

# An Introduction to Random Field Theory

Matthew Brett\*, Will Penny<sup>†</sup> and Stefan Kiebel<sup>†</sup>

\* MRC Cognition and Brain Sciences Unit, Cambridge UK;

<sup>†</sup> Functional Imaging Laboratory, Institute of Neurology, London, UK.

March 4, 2003

## 1 Introduction

This chapter is an introduction to the multiple comparison problem in functional imaging, and the way it can be solved using Random field theory (RFT).

In a standard functional imaging analysis, we fit a statistical model to the data, to give us model parameters. We then use the model parameters to look for an effect we are interested in, such as the difference between a task and baseline. To do this, we usually calculate a statistic for each brain voxel that tests for the effect of interest in that voxel. The result is a large volume of statistic values.

We now need to decide if this volume shows any evidence of the effect. To do this, we have to take into account that there are many thousands of voxels and therefore many thousands of statistic values. This is the multiple comparison problem in functional imaging. Random field theory is a recent branch of mathematics that can be used to solve this problem.

To explain the use of RFT, we will first go back to the basics of hypothesis testing in statistics. We describe the multiple comparison problem and the usual solution, which is the Bonferroni correction. We explain why spatial correlation in imaging data causes problems for the Bonferroni correction and introduce RFT as a solution. Finally, we discuss the assumptions underlying RFT and the problems that arise when these assumptions do not hold. We hope this chapter will be accessible to those with no specific expertise in mathematics or statistics. Those more interested in mathematical details and recent developments are referred to Chapter 15.

## 1.1 Rejecting the null hypothesis

When we calculate a statistic, we often want to decide whether the statistic represents convincing evidence of the effect we are interested in. Usually we test the statistic against the null hypothesis, which is the hypothesis that there is no effect. If the statistic is not compatible with the null hypothesis, we may conclude that there is an effect. To test against the null hypothesis, we can compare our statistic value to a *null distribution*, which is the distribution of statistic values we would expect if there is no effect. Using the null distribution, we can estimate how likely it is that our statistic could have come about by chance. We may find that the result we found has a 5% chance of resulting from a null distribution. We therefore decide to reject the null hypothesis, and accept the alternative hypothesis that there is an effect. In rejecting the null hypothesis, we must accept a 5% chance that the result has in fact arisen when there is in fact no effect, i.e. the null hypothesis is true. 5% is our expected *type I* error rate, or the chance that we take that we are wrong when we reject the null hypothesis.

For example, when we do a single t test, we compare the t value we have found to the null distribution for the t statistic. Let us say we have found a t value of 2.42, and have 40 degrees of freedom. The null distribution of t statistics with 40 degrees of freedom tells us that the probability of observing a value greater than or equal to 2.42, if there is no effect, is only 0.01. In our case, we can reject the null hypothesis with a 1% risk of type I error.

The situation is more complicated in functional imaging because we have many voxels and therefore many statistic values. If we do not know where in the brain our effect will occur, our hypothesis refers to the whole volume of statistics in the brain. Evidence against the null hypothesis would be that the whole observed *volume* of values is unlikely to have arisen from a null distribution. The question we are asking is now a question about the volume, or *family* of voxel statistics, and the risk of error that we are prepared to accept is the Family-Wise Error rate (FWE) — which is the likelihood that this family of voxel values could have arisen by chance.

We can test a family-wise null hypothesis in a variety of ways, but one useful method is to look for any statistic values that are larger than we would expect, if they all had come from a null distribution. The method requires that we find a threshold to apply to every statistic value, so that any values above the threshold are unlikely to have arisen by chance. This is often referred to as “height thresholding”, and it has the advantage that if we find voxels above threshold, we can conclude that there is an effect at these voxel locations — i.e. the test has localizing power. Alternative procedures based on cluster- and set-level inferences are discussed in section 4.

A height threshold that can control family-wise error must take into account the number of tests. We saw above that a single t statistic value from a null distribution with 40 degrees of freedom has a 1% probability of being greater than 2.42. Now imagine our experiment has generated 1000 t values with 40 degrees of freedom. If we look at any single statistic, then by chance it will have a 1% probability of being greater than 2.42. This means that we would expect 10 t values in our sample of 1000 to be greater than 2.42. So, if we see one or more t values above 2.42 in this family of tests, this is not good evidence against the *family-wise* null hypothesis, which is that all these values have been drawn from a null distribution. We need to find a new threshold, such that, in a family of 1000 t statistic values, there is a 1% probability of there being *one or more* t values above that threshold. The Bonferroni correction is a simple method of setting this threshold.

## 2 The Bonferroni correction

The Bonferroni correction is based on simple probability rules. Imagine we have taken our t values and used the null t distribution to convert them to probability values. We then apply a probability threshold  $\alpha$  to each of our  $n$  probability values; in our previous example  $\alpha$  was 0.01, and  $n$  was 1000. If all the test values are drawn from a null distribution, then each of our  $n$  probability values has a probability  $\alpha$  of being greater than threshold. The probability of *all* the tests being less than  $\alpha$  is therefore  $(1 - \alpha)^n$ . The family-wise error rate ( $P^{FWE}$ ) is the probability that one or more values will be greater than  $\alpha$ , which is simply:

$$P^{FWE} = 1 - (1 - \alpha)^n \quad (1)$$

Because  $\alpha$  is small this can be approximated by the simpler expression:

$$P^{FWE} \leq n\alpha \quad (2)$$

Using equation 2, we can find a single-voxel probability threshold  $\alpha$  that will give us our required family-wise error rate,  $P^{FWE}$ , such that we have a  $P^{FWE}$  probability of seeing any voxel above threshold in all of the  $n$  values. We simply solve equation 2 for  $\alpha$ :

$$\alpha = P^{FWE}/n \quad (3)$$

If we have a brain volume of 100,000 t statistics, all with 40 degrees of freedom, and we want a FWE rate of 0.05, then the required probability threshold for a single voxel, using the Bonferroni correction, would be  $0.05/100,000 = 0.0000005$ . The corresponding t statistic is 5.77. If any voxel t statistic is above 5.77, then we can conclude that a voxel statistic of this magnitude has only a 5% chance of arising anywhere in a volume of 100,000 t statistics drawn from the null distribution.

The Bonferroni procedure gives a *corrected* p value; in the case above, the uncorrected p value for a voxel with a t statistic of 5.77 was 0.0000005; the p value corrected for the number of comparisons is 0.05.

The Bonferroni correction is used for calculating FWE rates for some functional imaging analyses. However, in many cases, the Bonferroni correction is too conservative because most functional imaging data have some degree of spatial correlation - i.e. there is correlation between neighbouring statistic values. In this case, there are fewer *independent* values in the statistic volume than there are voxels.

## 2.1 Spatial correlation

Some degree of spatial correlation is almost universally present in functional imaging data. In general, data from any one voxel in the functional image will tend to be similar to data from nearby voxels, even after the modelled effects have been removed. Thus the errors from the statistical model will tend to be correlated for nearby voxels. The reasons for this include factors inherent in collecting the image, physiological signal that has not been modeled, and spatial preprocessing applied to the data before statistical analysis.

For PET data, much more than for fMRI, nearby voxels are related because of the way that the scanner collects and reconstructs the image. Thus, data that does in fact arise from a single voxel location in the brain will also cause some degree of signal change in neighbouring voxels in the resulting image. The extent to which this occurs is a measure of the performance of the PET scanner, and is referred to as the point spread function.

