

# Chapter Five

## An Empirical Bayesian Approach

In this chapter the problem of testing statistic images is reformulated as an image segmentation problem, to which techniques from image processing are applied. In particular, a Markov Random Field is used to convey prior belief regarding the contiguous nature of activated voxels.

The work described in this chapter was presented orally at Brain PET'93, the first International Symposium on Quantification of Brain Function, held in Akita, Japan. An abstract appears in the *Annals of Nuclear Medicine* (Holmes & Ford, 1993a), and a full paper in the conference proceedings, *Quantification of Brain Function: Tracer Kinetics and Image Analysis in PET* (Holmes & Ford, 1993b).

## 5.1. Introduction and Motivation

In a functional mapping experiment, we may have substantial prior belief or information regarding the shape and loci of the activated area. The simplest prior belief is that activated voxels will form contiguous regions. Single threshold approaches to testing statistic images from functional mapping experiments voxel-by-voxel do not make use of this prior belief. Rather, they rely on the smoothness of the statistic image to ensure that the set of voxels declared as “activated” form a few regions of contiguous voxels. As we have seen, statistic images frequently exhibit a high degree of noise, especially for statistics formed with variance estimates of few degrees of freedom. This can lead to isolated voxels being declared as “activated”, contrary to our prior beliefs.

To counter this, it is common to smooth statistic images, so called secondary smoothing. As discussed in §3.3.6.6., this increases the signal to noise ratio for signals greater in extent than the filter kernel, at the expense of resolution, and is not always desirable. If prior belief about the contiguous nature of activated regions can be built into a test, then such secondary smoothing may not be necessary. In effect, the prior would act as an “intelligent” smoothing.

### ***Markov random fields***

In the image processing literature, Markov Random Fields (MRFs) have been used successfully to express prior beliefs about the spatial coherence of images in problems of reconstruction, restoration, and segmentation, the latter being the problem of labelling pixels with one of a finite set of labels. Green (1990) successfully utilises a MRF to express such prior belief when reconstructing Single Photon Emission Computed Tomography images. Geman and Geman (1984) consider the task of restoring a discrete grey level image corrupted by the addition of Gaussian white noise, as a segmentation problem, using a discrete MRF to model prior belief about the local structure of the labelling of pixels by their true grey level.

We shall consider the testing scenario as a segmentation problem, where each voxel is to be labelled as “activated” or not, according to the evidence against the null hypothesis.

## 5.2. MRFs and Gibbs Distributions

We begin by briefly reviewing the necessary theory of Markov random fields. For further details, see one of the many papers reviewing random field models for image analysis. A general summary, richly decorated with examples and algorithms, and with extensive references, is that recently published by Dubes & Jain (1989). This also appears in *Advances in Applied Statistics*, a supplement to the *Journal of Applied Statistics*, which is devoted to statistics and images (Dubes & Jain, 1993).

### 5.2.1. Markov Random Fields

Consider a partition of a two-dimensional image space,  $\Xi \subset \mathbb{R}^2$ , into  $K$  square pixels  $\{V_k\}_{k=1}^K$ . To avoid the complication of edges, we shall take the image space  $\Xi$  to be the surface of a torus, as was considered for the Two-Stage simulation study described in §4.2.1. Let  $W = \{k\}_{k=1}^K$  be a set of indices for the pixels. As usual, a pixel will be referred to by its index.

#### Neighbours

The first order *neighbours* of a pixel are those pixels it shares a side with, whose centres are at most 1 pixel unit away (Euclidean distance). The second order neighbours of a pixel are those that touch it, either at a side or just a corner. The second order neighbours have centres at most  $\sqrt{2}$  pixel units away (fig. 70).

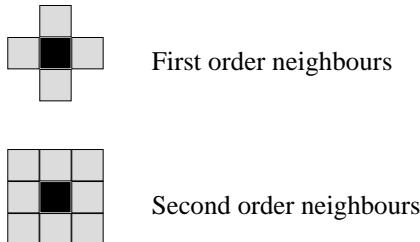


Figure 70  
First and second order neighbours in a two-dimensional image space,  
partitioned into square pixels.

For each pixel  $k \in W$ , define the *neighbourhood*  $\eta_k$ , to be the neighbouring pixels. The set  $\eta_W = \{\eta_k : k \in W\}$  is then the *neighbourhood system*.

#### Definition

Consider a random field defined on the lattice of pixel centres, with value at pixel  $k$  denoted by  $X_k$ . Let  $X$  be the vector of pixel values,  $X = (X_1, \dots, X_K)$ . The random field  $X$  is a *Markov Random Field* with respect to a neighbourhood system  $\eta_W$ , if and only if;

- a)  $\Pr(X_k = x_k | X_i = x_i, i \in W \setminus \{k\}) = \Pr(X_k = x_k | X_i = x_i, i \in \eta_k)$
- b)  $\Pr(X = \mathbf{x}) > 0$  for all possible configurations  $\mathbf{x} = (x_1, \dots, x_K)$

Here \ denotes set exclusion, so  $A \setminus B$  is the set of elements in  $A$  but not  $B$ .

