**Cluster Analysis Revisited**

FMRIB Technical Report TR00DF1

**David E. Flitney and Mark Jenkinson**

Oxford Centre for Functional Magnetic Resonance Imaging of the Brain (FMRIB),
Department of Clinical Neurology, University of Oxford, John Radcliffe Hospital,
Headley Way, Headington, Oxford, UK

**Abstract**

To complete the analysis of any FMRI experiment it is necessary to compensate for the multiple comparisons problem inherent in making a large number of simultaneous statistical tests. One method for accomplishing this is to employ cluster analysis to bolster one's confidence in any particular statistical result. The following review combines the well known works in this field to form the basis of an implementation of the Gaussian random field theory applied to the analysis of FMRI statistic images.

**keywords**: FMRI, Gaussian Random Fields, smoothness estimation, multiple comparisons, statistical significance.

# 1 Introduction

Cluster analysis is based on the observation that it is more likely that activations should form clusters rather than being spatially isolated events. Therefore, a SPM, thresholded at a particular statistic level, is further analysed to identify all the unique clusters of voxels. There is, however, the possibility that even these clusters might be a chance occurence. The theory of Gaussian random fields [1] allows us to model the probability that a cluster occurs as a chance process and so, to form our final result, we evaluate each cluster accordingly rejecting those clusters whose probability of being due to chance is unacceptably high.

# 2 Theory

## 2.1 Statistical inference

The set of voxels exceeding some threshold $u$ comprises $m$ clusters each of $n$ voxels. The expectations of $m$, $n$ and the total number of voxels above threshold, $N$, are related

$$E\{N\} = E\{m\} \cdot E\{n\}. \tag{1}$$

For our purposes we can calculate all the necessary information knowing $E\{m\}$ and this can be approximated by

$$E\{m\} \approx V(2\pi)^{-2}|\Lambda|^{1/2}(t^2-1)e^{-t^2/2}, \tag{2}$$

where $W$ is an estimate of the smoothness of the GRF, $D$ the number of dimensions in the GRF and $S$ is the number of *in-brain* voxels.

The probability of $c$ clusters in this *excursion set* is approximated by a Poisson

$$P(m=c) \approx \lambda(c, E\{m\}) \tag{3}$$

$$= \frac{E\{m\}^c \cdot e^{-E\{m\}}}{c!}. \tag{4}$$

The number of voxels $n$ comprising a cluster is distributed according to

$$P(n \geq k) \approx e^{-\beta k^{2/D}}, \tag{5}$$

where

$$\beta = \left[ \frac{\Gamma\left(\frac{D}{2}+1\right) E\{m\}}{(S \cdot \Phi(-u))} \right]^{\frac{2}{D}}. \tag{6}$$

We can calculate $P(n_{max} \geq k)$ [3] by calculating the product of $P(n = m)$ and $1 - P(n < k)$ summed over $m$

$$P(n_{max} \geq k) = \sum_{i=1}^{\infty} P(m = i) \cdot [1 - P(n < k)^i] \tag{7}$$

$$= 1 - \exp\left(-E\{m\} \cdot P(n \geq k)\right) \tag{8}$$

$$= 1 - \exp\left(-E\{m\} e^{-\beta k^{2/D}}\right). \tag{9}$$

## 2.2 Smoothness estimation

The smoothness estimator, $W$, needs to be measured from the residuals of the linear model fit. This estimator explains the spatial correlation structure of the SPM by assuming that it can be modeled as being due to a convolution of the (white) real signal with a Gaussian filter with an unknown, but determinable, width. Calculating the width of this Gaussian allows us to determine the effective resolution of the data.

### 2.2.1 Smoothness *a la* SPM

The current implementations of SPM (SPM99) use the smoothness estimation as described in Kiebel *et al* [5].

