

Statistical Inference and the Theory of Random Fields

J-B. Poline A.P. Holmes K.J. Worsley K.J. Friston

Contents

1	Introduction	2
2	Testing for the intensity of an activation in SPMs	3
2.1	Theory	3
2.2	Assumptions	5
2.3	Discussion	6
3	Testing for the significance of the spatial extent of an activation	6
3.1	Theory	6
3.2	Discussion	7
4	Testing for both peak height and spatial extent	9
4.1	Rationale	9
4.2	Method	9
4.3	Discussion	11
5	Testing for the significance of a set of regions	12
5.1	Theory	13
5.2	Power analysis	13
5.3	Discussion	14
6	Non parametric approaches: Statistical <i>Non</i> Parametric Mapping (SnPM)	15
6.1	Rationale and Method	15
6.2	Results and discussion	16
7	Discussion and conclusion	17
7.1	Which test should be used and when ?	17
7.2	Sensitivity and Specificity	18
7.3	How do we choose the parameters ?	18
7.4	Smoothness Variability	19
7.5	Smoothness estimation on the residuals	20
7.6	A priori hypothesis	22
7.7	Testing for commonalities	22
	References	23

1 Introduction

Statistical methods used to analyse functional neuroimaging data are essential for a proper interpretation of the results of experiments that ultimately aim at a better understanding of the neuroanatomy of human brain function. The analysis of functional imaging experiments often involves the formation of a Statistical Parametric Map (SPM). The conceptual idea of SPMs was first introduced by Friston *et al.* (1990). In such maps, the value at each position or voxel is a statistic that expresses evidence against a null hypothesis of no experimentally induced activation at that voxel. The construction of an SPM can be decomposed into three main steps :

- Spatial transformations: In the most general case, functional imaging experiments require the acquisition of data from several subjects, or several groups of subjects. Sophisticated techniques have been designed to normalise the anatomy of different brains into a standard stereotactic space (Friston *et al.*, 1996b; Ashburner & Friston, 1997). Spatial smoothing is also usually performed to allow for interindividual gyral variation and to improve the signal to noise ratio. Note that smoothing does not always improve the signal to noise ratio and the relationship between smoothing and sensitivity is discussed further in this chapter.
- Construction of an SPM. This is a key step because it requires the (generally non unique) modelling of effects of interest or of no interest for the experimental protocol analysed. The General Linear Model (GLM) offers the flexibility needed. This step is fully described in the previous chapter (Holmes & Friston, 1997), the output of which is a three dimensional (3D) statistic image or “map” formed of thousands of correlated Student t statistics.
- Statistical inference from the SPM. This step is the focus of this chapter.

Images contains a great number of voxels so that the SPM are not directly interpretable. An essential step was to find a way to correct for the multiple comparison problem. A difficulty with this correction lies in the non independence of voxel intensities due to both the initial resolution of images and to post processing smoothing. The non independence of voxels cannot be treated by “Bonferroni” procedures that treat voxels as if they were independent because they are much too stringent and would wipe out statistically reliable activation signals from the results.

Since the first attempts to analyse a voxel based activation map, a number of statistical techniques have been developed for the analysis of SPMs. Essential to the development of these techniques is (Gaussian) Random Field Theory that deals with the behaviour of stochastic processes defined over a space of any dimensions (D). Usually, D is 3 (analysis of a volume) but can be greater (e.g search over time, with potential application in *fMRI*, or search over scale space multi filtering strategy). In this chapter we review for the general reader some important tests (based on results from this field of mathematics) can be used for the assessment of significant activations in SPMs. These techniques have become increasingly important because they are general, require very little computation and provide an extensive characterisation of the different kinds of response expected in activation studies.

This chapter is organised in (almost) chronological order and we will show that this order also corresponds to the different kinds of risk of error associated with the different statistical tests described. These tests can be looked upon as a hierarchy of procedures with decreasing localising power but potentially increasing sensitivity.

We briefly review some important extensions to these statistical tests, and introduce alternative non parametric approaches which do not use Random Fields theory and are free from any assumptions (Holmes *et al.*, 1996). Finally we discuss relevant issues related to image smoothness.

2 Testing for the intensity of an activation in SPMs

2.1 Theory

Friston *et al.* (1991) proposed a procedure addressing the multiple comparison problem. Using very basic results on random processes (Cox & Miller, 1990), they derived a test for bi-dimensional (2D) processes that efficiently controlled for non independence in the data.

Building on this result, Worsley *et al.* (1992) used a mathematically more conventional procedure to extend the test in three or more dimensions.

We describe briefly how this was achieved emphasising assumptions about the volume or image to be analysed and critically assess validity of these correction procedures whilst proposing practical guidelines. Although results are available for different random fields (Worsley, 1994), we will concentrate on the use of the results established for Gaussian random fields. t -maps, usually generated by testing contrasts, are therefore transformed to Gaussianised t -maps using a voxel by voxel t -to- Z probability transformation such that $\Phi(Z) = \Psi(t)$, where $\Phi(\bullet)$ is the standard normal cumulative density function (CDF) and $\Psi(\bullet)$ the Student t -distribution with appropriate degrees of freedom).

To test for the significance for an activation intensity in a SPM, it is necessary to assess the probability that the maximum value in the map (Z_{max}) is greater than a given threshold t under the null hypothesis (when no activation is present). To approximate this probability Worsley *et al.* used the expected Euler characteristic $\mathbf{E}[\chi_t]$ of a binarised map thresholded at t . The Euler characteristic is a geometrical measure that, loosely speaking, counts the number of connected components minus the number of ‘‘holes’’ in volume of the image V . At high thresholds this characteristic simply counts the number of regions above t . Moreover, for such high thresholds, suprathreshold clusters are independent and the number of clusters C_t above t follows approximately a Poisson distribution (Adler, 1981, p.161) with mean $\mathbf{E}[\chi_t]$, i.e.

$$\Pr(C = x) = 1/x!(\mathbf{E}[\chi_t])^x e^{-\mathbf{E}[\chi_t]} = \Upsilon(x, \mathbf{E}[\chi_t]) \quad (1)$$

For high t , we have

$$\Pr(Z_{max} \geq t) \approx \Pr(\chi_t \geq 1) \approx 1 - e^{-\mathbf{E}[\chi_t]} \approx \mathbf{E}[\chi_t] \quad (2)$$

where χ_t is defined over a compact convex subset V of \mathfrak{R}^D whose boundary has zero Lebesgue measure (λ). The expected Euler characteristic is:

$$\mathbb{E}[\chi_t] = \lambda(V)|\mathbf{\Lambda}|^{1/2}(2\pi)^{-(D+1)/2}He_D(t)e^{-t^2/2} \quad (3)$$

where $\lambda(V)$ is the volume or image being analysed (V and $\mathbf{\Lambda}$ are measured with the same units. We make the simplification $V = \lambda(V)$ in the rest of the chapter). $He_D(t)$ is the Hermite polynomial of degree D in t ($He_0(t) = 1; He_1(t) = t; He_2(t) = t^2 - 1$). Notice that the ‘‘threshold’’ t here is not set by the user : the value of t is simply the local maxima or indeed any value that is tested, as opposed to the threshold used for spatial extent tests (see section (3)).

$\mathbf{\Lambda}$ is the variance covariance matrix of the partial derivative of the process in the D directions of space and is crucial for the assessment of $\mathbb{E}[\chi_t]$ and therefore to the calculation of p-values. In three dimensions (x, y, z), we have

$$\mathbf{\Lambda} = \begin{pmatrix} \text{var}\left(\frac{\partial X}{\partial x}\right) & \text{cov}\left(\frac{\partial X}{\partial x}, \frac{\partial X}{\partial y}\right) & \text{cov}\left(\frac{\partial X}{\partial x}, \frac{\partial X}{\partial z}\right) \\ \text{cov}\left(\frac{\partial X}{\partial x}, \frac{\partial X}{\partial y}\right) & \text{var}\left(\frac{\partial X}{\partial y}\right) & \text{cov}\left(\frac{\partial X}{\partial y}, \frac{\partial X}{\partial z}\right) \\ \text{cov}\left(\frac{\partial X}{\partial x}, \frac{\partial X}{\partial z}\right) & \text{cov}\left(\frac{\partial X}{\partial y}, \frac{\partial X}{\partial z}\right) & \text{var}\left(\frac{\partial X}{\partial z}\right) \end{pmatrix} \quad (4)$$

For an image (or volume) generated by white noise smoothed by a Gaussian point spread function (PSF) with dispersion Σ (leading to a Gaussian auto-covariance function with dispersion 2Σ), we have $\mathbf{\Lambda} = \frac{\Sigma^{-1}}{2}$. In most cases, the PSF can be assumed to be aligned with the coordinate axes of the volume analysed, giving null off-diagonal elements for $\mathbf{\Lambda}$ and Σ . In this case the Full Width at Half Maximum (FWHM) of the PSF relates to $|\mathbf{\Lambda}|$ with

$$|\mathbf{\Lambda}|^{-1/2} = (4 \ln(2))^{-D/2} \prod_{i=1}^D \text{FWHM}_i$$

leading to the definition of RESolution ELEMENTS (RESELS):

$$V|\mathbf{\Lambda}|^{1/2} = \text{RESELS}(4 \log_e 2)^{D/2},$$

where RESELS are equal to the volume of the search region divided by the product of the FWHMs of the PSF in each dimension (Worsley *et al.*, 1992). Note that even when the actual form of the PSF is unknown (and possibly not Gaussian), smoothness values are often described in terms of FWHM. The smoothness parameter is usually defined as $|\mathbf{\Lambda}|^{-1}$ such that it relates simply to the FWHM of a Gaussian kernel, and is assessed using the partial derivatives of the SPM, a valid procedure as long as the PSF of the SPM is aligned with the coordinate axes (Figure 1 for an illustration). Note that this estimation is itself subject to noise (see section 7).