Spatial preprocessing of functional data introduces spatial correlation. Typically, we will realign images for an individual subject to correct for motion during the scanning session (see Chapter 2), and may also spatially normalize a subject's brain to a template to compare data between subjects (see Chapter 3). These transformations will require the creation of new resampled images, which have voxel centres that are very unlikely to be the same as those in the original images. The resampling requires that we estimate the signal for these new voxel locations from the values in the original image, and typical resampling methods require some degree of averaging of

neighbouring voxels to derive the new voxel value (see Chapter 2).

It is very common to smooth the functional images before statistical analysis. A proportion of the noise in functional images is independent from voxel to voxel, whereas the signal of interest usually extends over several voxels. This is due both to the possibly distributed nature of neuronal sources and the spatially extended nature of the haemodynamic response (see Chapter 11). According to the matched filter theorem smoothing will therefore improve the signal to noise ratio. For multiple subject analyses, smoothing may also be useful for blurring the residual differences in location between corresponding areas of functional activation. Smoothing involves averaging over voxels, which will by definition increase spatial correlation.

## 2.2 The Bonferroni correction and independent observations

Spatial correlation means that there are fewer independent observations in the data than there are voxels. This means that the Bonferroni correction will be too conservative because the family-wise probability from equation 1 relies on the individual probability values being independent, so that we can use multiplication to calculate the probability of combined events. For equation 1, we used multiplication to calculate the probability that all tests will be below threshold with  $(1 - \alpha)^n$ . Thus the  $n$  in the equation must be the number of *independent* observations. If we have  $n$  voxels in our data, but there are only  $n_i$  independent observations, then equation 1 becomes  $P^{FWE} = 1 - (1 - \alpha)^{n_i}$ , and the corresponding  $\alpha$  from equation 3 is given by  $\alpha = P^{FWE}/n_i$ . This is best illustrated by example.

Let us take a single image slice, of 100 by 100 voxels, with a t statistic value for each voxel. For the sake of simplicity, let the t statistics have very high degrees of freedom, so that we can consider the t statistic values as being from the normal distribution - i.e. that they are Z scores. We can simulate this slice from a null distribution by filling the voxel values with independent random numbers from the normal distribution, which results in an image such as that in figure 1.

If this image had come from the analysis of real data, we might want to test if any of the numbers in the image are more positive than is likely by chance. The values are independent, so the Bonferroni correction will give an accurate threshold. There are 10,000 Z scores, so the Bonferroni threshold,  $\alpha$ , for a FWE rate of 0.05, is  $0.05/10000 = 0.000005$ . This corresponds to a Z score of 4.42. Given the null hypothesis (which is true in this case) we would expect only 5 out of 100 such images to have one or more Z scores

more positive than 4.42.

The situation changes if we add spatial correlation. Let us perform the following procedure on the image: break up the image into squares of 10 by 10 pixels; for each square, calculate the mean of the 100 values contained; replace the 100 random numbers in the square by the mean value<sup>1</sup>. The image that results is shown in figure 2.

We still have 10000 numbers in our image, but there are only 10 by 10 = 100 numbers that are independent. The appropriate Bonferroni correction is now  $0.05 / 100 = 0.0005$ , which corresponds to a Z score of 3.29. We would expect only 5 of 100 of such images to have a square of values greater than 3.29 by chance. If we had assumed all the values were independent, then we would have used the correction for 10,000 values, of  $\alpha = 0.000005$ . Because we actually have only 100 independent observations, equation 2, with  $n = 100$  and  $\alpha = 0.000005$ , tells us that we expect a FWE rate of 0.0005, which is one hundred times lower (i.e. more conservative) than the rate that we wanted.

### 2.3 Smoothing and independent observations

In the preceding section we replaced a square of values in the image with their mean in order to show the effect of reducing the number of independent observations. This procedure is a very simple form of smoothing. When we smooth an image with a smoothing kernel such as a Gaussian, each value in the image is replaced with a weighted average of itself and its neighbours. Figure 3 shows the image from figure 1 after smoothing with a Gaussian kernel of Full Width at Half Maximum (FWHM) of 10 pixels<sup>2</sup>. An FWHM of 10 pixels means that, at five pixels from the centre, the value of the kernel is half its peak value. Smoothing has the effect of blurring the image, and reduces the number of independent observations.

The smoothed image contains spatial correlation, which is typical of the output from the analysis of functional imaging data. We now have a problem, because there is no simple way of calculating the number of independent ob-

---

<sup>1</sup>Averaging the random numbers will make them tend to zero; to return the image to a variance of 1, we need to multiply the numbers in the image by 10; this is  $\sqrt{n}$ , where  $n$  is the number of values we have averaged.

<sup>2</sup>As for the procedure where we took the mean of the 100 observations in each square, the smoothed values will no longer have a variance of one, because the averaging involved in smoothing will make the values tend to zero. As for the square example, we need to multiply the values in the smoothed image by a scale factor to return the variance to one; the derivation of the scale factor is rather technical, and not relevant to our current discussion.

servations in the smoothed data, so we cannot use the Bonferroni correction. This problem can be addressed using random field theory.

### 3 Random field theory

Random field theory (RFT) is a recent body of mathematics defining theoretical results for smooth statistical maps. The theory has been versatile in dealing with many of the thresholding problems that we encounter in functional imaging. Among many other applications, it can be used to solve our problem of finding the height threshold for a smooth statistical map which gives the required family-wise error rate.

The way that RFT solves this problem is by using results that give the expected *Euler characteristic* (EC) for a smooth statistical map that has been thresholded. We will discuss the EC in more detail below; for now it is only necessary to note that the expected EC leads directly to the expected number of clusters above a given threshold, and that this in turn gives the height threshold that we need.

The application of RFT proceeds in stages. First we estimate the smoothness (spatial correlation) of our statistical map. Then we use the smoothness values in the appropriate RFT equation, to give the expected EC at different thresholds. This allows us to calculate the threshold at which we would expect 5% of equivalent statistical maps arising under the null hypothesis to contain at least one area above threshold.

#### 3.1 Smoothness and resels

Usually we do not know the smoothness of our statistical map. This is so even if the map resulted from smoothed data, because we usually do not know the extent of spatial correlation in the underlying data before smoothing. If we do not know the smoothness, it can be calculated using the observed spatial correlation in the images. For our example (figure 3), however, we *know* the smoothness, because the data were independent before smoothing. In this case, the smoothness results entirely from the smoothing we have applied. The smoothness can be expressed as the width of the smoothing kernel, which was 10 pixels FWHM in the x and y direction. We can use the FWHM to calculate the number of *resels* in the image. “Resel” was a term introduced by Worsley [14], and is a measure of the number of “resolution elements” in the statistical map. This can be thought of as *similar to* the number of independent observations, but it is not the same, as we will see below. A resel is simply defined as a block of values (in our case, pixels) that is the

same size as the FWHM. For the image in figure 3, the FWHMs were 10 by 10 pixels, so that a resel is a block of 100 pixels. As there are 10,000 pixels in our image, there are 100 resels. Note that the number of resels depends only on the smoothness (FWHM) and the number of pixels.