Condition (a) is known as the Markov property. This states that the probability of  $X$  having a certain value at a particular pixel, given the values of  $X$  elsewhere, is dependent only on the values of  $X$  in the neighbourhood of that pixel. The positivity condition (b), states that all combinations of pixel values are possible. If in addition, the probability of the field taking any value at a pixel, given the values at the neighbouring

pixels, is independent of the particular pixel under consideration (condition (c)), then the MRF is *homogeneous*.

$$\text{c) } \Pr(X_k = x_k | X_i = x_i, i \in \eta_k) = \Pr(X_{k+j} = x_k | X_{i+j} = x_i, i \in \eta_k) \\ \forall k \in W \& k + i \in W \& \forall x_k$$

Early work with MRFs was hampered because it was not known how to evaluate the joint probability distribution  $\Pr(X = x)$ , or even how to define the local conditional probabilities such that the joint probability distribution was valid. These problems were solved by the Hammersley-Clifford theorem, which identified MRFs with Gibbs random fields.

### 5.2.2. Gibbs Random Fields

Gibbs Random Fields (GRFs) originated in statistical physics, where it was desired to deduce the large scale properties of a lattice system from local models. Ising (1952) pioneered the approach for modelling the behaviour of ferromagnetic material by considering only the interaction of the “spins” of neighbouring atoms. Two neighbouring atoms of opposite spin were considered to have a positive *potential*. The state of the lattice was characterised by its *energy*, computed as the sum of the *potentials*. Configurations of low energy are therefore more stable than those of high energy.

The terminology of statistical physics is widely used in this branch of statistical image analysis, and the concepts are somewhat easier to understand if the application to ferromagnetic materials is kept in mind.

#### Cliques

A *clique* under a neighbourhood system is any set of pixels, all possible pairs of which are neighbours. For the second order neighbourhood system the cliques are all groupings of pixels with shapes as given in fig.71. For a first order neighbourhood system, only shapes 1,2 & 3 apply. Let  $C(\eta_W)$  be set of the cliques of  $W$  under neighbourhood system  $\eta_W$ .

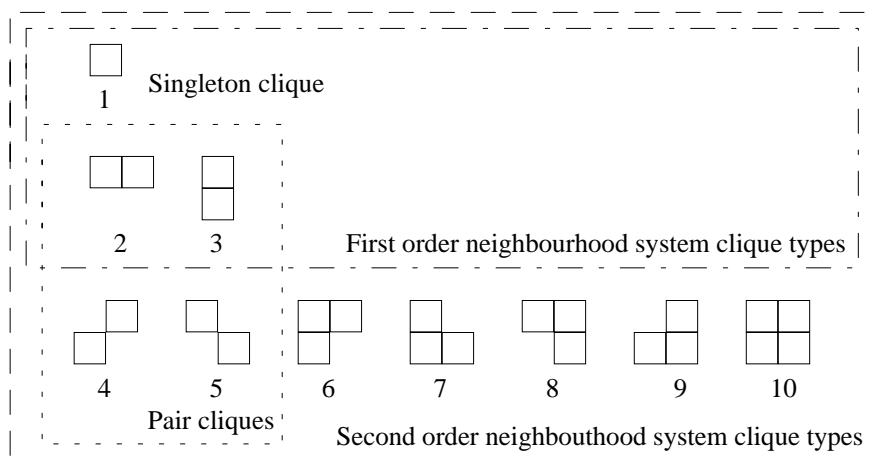


Figure 71

Clique types for first and second order neighbourhood systems of two-dimensional fields. Types 1–3 constitute the clique types for a first order neighbourhood system, types 1–10 the types for a second order system. Type 1 cliques are known as *singleton* cliques, types 2–5 as *pair* cliques.

### Gibbs Random Fields

A random field  $X$  is a Gibbs random field if and only if the probability (density) function has the following form:

$$\Pr(X = \mathbf{x}) = \frac{e^{-U(\mathbf{x})}}{Z}$$

Here,  $U(\mathbf{x})$  is called the *energy function*. The higher the energy of the configuration, the lower the probability. The denominator,  $Z$ , is the *partition function*, a normalising constant obtained by summing the numerator over all possible configurations  $\mathbf{x}$ . The partition function is usually not computable. For example, a small field of  $K = 64 \times 64$  pixels, each of which can take only two values, has  $2^{4096}$  possible configurations!

We can specify the energy function in terms of *potentials* for the individual cliques of a neighbourhood system. For clique  $c \in C(\eta_L)$  let  $V_c(\mathbf{x})$  be its potential, a function of the values of the pixels in the clique. Then, define the energy function as follows:

$$U(\mathbf{x}) = \sum_{c \in C(\eta_W)} V_c(\mathbf{x})$$

The Hammersley-Clifford theorem states that a random field  $X$  is a MRF with respect to the neighbourhood system  $\eta_W$ , if and only if  $X$  is a Gibbs distribution, with potentials defined on the cliques of that neighbourhood system. (See Besag (1974) for a proof.) Thus, it is usual to define a MRF through its representation as a GRF, by specifying clique potentials.

### Discrete $M$ colour GRF

Consider henceforth the discrete random field  $X$ , where each pixel can take values in  $\{0, 1, \dots, M-1\}$ , corresponding to  $M$  “colours”.

A simple scheme for constructing a GRF is described by Derin and Elliott (1987). They assign a potential  $V_c(\mathbf{x}) = -\zeta_t$  to cliques of type  $t$  when all the pixels in the clique have the same value, and  $+\zeta_t$  if any of the pixels are different ( $\zeta_t \geq 0$ ). Configurations containing cliques of clique type  $t$  with differing pixel values have higher energy and hence lower probability.

For the second order neighbourhood clique types (fig.71),  $X$  is defined as a GRF by taking the clique potential functions  $\zeta_t$ , as zero for the singleton cliques ( $\zeta_0 = 0$ ),  $\beta$  for the pair cliques where a side is shared ( $\zeta_t = \beta$  for  $t = 2, 3$ ),  $\beta/\sqrt{2}$  for the pair cliques where only a corner is shared ( $\zeta_t = \beta/\sqrt{2}$  for  $t = 4, 5$ ), and zero for the other cliques. The single parameter  $\beta$  specifies the dependency between a pixel and its neighbours, and characterises the strength of the field.

