

Developments in Random Field Theory

K.J. Worsley

March 4, 2003

Random field theory is used in the statistical analysis of SPM's whenever there is a spatial component to the inference. Most important is the question of detecting an effect or activation at an unknown spatial location. Very often we do not know in advance where to look for an effect, and we are interested in searching the whole brain, or part of it. This presents special statistical problems related to the problem of multiple comparisons, or multiple tests. Two methods have been proposed, the first based on the maximum of the T or F statistic, the second based on the spatial extent of the region where these statistics exceed some threshold value. Both involve results from random field theory (Adler, 1981).

1 The maximum test statistic

An obvious method is to select those locations where a test statistic Z (which could be a T , χ^2 , F or Hotelling's T^2 statistic) is large, that is, to threshold the image of Z at a height z . The problem is then to choose the threshold z to exclude false positives with a high probability, say 0.95. Setting z to the usual (uncorrected) $P = 0.05$ critical value of Z (1.64 in the Gaussian case) means that 5% of the unactivated parts of the brain will show false positives. We need to raise z so that the probability of finding any activation in the non-activated regions is 0.05. This is a type of multiple comparison problem, since we are testing the hypothesis of no activation at a very large number of voxels.

A simple solution is to apply a Bonferroni correction. The probability of detecting any activation in the unactivated locations is bounded by assuming that the unactivated locations cover the entire search region. By the Bonferroni inequality, the probability of detecting any activation is further bounded by

$$P(\max Z > z) \leq N P(Z > z), \quad (1)$$

where the maximum is taken over all N voxels in the search region. For a $P = 0.05$ test of Gaussian statistics, critical thresholds of 4-5 are common. This procedure is conservative if the image is smooth, although for fMRI data it often gives very accurate thresholds.

Random field theory gives a less conservative (lower) P -value if the image is smooth:

$$P(\max Z > z) \approx \sum_{d=0}^D \text{Resels}_d \text{EC}_d(z), \quad (2)$$

where D is the number of dimensions of the search region, Resels_d is the number of d -dimensional resels (resolution elements) in the search region, and $\text{EC}_d(z)$ is the d -dimensional Euler characteristic density. The approximation (2) is based on the fact that the left hand side is the exact expectation of the Euler characteristic of the region above the threshold z . The Euler characteristic counts the number of clusters if the region has no holes, which is likely to be the case if z is large. Details can be found in Worsley *et al.* (1996a).

The approximation (2) is accurate for search regions of any size or shape, even a single point, but it is best for search regions that are not too concave. Sometimes it is better to surround a highly convoluted search region, such as grey matter, by a convex hull with slightly higher volume but less surface area, to get a lower and more accurate P -value.

For large search regions, the last term ($d = D$) is the most important. The number of resels is

$$\text{Resels}_D = V/\text{FWHM}^D,$$

where V is the volume of the search region and FWHM is the effective Full Width at Half Maximum of a Gaussian kernel used to smooth the data (see Chapter 14). The corresponding EC density for a T statistic image with ν degrees of freedom is

$$\text{EC}_3(z) = \frac{(4 \log_e 2)^{\frac{3}{2}}}{(2\pi)^2} \left(\frac{\nu - 1}{\nu} z^2 - 1 \right) \left(1 + \frac{z^2}{\nu} \right)^{-\frac{1}{2}(\nu-1)}.$$

For small search regions, the lower dimensional terms $d < D$ become important. However the P -value (2) is not very sensitive to the shape of the search region, so that assuming a spherical search region gives a very good approximation.

Figure 1 shows the threshold z for a $P = 0.05$ test calculated by the two methods. If the FWHM is small relative to the voxel size, then the Bonferroni threshold is actually less than the random field one (2). In practice it is better to take the minimum of the two thresholds (1) and (2).

EC densities for F fields can be found in Worsley *et al.* (1996a), and for Hotelling's T^2 , see Cao and Worsley (1999a). Similar results are also available for correlation random fields, useful for detecting functional connectivity. See Cao and Worsley (1999b).

Extensions of the result (2) to scale space random fields are given in Worsley *et al.* (1996b). Here the search is over all spatial filter widths as well over location, so that the width of the signal is estimated as well as its location. The price to pay is an increase in critical threshold of about 0.5.