For a Gaussian random field the smoothness is defined as $W = |\mathbf{\Lambda}|^{-1/2D}$ where $D$ is the dimensionality of the field and $\mathbf{\Lambda}$ the covariance matrix of it's first partial derivatives. An unbiased estimator for the covariance of the partial derivatives is given by

$$\hat{\lambda}_{jk} = \frac{\nu - 2}{(\nu - 1)N} \sum_{i}^{N} \frac{1}{M} \sum_{t}^{M} \left( \frac{\partial S_{it}}{\partial x_j} \frac{\partial S_{it}}{\partial x_k} \right), \tag{10}$$

where $\nu$ is the number of degrees of freedom, $N$ the number of voxel positions ($i$) and $M$ the number of time points ($t$) in a FMRI time series. In fact we don't have to evaluate this equation for the whole covariance matrix. The off-diagonal elements will be zero so we only need to calculate the diagonal elements of $\mathbf{\Lambda}$

$$\hat{\lambda}_{jj} = \frac{\nu - 2}{(\nu - 1)N} \sum_{i}^{N} \frac{1}{M} \sum_{t}^{M} \left( \frac{\partial S_{it}}{\partial x_j} \right)^2. \tag{11}$$

The partial derivative is evaluated via the gradient operator

$$\nabla S_i = \frac{S_{i+} - S_{i-}}{2\delta d}, \tag{12}$$

for non-edge voxels, where $S_{i+}$ and $S_{i-}$ are $S_i$s neighbouring voxels in the dimension along which the derivative is currently being evaluated.

Now we can calculate the FWHM of the theoretical Gaussian responsible for the observed smoothness

$$FWHM_i = \sqrt{8 \cdot ln(2) \cdot W_{ii}}, \tag{13}$$

where $W_{ii}$ is the variance, $\sigma^2$, of the Gaussian point-spread-function computed from the covariance measure for the $i$th dimension

$$W_{ii} = \frac{1}{2\lambda_{ii}}. \tag{14}$$

A combined value for the FWHM is given by the geometric mean of the various $\mathrm{FWHM}_i$

$$FWHM = \left[\prod_{i=0}^{D} FWHM_i\right]^{1/D}. \tag{15}$$

#### 2.2.2 A more robust smoothness estimator

The estimator described in section 2.2.1 becomes increasingly inaccurate for sub voxel width filter sizes. As the filter width decreases equation 12 becomes increasingly inaccurate until the estimation in equation 14 is no longer valid. An alternative estimate [2] offers a more robust result when evaluating images with small spatial correlation. Consider the following definitions

$$S^2 = \frac{1}{N}\sum_{i}^{N}\frac{1}{M}\sum_{t}^{M} {S_{i,t}}^2 \tag{16}$$

$$(\nabla S)^2 = \frac{1}{N}\sum_{i}^{N}\frac{1}{M}\sum_{t}^{M} (S_{i,t} - S_{i-\mathbf{1},t})^2 \tag{17}$$

$$SS_- = \frac{1}{N}\sum_{i}^{N}\frac{1}{M}\sum_{t}^{M} (S_{i,t}S_{i-\mathbf{1},t}) \tag{18}$$

where **1** is a unit vector along the dimension currently under consideration. $S^2$ is the individual voxel variance, $(\nabla S)^2$ the variance of the difference between each voxel and its edgewise neighbours, and $SS_-$ is the correlation of two neighbouring voxels.

The approach outlined in 2.2.1 calculates the quantity

$$s = \sqrt{\frac{1}{2(\nabla S)^2}}. \tag{19}$$

Forman *et al* derive the alternate formula

$$s = \sqrt{\frac{-1}{4 \cdot ln\left(1 - \frac{(\nabla S)^2}{2S^2}\right)}}. \tag{20}$$

Mark Jenkinson [4] has provided an extensive derivation for this formula as well as pointing out another, computationally cleaner, variation on the same theme

$$s = \sqrt{\frac{-1}{4 \cdot ln\left(\frac{SS_-}{S^2}\right)}}. \tag{21}$$

Equations 20 and 21 remain accurate for relatively small spatial smoothnesses and it is equation 21 that you will find in the FSL code base.