For the discrete  $M$  colour GRF, we can deduce the local conditional probability structure in the usual way:

$$\begin{aligned} \Pr(X_k = x_k \mid X_i = x_i, i \in W \setminus \{k\}) &= \frac{\Pr(\mathbf{X} = \mathbf{x})}{\Pr(X_i = x_i, i \in W \setminus \{k\})} \\ &= \frac{\exp\left(\sum_{c \in C(\eta_W)} V_c(\mathbf{x})\right)}{\sum_{\{\mathbf{x}' : x'_i = x_i, i \in W \setminus \{k\}\}} \exp\left(\sum_{c \in C(\eta_W)} V_c(\mathbf{x}')\right)} \\ &= \frac{\exp\left(\sum_{\{c \in C(\eta_W) : i \in c\}} V_c(\mathbf{x})\right)}{\sum_{\{\mathbf{x}' : x'_i = x_i, i \in W \setminus \{k\}\}} \exp\left(\sum_{\{c \in C(\eta_W) : i \in c\}} V_c(\mathbf{x}')\right)} \quad (66) \end{aligned}$$

The denominator in each of the above expressions is the sum of the numerator for all possible labellings of  $x_k$  as one of  $\{0, 1, \dots, M-1\}$ , and as such is simply a normalising constant. The set  $\{\mathbf{x}' : x'_i = x_i, i \in W \setminus \{k\}\}$  is simply the set of  $M$  configurations that are the same as  $\mathbf{x}$ , save at pixel  $k$ , which takes all possible values  $0, 1, \dots, M-1$ . Note that the sums of clique potentials in eqn.66 are only for those cliques containing pixel  $k$ . Thus, the conditional probability for the value at voxel  $k$ , given the values elsewhere, depends only on the values of the field at the neighbours of pixel  $k$ , the Markov property. The partition function does not appear, so computation is straightforward.

## 5.3. Image Segmentation

To illustrate the techniques which we shall use in our Bayesian test for a statistic image from PET, consider the following simple image segmentation problem.

### 5.3.1. Segmentation

Consider a true image  $\mathbf{R}$  consisting of only  $M$  known grey levels. Label the grey levels with unique colour labels from the set  $\{0, 1, \dots, M-1\}$ , such that label  $x$  corresponds to a pixel grey value of  $g(x)$ . Call the labelling  $\mathbf{X}$ . Suppose that we observe  $\mathbf{R}$  with added Gaussian white noise of variance  $\sigma^2$ , and that from this noisy version,  $\mathbf{Y}$ , we wish to estimate the original colour labelling  $\mathbf{X}$ . The likelihood  $\Pr(\mathbf{Y} = \mathbf{y} | \mathbf{X} = \mathbf{x})$  of any particular labelling given the data is simply the product, over all pixels, of the univariate normal pixel likelihoods  $\Pr(Y_k = y_k | X_k = x_k)$ :

$$\Pr(Y_k = y_k | X_k = x_k) = \frac{\exp\left(-\frac{(y_k - g(x_k))^2}{2\sigma^2}\right)}{\sqrt{2\pi\sigma^2}} \quad (67)$$

#### *MRF Prior & posterior*

The maximum likelihood estimate  $\hat{\mathbf{x}}$ , simply labels each pixel with the label of the grey level nearest to the observed value. Expressing prior beliefs  $\Pr(\mathbf{X} = \mathbf{x})$  about the expected contiguous nature of the colour labelling via a discrete MRF with parameter  $\beta$ , as defined above, we obtain the posterior distribution for the labelling  $\mathbf{X}$ , given the data  $\mathbf{Y}$ , by Bayes theorem as:

$$\begin{aligned} \Pr(\mathbf{X} = \mathbf{x} | \mathbf{Y} = \mathbf{y}) &\propto \Pr(\mathbf{Y} = \mathbf{y} | \mathbf{X} = \mathbf{x}) \times \Pr(\mathbf{X} = \mathbf{x}) \\ &\propto \exp\left(-\sum_{k \in W} \left[ \ln(\sigma) + \frac{(y_k - g(x_k))^2}{2\sigma^2} \right] - U(\mathbf{x}) \right) \end{aligned} \quad (68)$$

Assuming  $\sigma$  is known, the posterior (eqn.68) is also a GRF (and hence a MRF), since the contribution of the likelihood is an extra term in the energy function corresponding to potentials for the singleton cliques. The constant of proportionality is thus the posterior partition function.

#### *MPM estimate*

A suitable estimate  $\tilde{\mathbf{x}}$  of the colour labelling  $\mathbf{x}$  is then the Maximum Posterior Marginal (MPM) estimate, which maximises the marginal posterior probabilities  $\Pr(X_k = x_k | \mathbf{Y} = \mathbf{y})$ . This therefore minimises the expected number of misclassified pixels under the posterior. A vague prior, with  $\beta=0$ , expresses no prior belief, and the MPM labelling is identical to the maximum likelihood estimate.

Unfortunately the marginal probabilities cannot be evaluated directly, because of the uncomputable posterior partition function. However realisations of GRFs can be generated using the Gibbs sampler, and from these realisations estimates of the marginal posterior probabilities obtained.