2 The maximum spatial extent of the test statistic

An alternative test can be based on the spatial extent of clusters of connected components of supra threshold voxels where $Z > z$ (Friston *et al.*, 1994). Typically z is chosen to be about 3 for a Gaussian random field. Once again the image must be a smooth stationary random field. The idea is to approximate the shape of the image by a quadratic with a peak at the local maximum. For a Gaussian random field, the spatial extent S is then approximated by the volume where the quadratic of height H above z cuts the threshold z :

$$S \approx cH^{D/2}, \tag{3}$$

where

$$c = \text{FWHM}^D (2\pi/z)^{D/2} (4 \log 2)^{-D/2} / \Gamma(D/2 + 1).$$

For large z , the upper tail probability of H is well approximated by

$$P(H > h) = P(\max Z > z + h) / P(\max Z > z) \approx \exp(-zh), \quad (4)$$

from which we conclude that H has an approximate exponential distribution with mean $1/z$. From this we can find the approximate P -value of the spatial extent S of a single cluster:

$$P(S > s) \approx \exp(-z(s/c)^{2/D}). \quad (5)$$

The P -value for the largest spatial extent is obtained by a simple Bonferroni correction for the expected number of clusters K :

$$P(\max S > s) \approx E(K) P(S > s), \quad \text{where} \quad E(K) \approx P(\max Z > z) \quad (6)$$

from (2).

We can substantially improve the value of the constant c by equating the expected total spatial extent, given by $V P(Z > z)$, to that obtained by summing up the spatial extents of all the clusters S_1, \dots, S_K :

$$V P(Z > z) = E(S_1 + \dots + S_K) = E(K) E(S).$$

Using the fact that

$$E(S) \approx c \Gamma(D/2 + 1) / z^{D/2}$$

from (3), and the expected number of clusters from (2), it follows that

$$c \approx \text{FWHM}^D z^{D/2} P(Z > z) / \{E C_D(z) \Gamma(D/2 + 1)\}.$$

Cao (1999) has extended these results to T , χ^2 and F fields, but unfortunately there are no theoretical results for non-smooth fields such as raw fMRI data.

3 Searching in small regions

For small pre-specified search regions such as the cingulate, the P -values for the maximum test statistic are very well estimated by (2), but the results in section 2 only apply to large search regions. Friston (1997) has proposed a fascinating method that avoids the awkward problem of pre-specifying a small search region altogether. We threshold the image of test statistics at z , then simply pick the nearest peak to a point or region of interest. The clever part is this. Since we have identified this peak based only on its spatial location and not based on its height or extent, there is now no need to correct for searching over all peaks. Hence, the P -value for its spatial extent S is simply $P(S > s)$ from (5), and the P -value for its peak height H above z is simply $P(H > h)$ from (4).

4 Estimating the FWHM

The only data-dependent component required for setting the above thresholds is Resels_D , and indirectly, the FWHM. The FWHM often depends on the location: fMRI data is considerably smoother in cortex than white matter (see Figure 2), and for VBM data FWHM varies considerably from one location to another. This means that the random field is not isotropic, so the above random field theory is not valid. Fortunately there is a simple way of allowing for this by estimating the FWHM separately at each voxel.

Let \mathbf{r} be the n -vector of least-squares residuals from the (possibly whitened) linear model fitted at each voxel, and let \mathbf{u} be the vector of normalised residuals $\mathbf{u} = \mathbf{r}/(\mathbf{r}'\mathbf{r})^{1/2}$. Let $\dot{\mathbf{u}}$ be the $n \times 3$ spatial derivative of \mathbf{u} in the three orthogonal directions of the voxel lattice. The estimated FWHM is

$$\widehat{\text{FWHM}} = (4 \log 2)^{1/2} |\dot{\mathbf{u}}' \dot{\mathbf{u}}|^{-1/(2D)} \quad (7)$$

and the estimated Resels_D is

$$\widehat{\text{Resels}}_D = \sum_{\text{volume}} \widehat{\text{FWHM}}^{-D} v,$$

where summation is over all voxels in the search region and v is the volume of a single voxel (Worsley *et al.*, 1998). The extra randomness added by estimating Resels_D can be ignored if the search region is large.

However, spatially varying FWHM can have a strong effect on the validity of the P -value for spatial extent. If the cluster is in a region where FWHM is large, then its extent will be larger by chance alone, and so its P -value will be too small. In other words, clusters will look more significant in smooth regions than in rough regions of the image. To correct for this, we simply replace cluster volume by cluster resels, defined as

$$\tilde{S} = \sum_{\text{cluster}} v \widehat{\text{FWHM}}^{-D}$$

where summation is over all voxels in the cluster.

