Statistical Considerations in Magnetic Resonance Imaging of Brain Function

Brian D. Ripley and Jonathan Marchini
Department of Statistics, University of Oxford
1 South Parks Road, Oxford OX1 3TG, UK
http://www.stats.ox.ac.uk/~ripley

Abstract

Magnetic Resonance Imaging is an important non-invasive 3D imaging technique most used to study human bodies and particularly brains. We discuss some of the statistical issues, and illustrate these with a study of segmentation.

Magnetic Resonance Imaging (MRI) [9] is an important non-invasive 3D imaging technique most used to study human bodies and particularly brains. We are currently working with Oxford collaborators on two scientific problems using MR brain images.

(A) Studies of the change in tissue state and neurological function after traumatic events such as a stroke or tumour growth and removal. The aim here is to identify tissue as normal, impaired or dead, and to compare images from a patient taken over a period of several months.

(B) Direct studies of brain function by presenting subjects with stimuli and observing which areas of the brain 'light up'. Functional MRI effectively only observes oxygenated blood supply, and the challenge is to identify accurately the small areas of change amongst a large 3D spatial extent that are signal and not noise.

Both areas present enormous statistical challenges: only (A) is discussed here, but some work on (B) will be presented at the meeting.

1 MRI Data

Figure 1 shows MRI data on a brain slice of a 'normal' subject (JM). A single slice has a spatial resolution of about 1 mm and there are about 27 000 pixels wholly inside the skull. This is one of a series of vertically stacked sections that can be taken at about 3 mm spacing. Thus the pixels should be thought of representing a voxel in a fairly thin slice.

We show 'T1' and 'T2' images, the conventional MRI measurements[1]. Other measurements are available using spectroscopic information, but for our early studies these are used to indicate 'ground truth'.

[1] 'T1' is the time constant in the relaxation time of the recovery of longitudinal magnetisation: 'T2' is that for the loss of transverse magnetisation.

An annotated scatterplot of these data is shown in Figure 2. The initial goal is to segment these data into the conventional anatomical classes of 'white matter', 'grey matter', 'CSF' (cerebro–spinal fluid) and 'skull'. Although our procedures are unsupervised pattern recognition problems, we use the knowledge indicated in Figure 2 both to roughly locate the clusters and, of course, to label them.

Ideally the scatterplot would show three compact clusters of white and grey matter and CSF well away from the points representing the skull. This is clearly not what was observed, and the extent of the spread surprised our collaborators. In discussions we identified three possible causes for the spread of the clusters.

(a) 'Mixed voxel' / 'partial volume' effects. These are two names for the same phenomenon, that the tissue within a voxel is not all of one class.

(b) A 'bias field' in which the mean intensity from a tissue type varies across the image. This effect is thought to vary approximately multiplicatively and to consist of a radial component plus a linear component, the latter varying from day to day. (Because the subject is resting his head on a support within a cylindrical magnet, the centre of the magnet is in the upper part of Figure 1.)

(c) The 'point spread function'. Because of bandwidth limitations in the Fourier domain in which the image is acquired, the true observed image is convolved with a spatial point spread function of 'sinc' ($\sin x/x$) form. The effect can sometimes be seen at sharp interfaces (most often the skull / tissue interface) as a rippling effect, but is thought to be small.

For a reviews of the segmentation of MRI images see [2, 3]. These are however, catalogues of pattern recognition methods, and do not consider the special issues of classifying spatially structured datasets, let alone the particular issues raised by MRI data.

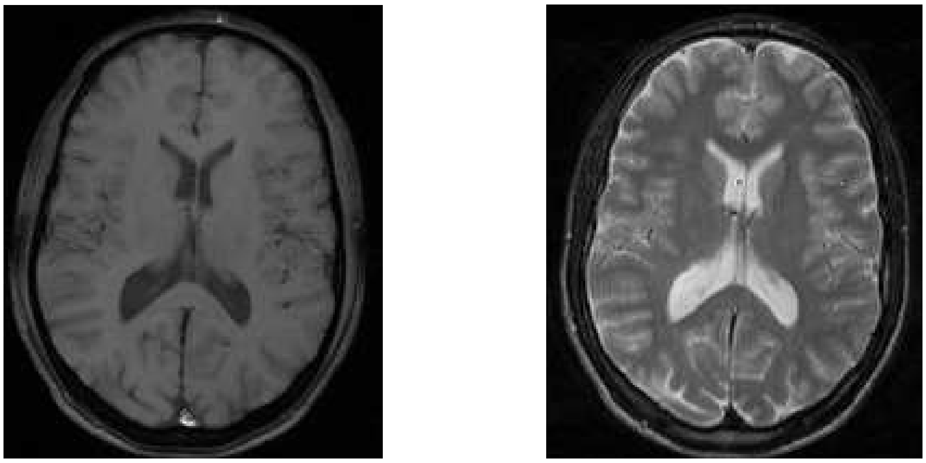

Figure 1: T1 (left) and T2 (right) MRI sections of a 'normal' human brain. This slice is of 172×208 pixels.