### 3.2 The Euler characteristic

The Euler characteristic is a property of an image after it has been thresholded. For our purposes, the EC can be thought of as the number of blobs in an image after thresholding. For example, we can threshold our smoothed image (figure 3) at  $Z=2.5$ ; all pixels with  $Z$  scores less than 2.5 are set to zero, and the rest are set to one. This results in the image in figure 4.

There are three white blobs in figure 4, corresponding to three areas with  $Z$  scores higher than 2.5. The EC of this image is therefore 3. If we increase the  $Z$  score threshold to 2.75, we find that the two central blobs disappear – because the  $Z$  scores were less than 2.75 (figure 5).

The area in the upper right of the image remains; the EC of the image in figure 5 is therefore one. At high thresholds the EC is either one or zero. Hence, the average or expected EC, written  $\mathbf{E}[EC]$ , corresponds (approximately) to the probability of finding an above threshold blob in our statistic image. That is, the probability of a family wise error is approximately equivalent to the expected Euler Characteristic,  $P^{FWE} \approx \mathbf{E}[EC]$ .

It turns out that if we know the number of resels in our image, it is possible to calculate  $\mathbf{E}[EC]$  at any given threshold. For an image of two dimensions  $\mathbf{E}[EC]$  is given by Worsley [14]. If  $R$  is the number of resels,  $Z_t$  is the  $Z$  score threshold, then:

$$\mathbf{E}[EC] = R(4\log_e 2)(2\pi)^{-\frac{3}{2}} Z_t e^{-\frac{1}{2}Z_t^2}; \quad (4)$$

Figure 6 shows  $\mathbf{E}[EC]$  for an image of 100 resels, for  $Z$  score thresholds between zero and five. As the threshold drops from one to zero,  $\mathbf{E}[EC]$  drops to zero; this is because the precise definition of the EC is more complex than simply the number of blobs [9]. This makes a difference at low thresholds but is not relevant for our purposes because, as explained above, we are only interested in the properties of  $\mathbf{E}[EC]$  at high thresholds ie. when it approximates  $P^{FWE}$ .

Note also that the graph in figure 6 does a reasonable job of predicting the EC in our image; at a  $Z$  threshold of 2.5 it predicted an EC of 1.9, when we observed a value of 3; at  $Z=2.75$  it predicted an EC of 1.1, for an observed EC of 1.

We can now apply RFT to our smoothed image (figure 3) which has 100 resels. For 100 resels, equation 4 gives an  $E[EC]$  of 0.049 for a  $Z$  threshold of 3.8 (c.f. the graph in figure 6). If we have a two dimensional image with 100 resels, then the probability of getting one or more blobs where  $Z$  is greater than 3.8, is 0.049. We can use this for thresholding. Let  $x$  be the  $Z$  score threshold that gives an  $E[EC]$  of 0.05. If we threshold our image at  $x$ , we can conclude that any blobs that remain have a probability of less than or equal to 0.05 that they have occurred by chance. From equation 4, the threshold,  $x$ , depends only on the number of resels in our image.

### 3.3 Random field thresholds and the Bonferroni correction

The random field correction derived using the EC is not the same as a Bonferroni correction for the number of resels. We stated above that the resel count in an image is not exactly the same as the number of independent observations. If it was the same, then instead of using RFT, we could use a Bonferroni correction based on the number of resels. However, these two corrections give different answers. For  $\alpha = 0.05$ , the  $Z$  threshold according to RFT, for our 100 resel image, is  $Z=3.8$ . The Bonferroni threshold for 100 independent tests, is  $0.05/100$ , which equates to a  $Z$  score of 3.3. Although the RFT maths gives us a correction that is similar in principle to a Bonferroni correction, it is not the same. If the assumptions of RFT are met (see section 4) then the RFT threshold is more accurate than the Bonferroni.

### 3.4 Random fields and functional imaging

Analyses of functional imaging data usually lead to three dimensional statistical images. So far we have discussed the application of RFT to an image of two dimensions, but the same principles apply in three dimensions. The EC is the number of 3D blobs of  $Z$  scores above a certain threshold and a resel is a cube of voxels of size (FWHM in  $x$ ) by (FWHM in  $y$ ) by (FWHM in  $z$ ). The equation for  $E[EC]$  is different in the 3D case, but still depends only on the resels in the image.

For the sake of simplicity, we have only considered a random field of  $Z$  scores, i.e. numbers drawn from the normal distribution. There are now equivalent results for  $t$ ,  $F$  and  $\chi^2$  random fields [10]. For example, SPM99 software uses formulae for  $t$  and  $F$  random fields to calculate corrected thresholds for height.

As noted in section 3.1, we usually do not know the smoothness of a statistic volume from a functional imaging analysis, because we do not know

the extent of spatial correlation before smoothing. We cannot assume that the smoothness is the same as any explicit smoothing that we have applied and will need to calculate smoothness from the images themselves. In practice, smoothness is calculated using the residual values from the statistical analysis as described in [6] and [13].

### 3.5 Small volume correction

We noted above that the results for the expected Euler characteristic depend only on the number of resels contained in the volume of voxels we are analyzing. This is not strictly accurate, although it is a very close approximation when the voxel volume is large compared to the size of a resel [9]. In fact,  $E[EC]$  also depends on the shape and size of the volume. The shape of the volume becomes important when we have a small or oddly shaped region. This is unusual if we are analyzing statistics from the whole brain, but there are often situations where we wish to restrict our search to a smaller subset of the volume, for example where we have a specific hypothesis as to where our signal is likely to occur.

The reason that the shape of the volume may influence the correction is best explained by example. Let us return to the 2D image of smoothed random numbers (figure 3). We could imagine that we had reason to believe that signal change will occur only in the centre of the image. Our search region will not be the whole image, but might be a box at the image centre, with size 30 by 30 pixels (see figure 7).

The box contains 9 resels. The figure shows a grid of X-shaped markers; these are spaced at the width of a resel, i.e. 10 pixels. The box can contain a maximum of 16 of these markers. Now let us imagine we had a more unusually shaped search region. For some reason we might expect that our signal of interest will occur within a frame 2.5 pixels wide around the outside of the image. The frame contains the same number of voxels as the box, and therefore has the same volume in terms of resels. However, the frame contains many more markers (32), so the frame is sampling from the data of more resels than the box. Multiple comparison correction for the frame must therefore be more stringent than for the box.

In fact,  $E[EC]$  depends on the volume, surface area, and diameter of the search region [9]. These parameters can be calculated for continuous shapes for which formulae are available for volume, surface area and diameter, such as spheres or boxes. Otherwise, the parameters can be estimated from any shape that has been defined in an image. Restricting the search region to a small volume within a statistical map can lead to greatly reduced thresholds for given FWE rates. For the figure below (figure 8), we assumed a statistical

analysis that resulted in a  $t$  statistic map with 8mm smoothness in the X, Y and Z directions. The  $t$  statistic has 200 degrees of freedom, and we have a spherical search region. The graph shows the  $t$  statistic value that gives a corrected  $p$  value of 0.05 for spheres of increasing radius.

For a sphere of zero radius, the threshold is simply that for a single  $t$  statistic (uncorrected = corrected  $p = 0.05$  for  $t = 1.65$  with 200 degrees of freedom). The corrected  $t$  threshold increases sharply as the radius increases to  $\approx 10mm$ , and less steeply thereafter.

### 3.6 Uncorrected $p$ values and regional hypotheses

When making inferences about regional effects (e.g. activations) in SPMs, one often has some idea about where the activation should be. In this instance a correction for the entire search volume is inappropriate.