There is one remaining problem: since the above summation is over a small cluster, rather than a large search region, the randomness in estimating FWHM now makes a significant contribution to the randomness of \tilde{S} , and hence its P -value. Worsley (2002) suggests allowing for this by the approximation

$$\tilde{S} \approx \tilde{c} H^{D/2} \prod_{k=1}^{D+1} X_k^{p_k}, \quad (8)$$

where X_1, \dots, X_{D+1} are independent χ^2 random variables. The degrees of freedom of X_k is $\nu - k + 1$ where $\nu = n - p$ and p is the number of regressors in the linear model, raised to the power $p_k = -D/2$ if $k = 1$ and $p_k = 1/2$ if $k > 1$. Again the constant \tilde{c} is chosen so that the expected total resels of all clusters matches the probability of exceeding the threshold times the volume of the search region

$$\tilde{c} \approx z^{D/2} \text{P}(Z > z) / \{\text{EC}_D(z) \Gamma(D/2 + 1)\}.$$

Combining this with the approximate distributions of spatial extents for T , χ^2 and F fields from Cao (1999) requires no extra computational effort. H is replaced by a Beta random

Proportion of true + in image	1	0.1	0.01	0.001	0.0001
FDR threshold	1.64	2.56	3.28	3.88	4.41
Number of voxels in image	1	10	100	1000	10000
Bonferroni threshold	1.64	2.58	3.29	3.89	4.42
Number of resels in image	0	1	10	100	1000
Random fields threshold	1.64	2.82	3.46	4.09	4.65

Table 1: Examples of the thresholds of False Discovery Rate, Bonferroni and random field methods for thresholding an image.

variable in (8), multiplied by powers of yet more χ^2 random variables, with appropriate adjustments to \tilde{c} .

In practice, the distribution function of \tilde{S} is best calculated by first taking logarithms, so that $\log \tilde{S}$ is then a sum of independent random variables. The density of a sum is the convolution of the densities, whose Fourier transform is the sum of the Fourier transforms. It is easier to find the upper tail probability of $\log \tilde{S}$ by replacing the density of one of the random variables by its upper tail probability *before* doing the convolution. The obvious choice is the exponential or Beta random variable, since its upper tail probability has a simple closed form expression. This method has been implemented in the `stat_threshold.m` function of `fmristat`, available from <http://www.math.mcgill.ca/keith/fmristat>.

5 False Discovery rate

A remarkable breakthrough in multiple testing was made by Benjamini and Hochberg in 1995 who took a completely different approach. Instead of controlling the probability of ever reporting a false positive, they devised a procedure for controlling the *false discovery rate* (FDR), the expected proportion of false positives amongst those voxels declared positive (the *discoveries*) - see Figure 3. The procedure is extremely simple to implement. Simply calculate the uncorrected P -value for each voxel and order them so that the ordered P -values are $P_1 \leq P_2 \leq \dots \leq P_N$. To control the FDR at α , find the largest value k so that $P_k < \alpha k/N$. This procedure is conservative if the voxels are positively dependent, which is a reasonable assumption for most unsmoothed or smoothed imaging data. See Genovese *et al.* (2002) for an application of this method to fMRI data, and for further references.

The resulting threshold, corresponding to the value of Z for P_k , depends on the amount of signal in the data, not on the number of voxels or the smoothness. Table 1 compares thresholds for the FDR, Bonferroni and random field methods. Thresholds of 2-3 are typical for brain mapping data with a reasonably strong signal, quite a bit lower than the Bonferroni or random field thresholds.

But we must remember that the interpretation of the FDR is quite different. False positives will be detected; we are simply controlling them so that they make up no more than α of our discoveries. On the other hand, the Bonferroni and random field methods control the probability of *ever* reporting a false discovery (see Figure 3).

