of a continuous strictly stationary standard Gaussian random field, and the threshold plane. The details are as follows:

Adler (1981, p.158) reports Nosko's result for two dimensions, stating that the (asymptotic) distribution of the square root of the volume enclosed by the excursion of the field and the threshold plane, is exponential with mean $\pi^{1/2}/(u^{3/2}\,|\Lambda|^{1/2})$. This can be generalised to D dimensions using the results for the shape and peak height of an

[54]This statistic is considered for selection of ROI in the Two-Stage approach described in the next chapter, where additional description is given.

excursion of such a field above a high threshold, as discussed in §3.5.2. Doing so (Worsley, private communication) reveals that the "excess mass", V_u, to the power $2/(D+2)$, has an exponential distribution with mean

$$\mathrm{E}[V_u^{2/(D+2)}] = \left(\frac{2\pi}{u\,|\Lambda|}\right)^{\frac{D}{D+2}} \frac{1}{u\,\Gamma(D/2+2)^{2/(D+2)}}$$

Thus, $V_{\max}$ can be considered instead of $S_{\max}$ in the framework presented by Friston *et al*. (1994). A low threshold correction for this similar to Friston's correction for the size of a component of the excursion set could be considered to make this result more accurate for the low thresholds at which the test would be applied. This remains to be attempted. The usual reservations for applying results for continuous fields to statistic images continue to apply.

3.6. Example–“V5” Study

3.6.1. Approaches on *t*-statistic images

Statistic images

To illustrate the differences in power between the methods discussed in this chapter, consider analysing the “V5” study data. We shall use the *t*-statistic image formed from subject difference images, where global changes have been removed by proportional scaling (§2.3.1.1.). The AC-PC plane of the resulting *t*-statistic image, depicted previously in §2.6.1., is shown again below (fig.48).

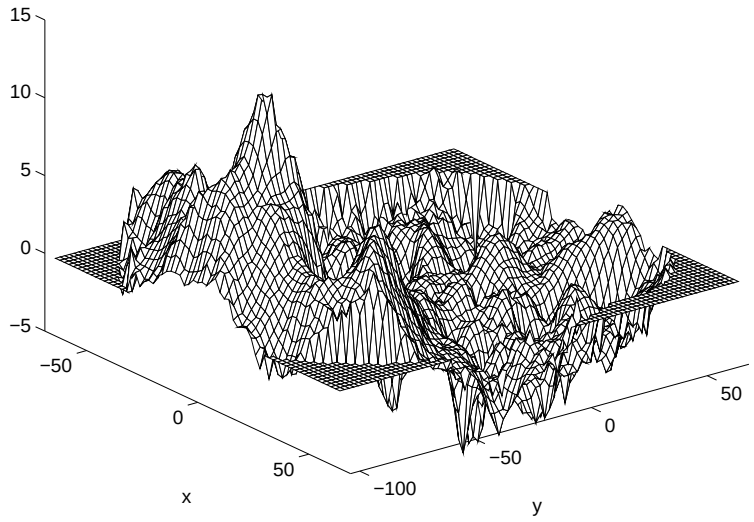

Figure 48

Mesh plot of *t*-statistic image ***T*** for “V5” study. (Eqn.21) Each voxel statistic is distributed as a Student’s *t* with 11 degrees of freedom under the hypothesis of no activation at that voxel, $H_k: \mu_k = 0$. $\Delta_{ik} \sim N(\mu_k, \sigma_k^2)$ is assumed. The AC-PC plane is shown.

Bonferroni approach

Unadjusted and Bonferroni adjusted *p*-values for this *t*-statistic image were presented in §2.6.1. For completeness, they are repeated here. The unadjusted *p*-value for H_k , P_k, is found by referring T_k to the CDF of a Student's *t*-distribution with 11 degrees of freedom. Bonferroni adjusted *p*-values are computed as $\tilde{P}_k = \min\{K P_k, 1\}$, and are shown below (fig.50). The number of intracerebral voxels is K=77189.