If the hypothesized region contained a single voxel, then inference could be made using an uncorrected  $p$ -value (as there is no extra search volume to correct for). In practice, however, the hypothesized region will usually contain many voxels and can be characterised, for example, using spheres or boxes centred on the region of interest, and we must therefore use a  $p$ -value that has been appropriately corrected. As described in the previous section, this corrected  $p$  value will depend on the size and shape of the hypothesized region, and the smoothness of the statistic image.

Some research groups have used uncorrected  $p$  value thresholds, such as  $p < 0.001$ , in order to control FWE when there is a regional hypothesis of where the activation will occur. This approach gives unquantified error control, however, because any hypothesized region is likely to contain more than a single voxel. For example, for 8mm smoothness, a spherical region with a radius greater than 6.7mm will require an uncorrected  $p$  value threshold of less than 0.001 for a FWE rate  $\leq 0.05$ . For a sphere of radius 15mm, an uncorrected  $p$  value threshold of 0.001 gives an  $E[EC]$  of 0.36, so there is approximately a 36% chance of seeing one or more voxels above threshold even if the null hypothesis is true.

## 4 Discussion

In this chapter we have focussed on voxel-level inference based on height thresholds to ask the question: is activation at a given voxel significantly non-zero? More generally, however, voxel-level inference can be placed in a larger framework involving cluster-level and set-level inference. These require height and spatial extent thresholds to be specified by the user. Corrected

p-values can then be derived that pertain to; (i) the number of activated regions (i.e. number of clusters above the height and volume threshold) - set level inferences, (ii) the number of activated voxels (i.e. volume) comprising a particular region - cluster level inferences and (iii) the p-value for each voxel within that cluster - voxel level inferences. These p-values are corrected for the multiple dependent comparisons and are based on the probability of obtaining  $c$ , or more, clusters with  $k$ , or more, voxels, above a threshold  $u$  in an SPM of known or estimated smoothness. Chapter 16, for example, describes cluster-level inference in a PET auditory stimulation study. Set-level inferences can be more powerful than cluster-level inferences and cluster-level inferences can be more powerful than voxel-level inferences. The price paid for this increased sensitivity is reduced localizing power. Voxel-level tests permit individual voxels to be identified as significant, whereas cluster and set-level inferences only allow clusters or sets of clusters to be declared significant. It should be remembered that these conclusions, about the relative power of different inference levels, are based on distributed activations. Focal activation may well be detected with greater sensitivity using voxel-level tests based on peak height. Typically, people use voxel-level inferences and a spatial extent threshold of zero. This reflects the fact that characterizations of functional anatomy are generally more useful when specified with a high degree of anatomical precision.

There are two assumptions underlying RFT. The first is that the error fields are a reasonable lattice approximation to an underlying random field with a multivariate Gaussian distribution. The second is that these fields are continuous, with a twice-differentiable autocorrelation function. A common misconception is that the autocorrelation function has to be Gaussian. But this is not the case.

If the data have been sufficiently smoothed and the General Linear Models correctly specified (so that the errors are indeed Gaussian) then the RFT assumptions will be met. One scenario in which the assumptions may not be met, however, is in a random effects analysis (see Chapter 12) with a small number of subjects. This is because the resulting error fields will not be very smooth and so may violate the ‘reasonable lattice approximation’ assumption. One solution is to reduce the voxel size by sub-sampling. Alternatively one can turn to different inference procedures. One such alternative is the nonparametric framework described in Chapter 16 where, for example, inference on statistic images with low degrees of freedom can be improved with the use of pseudo-t statistics which are much smoother than the corresponding t statistics.

Other inference frameworks are the False Discovery Rate (FDR) approach and Bayesian Inference. Whilst RFT controls the family-wise error, the prob-

ability of reporting a false positive anywhere in the volume, FDR controls the proportion of false positives amongst those that are declared positive. This very different approach is discussed in the next chapter. Finally, Chapter 17 introduces Bayesian inference where, instead of focussing on how unlikely the data are under a null hypothesis, inferences are made on the basis of a posterior distribution which characterises our parameter uncertainty without reference to a null distribution.

## 4.1 Bibliography

The mathematical basis of RFT is described in a series of peer-reviewed articles in statistical journals [10, 8, 12, 1, 11]. The core paper for RFT as applied to functional imaging is [9]. This provides estimates of  $p$ -values for local maxima of Gaussian,  $t$ ,  $\chi^2$  and  $F$  fields over search regions of any shape or size in any number of dimensions. This unifies earlier results on 2D [4] and 3D [14] images.

The above analysis requires an estimate of the smoothness of the images. In [7], Poline *et al.* estimate the dependence of the resulting SPMs on the estimate of this parameter. Whilst the applied smoothness is usually fixed, [15] propose a scale-space procedure for assessing significance of activations over a range of proposed smoothings. In [6], the authors implement an unbiased smoothness estimator for Gaussianised  $t$ -fields and  $t$ -fields. Worsley *et al.* [13] derive a further improved estimator, which takes into account some non-stationarity of the statistic field.

Another approach to assessing significance is based, not on the height of activity, but on spatial extent [3], as described in the previous section. In [5], the authors consider a hierarchy of tests that are regarded as voxel-level, cluster-level and set-level inferences. If the approximate location of an activation can be specified in advance then the significance of the activation can be assessed using the spatial extent or volume of the nearest activated region [2]. This test is particularly elegant as it does not require a correction for multiple comparisons.

More recent developments in applying RFT to neuroimaging are described in the following chapter. Finally, we refer readers to an online resource, <http://www.mrc-cbu.cam.ac.uk/Imaging/randomfields.html>, from which much of the material in this chapter was collated.

## References

- [1] J. Cao and K.J. Worsley. The detection of local shape changes via the geometry of Hotelling's  $T^2$  fields. *Annals of Statistics*, 27(3):925–942, 1999.
- [2] Karl J. Friston. Testing for anatomically specified regional effects. *Human Brain Mapping*, 5:133–136, 1997.
- [3] K.J. Friston, K.J. Worsley R.S.J. Frackowiak, J.C. Mazziotta, and A.C. Evans. Assessing the significance of focal activations using their spatial extent. *Human Brain Mapping*, 1:214–220, 1994.
- [4] K.J. Friston, C.D. Frith, P.F. Liddle, and R.S.J. Frackowiak. Comparing functional (PET) images: The assessment of significant change. *Journal of Cerebral Blood Flow and Metabolism*, 11:690–699, 1991.
- [5] K.J. Friston, A. Holmes, J.-B. Poline, C.J. Price, and C.D. Frith. Detecting activations in PET and fMRI: Levels of inference and power. *NeuroImage*, 40:223–235, 1995.
- [6] Stefan J. Kiebel, Jean-Baptiste Poline, Karl J. Friston, Andrew P. Holmes, and Keith J. Worsley. Robust smoothness estimation in statistical parametric maps using standardized residuals from the general linear model. *NeuroImage*, 10:756–766, 1999.
- [7] J.-B. Poline, K. J. Friston, K. J. Worsley, and R. S. J. Frackowiak. Estimating smoothness in statistical parametric maps: Confidence intervals on  $p$ -values. *J. Comput. Assist. Tomogr.*, 19(5):788–796, 1995.
- [8] D.O. Sigmund and K.J. Worsley. Testing for a signal with unknown location and scale in a stationary Gaussian random field. *Annals of Statistics*, 23:608–639, 1995.
- [9] K. J. Worsley, S. Marrett, P. Neelin, A. C. Vandal, K. J. Friston, and A. C. Evans. A unified statistical approach for determining significant voxels in images of cerebral activation. *Human Brain Mapping*, 4:58–73, 1996.
- [10] K.J. Worsley. Local maxima and the expected Euler characteristic of excursion sets of  $\chi^2$ ,  $F$  and  $t$  fields. *Advances in Applied Probability*, 26:13–42, 1994.