A very similar formula to (3) was established by V.P.Nosko (1969), formally proved by A.M.Hasofer (1976) and reported in Adler (1981, p.133). It gives an asymptotic result for the expected number $M_t(V)$ of maxima above a level t in V as:

$$\mathbb{E}[M_t(V)] = V |\mathbf{\Lambda}|^{1/2}(2\pi)^{-(D+1)/2}t^{D-1}e^{-t^2/2} [1 + O(1/t)] \quad (5)$$

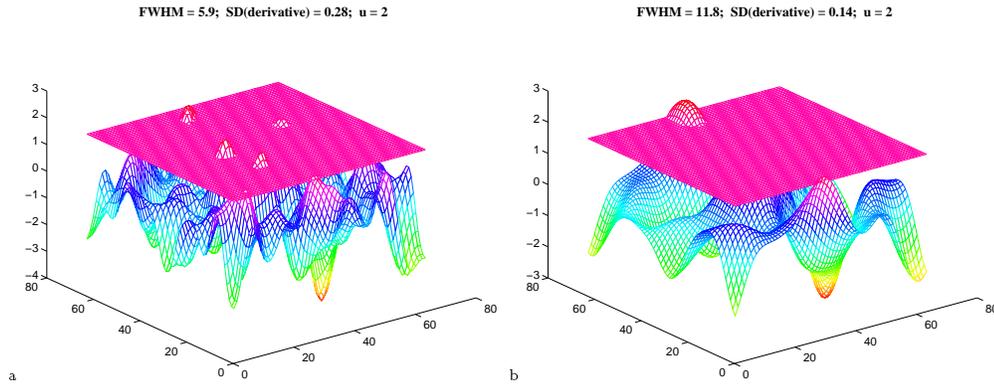


Figure 1: 2D illustration of the relationship between the variance (or standard deviation) of the derivative of the SPM and the width of the kernel. (a) little smoothing and (b) large smoothing. FWHM is in pixel (identical in x and y).

Clearly, for high t , equations (3) and (5) give similar results. In the current implementation of SPM software (that we will denote as SPM* to avoid confusion with the statistical maps themselves) equation (5) is used as it gives slightly more conservative results.

2.2 Assumptions

For equations (2) and (3) to hold, several assumptions have to be made: a discrete SPM approximates a continuous, zero-mean, unit variance, homogeneous, smoothed Gaussian Random Field (GRF); the threshold t is high; and the volume V is large compared to the resolution of the map.

1. The zero-mean and unit variance conditions hold under the null hypothesis, provided the statistical model used is correct.
2. The homogeneity condition implies that both the statistical and the spatial characteristics of the volume are constant with position. This might be of concern with 3D acquisition but is still a very reasonable first approximation. However, effects of nonstationarity of the PSF due to both instrumental or physiological factors are still need to be assessed.
3. Strictly speaking, a random field is Gaussian if the joint distribution of any subset of points has a multivariate Gaussian distribution. This implies that at the univariate level each point or position should have a Gaussian distribution. In SPM*, this is ensured by the voxel by voxel t -to- Z transformation. At the multivariate level the condition is satisfied when the degrees of freedom (df) of the t -statistic are large enough (e.g. $df \geq 30$). In general, it is difficult to address the validity of this assumption for lower degrees of freedom and results provided by Worsley (1994) for t -fields should be used when working with low df . However, spatial or temporal smoothing, usually applied to increase signal to noise ratio, minimises the risk of breaching the multivariate Gaussian assumption. Worsley (1993) showed that with $df \leq 3$ singularities will almost certainly occur in any continuous random t -field.

4. *How high should t be ?*

Although equations (3) and (5) only hold for high t , these results are used in such a way that for small t , the probability computed by equation (2) is high and therefore not generally of interest.

5. A discrete lattice should approximate a continuous smoothed random field well and allow for a good estimation of the smoothness of an SPM. The estimation can be obtained when sampling is high compared to the resolution of an SPM. For Gaussian PSFs, we found that a good smoothness estimation was obtained when the FWHM in any direction is at least 2 or 3 times greater than the voxel size. In PET, the resolution of the original image volumes usually ensures that it is the case even when post-processing smoothing is fairly low (≈ 6 mm FWHM). In *fMRI* where the resolution of the data is high, when the method is applied to raw data the assessment of the smoothness fails. However, because we are dealing with an underlying biological signal that is smooth we can get around the problem simply by undersampling the *fMRI* maps or to apply a small amount of spatial filtering to ensure validity of the assumption.

6. Results presented by Worsley *et al.* (1992) and Friston *et al.* (1991)) are accurate for search volumes V that are large compared to the resolution of the SPM (practically at least three or four times the FWHM of the SPM). Results accurate for any volume have been developed by Worsley *et al.* (1995b). When analysing the brain volume, current estimations based on the equations described above are accurate enough.

2.3 Discussion

The significance test we have presented provides strong control over the type I error at the *voxel level*. Strong control over type I error is obtained if the probability of falsely rejecting any hypothesis is less than the given level α , regardless of the truth of other hypotheses. See Hochberg & Tamhane (1987) for rigorous definitions. Note that rejecting the null hypothesis at any voxel also rejects the so called “omnibus hypothesis” (is there any signal in the entire volume?). As noted above, this test has been extended to other types of random field by Worsley *et al.* (1994) e.g. χ^2 , F and t -fields. F -fields have potentially important applications for model selection at the *voxel level*. An example of their application is the “non linear” regression that tests for the best model that describes the relationship between the measured brain response (regional perfusion) and a non categorical parameter (e.g. rate of presentation (Buechel & Friston, 1997)).

3 Testing for the significance of the spatial extent of an activation

3.1 Theory

The previous described procedure tests for the significance of a simple increase in intensity of an activation in an SPM. Early work using Monte Carlo simulations suggested that

using information about the spatial properties of potential brain signals, and testing for the significance of the spatial extent of activated regions above a given threshold, could greatly improve the sensitivity of functional imaging experiments (Poline & Mazoyer, 1993). We present here a theoretical test that again uses results from Gaussian Random Field theory (Friston *et al.*, 1994).

The first step is to start with asymptotic results for the distribution (and expectation) of an area n_u of regions defined by thresholding an SPM at u . These results derived by Nosko (1976; 1969; 1970) are also reported by Adler (1981, p.158). They show that at high threshold u the conditional distribution of n_u is such that

$$\lim_{u \rightarrow \infty} P\left(|\Lambda|^{1/2} (2\pi)^{-1} u^2 n_u^{2/D} > v \mid n_u > 0\right) = \exp(-v)$$

However, this approximate distribution significantly overestimates the area n_u . To correct for this overestimation, Friston *et al.* used the fact that the expected area $E[n_u]$ can also be derived from results previously described in section 2 :

$$E[n_u] = V\Phi(-u)/E[M_u(V)]$$

Where $E[M_u(V)]$ is the number of expected regions above u given by equation (5) and $V\Phi(-u)$ the number of expected voxels above u . The corrected distribution for n_u then becomes :

$$\Pr(n_u > v) = \exp(-\beta v^{2/D}) \tag{6}$$

with

$$\beta = \left(\frac{(D/2 - 1)E[M_u(V)]}{V\Phi(-u)} \right)^{2/D}$$

and $\Phi(\bullet)$ the standard normal CDF.