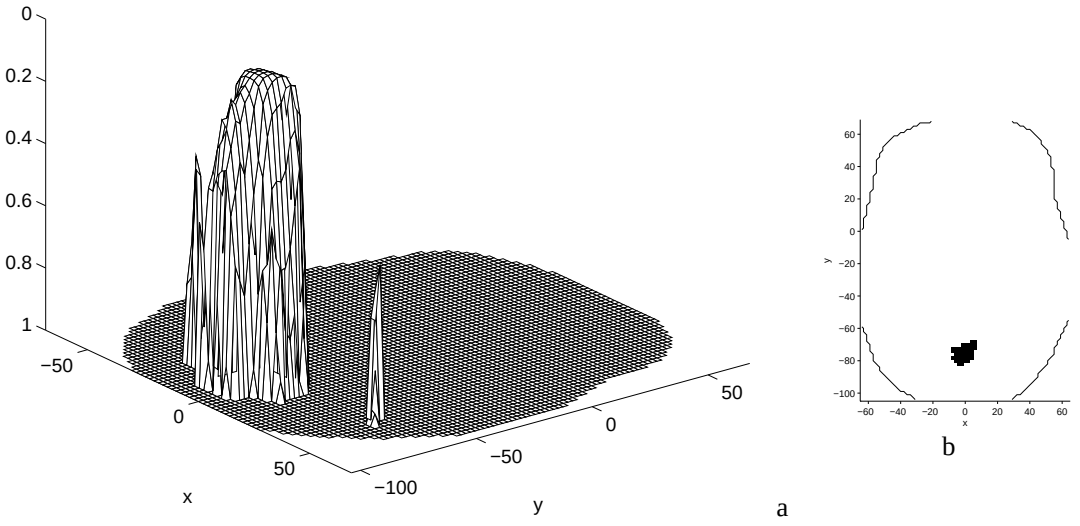

Figure 49

(a) Mesh plot of (one-sided) Bonferroni adjusted *p*-values, computed from the *t*-statistic image of fig.48. The *p*-value axis is graduated in reverse. Voxels outside the intracerebral volume have been removed. (b) Voxels with adjusted *p*-value below level 0.05. The outline of the intracerebral area is superimposed. The AC-PC plane is shown.

Worsley's expected Euler characteristic approach for t-fields

Consider applying Worsley's "Euler characteristic" method to this data (Worsley *et al.*, 1992, discussed in §3.3.1.), using the result for *t*-fields (Worsley 1994a, summarised in appendix D:3). An estimate is required for $\mathbf{\Lambda}$, the variance-covariance matrix of partial derivatives of the subject difference images, under the null hypothesis. (Recall the discussion regarding estimation of smoothness, §3.3.5.) In practice it is common to estimate the smoothness within the statistic image. This estimate of the variance-covariance matrix of partial derivatives of the *t*-field, $\mathbf{\Lambda}_T$, is related to $\mathbf{\Lambda}$ by $\mathbf{\Lambda}_T = \lambda_N \mathbf{\Lambda}$ (appendix G).

The variances of the partial derivatives of the *t*-field are 0.1266, 0.1068 & 0.1084 in the X, Y & Z directions respectively. Here N =12 and λ_{12}=1.389 (4dp). Assuming the off diagonal elements of $\mathbf{\Lambda}$ are zero gives estimate (to 4dp):

$$\hat{\mathbf{\Lambda}} = \begin{pmatrix} 0.0911 & 0 & 0 \\ 0 & 0.0769 & 0 \\ 0 & 0 & 0.0781 \end{pmatrix}$$

For strictly stationary standard Gaussian fields with Gaussian PRF, this corresponds to FWHM of 5.52mm×6.01mm×5.96mm (2dp). The K voxels are of dimension 2mm× 2mm×4mm, and therefore constitute a volume of 1 235 024mm^3. For three dimensional t-fields with 11 degrees of freedom and variance-covariance matrix of partial derivatives $\hat{\Lambda}$, defined over this volume, the expected Euler characteristic is equal to $\alpha = 0.05$ for threshold $u_\alpha = 14.1779$. This critical threshold is exceeded by T_k at a mere 57 voxels, none of which are in the AC-PC plane (fig.50b). The expected Euler characteristic gives adjusted p-values for high values of the t-statistic image, where the expected Euler characteristic is greater than zero (fig.50a). In this case, the expected Euler characteristic falls below zero at threshold $u = 9.3518$, so all voxels with value T_k below this, have adjusted p-value of zero.

The adjusted p-values are larger than those from a conservative Bonferroni approach, illustrating the excessive conservativeness of random field methods for noisy statistic images, and for F and t-statistic images with low degrees of freedom (denominator degrees of freedom in the case of F-fields).