6 Conclusion

The idea of using a hypothesis test to detect activated regions does contain a fundamental flaw that all experimenters should be aware of. Think of it this way: if we had enough data, T statistics would increase (as the square root of the number of scans or subjects) until *all* voxels were ‘activated’! In reality, *every* voxel must be affected by the stimulus, perhaps by a very tiny amount; it is impossible to believe that there is never any signal at all. So thresholding simply excludes those voxels where we don’t yet have enough evidence to distinguish their effects from zero. If we had more evidence, perhaps with better scanners, or simply more subjects, we surely would be able to do so. But then we would probably not want to detect activated regions. As for satellite images, the job for statisticians would then be signal *enhancement* rather than signal detection. The distinguishing feature of most brain mapping data is that there is so little signal to enhance. Even with the advent of better scanners this is still likely to be the case, because neuroscientists will surely devise yet more subtle experiments that are always pushing the signal to the limits of detectability.

REFERENCES

- Adler, R.J. (1981). *The Geometry of Random Fields*. New York: Wiley.
- Cao, J. (1999). The size of the connected components of excursion sets of χ^2 , t and F fields. *Advances in Applied Probability*, **31**:577-593.
- Cao, J. & Worsley, K.J. (1999a). The detection of local shape changes via the geometry of Hotelling’s T^2 fields. *Annals of Statistics*, **27**:925-942.
- Cao, J. & Worsley, K.J. (1999b). The geometry of correlation fields, with an application to functional connectivity of the brain. *Annals of Applied Probability*, **9**:1021-1057.
- Friston, K.J., Worsley, K.J., Frackowiak, R.S.J., Mazziotta, J.C. & Evans, A.C. (1994). Assessing the significance of focal activations using their spatial extent. *Human Brain Mapping*, **1**:214-220.
- Friston, K.J. (1997). Testing for anatomically specified regional effects. *Human Brain Mapping*, **5**:133-136.
- Genovese, C.R., Lazar, N.A. & Nichols, T.E. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *NeuroImage*, **15**:772-786.
- Worsley, K.J., Marrett, S., Neelin, P., Vandal, A.C., Friston, K.J. & Evans, A.C. (1996a). A unified statistical approach for determining significant signals in images of cerebral activation. *Human Brain Mapping*, **4**:58-73.
- Worsley, K.J., Marrett, S., Neelin, P. & Evans, A.C. (1996b). Searching scale space for activation in PET images. *Human Brain Mapping*, **4**:74-90.
- Worsley, K.J., Andermann, M., Koulis, T., MacDonald, D. & Evans, A.C. (1999). Detecting changes in nonisotropic images. *Human Brain Mapping*, **8**:98-101.

Worsley, K.J. (2002). Non-stationary FWHM and its effect on statistical inference for fMRI data. *NeuroImage*, **15**:S346.

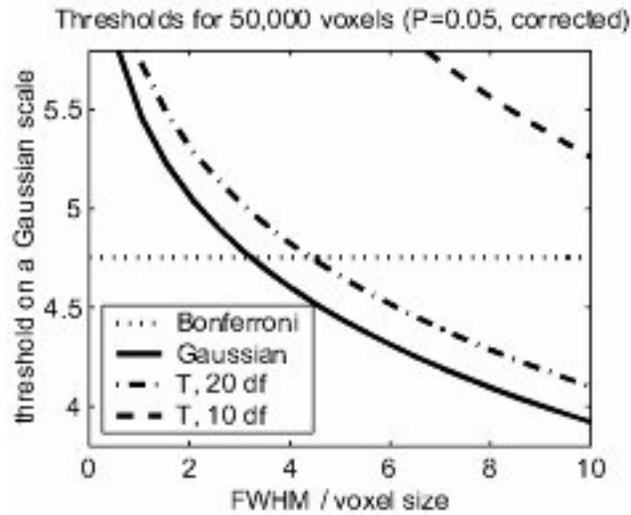


Figure 1: Thresholds for a volume with $N=50,000$ voxels ($P = 0.05$, corrected). Note that if the FWHM is less than 3.2 voxels, then the Bonferroni method is better than the random field method for a Gaussian statistic. For T statistics with $\nu = 20$ df, this limit is higher (4.4), and much higher (14.4 - off scale) with $\nu = 10$ df.

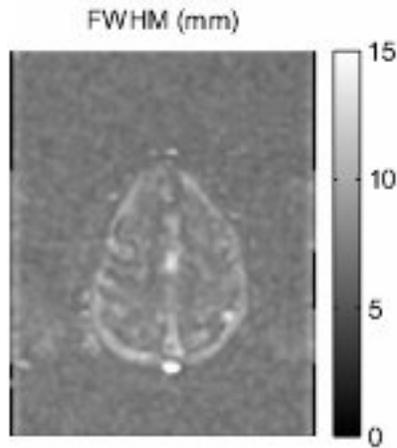


Figure 2: The estimated FWHM for one slice of fMRI data. Note the ~ 6 mm FWHM outside the brain due to smoothing imposed by motion correction. The FWHM in cortex is much higher, ~ 10 mm, while white matter is lower ~ 6 mm.