### 5.3.2. The Gibbs sampler

The Gibbs sampler has radically changed the face of Bayesian inference over the past few years, enabling complex posterior distributions to be dealt with easily. In the current context, a sequence of random fields  $\{X^0, X^1, \dots, X^q, \dots, X^Q\}$  is generated from an arbitrary initial colour labelling  $X^0$ . Here, take the initial colour labelling as the Maximum Likelihood estimate,  $X^0 = \hat{x}$ . Each field in the sequence is generated from the previous one by visiting each pixel in turn, computing the local conditional posterior probabilities for each of the  $M$  colour labels (given the current labels of the neighbouring pixels and the data), and then choosing a new colour label for pixel  $k$  according to these probabilities. The sequence  $\{X^0, X^1, \dots\}$  can be proved to be a Markov Chain, with equilibrium distribution  $\Pr(X = x | Y = y)$ , the posterior distribution. Hence,  $\Pr(X^q = x) \xrightarrow{q} \Pr(X = x | Y = y)$ . So, for large  $q_B$ ,  $X^q$  may be regarded as a realisation of the posterior distribution for any  $q \geq q_B$ . Frequently  $q_B$  is called the “burn in” period. See Smith & Roberts (1993) and Besag & Green (1993) for a rigorous discussion of the method.

The set  $S = \{X^{q_B}, X^{q_B+t}, X^{q_B+2t}, \dots, X^{q_B+nt}\}$  can be regarded as an independent random sample of size  $n$  from the posterior distribution, provided the spacing  $t$  is chosen to overcome the serial correlation between successive labellings. The natural estimates of the marginal posterior probabilities of each colour label at pixel  $k$ ,  $\Pr(X_k = x_k | Y = y)$ , are the proportions of labellings in  $S$  with pixel  $k$  labelled that colour. Dependence between the elements of  $S$  does not bias these estimates, so  $t$  may be taken as 1.

### 5.3.3. Example

Figure 72a depicts a grey scale scene,  $r$ , defined on a  $64 \times 64$  pixellation of the surface of a torus, with 3 grey levels  $\{-2, 0, +2\}$ . This scene is a realisation of a GRF  $X$ , with clique potentials characterised by  $\beta=0.4$  in the scheme described above. (The colour labels are  $\{2, 0, 1\}$  respectively:  $g(0) = 0$ ;  $g(1) = +2$ ;  $g(2) = -2$ .) The realisation was generated by running 50 full sweeps of the Gibbs Sampler from an initial random colour labelling.

Note that although the image is mainly coherent, isolated pixels are present. Regarding this image as our unobservable image  $\mathbf{r}$ , we add standard Gaussian white noise to it to generate our observed image  $\mathbf{y}$ , depicted in Figure 72b. From  $\mathbf{y}$ , an estimate of  $\mathbf{x}$ , and hence of  $\mathbf{r}$  is required.

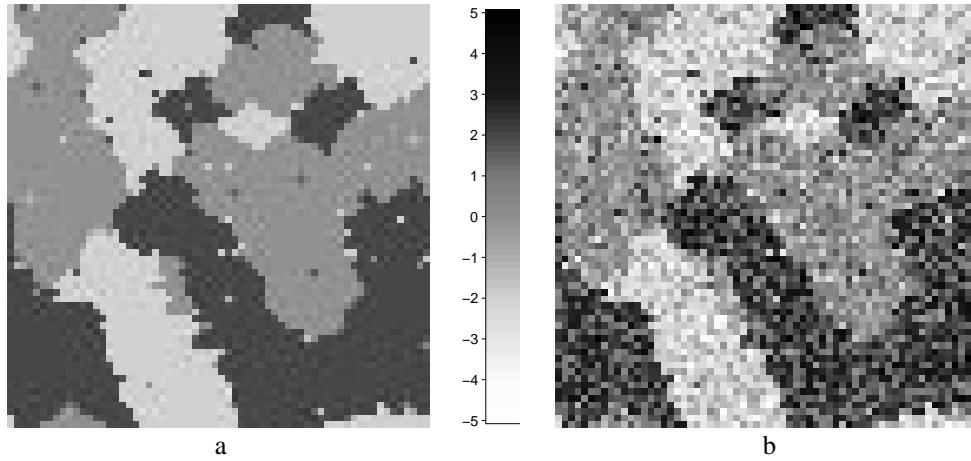


Figure 72

(a) Example grey level scene,  $\mathbf{r}$ , a realisation of a three colour GRF with clique potential functions parameterised by  $\beta = 0.4$ . (b) The grey scale scene corrupted by the addition of Gaussian white noise of unit variance, to give an example observable image  $\mathbf{y}$ .

The maximum likelihood labelling is depicted in fig.73, with the appropriate grey levels. This segmentation is rather noisy, and misclassifies 891 of the pixels ( $\approx 22\%$ ).



Figure 73

Maximum likelihood estimate of the true scene  $\mathbf{r}$ , obtained from the corrupted scene  $\mathbf{y}$ , given the grey levels corresponding to the three colour labels, and the variance of the Gaussian white noise process.

Including prior belief gives a vast improvement. Figure 74 depicts estimates of the MPM estimates  $\tilde{\mathbf{x}}$ , for priors with strengths parameterised by  $\beta=0$ ,  $\beta=0.2$ ,  $\beta=0.4$  &  $\beta=0.6$ . The marginal posterior probabilities in each case were estimated using the Gibbs sampler as described above, with 20 iterations burn in and the 2000 subsequent iterations as sample. The case  $\beta=0$  gives a vague prior, where the posterior is proportional to the likelihood. The MPM estimate  $\tilde{\mathbf{x}}$  in this case is therefore equivalent to the maximum

likelihood estimate  $\hat{x}$ . The estimated MPM estimate for this vague prior (fig.74a) differs from the true ML at only 28 pixels ( $\approx 0.7\%$ ). This demonstrates the validity of the computer programs used. However, in this case of vague prior, successive realisations from the Gibbs sampler are independent realisations of the posterior, which is not the case for non-vague priors. Thus, the number of errors in estimation of  $\hat{x}$  is not indicative of the number of errors in estimating the MPM estimates for non-vague priors.