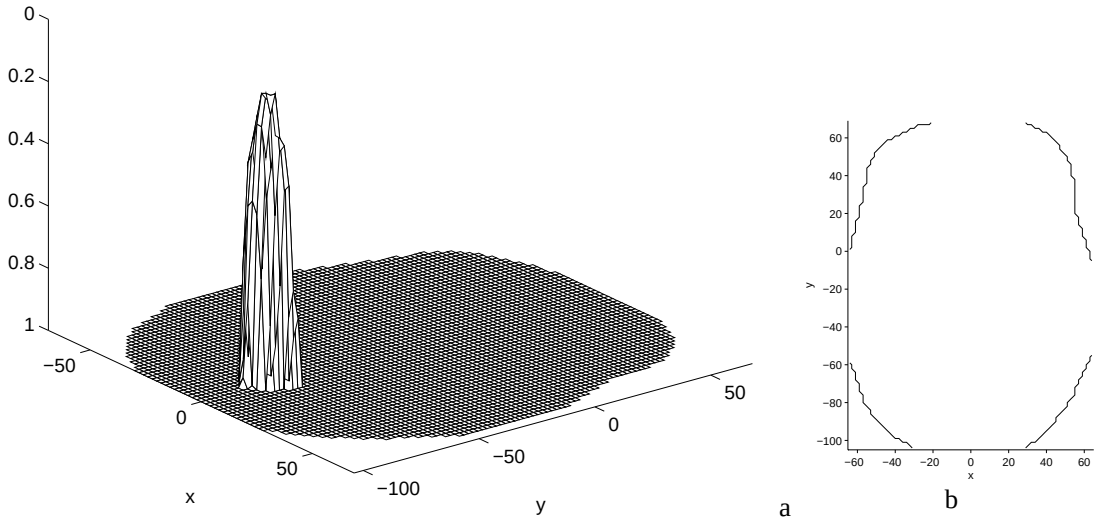

Figure 50

(a) Mesh plot of (one-sided) adjusted p-values, computed for high voxel values T_k (fig.48), as the expected Euler characteristic of a matching continuous t-field of 11 degrees of freedom, thresholded at T_k. Voxels outside the intracerebral volume have been removed. (b) Voxels with adjusted p-value below level 0.05. The outline of the intracerebral area is superimposed. The AC-PC plane is shown.

3.6.2. Approaches on Gaussianised *t*-statistic image

Gaussianised t-statistic image

Gaussianising the *t*-statistic image by replacing each voxel value with a Gaussian variate with the same probability of being exceeded, gives the Gaussian statistic image $\boldsymbol{Z} = (Z_1,\ldots,Z_K)$, of fig.51.

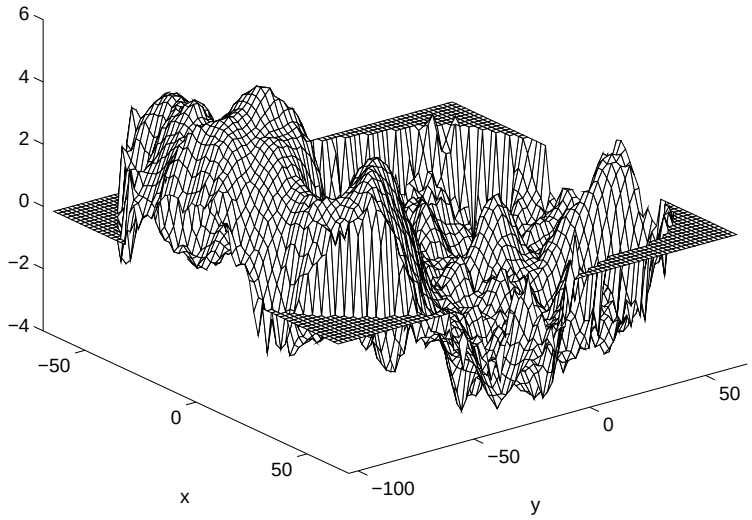

Figure 51
Gaussianised *t*-statistic image, formed from the *t*-statistic image of fig,48 as $Z_k = \Phi^{-1}(F_T(T_k))$.

The estimated variance-covariance matrix of partial derivatives of this field, estimated within the image, and assuming that the off diagonal elements are zero, is (to 4dp):

$$\hat{\boldsymbol{\Lambda}}_Z = \begin{pmatrix} 0.0448 & 0 & 0 \\ 0 & 0.0401 & 0 \\ 0 & 0 & 0.0380 \end{pmatrix}$$

If the PRF is assumed to be Gaussian, then this corresponds to FWHM of 7.87mm×8.32mm×8.54mm (2dp).