- [11] K.J. Worsley. Estimating the number of peaks in a random field using the Hadwiger characteristic of excursion sets, with applications to medical images. *Annals of Statistics*, 23:640–669, 1995.
- [12] K.J. Worsley. The geometry of random images. *Chance*, 9(1):27–40, 1996.
- [13] K.J. Worsley, M. Andermann, T. Koulis, D. MacDonald, and A.C. Evans. Detecting changes in nonisotropic images. *Human Brain Mapping*, 8:98–101, 1999.
- [14] K.J. Worsley, A.C. Evans, S. Marrett, and P. Neelin. A three dimensional statistical analysis for CBF activation studies in the human brain. *Journal of Cerebral Blood Flow and Metabolism*, 12:900–918, 1992.
- [15] K.J. Worsley, S. Marrett, P. Neelin, A.C. Vandal, K.J. Friston, and A.C. Evans. Searching scale space for activation in PET images. *Human Brain Mapping*, 4:74–90, 1995.

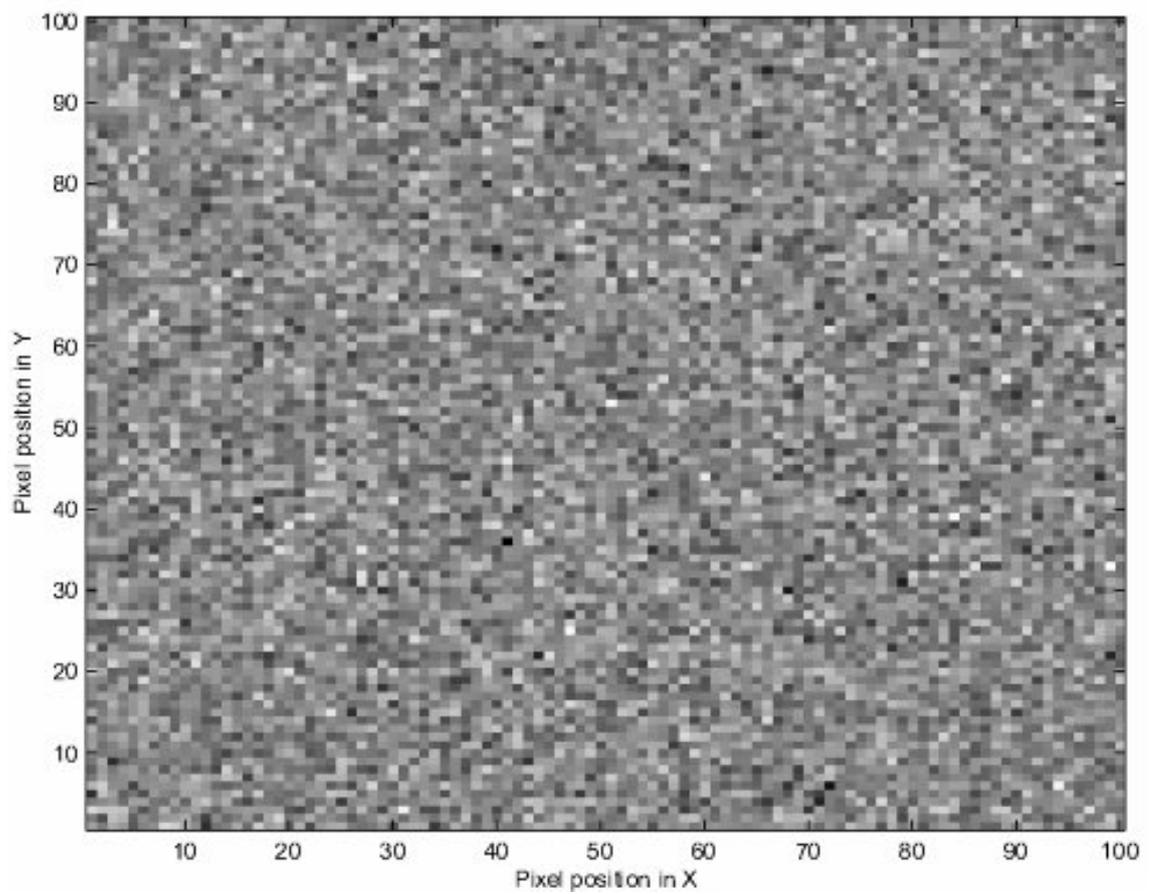


Figure 1: Simulated image slice using independent random numbers from the normal distribution. Whiter pixels are more positive.

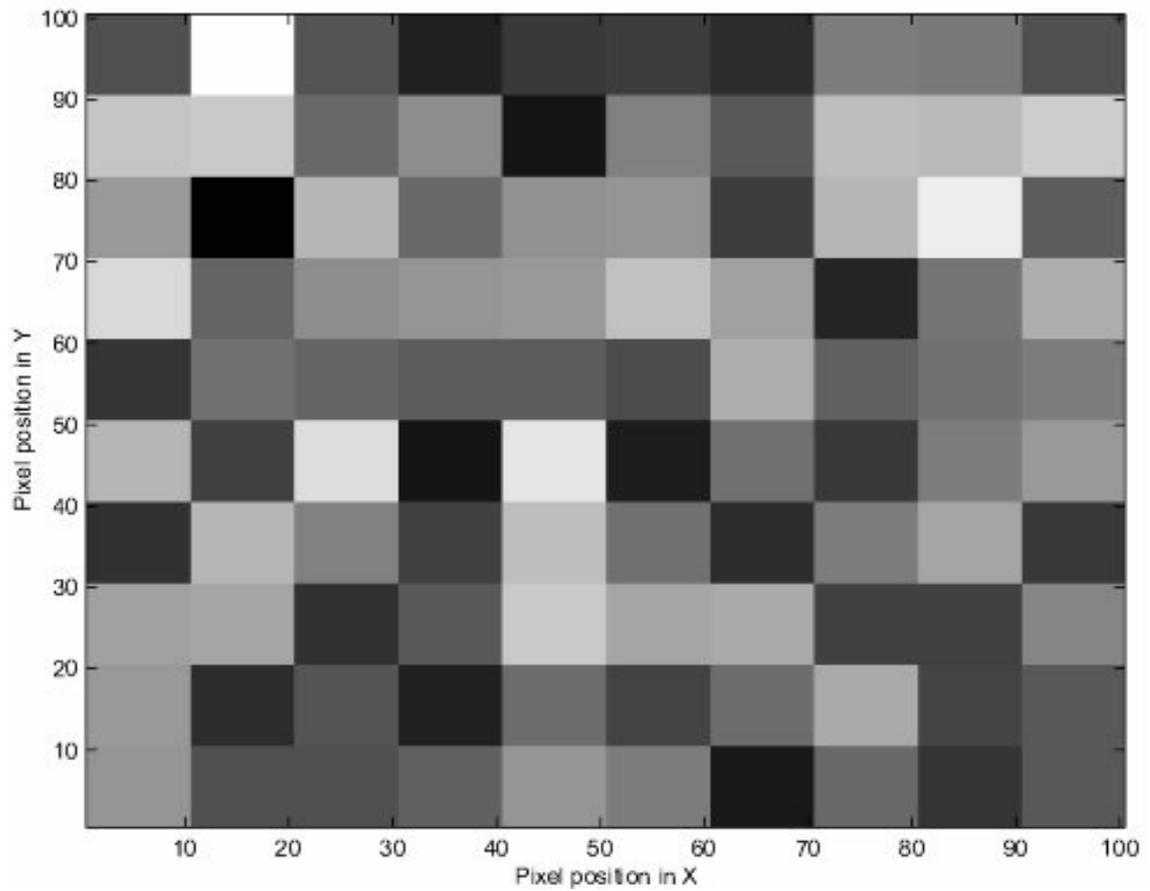


Figure 2: Random number image from figure 1 after replacing values in the 10 by 10 squares by the value of the mean within each square