This formula establishes the distribution of an area n_u *given the occurrence of a region above u* . The parameter we are primarily interested in is the maximum value of n_u , $n_{u_{\max}}$, in V . The probability of having a maximum value of n_u greater than v is simply one minus the probability that all the $M_u(V)$ supra threshold regions in V have areas less than v , times the probability of having $M_u(V)$ regions. Using equations 1 and 6 we obtain:

$$\begin{aligned} \Pr(n_{u_{\max}} > v) &= \sum_{i=1}^{\infty} \Pr(M_u(V) = i) \left(1 - \Pr(n_u < v)\right)^i \\ &= 1 - \exp(-E[M_u(V)] \Pr(n_u \geq v)) \\ &= 1 - \exp\left(-E[M_u(V)] \exp\left(-\beta v^{2/D}\right)\right) \end{aligned} \tag{7}$$

with β as defined above. For a full development of these equations see Friston *et al.* (1994)

3.2 Discussion

3.2.1 Improved sensitivity

Generally, as expected, the test provides an improved sensitivity compared to intensity testing alone, although this is not necessarily the case. A power analysis (Friston *et al.*,

1994), shows that if the underlying signal to be detected is wider than the resolution of the SPM, power increases with a low threshold (assuming a Gaussian shaped underlying signal). However, if signal width is smaller than the noise PSF, power increases with high values of u and the optimal sensitivity is found for the intensity test. As all kinds of signal are potentially present in an SPM, it seems that the optimal procedure is to use either a series of thresholds or combined criteria. The next section deals specifically with this question.

3.2.2 The loss of voxelwise control over the risk of error

It is essential to note that the new extent test does not provide control of the risk of error at the *voxel level* and therefore individual voxels cannot be declared as “significantly activated” within a region. The localising power of the extent test has moved from the voxel level to the region (cluster) level. The localising power depends on u since high thresholds provide better localisation and greater insurance that non-activated parts of the brain are not grouped with activated regions by the thresholding process. Although nothing can be said at the voxel level, the interpretation of results will clearly be different depending on threshold (the higher the threshold, the greater is the chance that most of the voxels in the cluster are part of a underlying signal). Indeed, an essential parameter is the expected number of voxels in noise $E[n_u]$ that should be compared to the observed number of voxels forming a supra-threshold region. This comparison will help quantify the regional specificity of the test. Another good indicator is the probability of occurrence based on the voxel by voxel test as computed in section 2 for voxels with an intensity u .

3.2.3 How high should t be to insure the validity of analysis ?

It is difficult to generalise since the magnitude of t that guarantees validity depends on the smoothness ($|\Lambda|^{-1}$). However, in most PET studies, t values between 2.5 and 3 can be used safely as demonstrated by simulations. In *fMRI* experiments that generally have higher spatial resolution, safe values should be higher (≥ 3).

3.2.4 Effect of smoothing on detection

Interestingly, smoothing has an opposite effect on the sensitivity of the extent test compared to the voxel intensity test described in section 2. This is because as smoothing increases the probability that Z_{\max} crosses the level t *by chance* decreases. Clearly, when smoothing increases the probability that a large region occurs above u *by chance* increases as well. This is illustrated in figure 12 which plots, for fixed values of t or area n , the probability of occurrence by chance (noise only case) as a function of smoothness. It is usually the case that greater smoothing improves the detection of significant activation at the voxel level while relatively small degrees of low pass filtering tend to improve the sensitivity of spatial extent detection. This last observation is only *generally* true and results depend on the shape of the activated area : for instance, filters that are too large will wipe out peaky signals.

Although no assumption has been made about the shape of the spatial autocovariance

function of an SPM, because of the nature of this extent test, it is likely to be more sensitive to non stationarity of the PSF than the intensity test. In terms of implementation, a connectivity scheme has to be chosen for $D \geq 2$. We recommend an 18-connectivity scheme for $D = 3$ and a 4-connectivity scheme for $D = 2$.

4 Testing for both peak height and spatial extent

4.1 Rationale

The sections above showed that sensitivity to Gaussian signals depends on the choice of intensity thresholds, u , wide signals being best detected with low thresholds, sharp signals with high thresholds. Not only is it generally impossible to predict which test would be best for a particular analysis, but, because of the complexity of the underlying anatomy of the brain, several kinds of signal (wide or sharp) might occur simultaneously. It is also not valid to use both tests without correcting for the implicit multiple comparison involved. If the two tests were independent, a simple ‘‘Bonferroni’’ correction would be appropriate. However, the maximum intensity and spatial extent of a region above u are not independent and such corrections would lead to an over conservative test. In the next section we develop a test based on both spatial extent and peak intensity of regions above u .

4.2 Method

In this section we describe a combined test based on two parameters (peak height and spatial extent).

First, we derive an approximation for the probability that a given cluster will have a spatial extent S greater than s_0 , and maximum intensity or peak height H greater than h_0 , using results from Gaussian random field theory (Poline *et al.*, 1996a). The derivation of this result is based on modelling the shape of a region above u (near a local maximum) as an inverted paraboloid. The first terms of the Taylor expansion of the processes’ second derivative are then used to obtain an approximate distribution for the conditional distribution of n_u , *knowing* the height h_u above u . We use this approximation and the known marginal distribution of h_u (h_u has an approximate exponential distribution with mean $1/u$ (Adler, 1981, Ch.6)), to get an approximate conjoint distribution :

$$P(n_u \geq s_0, h_u \geq h_0) \approx \int_{h=h_0}^{\infty} \Psi_{\nu} \left\{ \nu ac |\mathbf{\Lambda}|^{-1/2} u^{-D/2} h^{D/2} / s_0 \right\} u e^{-uh} dh \quad (8)$$

where Ψ_{ν} is one minus the χ^2 cumulative distribution function with degrees of freedom $\nu = 4u^2/D$, given by

$$\Psi_{\nu}(x) = \int_x^{\infty} \frac{t^{\nu/2-1} e^{-t/2}}{2^{\nu/2}, (\nu/2)} dt.$$

Figure 2 shows the match between the theoretical approximation and the conjoint distribution derived with simulations of white noise convolved with a Gaussian PSF.

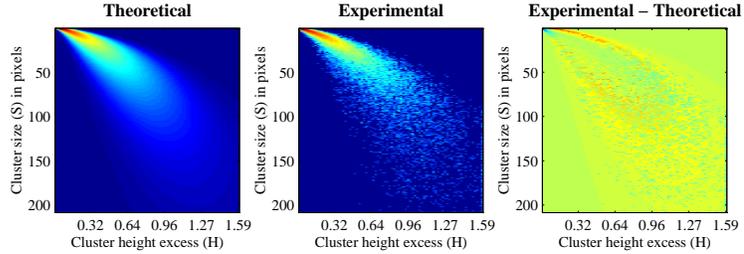


Figure 2: Left: Theoretical (predicted) bivariate distribution of spatial extent and peak height for regions occurring above an image threshold of $t = 3$ in a $64 \times 64 \times 32$ volume ($128 \times 128 \times 64 \text{ mm}^3$) with resolution 17.5 mm in x and y , and 12.5 mm in z . Data intensity is presented in a log scale to increase the visibility of the tail of the distribution. Middle: Observed bivariate distribution of spatial extent peak height under the same conditions as above. Right: Difference between the two.

Second, a way of combining the spatial extent and the maximum intensity is chosen in order to select events (an occurrence of a cluster) that will be rejected at a given risk of error under the null hypothesis of pure noise. We note that there are an almost infinite number of possibilities for this step: in a two-parameter testing procedure a statistical threshold becomes a curve in a plane.

For our proposed combined test, the risk of error is simply defined as the minimum of the risk for spatial extent n_u and the risk for maximum peak height H . This gives a rejection area defined by

$$\min\{\Pr(n_u \geq s_0), \Pr(H \geq h_0)\} = \text{constant}$$

which leads to the probability of rejection of a given cluster :

$$\Pr_{joint} = \Pr(n_u \geq s_0) + \Pr(H \geq h_0) - \Pr(n_u \geq s_0, H \geq h_0) \quad (9)$$

We then use \Pr_{joint} the probability that the spatial extent and peak height probability of a single cluster falls in the rejection area to compute the probability that at least one cluster is rejected in the volume V . If k clusters occur in the volume V , the probability that at least one of them will be rejected is simply

$$\Pr(\text{rejection} \mid C = k) = (1 - (1 - \Pr_{joint})^k).$$

Summing over k , weighted by the probability that $C = k$, we get

$$\begin{aligned} \Pr(\text{rejection}) &= \sum_{k=0}^{\infty} (1 - (1 - \Pr_{joint})^k) \mathbb{E} [M_t(V)]^k e^{-\mathbb{E}[M_t(V)]} / k! \\ &= 1 - e^{-\mathbb{E}[M_t(V)] \Pr_{joint}}. \end{aligned} \quad (10)$$

Simulations in pure noise for various values of u and various resolutions show that the conjoint test protects against type I risk of error α (except if the threshold is very low ($u=2$) and the α -risk greater than 0.15). Figure 3 shows the expected versus observed risk of error in 3D for various thresholds.