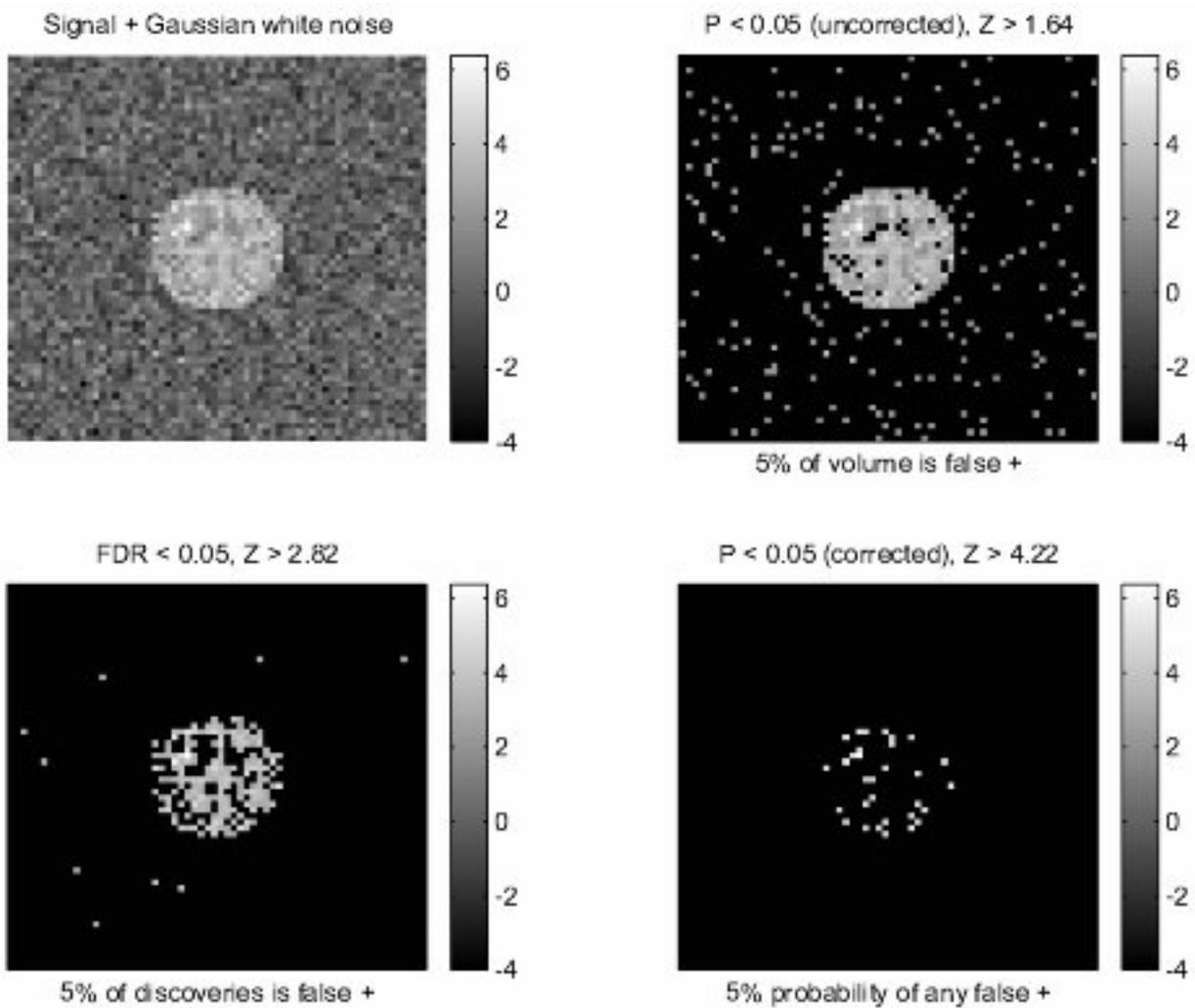


Figure 3: Illustration of the difference between False Discovery Rate and Bonferroni/random field methods for thresholding an image.