As can be seen, the incorporation of prior knowledge about the spatial coherence of the image affords considerable improvement over the maximum likelihood segmentation, although too strong a prior makes the segmentation “too smooth”. The numbers of misclassified pixels are 235 (5.7%), 124 (3.0%) & 142 (3.5%).

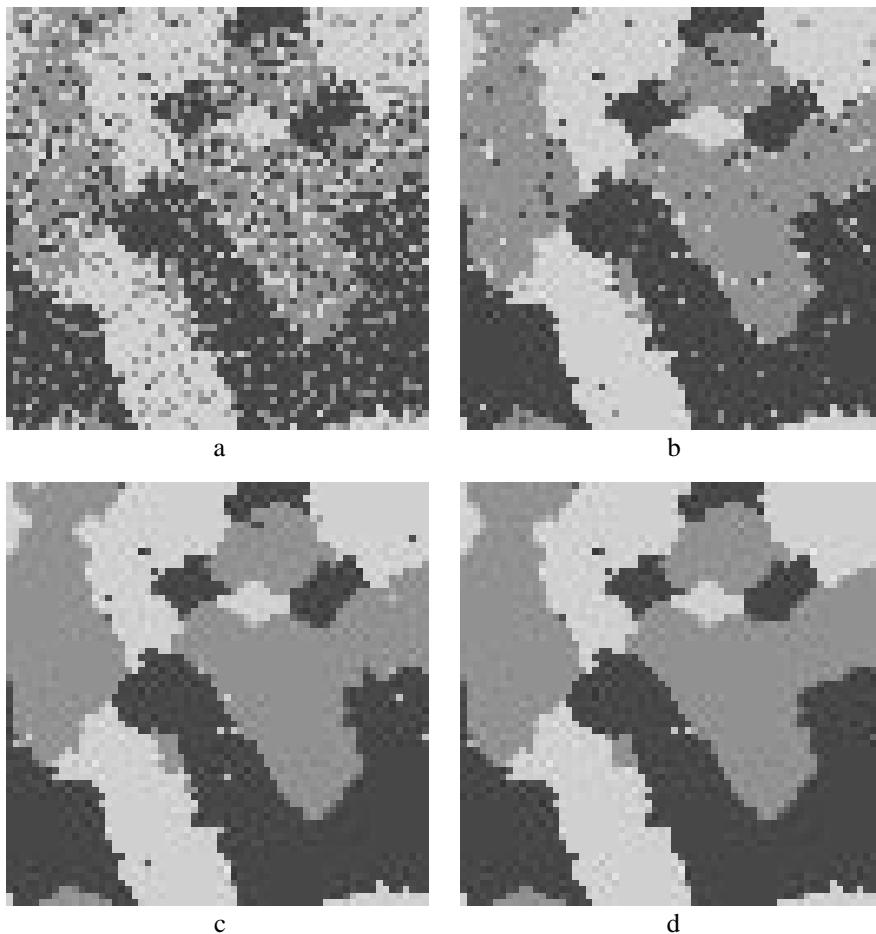


Figure 74  
Estimates of the Maximum Posterior Marginal estimates of the true scene  $r$ ,  
for prior strengths of (a)  $\beta=0$ ; (b)  $\beta=0.2$ ; (c)  $\beta=0.4$ ; & (d)  $\beta=0.6$ .

The prior effectively acts as an intelligent filter of the colour labelling. Indeed, this Bayesian approach can be viewed in classical terms as a penalised likelihood method, the MRF providing a functional form of the “penalty” of labelling a pixel differently to its neighbours. It is this property that motivates the use of a MRF as a prior for testing statistic images.

## 5.4. A Bayesian Segmentation Test

Consider a Gaussian statistic image  $\mathbf{Y}$ ,  $Y_k \sim N(\mu_k, 1)$ . The omnibus null hypothesis  $H_W$ , is the intersection of the pixel hypotheses  $H_k: \mu_k = 0$ ,  $k \in W$ . The Bonferroni method for testing a Gaussian statistic image  $\mathbf{Y}$ , tests each pixel individually, correcting for the number of comparisons using the Bonferroni inequality. For a two-sided test at level  $\alpha$  this results in a procedure where the statistic image is thresholded above at  $\Phi^{-1}(1-\alpha/(2K))$  and below at  $\Phi^{-1}(\alpha/(2K))$  (§3.2.1.). High pixels give significant evidence of an increase in the test statistic, indicating positive activation, and low pixels a significant decrease, indicating negative activation.

### *Segmentation formulation of Bonferroni test*

We can view the test as an image segmentation problem with three labels: “no activation”, “positive activation” and “negative activation”. The “no activation” label corresponds to a true statistic or “grey level” of zero, and we observe this with an added Gaussian error. The two alternative labellings do not correspond to any particular true statistic value, but if we assign artificial alternative true values  $2\Phi^{-1}(\alpha/(2N))$  and  $2\Phi^{-1}(1-\alpha/(2N))$  to “negative activation” and “positive activation” respectively, then a maximum likelihood segmentation of the statistic image will give the same labelling of the pixels as would the Bonferroni test.