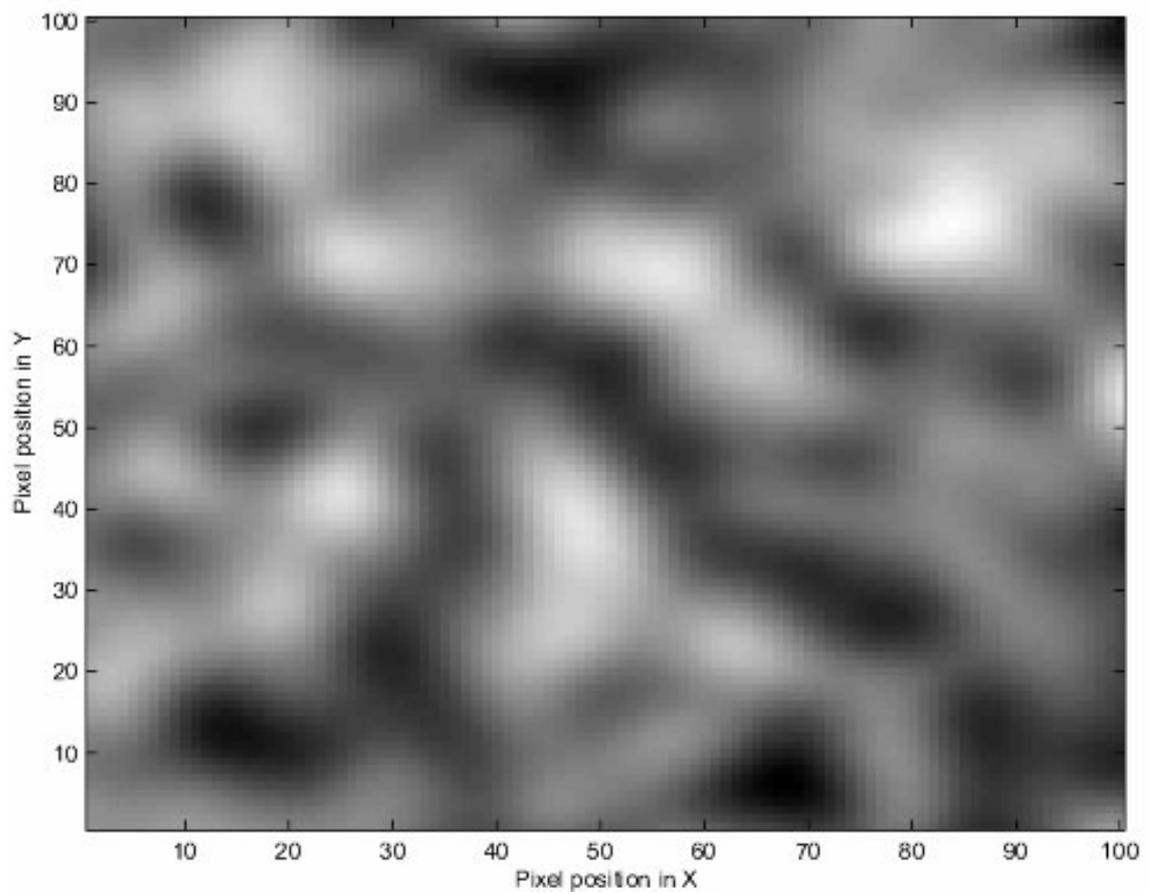


Figure 3: Random number image from figure 1 after smoothing with a Gaussian smoothing kernel of full width at half maximum of 10 pixels

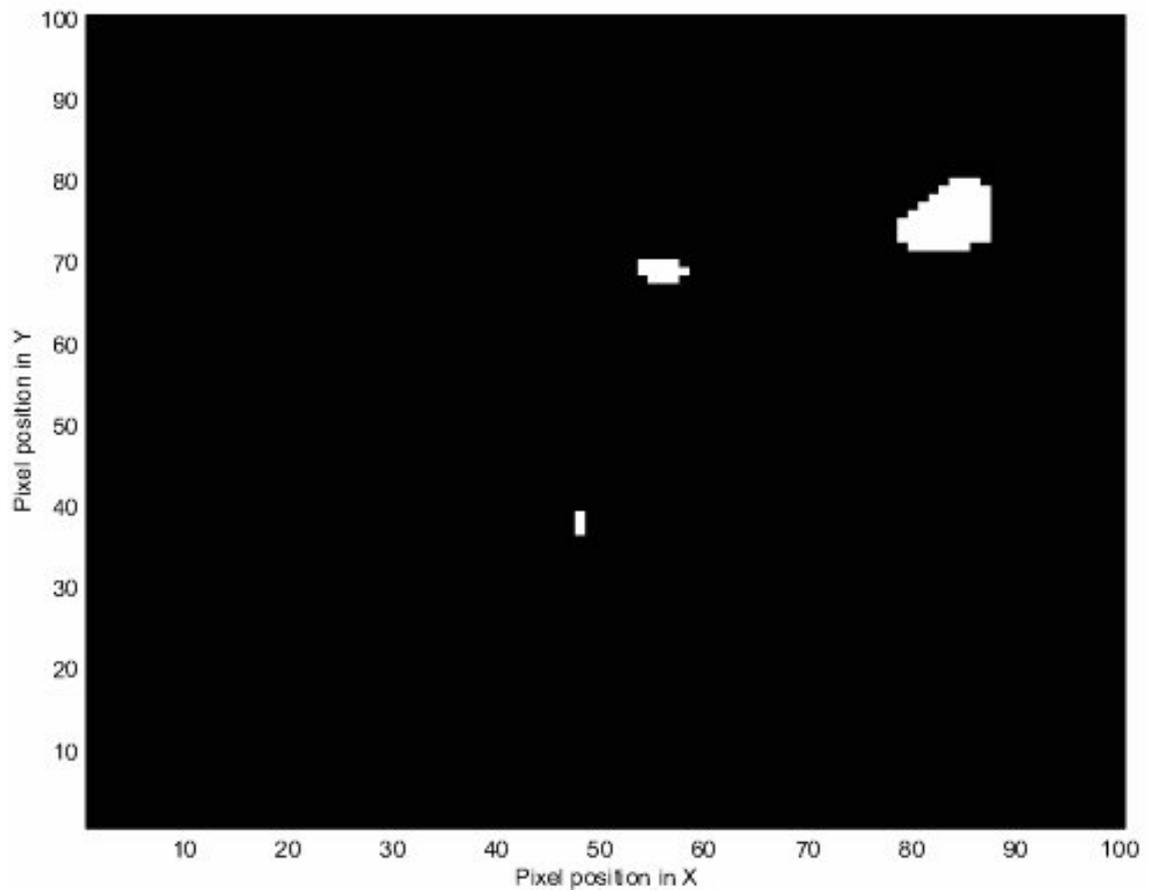


Figure 4: Smoothed random number image from figure 3 after thresholding at  $Z = 2.5$ . Values less than 2.5 have been set to zero (displayed as black). The remaining values have been set to one (displayed as white).