Worsley's expected Euler characteristic approach for Gaussian fields

For three dimensional Gaussian fields with variance-covariance matrix of partial derivatives $\hat{\Lambda}_Z$ as above, and defined over this volume, the expected Euler characteristic (eqn.36) is equal to $\alpha = 0.05$ for threshold $u_\alpha = 4.8279$. This critical threshold is exceeded by Z_k at 453 voxels, 44 of which are in the AC-PC plane (fig.52b). The expected Euler characteristic gives *p*-values for high values of the Gaussianised *t*-statistic image (fig.52a). In this case, the expected Euler characteristic falls below zero at threshold $u = 4.0746$, so all voxels with value T_k less than this have adjusted *p*-value of zero. The adjusted *p*-values are smaller than those from the Bonferroni approach.

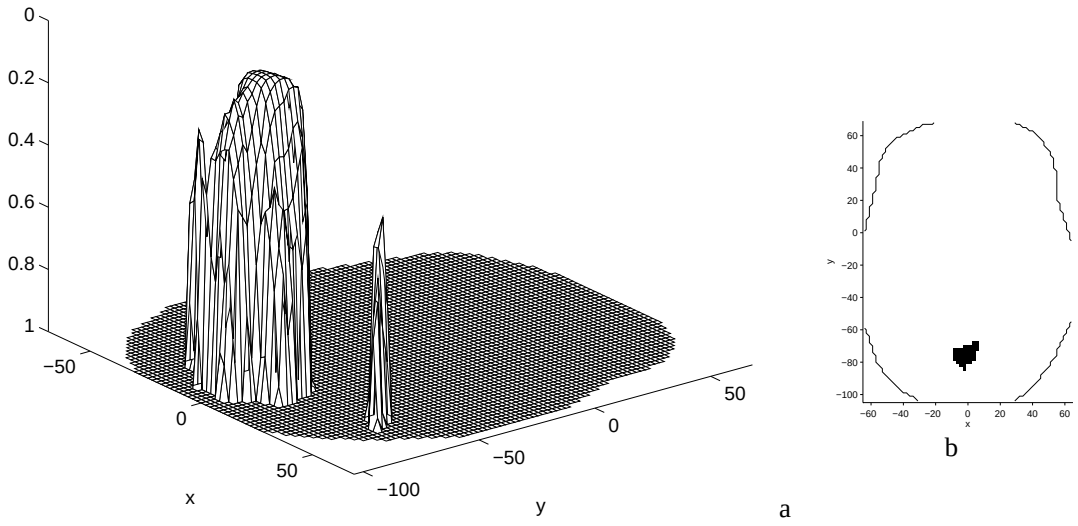

Figure 52

(a) Mesh plot of (one-sided) adjusted *p*-values, computed for high voxel values Z_k (fig.51), as the expected Euler characteristic of a matching continuous Gaussian random field, thresholded at Z_k. Voxels outside the intracerebral volume have been removed. (b) Voxels with adjusted *p*-value below level 0.05. The outline of the intracerebral area is superimposed. The AC-PC plane is shown.

Friston's "Bonferroni" approach for two dimensional Gaussian fields

Consider applying Friston's "Bonferroni" approach (§3.3.2.), to the AC-PC plane of the Gaussianised *t*-statistic ***Z***. Since this approach is usually applied plane by plane at level α (as implemented in SPM software prior to SPM94), we shall do likewise.

Usually the smoothness is estimated plane by plane, since the end planes are rougher than those in the middle. The sample variances of the numerical partial derivatives of the AC-PC plane of the Gaussianised *t*-statistic are 0.0358 and 0.0351 in the X and Y directions respectively, estimated within the image to 4dp. Assuming a Gaussian PRF, this corresponds to a FWHM of 8.80mm×8.89mm (2dp). Assuming isotropy, the pooled sample variance of the partial derivatives in both axial dimensions, 0.0354 (4dp), corresponds to a Gaussian PRF with estimated variance-covariance matrix $\hat{\mathbf{\Lambda}} = (2\hat{s})^{-2}\,\mathbf{I}_2$ giving $\hat{s} = 14.1080$ (4dp). The number of intracerebral voxels in the AC-PC plane is $K_{P8} = 4884$, with faces parallel to the X-Y plane of area $\lambda(\Omega_{P8}) = 19\,536\text{mm}^2$.