Formulated as a segmentation problem thus, prior belief can be included. This leads to a *Bayesian segmentation “test”*, as a segmentation of the statistic image with the (artificial) grey levels for the labellings as given above. To evaluate this idea a simulation study was carried out.

## 5.5. Simulation Study

### 5.5.1. Simulation methods

#### *Null statistic images*

Null statistic images were generated in exactly the same manner as the subject difference images of the Two-Stage simulation, described in §4.2.2. Briefly, Gaussian white noise fields were generated on a  $64 \times 64$  pixellation of the surface of a torus. These were smoothed with an isotropic Gaussian filter kernel of standard deviation  $5/\sqrt{8\ln(2)}$ , corresponding to a FWHM of 5 pixels (10mm for 2mm square pixels). The filter kernel was implemented as a moving average filter, with weights computed by evaluating the kernel on a regular  $17 \times 17$  array of points 1 unit (pixel) apart. The variance of the initial white noise process was chosen so that the null statistic image had pixel values with unit variance.

#### *Signal*

Departures from the null hypothesis were simulated by adding a signal image to the simulated null statistic images. Two signal images were considered. The first was a zero image, used to examine the true level of the test. The second was the focal signal used in the Two-Stage simulation (§4.2.2.). This is a centrally located isotropic Gaussian kernel of standard deviation  $5/\sqrt{2\ln(2)}$ , corresponding to a centrally located Gaussian point response function with FWHM of 5 pixels, convolved with itself. The signal was scaled to have maximum height, or amplitude, of 4.5.

#### *Tests*

The artificial segmentation formulation of the Bonferroni test using estimated posterior probabilities for prior strengths of  $\beta = 0, 1, 2, 3$ , was compared with the results of the actual Bonferroni test. For each combination of signal and prior strength the two tests were compared on separate sets of simulated statistic images. With vague prior the significance labelling from the two tests should be the same, however there will be sampling error due the estimation of the marginal posterior probabilities. A burn in period of 10 iterations of the Gibbs sampler was adopted, and the subsequent 1000 realisations used to estimate the marginal posterior probabilities. Clearly it is desirable to estimate these probabilities fairly accurately, especially those near 0.5 where the marginal posterior probabilities for two labellings are close. However, it is the properties of the resultant test that are of interest, rather than the niceties of Markov Chain Monte-Carlo, and small symmetrical errors shouldn't impair the qualities of the test.

## 5.5.2. Results

### Zero signal

The results of the simulation study for the pure noise test images are given below (table 75). The  $p$ -values given are for the null hypothesis that both tests have the same size, against a two sided alternative, computed using McNemars test. Both tests are ultra-conservative, the test images being very smooth. Although there is little evidence, it appears that, as the prior strength is increased, the Bayesian segmentation test becomes progressively more conservative than the Bonferroni test. A 95% confidence interval for the true size of the level  $\alpha = 0.05$  Bonferroni test is (0.0092, 0.0148), computed to 4dp using the normal approximation to the Binomial.

1 = “reject” $H_w$		Bonferroni		
Zero signal		0	1	
MPM $\beta=0.0$	0	984	0	984
$p=1$	1	1	15	16
MPM $\beta=0.1$	0	986	3	989
$p=0.25$	1	0	11	11
MPM $\beta=0.2$	0	988	9	997
$p \approx 0.0039$	1	0	3	3
MPM $\beta=0.3$	0	993	6	999
$p \approx 0.0312$	1	0	1	1
Bonferroni		3952	48	

Table 75  
Summary of simulation results for zero signal

### Focal signal

The results of the simulation for the centrally located focal signal, scaled to have a maxima of 4.5, are given in table 76. Increasing the strength of the prior makes the Bayesian MPM test less powerful than the Bonferroni method.

1 = “reject” $H_w$		Bonferroni		
Focal signal		0	1	
MPM $\beta=0.0$	0	172	2	174
$p=1$	1	1	325	326
MPM $\beta=0.1$	0	173	31	204
$p \approx 9 \times 10^{-10}$	1	0	296	296
MPM $\beta=0.2$	0	190	57	247
$p \approx 1 \times 10^{-17}$	1	0	253	253
MPM $\beta=0.3$	0	182	81	263
$p \approx 8 \times 10^{-25}$	1	0	237	237
Bonferroni		718	1282	

Table 76  
Summary of simulation results for focal signal.  
All rejections of the omnibus hypothesis  $H_w$  are considered.

## 5.6. PET Example

Figure 77 shows the AC-PC plane of the  $t$ -statistic image for the “V5” study subject difference images, after transformation to have a standard Gaussian distribution under the null hypotheses. The paired  $t$ -statistic was constructed as described in §2.3.1., and was depicted previously in §2.6.1. This  $t$ -statistic image was Gaussianised by replacing each pixel  $t$  value with the standard Normal ordinate with the same extremum probability. (Recall the discussion of “transform functions” of §3.3.3., and see appendix E for computational details.)

Let this two dimensional image be  $\mathbf{y}$ . The AC-PC plane is discretised into  $K = 65 \times 87$  square pixels of side 2mm. For consistency with the previous experiments in this chapter, periodic boundary conditions were assumed: The rectangular image space was considered as the unfolded surface of a torus. The top and bottom, and left and right of the image space are taken to abut, respectively. The set  $W$  was taken to be the set of (indices of) all  $K$  voxels, rather than the subset corresponding to the intracerebral area as in other chapters. The voxel hypotheses  $H_k: \mu_k = 0$ , were tested against two-sided alternatives, where the pixel values  $y_k$  are assumed to be drawn from a Gaussian distribution with mean  $\mu_k$  and unit variance.