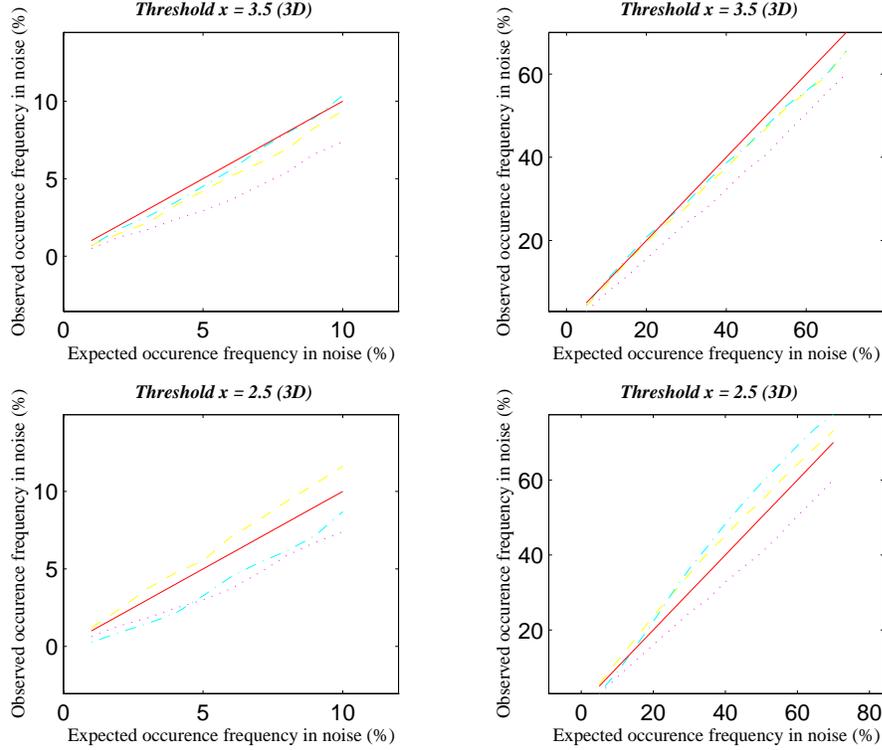


Figure 3: Expected versus observed risk of error with two thresholds u in 3D volumes ($64 \times 64 \times 32$ voxels or $128 \times 128 \times 64$ mm³) at a fixed resolution (FWHM_{xy} = 14.1 mm and FWHM_z = 11.8 mm). Top: high intensity threshold ($t = 3.5$). Bottom: low threshold ($t = 2.5$). On the left: risk of error between 1% and 10%. On the right: risk of error varying between 10% and 70%. The dashed line shows the results from the spatial extent test, the dotted line from the peak height test and the dot and dashed line from the combined test. The solid line corresponds to the $y = x$ line. Results were assessed using 3×10^3 simulations.

The sensitivity of the combined test was assessed for 3 simulated signals : a sharp signal, an extended signal and a signal with approximately the same probability of being detected by either the intensity test (in section 2) or the spatial extent test (section 3). Results (presented in figure 4) show that the conjoint test should generally increase the overall sensitivity of analyses, as well as increasing their validity by correcting for the implicit multi-testing procedure.

4.3 Discussion

As for the spatial extent test, the risk of error is determined at the region level. In fact, the two tests are conceptually very similar but the conjoint test is more general. Note that it is always possible to know whether a supra-threshold region is unlikely to occur because of its size or its height, giving further information on the type of regional activation observed.

Also, we note that to derive equation (9) slightly stronger hypotheses were required. It

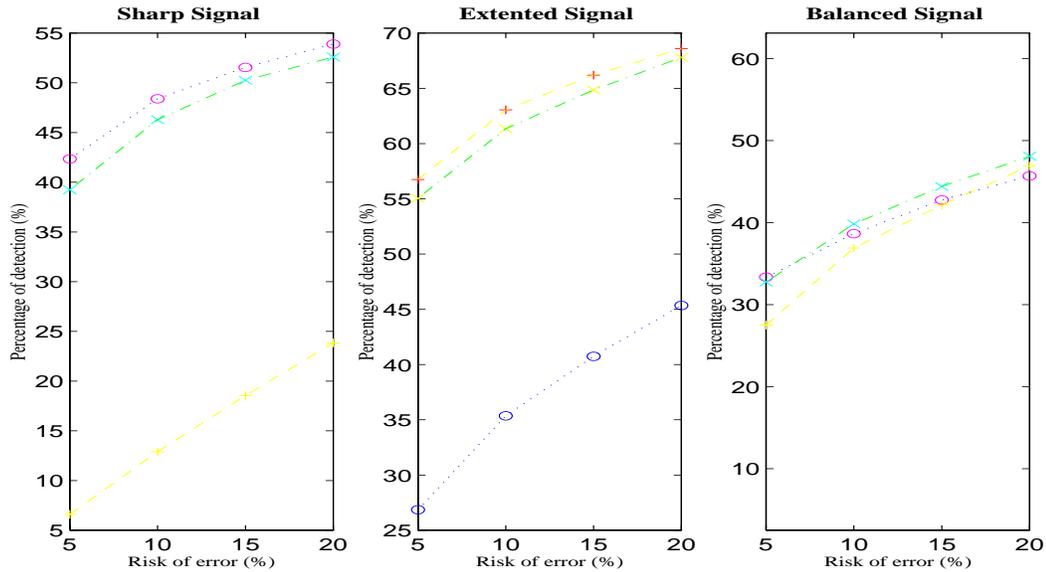


Figure 4: Percentage of detected signal versus risk of error for three types of signal in 3×10^3 3D volumes ($64 \times 64 \times 32$ voxels or $128 \times 128 \times 64$ mm³). Left: sharp peak. Middle: extended signal. Right: “balanced” signal. The dashed line shows the results from the spatial extent test, the dotted line from the peak height test and the dot and dashed line from the combined test. Volume resolution was FWHM_{xy}= 14.1 mm FWHM_z= 11.8 mm and threshold $t = 3$.

is assumed that the PSF, resulting from both the image reconstruction apparatus and post processing filtering, can be modeled by a Gaussian function. The robustness of the conjoint test with regard to this assumption remains to be evaluated.

The conjoint test may prove to be an interesting alternative to multi-filtering strategies, methods not presented in this short review (Poline & Mazoyer, 1994; Worsley *et al.*, 1996). These strategies have ultimately a similar goal that is to detect signals of various sizes in one statistically valid procedure. The conjoint (or bivariate) test however has the advantage of requiring fewer computations and should preserve better the spatial resolution large signals, an important feature for the analysis of *fMRI* data.

5 Testing for the significance of a set of regions

This section extends the previous tests and describes a new level of inference that is, in general, more sensitive but has less localising power. The test is based on the number of supra-threshold regions of size greater than k_u compared to the expected number of such regions. Control of the risk of error at the region level cannot be obtained, and control over the risk of error now has to be considered at the *set-level*.

We first review the operational equations and then report power analyses (Friston *et al.*, 1995).

5.1 Theory

Let C_{nu} be the number of regions defined with a threshold u and area greater than n occurring in V . To test for this number we compute the probability of having C_{nu} regions *or more* of size n_u *or more* in V . This is also one minus the probability of obtaining less than C_{nu} regions with size greater than n_u :

$$\begin{aligned} \Pr(C_{nu} \geq c) &= \sum_{i=0}^{c-1} \sum_{j=i}^{\infty} \Pr(C_{\odot u} = j) \binom{j}{i} \Pr(n_u \geq n)^i \Pr(n_u < n)^{j-i} \\ &= 1 - \sum_{i=0}^{c-1} \Upsilon(\mathbb{E}[\chi_u] \Pr(n_u \geq n)) \end{aligned} \quad (11)$$

Where $\Pr(C_{\odot u} = j)$ is the probability of getting j regions above u of any size (\odot denotes any value here) in V , given equation 1 that also defines $\Upsilon(\odot)$. The second equality above can be seen directly by noting that the number of cluster of size n_u or more is a restriction of the process defined by the number of cluster (with any size) and therefore also follows a Poisson law. The mean of this process is simply the mean of the original process times the probability $\Pr(n_u \geq n)$.

The equation is very general and reduces to the intensity test (section 2) and to the spatial extent test (section 3) with appropriate parameters. If $n = 0$ and $c = 1$ then (11) reduces to the probability found in section 2 (probability of having at least one cluster of unspecified size). If $c = 1$ and n_u is left unspecified, then the test reduces to the spatial extent test.