Substituting these values into Friston's false positive expression (eqn.39), and solving for c such that the false positive probability is α/K_{P8}, for $\alpha = 0.05$, gives a critical threshold of $c_\alpha = 3.9943$ (4dp) for this plane. This threshold is exceeded by Z_k at 229 voxels in the AC-PC plane (fig.53b). Eqn.40 gives an expression for the *p*-value for the maxima in a plane. For voxels with high values Z_k, such that this *p*-value is positive, this gives adjusted *p*-values (fig.53a). In this case, the *p*-value expression falls below zero at threshold $c = 3.0835$, so voxels with values less than this have adjusted *p*-value of zero. The adjusted *p*-values are dramatically smaller than all those seen thus far, because the method is applied plane by plane with no correction for multiple comparisons over planes.

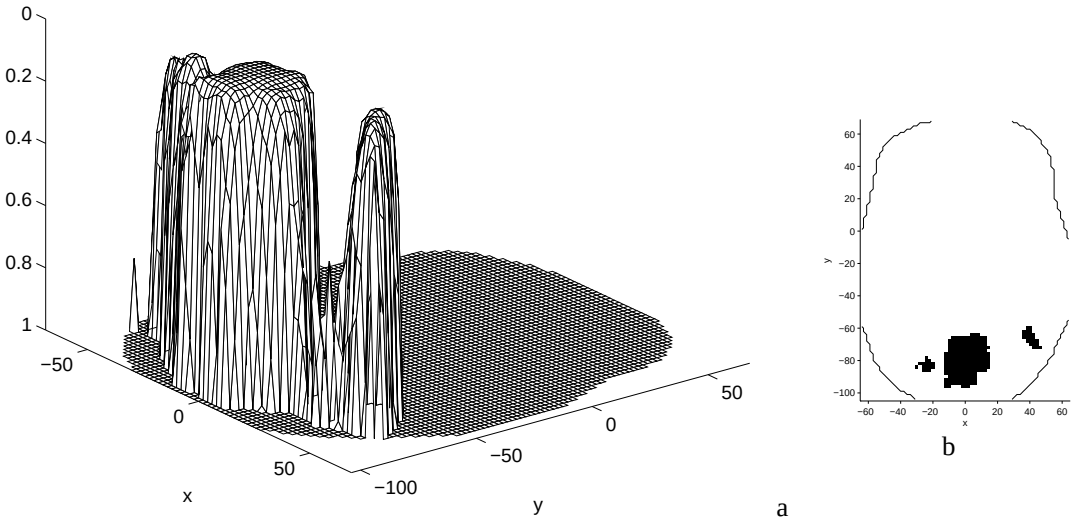

Figure 53

(a) Mesh plot of (one-sided) adjusted *p*-values, computed for high voxel values Z_k (fig.51), using Friston's false positive expression and a Bonferroni correction for the number of intracerebral voxels in the AC-PC plane (eqn.40). Voxels outside the intracerebral volume have been removed. (b) Voxels with adjusted *p*-value below level 0.05.

Friston's suprathreshold cluster size test

Now, consider applying Friston's suprathreshold cluster size test to the Gaussianised *t*-statistic. The SPM94 software uses this approach.

The method was described in §3.5.2. Assume that, under the omnibus null hypothesis, the Gaussianised *t*-statistic is a strictly stationary discrete standard Gaussian random field, and that the distribution of the maximum suprathreshold cluster size is distributed approximately the same as that of a matching continuous field. For $D = 3$ dimensions, volume $\lambda(\Omega) = 1\ 235\ 024\text{mm}^3$, and variance-covariance matrix of partial derivatives $\hat{\Lambda}_Z$; eqn.60 gives the critical suprathreshold cluster sizes for a level $\alpha = 0.05$ test. For thresholds of $\Phi^{-1}(1-\eta)$ with $\eta = 0.001$ & 0.0001, this gives critical suprathreshold cluster sizes of $s_\alpha = 657.0374\text{mm}^3$ & 228.5231mm^3 respectively.