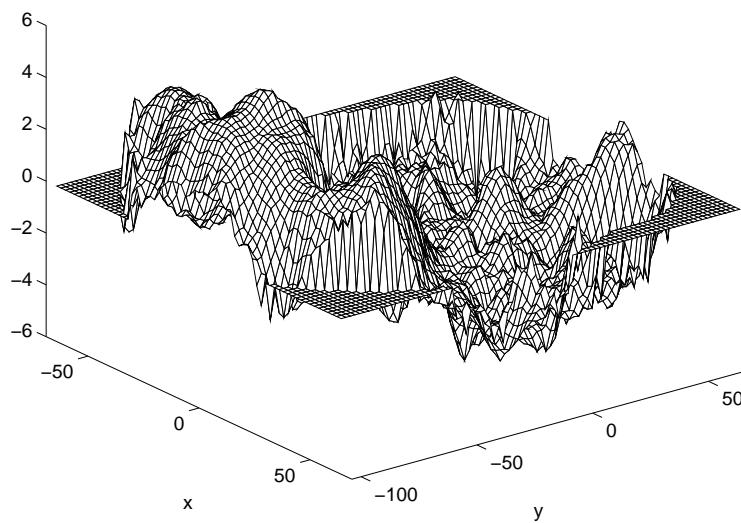


Figure 77

Mesh plot of “Gaussianised”  $t$ -statistic for the “V5” study, computed using the proportional scaling approach and paired  $t$ -statistic, described in §2.3.1. The X and Y axes are graduated in millimetres, according to the Talairach system. The AC-PC plane is shown.

### Bonferroni assessment

A two-sided Bonferroni assessment of the AC-PC plane of the Gaussianised  $t$ -statistic image, correcting for the number of pixels in the plane, leads to the rejection of  $H_k$  at 140 pixels ( $\approx 2.5\%$ ), shown in figure 78. The small regions of negative activation were considered to be artefacts. More primary (or secondary) smoothing would have removed these artefacts, but would also have smoothed out the signal.

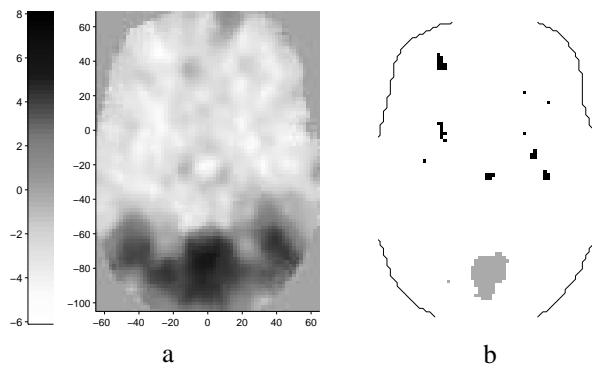


Figure 78

(a) Image of AC-PC plane of Gaussianised  $t$ -statistic for the "V5" study, shown as a mesh plot in fig.77. (b) Results of Bonferroni assessment of this plane, correcting for the number of pixels in the plane. Pixels with values in the upper tail of the null distribution are shown painted grey, those in the lower tail black. The outline of the intracerebral area is shown for orientation.

### **Bayesian MPM segmentation "test"**

The Gaussianised  $t$ -statistic,  $y$ , was assessed using the Bayesian MPM segmentation test, with prior strengths parameterised by  $\beta = 0.1, 0.2, \& 0.3$ . The artificial alternative grey values for the "positive activation" and "negative activation" labels were set at  $\pm 2\Phi^{-1}(1-\alpha/2K)$ , for level  $\alpha = 0.05$ . Once again, the posterior probabilities were estimated from 1000 successive realisations from the Gibbs sampler, after a burn in of 10 iterations. The results are shown in figure 79. As can be seen, the smaller regions are successively eliminated as the prior strength is increased, and the shape of the large activated area is rounded off.

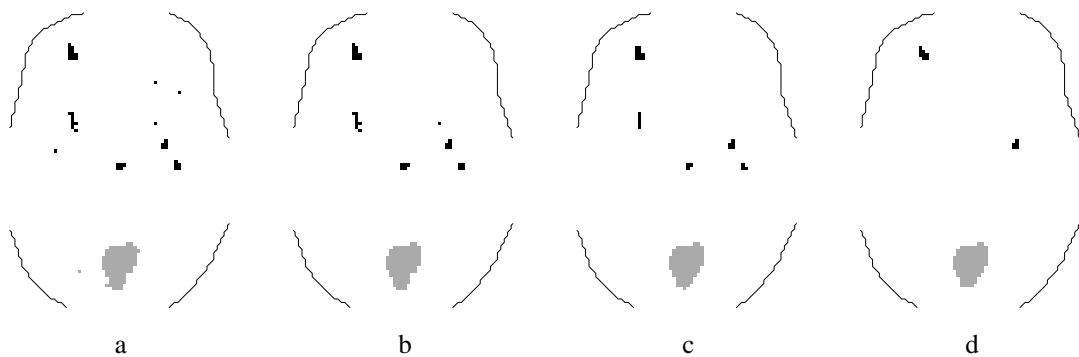


Figure 79

Results of Bayesian MPM segmentation assessments of the AC-PC plane of the Gaussianised "V5" study  $t$ -statistic. The priors used are parameterised by (a)  $\beta=0.0$  (vague prior), (b)  $\beta=0.1$ , (c)  $\beta=0.2$ , & (d)  $\beta=0.4$ . Pixels labelled "positively activated" are shown grey, those labelled "negatively activated" as black. The outline of the intracerebral area is shown for orientation. The large region of positive activation corresponds to the primary visual cortex.