5.2 Power analysis

We use a simulated “activation signal” that can be modeled mathematically and is physiologically plausible. Brain signals are modeled by a Gaussian random process (therefore distributed) of a certain width f (expressed as a proportion of the smoothness of noise) and height (variance σ^2). Using this model, we can compute the probability of the alternative hypothesis depending on the parameters σ . The smoothness under the alternative hypothesis are (Friston *et al.*, 1994):

$$\begin{aligned} u^* &= u (1 + \sigma^2)^{-1/2} \\ |\Lambda^*|^{-1/2} &= |\Lambda|^{-1/2} \left[\frac{(1 + \sigma^2)}{(1 + \sigma^2/(1 + f^2))} \right]^{1/2} \end{aligned}$$

For a given risk of error, α , given by $\Pr_{|\Lambda|}(C_{nu} \geq x)$, the sensitivity of the test is simply the probability $\Pr_{|\Lambda^*|}(C_{nu^*} \geq x)$. Using this model, we simply vary the parameters u , $|\Lambda|$, and n to assess the power of the different tests. Traditionally, sensitivity is plotted against the risk α : these plots are called Receiving Operator Curves (ROC). Figure 5 shows the result of this power analysis.

In figures 6 and 7 we illustrate the use of such tests in a PET dataset (a verbal fluency experiment). In this case $u = 3.2$, and the spatial extent threshold was the expected value given the smoothness and the volume analysed (8 voxels). The most sensitive test

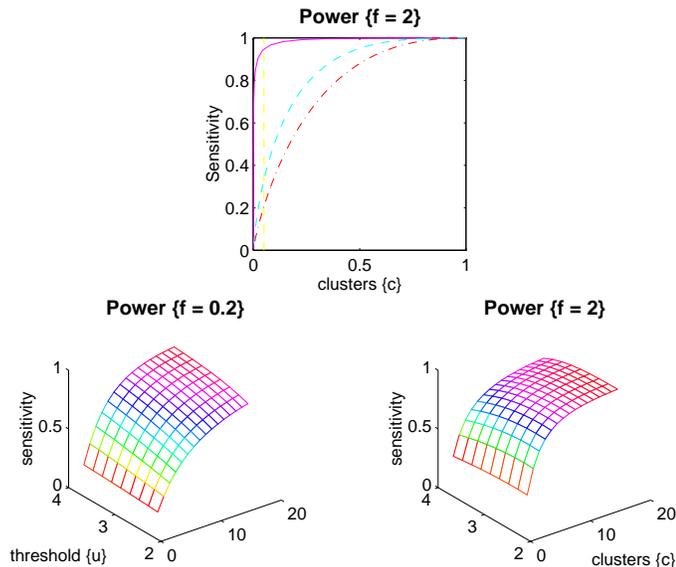


Figure 5: Top panel: ROC curve for set-level inference with $u = 2.8$ and $n = 16$ voxels. $\Pr_{|\Lambda|}(C_{nu} \geq x)$ where $|\Lambda|$ corresponds to a FWHM of 3 voxels and the volume $V = 64^3$. Signal amplitude $\sigma = 0.3$ and width $f = 2$. The dashed and broken lines corresponds to the equivalent cluster and voxel-level ROC curves respectively. Lower panels: 3-dimensional plot of power ($\alpha = 0.05$) as a function of cluster number c and threshold u , for the same smoothness, volume V and σ . Left: $f = 0.2$ and right $f = 2$.

was that at the set level of inference, but note that amongst 8 clusters in the SPM, one or two are expected to occur by chance (expected number : 1.4). For the second region listed in figure 7 the conjoint test was much more significant than the test on intensity (because of the cluster size) and performed approximately as well for the other clusters.

5.3 Discussion

Clearly, results obtained with the set level of inference should be interpreted with caution when reporting the anatomical localisation of regions forming the significant set. However, if the number of observed regions (above u and of size greater than n_u) is much greater than the predicted number, (e.g. 0.5 regions expected, 5 observed) then it makes sense to report all the clusters if only descriptively. Conversely, if 5 clusters are observed but 2.5 are expected by chance, it is difficult to elaborate on the regional specificity of the results, and the set-level of inference gives information that is only slightly more precise than an omnibus test, thus providing very little regional information.

The set level of inference can be extended using a conjoint probability for both peak height and spatial extent. Simply, $\Pr(n_u \geq n)$ is changed for $\Pr_{joint} = \alpha_{joint}$ in equation (11). This maneuver will not add another parameter (peak height) : the set will simply be formed by clusters that have a probability less than a chosen value pf α_{joint} (for instance $\alpha_{joint} = 0.4$, either because of the height or the extent above u).

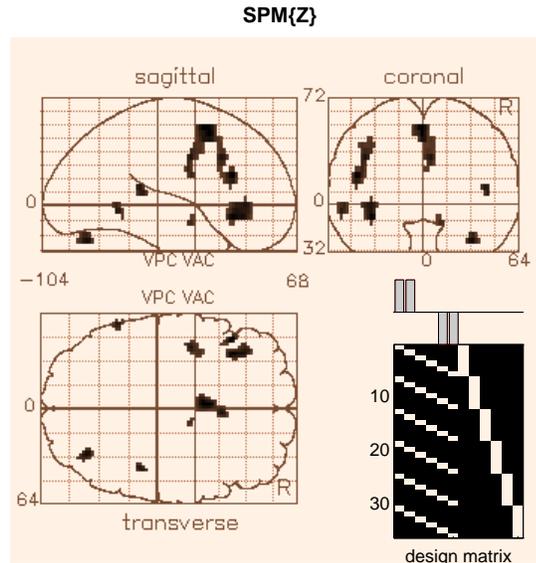


Figure 6: “Glass brain” view of an SPM of a verbal fluency experiment showing activation in the frontal gyrus. The design matrix of the experimental model is shown in the bottom right corner. Figure 7 presents the statistical results associated with this SPM.

6 Non parametric approaches: Statistical *Non* Parametric Mapping (SnPM)

6.1 Rationale and Method

Recently, non parametric multiple comparisons procedures have been introduced for the assessment of functional mapping experiments, based on randomisation or permutation test theory (Holmes *et al.*, 1996). By considering appropriate permutations of the labeling of scans (labeling as “rest” & “active”, or by some associated covariate such as scan score), and computing statistic images for each labeling, a permutation distribution for the entire statistic image can be obtained. From this null distribution of the statistic image, given the data and appropriate null hypothesis, the permutation distribution of any statistic summarising the statistic image can be found. Summarising each statistic image by its maximum statistic gives the permutation distribution for Z_{\max} , the $100(1 - \alpha)^{\text{th}}$ percentile of which is the appropriate critical threshold for a single threshold test at level α . Summarising each statistic image by the size of the largest cluster of voxels with values above a pre-specified threshold gives the permutation distribution of S_{\max} , and appropriate critical suprathreshold cluster sizes. Strong control over experimentwise type I error is maintained (at the appropriate level) in both cases.

In addition to the usual attractions of non parametric method, namely minimal assumptions, guaranteed validity and exactness, flexibility and intuitiveness, the approach is especially attractive for small data sets such as those from single subject PET studies. Statistic images with low degrees of freedom exhibit high (spatial) frequency noise therefore the statistic image is rough. The properties of such statistic images are not well approximated by continuous random fields with the same distributions. Continuous fields

P values & statistics:

set-level (c)	cluster-level (k,Z)	voxel-level (Z)	location (mm)
0.000 (8)	0.028 (27, 4.78)	0.018 (4.78)	-46 24 20
	0.006 (126, 4.68)	0.027 (4.68)	-2 8 48
		0.173 (4.20)	4 16 32
		0.889 (3.46)	4 14 44
	0.031 (76, 4.61)	0.037 (4.61)	-36 24 -8
		0.154 (4.24)	-36 32 0
	0.129 (26, 4.33)	0.108 (4.33)	32 -74 -24

Height threshold (u) = 3.20, p = 0.001	Volume (S) = 53132 voxels or 625 Resels
Extent threshold (k) = 8 voxels	Degrees of freedom due to error = 25
Expected voxels per cluster, E(n) = 8.2	Smoothness = 9.8 11.2 12.5 mm (FWHM)
Expected number of clusters, E(m) = 1.4	= 4.1 4.8 5.3 (voxels)

Figure 7: This table present an example of the SPM* statistical results with the set, cluster, and voxel level of inference (n_u is denoted k in this table). Note the relative sensitivity of this three tests and the loss of regional specificity.

have features smaller than the voxel dimensions, leading to critical thresholds for single threshold tests that are conservative for lattice representations of the continuous field. An extreme example is a 3-dimensional strictly stationary continuous random t -field with 3 degrees of freedom, which almost certainly has a singularity (Worsley, 1993).