For $\eta = 0.001$, there are 14 suprathreshold clusters, all consisting of one or two voxels, except for one of 4146 voxels, corresponding to a volume of $66\ 336\text{mm}^3$. This is well above the critical size, and indicates significant evidence against the omnibus hypothesis for that region. Of the 4146 voxels, 573 are in the AC-PC plane (fig.54b). Eqn.59 gives the *p*-value for this suprathreshold cluster as $<1\times10^{-10}$. Let the adjusted *p*-value at suprathreshold voxels be the *p*-value of the size of the cluster containing that voxel, and zero for voxels with sub-threshold value. For the Gaussianised *t*-statistic image thresholded at probability threshold $\eta = 0.001$ this gives the adjusted *p*-value image of fig.54a.

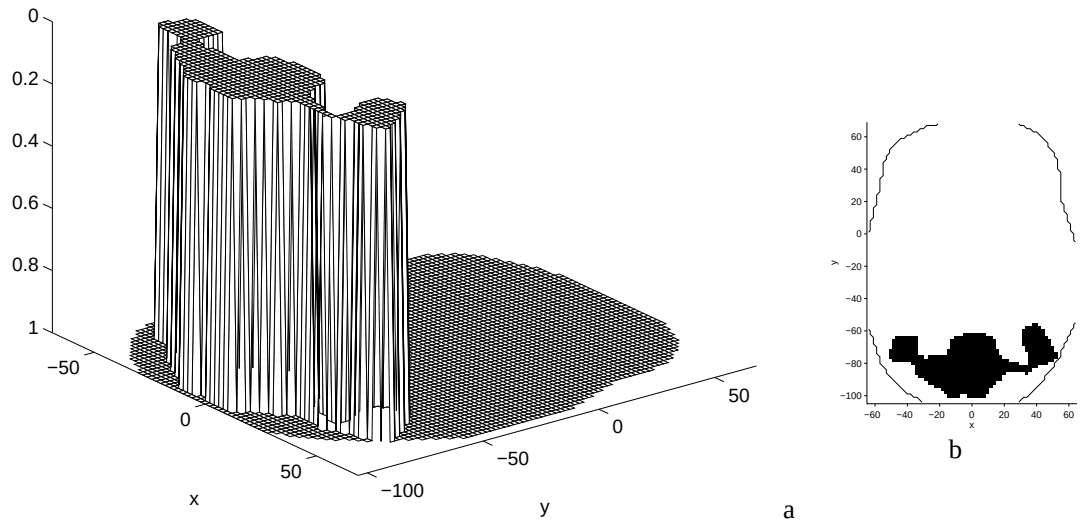

Figure 54

(a) Mesh plot of (one-sided) adjusted *p*-values, computed for voxels with values Z_k greater than the threshold $-\Phi^{-1}(0.001)$, as the *p*-value for the size of the suprathreshold cluster containing that cluster (eqn.59). For voxels with subthreshold values Z_k, the adjusted *p*-value is set to zero. Voxels outside the intracerebral volume have been removed. (b) Voxels with adjusted *p*-value below level 0.05. The outline of the intracerebral area is superimposed. The AC-PC plane is shown.

For upper tail probability threshold $\eta = 0.0001$, there are 7 suprathreshold clusters. Four of these are significant, with sizes of 34320mm^3, 2464mm^3, 720mm^3 & 272mm^3. Clearly these clusters are subsets of the significant cluster found for the lower threshold of $\eta = 0.001$. The four clusters constitute 2361 voxels, 324 of which are in the AC-PC plane (fig.55b). The corresponding adjusted *p*-value image is shown in fig.55a.