## 5.7. Conclusions

In this chapter an attempt has been made to include prior belief regarding the contiguous nature of activated pixels, into the testing procedure. This prior belief was expressed through a discrete MRF, motivated by the usefulness of this model in digital image processing.

### ***Summary of the approach***

The exact method proposed is based on the simple segmentation problem, where a grey scale scene is sought from an observation corrupted by added Gaussian white noise. This Bayesian segmentation requires knowledge of the true grey levels, and the variance of the white noise process.

For a standard Gaussian statistic image, we wish to label each pixel (voxel) as “not activated”, “negatively activated” or “positively activated”. Simple threshold tests, the simplest of which is the Bonferroni, achieve this by thresholding the statistic image above and below, and labelling suprathreshold pixels (voxels) accordingly. Assigning “artificial” grey levels to the activated labels that are twice the appropriate threshold, so that the likelihoods of the labellings cross at the thresholds, results in a ML segmentation that duplicates the results of the test. The Bayesian MPM segmentation test proposed here simply seeks to include a prior into this segmentation.

Although specific statistic values are specified for pixels labelled as “positively activated” and “negatively activated”, the alternative hypothesis is still composite, since very extreme values will still be nearest to the grey level for the appropriate labelling.

### ***Extensions***

As presented here, the thresholds for a two-sided Bonferroni test have been used to set the thresholds. Clearly thresholds from any valid test could be used. Extension of the approach to three dimensions is trivial, as is modification for a one sided test.

### ***Conservativeness***

The Bonferroni method is conservative for testing multiple hypotheses that are dependent. The addition of prior beliefs to the Bonferroni method via the Bayesian MPM segmentation test formulated above, results in a less powerful test. As mentioned above, a Bayesian MPM segmentation test could be constructed using thresholds for any valid voxel-by-voxel test employing a fixed threshold. However, as demonstrated for the Bonferroni approach, the Bayesian MPM segmentation test is less powerful than the thresholding test on which it is based.

The inclusion of prior belief via a MRF biases the labelling of each pixel towards that of the majority of its neighbours. Regions of activation are convex, so pixels on the boundary of this region will have less activated neighbouring pixels than activated ones. Thus, the inclusion of prior belief will weaken their significance, reducing the size of the activated region. However, the resulting labellings are more contiguous, as illustrated in the PET example.

### ***Tuning***

To overcome this conservativeness, the Bayesian MPM segmentation test could be “tuned”, by setting the artificial statistic levels for the “activated” labellings closer to zero so that a test with size approximately equal to the desired level was obtained.

For instance, a Bayesian MPM test based on a simple pixel threshold of  $\Phi^{-1}(1-\alpha'/2)$ , with  $\alpha'=0.0001$ , gives 31 rejections of the omnibus null hypothesis over 500 zero signal simulations at prior strength  $\beta=0.2$ . An approximate 95% CI for the size of the test is therefore (0.0409,0.0831), which includes 0.05.

Such tuning requires large numbers of null statistic images, which can only be obtained by simulation. The robustness of tests based on simulated null statistic images is still in doubt, as was discussed with regard to suprathreshold cluster size tests in §3.5.1., and we shall not pursue this line of investigation.

### *Contiguous labellings*

For (expected) diffuse activations it is best to smooth and use a test method taking into account the spatial redundancy due to smoothness. For extremely focal activations, or statistic images with a degree of high frequency noise, the Bayesian approach incorporates our prior belief well, and works as an intelligent smoothing of the pixel labelling. Although the Bayesian test formulated here is conservative, it illustrates the usefulness of MRFs for conveying prior belief.

With hindsight, the simulation experiment appears inadequate for assessing the promise of the Bayesian approach, since the simulated statistic images were very smooth.

### *Bayesian rigour*

The use of Bayesian tools in this context is merely a means to an end, that end being a contiguous pixel labelling. This is not a rigorous empirical Bayesian approach. The model used is clearly an oversimplification. A statistic image from PET clearly does not consist of three levels for “negative activation”, “no activation” and “positive activation” to which white Gaussian noise is added.

The noise process is smooth. Considering the segmentation of a discrete grey level scene to which smooth noise has been added, the posterior is not a MRF as specified in eqn.68, since the likelihood  $\Pr(\mathbf{Y} = \mathbf{y} | \mathbf{X} = \mathbf{x})$  is not the product of the univariate likelihoods of eqn.67. If the noise process can be modelled as a (continuous) MRF, say a Gaussian Markov random field, then the posterior can be computed, and is a MRF on the neighbourhood system that is the union of those for the prior and the noise.

This modification for smooth noise is still inappropriate for statistic images, since the “true” statistic image is not discrete. In short, the segmentation model is too simplistic for a rigorous empirical Bayesian approach to segmenting statistic images from PET.

### *Summary*

A simple attempt to embody prior belief regarding the contiguous nature of activated pixels into a significance test for statistic images has been presented. A MRF prior has been applied to a simple threshold method, viewed as a segmentation problem. The resulting Bayesian MPM segmentation “test” is less powerful than the thresholding test upon which it is based, but the identified activated regions are more contiguous. Given the operational complexities of the method, and the difficulty of “tuning”, the proposed method is perhaps consigned to the large pile of ideas that “didn’t quite make it”.