The noise in low degree of freedom statistic images results from variability of the residual variance estimate. In PET it is reasonable to assume that the residual variability is approximately constant over small localities, suggesting that variance estimates could be locally pooled. A weighted local pooling of variance estimates is a smoothing of the estimated variance image (since the degrees of freedom are the same at every voxel). An example of a smoothed variance image for a PET dataset is shown in figure 8a, where weights from an isotropic three-dimensional Gaussian kernel of FWHM 12mm were used (the kernel was truncated at the edges of the intracerebral volume). Clearly variance estimates at proximate voxels are not independent. A theoretical distribution for such smoothed variance images has proved elusive, thus precluding further parametric analysis. The “pseudo” t -statistic image formed with such a variance image is shown in figure 8b, and is much smoother than the original variance map (not shown). Figure 9 illustrates the results obtained with “pseudo” t -statistic.

6.2 Results and discussion

The ability to consider statistic images constructed with smoothed variance estimates appears to makes the non parametric approach considerably more powerful than the parametric approaches discussed. Non-parametric results for a PET data set are shown in figure 9a. 1000 permutations (including the actual allocation) of the $12!$ possible permutations of scan scores were considered, and the (approximate) permutation distri-

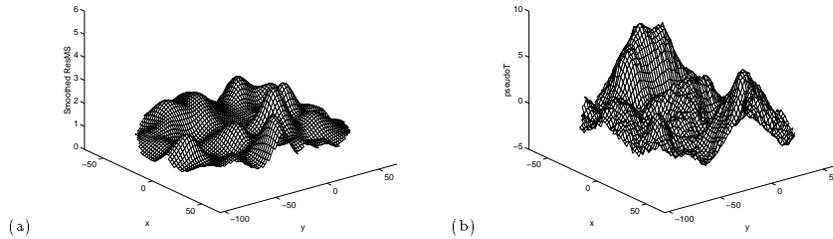


Figure 8: Statistic images for PET dataset: (a) Mesh plot (intercommisural plane) of estimated variance image smoothed with an isotropic Gaussian kernel of FWHM 12mm, truncated at the edge of the intracerebral volume. (b) Mesh plot of “pseudo” t -statistic computed with smoothed variance estimate.

bution of the maximum “pseudo” t -statistic computed. The resulting single threshold test identifies many more significant voxels than the parametric single threshold test using the expected Euler characteristic on the “Gaussianised” t -statistic (figure 9b).

Using raw t -statistic images, the non-parametric approach on the whole agrees largely with parametric approaches, which is a comforting observation. Disadvantages of the non-parametric approach are a greater need for computer resources and a possible limitation when dealing with too small a number of relabellings.

Alternatively, the variance estimate can be improve by including more scans in the SPM* analysis, taken from other subjects, while tests of the appropriate statistical contrast include only the actual subjects “of interest”. This procedure assumes that physiological and instrumental noise variances are similar across pooled subjects (and that experimental effects have been removed using multilinear regression (Holmes & Friston, 1997). This assumption would not usually apply for patient studies. An example of such an analysis with normal subjects is given in (Poline *et al.*, 1996b).

7 Discussion and conclusion

7.1 Which test should be used and when ?

As the nature of a signal is unknown, it is impossible to predict which would be the best procedure to use for a given data set. Although, strictly speaking, it might not be valid to use several tests concurrently, the complex relationship between them and their nested aspect, should ensure that the risk of error is not excessively increased by the multi-testing procedure. In future, Monte Carlo simulation will assess the extent of departure from the α risk of false positives chosen by the experimenter. We summarise the tests described above (figure 10) by a schematic uni-dimensional graph. Figure 11 gives an overview of the characteristics of the tests. For completeness we have added “omnibus” tests in this figure that give a probability value for the general overall pattern of the SPM although they are not described in this paper (see Worsley *et al.* (1995a) and Friston *et al.* (1996a)).

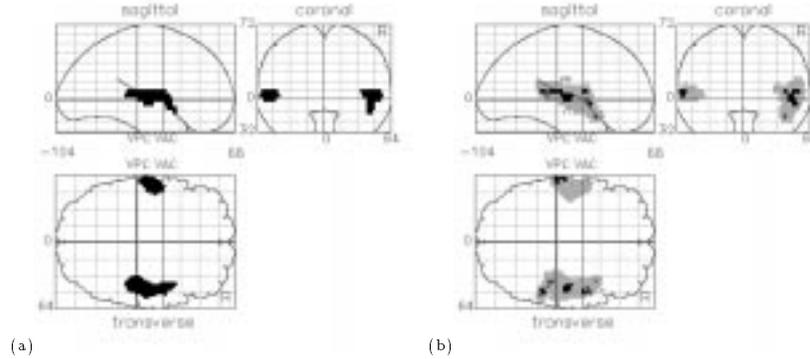


Figure 9: (a) “Glass brain” views of the significant voxels at $\alpha = 0.05$ from a non parametric single threshold test using “pseudo” t -statistic images. (b) Orthogonal “glass brain” views of the significant voxels at $\alpha = 0.05$ for the same PET dataset using the parametric approach. The smoothness of the “Gaussianised” t -statistic image was estimated at $16.4 \times 17.5 \times 13.5$ mm, equivalently 273 resels, for 66689 intracerebral voxels. Voxels above the critical threshold $u_{\alpha=0.05}$ are shown black. Supratherreshold clusters of voxels were identified using a primary threshold of $\Phi^{-1}(1 - 0.001)$, identifying two significantly large clusters of voxels, shown translucent grey.

7.2 Sensitivity and Specificity

In statistical analyses the risk of error is usually chosen to be 5%. We emphasise that this is an *arbitrary* threshold that may be too stringent on some occasion. In any case, *a failure to reject the null hypothesis is never proof that the alternative hypothesis is untrue*. In other words, we can never be sure that a region is *not* activated. We therefore recommend the discussion of results that do not reach the 5% level but are improbable under the hypothesis of noise only (risk of error of 5 to 20% for instance).

7.3 How do we choose the parameters ?

The more parameters used by the tests, the more difficult it is to choose a priori optimal values for them *a priori*. Currently, three parameters must be chosen : the FWHM of the Gaussian kernel used for smoothing (affecting directly $|\mathbf{\Lambda}|$), the threshold u (for the cluster level of inference) and the area n used in the set level of inference. An obvious way of proceeding is to acquire experience by analysing standard data sets and then fixing the parameters to some appropriate values.

However, this procedure would require that the volume analysed (V) and the smoothness parameter remain identical from one study to another. As this is not generally the case, we suggest setting u and n using *statistical* thresholds. For instance, u can be set such that, *given* V and $|\mathbf{\Lambda}|^{-1/2}$, we have $\Pr(Z_{max} \geq u) = \xi$, where ξ will depend on the regional specificity required for the experiment (e.g. for high regional specificity: $\xi = 0.75$). In an analogous way, n can be set using the expected area above u ($\mathbf{E}[n_u]$) as a reference : $n_u = \zeta \mathbf{E}[n_u]$ where $\zeta \leq 1$ for a moderate regional specificity. Future versions of SPM* software will provide default values based on a desired regional specificity. Note that

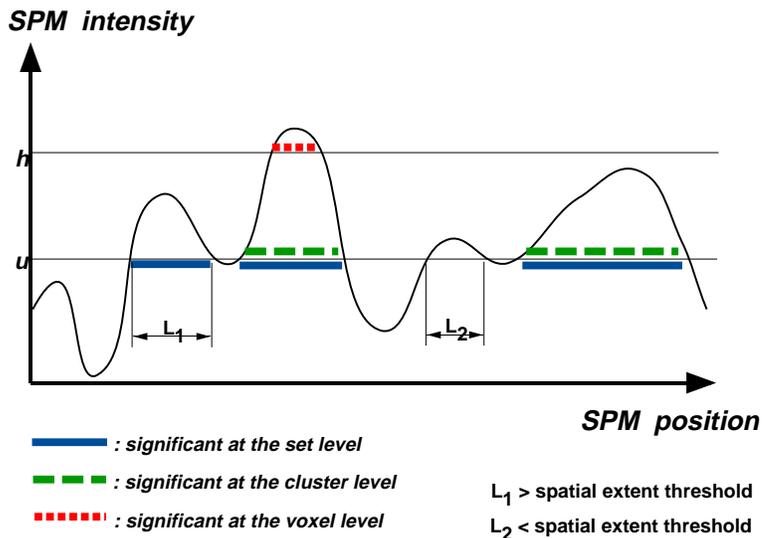


Figure 10: Symbolic representation of the intensity test, the cluster size (or conjoint) test and of the test over a set of region.

repeatedly trying different parameters will invalidate the confidence level to an unknown degree and should therefore, be avoided. Note also that no correction is made for the number of contrasts performed: the risk of error is set per contrast.