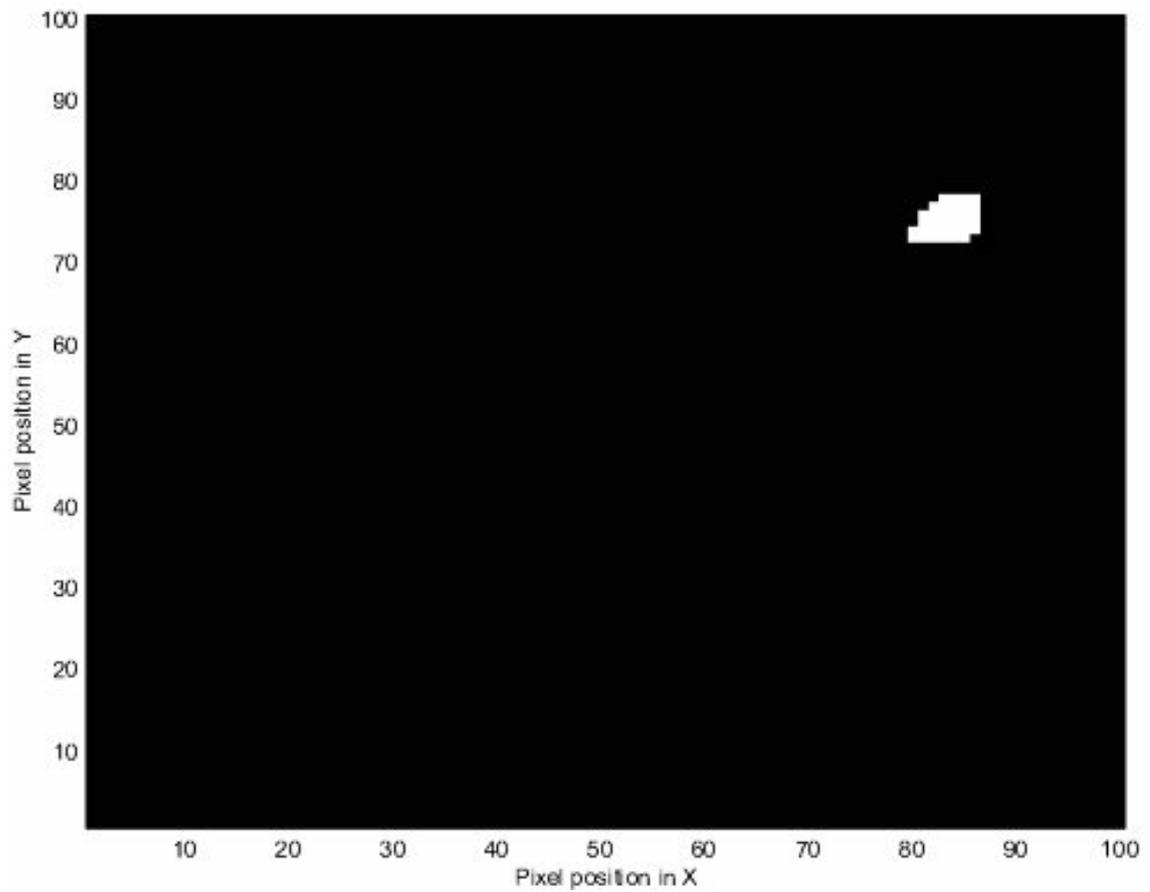


Figure 5: Smoothed random number image from figure 3 after thresholding at  $Z = 2.75$ . Values less than 2.75 have been set to zero (displayed as black). The remaining values have been set to one (displayed as white).