### 2.3 Calculating the smoothness of $Z_0$

The estimate of the covariance matrix for the residual field, $\hat{\Lambda}_\epsilon$, needs to be transformed to $Z_0$ for calculating the probabilities of clusters found in the $Z$-statistic image. The estimate of the covariance matrix for the $Z_0$ field, $\hat{\Lambda}_{Z_0}$ can be computed by

$$\hat{\Lambda}_{Z_0} = \lambda_\nu \cdot \hat{\Lambda}_\epsilon, \tag{22}$$

where

$$\lambda_\nu = \int_{-\infty}^{\infty} \frac{(t^2+\nu-1)^2}{(\nu-1)(\nu-2)} \frac{T_\nu(t)^3}{p(t)^2} dt, \tag{23}$$

where $T_\nu$ is the PDF of a $t$-distribution with $\nu$ degrees of freedom and

$$p(t) = \phi(\Phi^{-1}(1 - \Phi_\nu(t))), \tag{24}$$

where $\phi(z)$ is the PDF of the standard normal distribution.

# 3 Implementation issues

## 3.1 Calculating $\lambda_\nu$

Equation 23 is evaluated numerically to yield a look up table of values for different degrees of freedom. To aid calculation we change the limits as follows

$$\lambda_\nu = \frac{1}{(\nu-1)(\nu-2)} \int_{-\infty}^{\infty} \frac{(t^2+\nu-1)^2 T_\nu(t)^3}{p(t)^2} dt, \tag{25}$$

let $w = 1/t$ so that $dt = -w^{-2}dw$ giving

$$= \frac{2}{(\nu-1)(\nu-2)} \left( \int_0^1 \frac{(t^2+\nu-1)^2 T_\nu(t)^3}{p(t)^2} dt + \int_0^1 \frac{(1/w^2+\nu-1)^2 T_\nu(1/w)^3}{p(1/w)^2} dw \right) \tag{26}$$

## 3.2 Standardization of the residuals

A pre-requisite to much of the analysis is that the residual time series be standardized. The standardizing $z$ transform, $z = \frac{x-\mu}{\sigma}$, is applied to each time series

$$z = \frac{x - \bar{x}}{s}, \tag{27}$$

where the computational formula for $s$ is

$$s = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{N}}{N-1}}. \tag{28}$$

## 3.3 Error function

The error function, $\Phi(x)$, is defined as

$$\Phi(x) = \frac{2}{\sqrt{\pi}} \int_{t=0}^{x} e^{-t^2} dt. \tag{29}$$

A suitable approximation can be found in **libprob**.

## 3.4 Gamma function

The gamma function, $\Gamma(x)$, is defined as

$$\Gamma(x) = \int_{u=0}^{\infty} u^{x-1} e^{u} du. \tag{30}$$

Once again **libprob** contains a suitable numeric evaluation.

### 3.5 *t*-distribution

The density function of the $t$-distribution is parameterised by the number of degrees of freedom, $n$, and is given by

$$f(t) = \frac{\Gamma\left[(n+1)/2\right]}{\sqrt{n\pi}\,\Gamma(n/2)} \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2} \tag{31}$$

# References


[1] R. J. Adler and A. M. Hasofer. *The geometry of random fields*. Wiley, New York, 1981.

[2] S. D. Forman, J. D. Cohen, M. Fitzgerald, W. F. Eddy, M. A. Mintum, and D. C. Noll. Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): Use of a cluster-size threshold. *Magnetic Resonance in Medicine*, 33:636–647, 1995.

[3] K. Friston, K. Worsley, R. Frackowiak, J. Mazziotta, and A. Evans. Assessing the significance of focal activations using their spatial extent. *Human Brain Mapping*, 1:214–220, 1994.

[4] M. Jenkinson. Estimation of smoothness from the residual field. Internal Technical Report TR00MJ3, Oxford Centre for Functional Magnetic Resonance Imaging of the Brain, Department of Clinical Neurology, Oxford University, Oxford, UK, 2000.

[5] S. J. Kiebel, J.-B. Poline, K. J. Friston, A. P. Holmes, and K. J. Worsley. Robust smoothness estimation in statistical parametric maps using standardized residuals from the general linear model. *NeuroImage*, 10:756–766, 1999.