7.4 Smoothness Variability

It is essential to note that in general, the value of $|\mathbf{\Lambda}|$ and the variance of an SPM are the only values that need to be estimated when assessing the significance of activation in SPMs (there is no error of measure on the volume V). An error on the assessment of $\mathbf{\Lambda}$ directly influences the estimation of the significance of results. Using the frequency (spectral) representation of the process (i.e. the SPM, denoted X) we were able to derive the variance of the estimate of the smoothness. The principle of this computation is the following. We first assume that the PSF of the SPM is known and use this to compute the variance covariance matrix of the vector \underline{U} :

$$\underline{U} = (U_1, U_2, \dots, U_{D+1}) = \left(\text{var}[\widehat{X}(\underline{x})], \text{var}\left[\frac{\partial \widehat{X}}{\partial x_1}\right], \text{var}\left[\frac{\partial \widehat{X}}{\partial x_2}\right], \dots \right)$$

where $\underline{x} = (x_1, x_2, \dots, x_D)$ are the D dimensions of the space and $\widehat{}$ signifies that the variance is only estimated from the data (as the sum of squares divided by the number of data points). The smoothness estimation is a simple function of the vector \underline{U} , say $f(\underline{U})$ and once the variance covariance matrix of its component have been found we use the Taylor expansion to obtain an estimate of the variance of $f(\underline{U})$ (where $f(\underline{U}) = (\text{var}[\widehat{X}(\underline{x})])^D |\mathbf{\Lambda}|^{-1}$ in D dimensions):

$$\text{var}[f(\underline{U})] = \sum_{i,j \in (1, \dots, D+1)} \frac{\partial f(\underline{U})}{\partial U_i} \frac{\partial f(\underline{U})}{\partial U_j} \text{cov}[\widehat{U}_i, \widehat{U}_j]$$

For instance, using this approximation, we found that the standard deviation of the smoothness estimation $\sigma_{\mathbf{\Lambda}}$ was around 25% of the smoothness value (using common

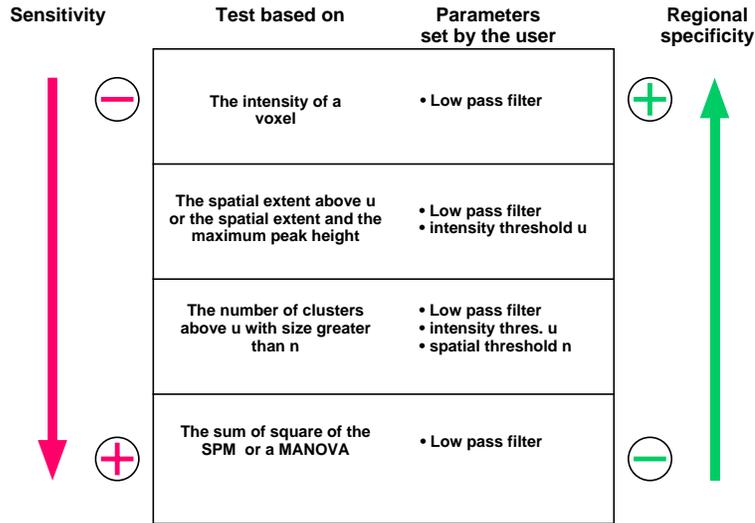


Figure 11: Overview of the characteristics of the hierarchy of test proposed for the inference in SPM. Note that the sensitivity is only *generally* increasing and in some occasions more regionally specific tests will also be more sensitive.

values for PET experiment). Figure 12 shows the effect of this uncertainty on the p-values obtained with the intensity or the cluster size tests ¹.

7.5 Smoothness estimation on the residuals

Previously (SPM*95), smoothness is assessed on Gaussianised t -maps (G-tm) that are not generally free of physiological signal. This technique has two major drawbacks. First, the estimation is not stable (the variance of the estimate being far from negligible (Poline *et al.*, 1995)), and second, the signal in the G-tm will bias any estimation. A rigorous method that overcomes these drawbacks based on previously derived theoretical results is presented here (Worsley *et al.*, 1992), which is implemented in the new versions of SPM*. To free the smoothness estimation from signal introduced by an experimental design, we propose using the residual processes that are left after removing the effects modeled in the design matrix. We make the assumption that the smoothness of these fields will approximate the smoothness of the component processes of the t -field under the null hypothesis. The residual fields are defined by

$$R_i(\underline{x}) = Y_i(\underline{x}) - \hat{Y}_i(\underline{x}) = Y_i(\underline{x}) - D\hat{\beta}(\underline{x})$$

where \underline{x} is a location in space, i indexes the i th observation, D (denoted X in the previous chapter (Holmes & Friston, 1997)) is the design matrix of the experiment, the $\hat{\beta}$ are the estimated effects, Y_i are the original values (scans) and \hat{Y}_i the fitted values. The R_i are free from all linear effects explicitly modeled in the analysis. We first demonstrate (using simulated stationary Gaussian smoothed processes) that smoothnesses of residual fields R_i and of original fields Y_i , are equivalent and that this holds whatever the degrees of

¹Note that the variance of the process (the SPM) is often known but is assessed in a more general case and therefore the the estimation of $\text{var}[\hat{X}(\underline{x})]$ can be “included” in the smoothness estimation. In other words, the smoothness estimation has generally to include the map variance estimation.

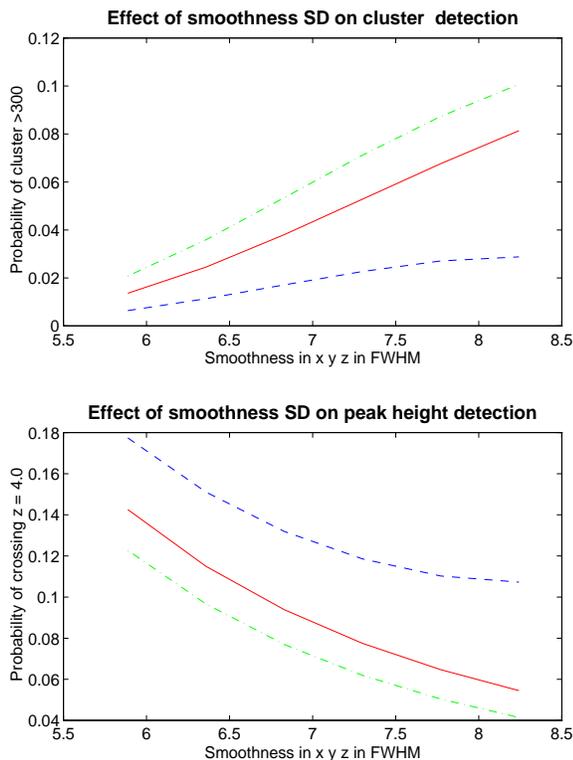


Figure 12: Upper panel : Variation of cluster size probability for a 3D process (or SPM) with the smoothness defined as the FWHM of a Gaussian kernel (defined in *pixels*) and the variation of the smoothness estimate (dashed line : $+2\sigma_{\Lambda}$, dotted and dashed line : $-2\sigma_{\Lambda}$). Threshold for the cluster definition was 2.8, size of the SPM : $5 \cdot 10^4$ pixels. Lower panel : Variation of the Z value probability for 2D data for $Z = 4$ with the smoothness value and with the variation σ_{Λ} of its estimate (dashed line : $+2\sigma_{\Lambda}$, dotted and dashed line : $-2\sigma_{\Lambda}$).

freedom (df) in R_i . The smoothness of 36 R_i using noise only random fields (8.2,8.2,5.9 FWHM in (x,y,z)) was assessed for a series of design matrices of decreasing rank, giving 7, 15, and 25 df . We then used results derived by Worsley *et al.* to relate the smoothness $|\Lambda|$ of the original component fields (estimated with R_i) to the smoothness $|\Lambda^\dagger|$ of the Gaussianised t fields with $|\Lambda^\dagger| = \lambda_{n'} \Lambda \sigma^2$ with σ^2 the variance of the original processes and $\lambda_{n'}$ a correction factor derived by Worsley *et al.* (1992), that depends on the number of independent residual fields n' .