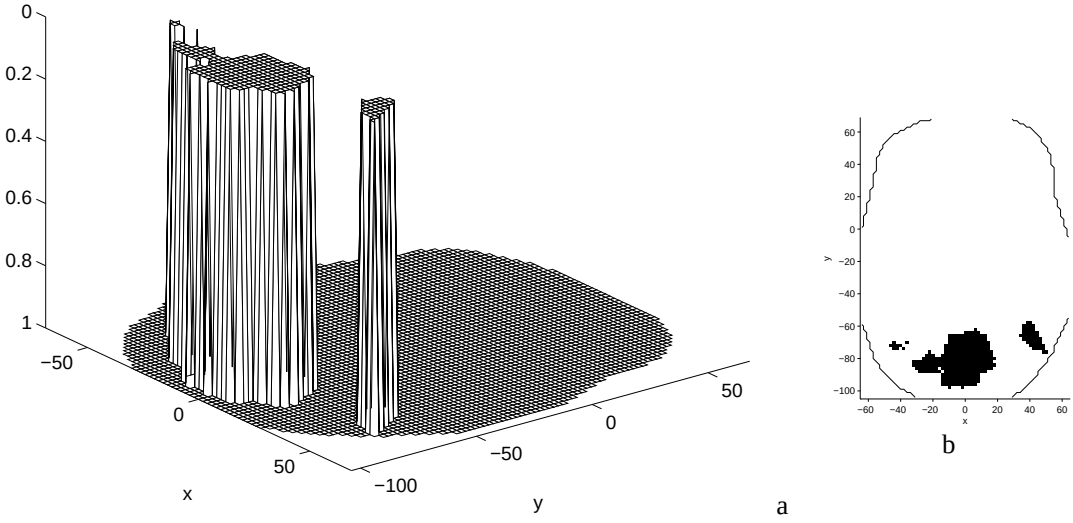

Figure 55

(a) Mesh plot of (one-sided) adjusted *p*-values, computed for voxels with values Z_k greater than the threshold $-\Phi^{-1}(0.0001)$, as the *p*-value for the size of the suprathreshold cluster containing that cluster (eqn.59). For voxels with subthreshold values Z_k, the adjusted *p*-value is set to zero. Voxels outside the intracerebral volume have been removed. (b) Voxels with adjusted *p*-value below level 0.05. The outline of the intracerebral area is superimposed. The AC-PC plane is shown.

3.6.3. Secondary smoothing

Finally, consider secondary smoothing, smoothing the statistic image and then normalising the variance. For this example, we shall use a Gaussian filter kernel of FWHM 10mm×10mm×6mm. The variance-covariance matrix, $\mathbf{\Sigma}$, of the kernel is therefore (appendix B:4):

$$\mathbf{\Sigma} = \begin{pmatrix} 10^2 & 0 & 0 \\ 0 & 10^2 & 0 \\ 0 & 0 & 12^2 \end{pmatrix} \frac{1}{8\ln(2)}$$

As discussed in §3.3.6.6., the Gaussianised *t*-statistic is smoothed with a discretisation of this kernel (appendix B:2), and the resulting image normalised by division by $\sqrt{c}$, where $c = 1/\sqrt{\left|2\mathbf{\Lambda}_Z\mathbf{\Sigma} + \mathbf{I}_3\right|}$. For the estimated variance-covariance matrix of partial derivatives of the Gaussianised *t*-statistic image $\hat{\mathbf{\Lambda}}_Z$, this gives $c = 0.3237$.

Secondary smoothing of the Gaussianised *t*-statistic image gives the new Gaussian statistic image, $\boldsymbol{SZ} = (SZ_1,\ldots,SZ_K)$, depicted in fig.56. Compare this with the Gaussianised *t*-statistic of fig.51 (note the different z-axis scales). The effect of secondary smoothing is dramatic.

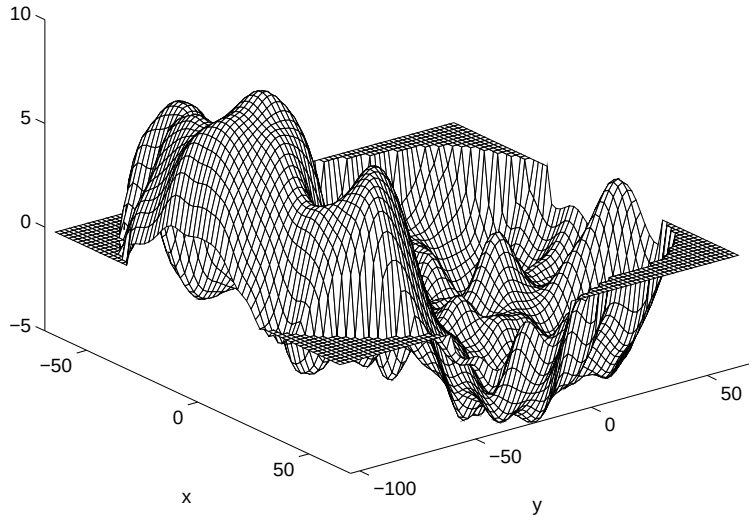

Figure 56
AC-PC plane of secondary smoothed Gaussianised *t*-statistic for the “V5” study.