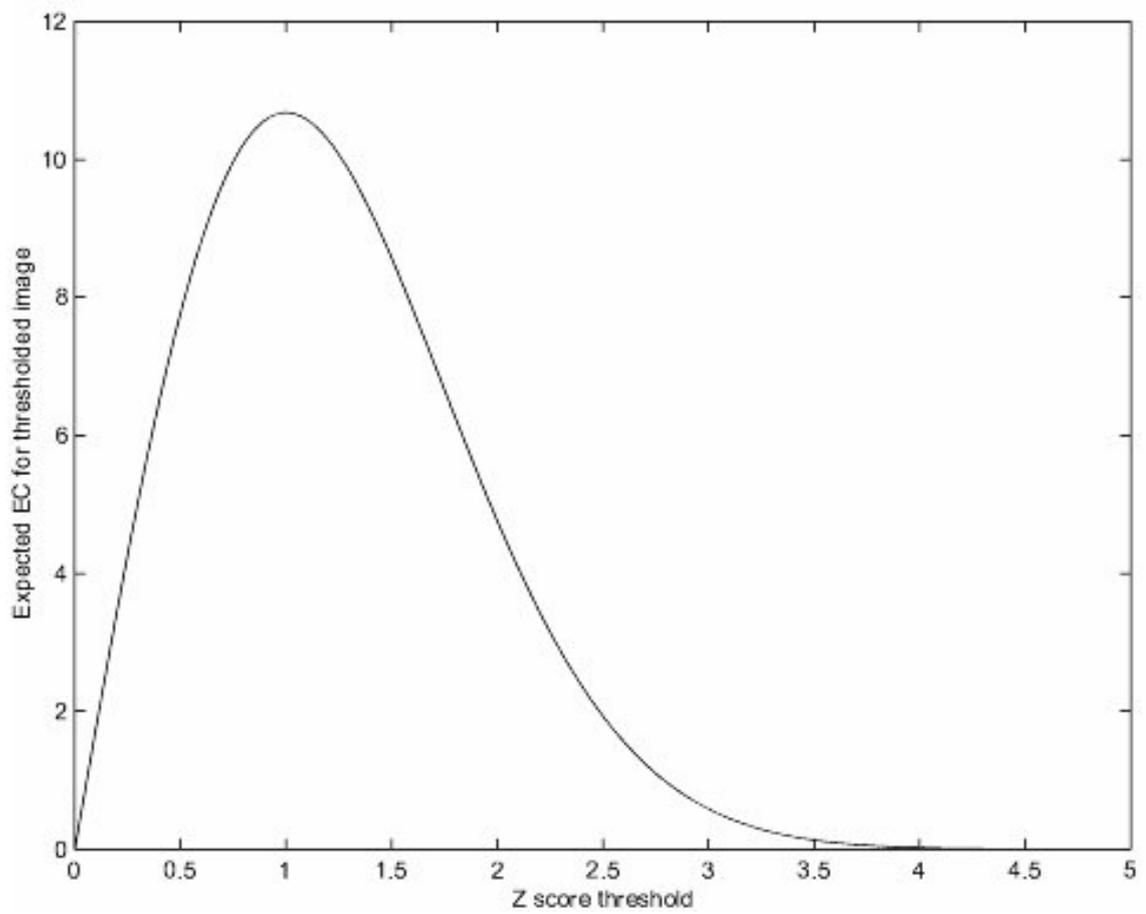


Figure 6: Expected EC values for an image of 100 resels

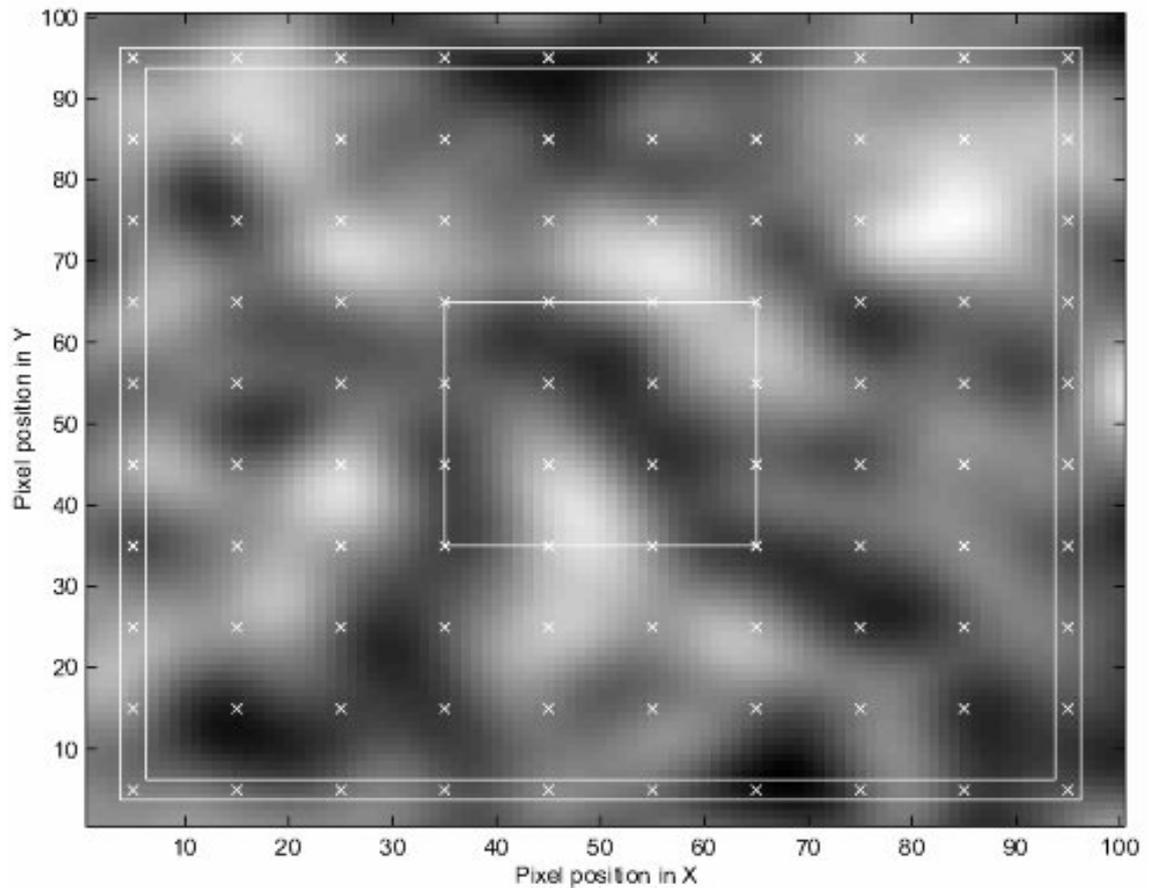


Figure 7: Smoothed random number image from figure 3 with two example search regions: a box (centre) and a frame (outer border of image). X-shaped markers are spaced at one resel widths across the image.

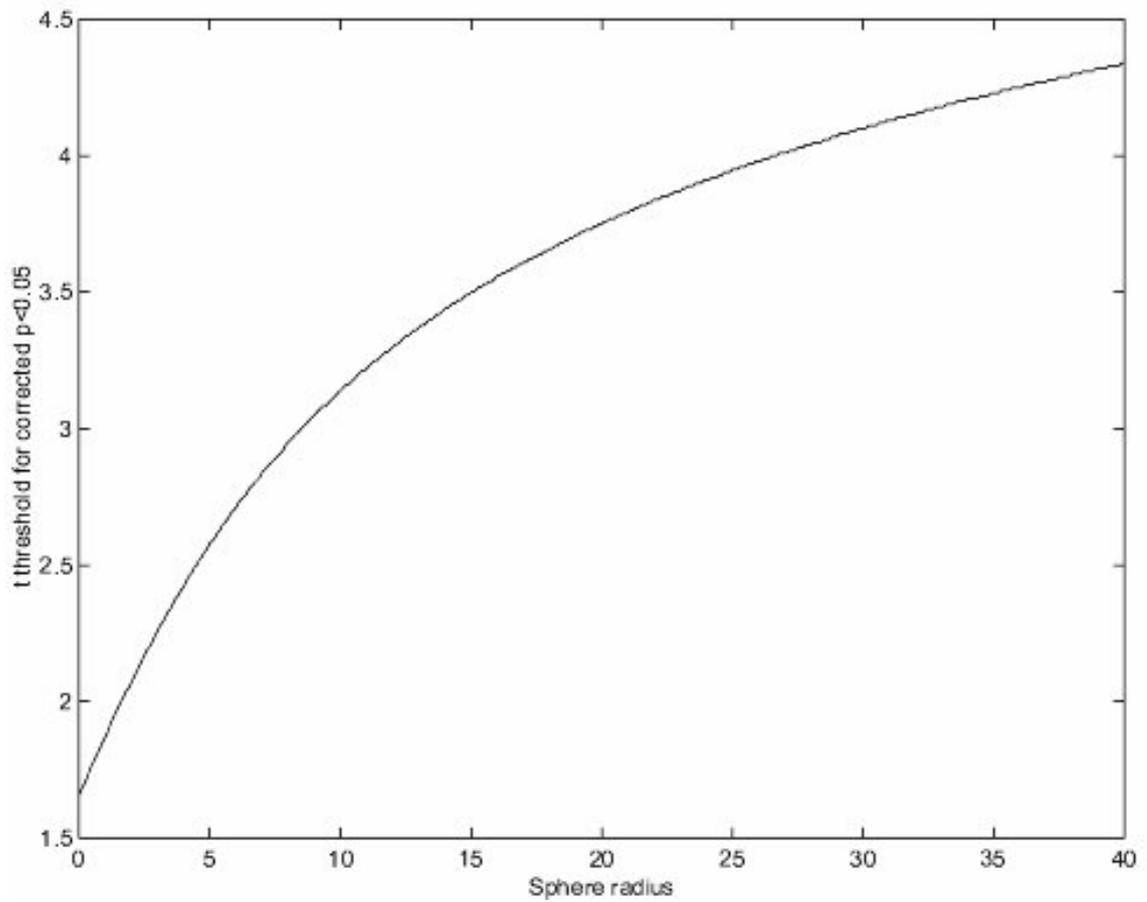


Figure 8: t threshold giving a FWE rate of 0.05, for spheres of increasing radius. Smoothness was 8mm in X Y and Z directions, and the example analysis had 200 degrees of freedom.