Table 1 presents the theoretical (Theo) smoothness values for the Gaussianised t -fields (i.e. true values corrected by $\lambda_{n'}$), the values estimated using 36 R_i (Res) and the values estimated using t -maps (Gt-m) (with random orthogonal contrasts). It is seen that the values assessed on t -maps and on the residual maps are good estimates. However, we also demonstrated that the smoothness estimate can be biased under an alternative hypothesis, by assessing its value using the Gt-m of a dataset in which half of the images contain a cubic signal (size $17 \times 17 \times 9$ voxels, and magnitude sets to 0.3 noise SD). The contrast used to create the t -fields (25 df) tested for the main activation effect. These simulations show that the discrepancy between the theoretical value and the Gt-m estimate is important; Theo : (8.1, 8.1, 5.8), Gt-m : (10.1, 10.4, 7.1) FWHM in (x,y,z)).

df	7			15			25		
	Theo.	Res.	Gt-m.	Theo.	Res.	Gt-m.	Theo.	Res.	Gt-m.
x	7.5	7.6(.1)	7.9(.2)	7.9	8.0(.1)	7.8(.2)	8.1	8.1(.1)	8.3(.2)
y	7.5	7.7(.1)	7.7(.1)	7.9	8.0(.1)	8.2(.2)	8.1	8.2(.1)	8.1(.2)
z	5.4	5.6(.1)	5.3(.1)	5.7	5.8(.1)	5.6(.2)	5.8	5.8(.1)	5.8(.2)

Table 1: Theoretical vs estimated (on residuals and Gt-m) smoothness values in pixels FWHM. Res. : (SDM) 36 processes and Gt-m (SD) over random contrasts

Assessing the smoothness of the residuals fields provides a much better estimate: (8.1, 8.0, 5.8)). Recent work has further refined this method and SPMs smoothness is now assessed using normalised residuals.

7.6 A priori hypothesis

It should be noted that with a priori hypotheses about the localisation of signal, i.e if a precise position (x,y,z) is tested, it is possible to use non corrected p-values. If the precise localisation is not known, but a larger circumscribed region is interrogated for the occurrence of an activation, the p-value should be corrected for that volume (for instance using the results derived by Worsley (1995b) for small regions). More often than not several hypotheses about the localisation are possible and therefore a correction (Bonferroni) should be made for the number of regions to be interrogated to ensure the validity of the statistical procedures.

7.7 Testing for commonalities

Often, the question of experimental interest relates to the commonalities between two SPMs. A simple way to deal with this problem is to look for voxels that have a low probability of occurrence in both SPMs. If the components of the design matrix (see the chapter by Holmes et al., for description of the design matrix) used to produce the SPMs are orthogonal then the resulting p-values in the conjoint map are the product of the p-values in the original SPMs.

Conclusion

We have presented the tests used to assess significance of SPMs and have discussed the parameters that influence the output of these tests.

Acknowledgement We thank our colleagues and especially S.Kiebel for providing some of the results presented here. JBP is funded by an European Union grant, Human Capital and Mobility, no ERB4001GT932036. APH and KJF are funded by the Wellcome Trust. We are deeply in debt to P.Fletcher, J.Ashburner and J.Greene for their help.

References

- Adler, R. 1981. *The Geometry of Random Fields*. New York: Wiley.
- A.M., Hasofer. 1976. The Mean Number of Maxima Above High Levels in Gaussian Random Fields. *Journal of Applied Probability*, **13**, 377–379.
- Ashburner, J., & Friston, K.J. 1997. Spatial Transformation of Images. *Chap. 2 of: SPM short course notes*. <http://www.fil.ion.ucl.ac.uk/spm/course/notes.html>: Wellcome Department of Cognitive Neurology.
- Buechel, C., & Friston, K.J. 1997. Effective Connectivity in Neuroimaging. *Chap. 6 of: SPM short course notes*. <http://www.fil.ion.ucl.ac.uk/spm/course/notes.html>: Wellcome Department of Cognitive Neurology.
- Cox, D.R., & Miller, H.D. 1990. *The Theory of Stochastic processes*. London: Chapman and Hall.
- Friston, K.J., Frith, C.D., Liddle, P.F., Dolan, R.J., Lammertsma, A.A., & Frackowiak, R.S.J. 1990. The Relationship Between Global and Local Changes in PET Scans. *Journal of Cerebral Blood Flow and Metabolism*, **10**, 458–466.
- Friston, K.J., Frith, C.D., Liddle, P.F., & Frackowiak, R.S.J. 1991. Comparing Functional (PET) Images: The Assessment of Significant Change. *Journal of Cerebral Blood Flow and Metabolism*, **10**, 690–699.
- 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., Poline, J.-B., Holmes, A.P., Price, C.J., & Frith, C.D. 1995. Detecting Activations in PET and fMRI: Levels of Inference and Power. *NeuroImage*, **4**, 223–235.
- Friston, K.J., Poline, J.-B., Strother, S., Holmes, A.P., Frith, C.D., & Frackowiak, R.S.J. 1996a. A Multivariate Analysis of PET Activation Studies. *Human Brain Mapping*, **4**, 140–151.
- Friston, K.J., Ashburner, J., Frith, C.D., Poline, J.-B., Heather, J.D., & Frackowiak, R.S.J. 1996b. Spatial Registration and Normalization of Images. *Human Brain Mapping*, **2**, 165–189.
- Hochberg, Y., & Tamhane, A.C. 1987. *Multiple Comparisons Procedures*. John Wiley & Sons.
- Holmes, A.P., & Friston, K.J. 1997. *Statistical Models and Experimental Design*. <http://www.fil.ion.ucl.ac.uk/spm/course/notes.html>: Wellcome Department of Cognitive Neurology. Chap. 3.
- Holmes, A.P., Blair, R.C., Watson, J.D.G., & Ford, I. 1996. Non-Parametric Analysis of Statistic Images from Functional Mapping Experiments. *Journal of Cerebral Blood Flow and Metabolism*, **16**, 7–22.
- Poline, J.-B., & Mazoyer, B.M. 1993. Analysis of Individual Positron Emission Tomography Activation Maps by Detection of High Signal-to-Noise Ratio Pixel Clusters. *Journal of Cerebral Blood Flow and Metabolism*, **13**, 425–437.
- Poline, J.-B., & Mazoyer, B.M. 1994. Enhanced Detection in Brain Activation Maps Using a Multi Filtering Approach. *Journal of Cerebral Blood Flow and Metabolism*, **14**, 639–641.
- Poline, J.-B., Worsley, K.J., Holmes, A.P., Frackowiak, R.S.J., & Friston, K.J. 1995. Estimating Smoothness in Statistical Parametric Maps: Variability of p -values. *Journal of Computed Assisted Tomography*, **19**(5), 788–796.
- Poline, J.-B., Worsley, K.J., Evans, A.C., & Friston, K.J. 1996a. Combining Spatial Extent and Peak Intensity to Test for Activations in Functional Imaging. *NeuroImage*, **In Press**.
- Poline, J.-B., Vandenberghe, R., Holmes, A.P., Friston, K.J., & Frackowiak, R.S.J. 1996b. Reproducibility of PET Activation Studies: Lessons from a Multi-centre European Experiment. *NeuroImage*, **4**, 34–54. On behalf of the European Union concerted action on functional imaging.
- V.P., Nosko. 1969. The Characteristics of Excursions of Gaussian Homogeneous Random Fields Above a High Level. *Pages 216–222 of: Proc. USSR-Japan Symp. on Probability*.
- V.P., Nosko. 1970. *On shines of Gaussian random fields*. Tech. rept. Vestnik Moscov. Univ. Ser. I Mat. Meh. In Russian.

- V.P., Nosko. 1976. Local Structure of Gaussian Random Fields in the Vicinity of High Level Shines. *Soviet Mathematics Doklady*, **10**, 1481–1484.
- Worsley, K.J. 1993. Instability of Localisation of Cerebral Blood Flow Activation Foci with Parametric Maps. *Journal of Cerebral Blood Flow and Metabolism*, **13**(6), 1041–1042. In reply to S.F. Taylor, S.M. Minoshima and R.A. Koeppe.
- Worsley, K.J. 1994. Local Maxima and the Expected Euler Characteristic of Excursion Sets of χ^2 , F , and t Fields. *Advances in Applied Probability*, **26**, 13–42.
- Worsley, K.J., Evans, A.C., Marrett, S., & Neelin, P. 1992. A Three-Dimensional Statistical Analysis for CBF Activation Studies in Human Brain. *Journal of Cerebral Blood Flow and Metabolism*, **12**, 900–918.
- Worsley, K.J., Poline, J.-B., Frackowiak, R.S.J., & Friston, K.J. 1995a. A Test for Distributed, Non Focal Brain Activation. *NeuroImage*, **2**, 183–194.
- Worsley, K.J., Marrett, S., Neelin, P., Friston, K.J., & Evans, A. 1995b. A Unified Statistical Approach for Determining Significant Signals in Images of Cerebral Activation. *Pages 327–333 of: T.Jones, V.Cunningham, R.Myers, & D.Bailey (eds), Quantification of Brain Function using PET*. San Diego: Academic Press.
- Worsley, K.J., Marrett, S., Neelin, P., & Evans, A.C. 1996. A Three-Dimensional Statistical Analysis for CBF Activation Studies in Human Brain. *Human Brain Mapping*, **In press**.