Estimating smoothness for secondary smoothed statistic images

Assuming the Gaussianised *t*-statistic is a strictly stationary discrete standard Gaussian (zero mean, unit variance) random field, under the omnibus null hypothesis, the variance-covariance matrix of partial derivatives of the secondary smoothed Gaussianised *t*-statistic image is $\mathbf{\Lambda}_{SZ} = (2\mathbf{\Sigma} + \mathbf{\Lambda}_Z^{-1})^{-1}$ (appendix C:8). Here, this gives:

$$\hat{\mathbf{\Lambda}}_{SZ} = \begin{pmatrix} 0.0171 & 0 & 0 \\ 0 & 0.0164 & 0 \\ 0 & 0 & 0.0254 \end{pmatrix}$$

This corresponds to a Gaussian PRF of FWHM 12.73mm×13.01mm×10.44mm.

It is interesting to note that the estimated variances of the numerical partial derivatives within the secondary smoothed Gaussianised *t*-statistic image corresponds to a Gaussian PRF of FWHM 7.11mm×7.53mm×7.00mm. This is much lower than the theoretical one derived from the Gaussianised *t*-statistic, because the null hypothesis is not true. The signal in the Gaussianised *t*-statistic image is magnified by the normalisation, resulting in a statistic image whose voxel level variance is much greater than the hypothesised unit variance. This gives large variances and covariances for the partial derivatives (appendix C:9), leading to an underestimate of the smoothness.

Bonferroni assessment of the secondary smoothed Gaussianised t-statistic image

To illustrate the power of secondary smoothing, consider a Bonferroni analysis of the secondary smoothed Gaussianised *t*-statistic image.

The null hypothesis at voxel k is H_k: $\mu_k = 0$, where μ_k is the mean of the assumed Gaussian distribution at that voxel, $SZ_k \sim N(\mu_k,1)$. The unadjusted *p*-value for voxel k is then $P_k = 1-\Phi(SZ_k)$. Bonferroni adjusted *p*-values are computed as $\tilde{P}_k = \min\{KP_k, 1\}$. The Bonferroni adjusted *p*-values for the AC-PC plane are shown below (fig.57). Compare these with the corresponding figure for the *t*-statistic (fig.50). (The *p*-values for the *t*-statistic image and the Gaussianised *t* are identical, by construction of the Gaussianised *t*.)

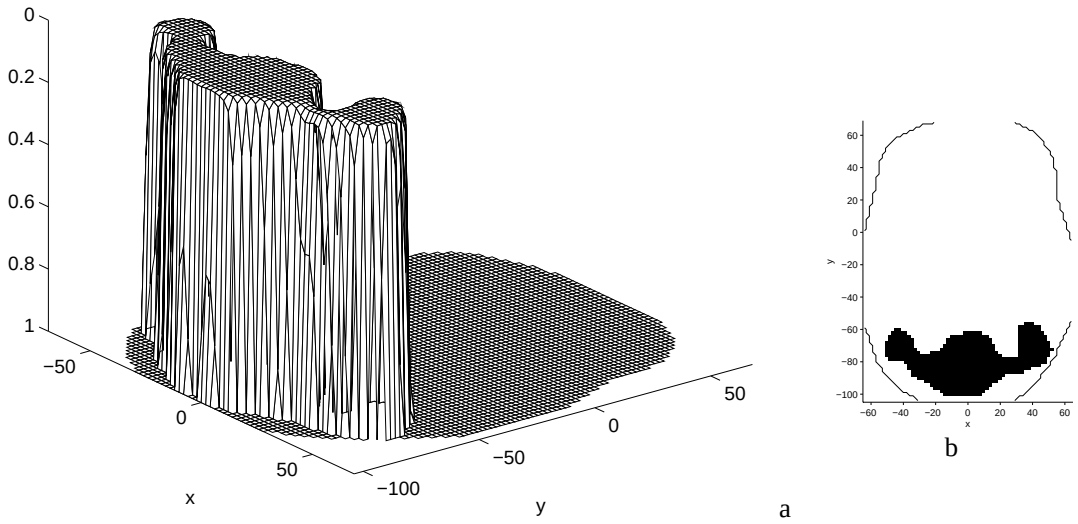

Figure 57

(a) Mesh plot of Bonferroni single step adjusted one-sided *p*-values, computed from the secondary smoothed Gaussianised *t*-statistic image of fig.56. Voxels outside the intracerebral volume have been removed. (b) Voxels with adjusted *p*-value below level 0.05. The outline of the intracerebral area is superimposed. The AC-PC plane is shown.