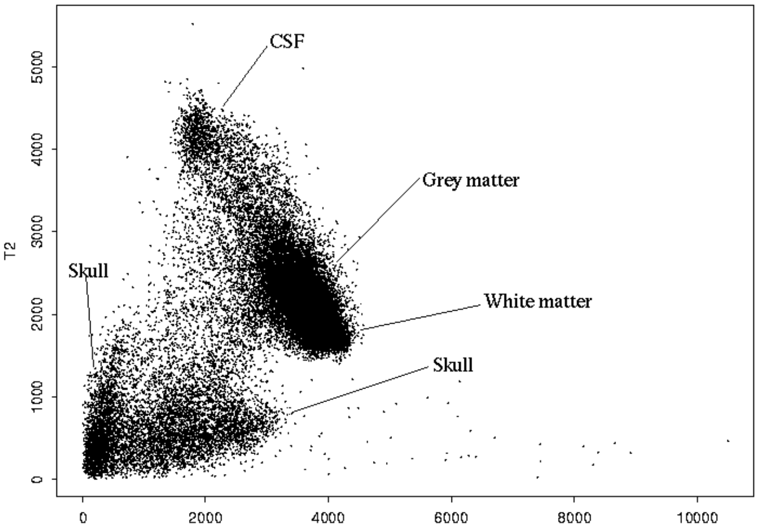

Figure 2: An annotated T1–T2 plot of the brain slice shown in Figure 1.

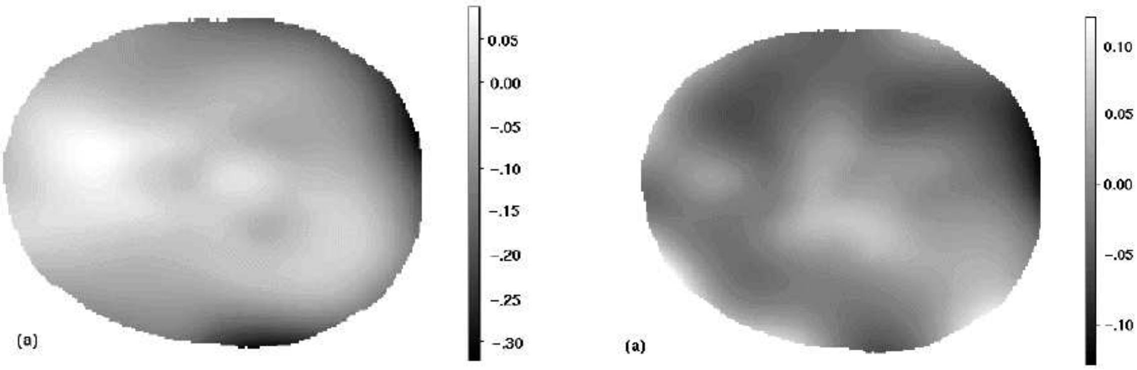

Figure 3: Fitted 'bias fields' for T1 (left) and T2 (right) images, excluding the skull. Note that the variation in intensity is correlated with tissue type apparent in Figure 1: it seems that the response of CSF has a systematic spatial component.

2 Bias fields

These are not new problems and there is an extensive literature on bias field correction. One approach [15, 5] uses a stochastic process prior for the bias field, and is thus another re-invention of the ideas known as *kriging* in the geostatistical literature [11, Chapter 4]. Based on experience with the difficulty of choosing the degree of smoothing and the lack of resistance to outliers (kriging is based on assumptions of Gaussian processes) we prefer methods with more statistical content and control.

Our basic model is

$$\log Y_{ij} = \mu + \beta_{\text{class}(ij)} + s(i,j) + \epsilon_{ij} \tag{1}$$

for the intensity at voxel (i,j), studied independently for each of the T1 and T2 responses. Here $s(x,y)$ is a spatially smooth function.

Of course, (1) depends on the classification, which will itself depend on the predicted bias field. This circularity is solved by iterative procedure, starting with no bias field.

To fit (1) if the classification were known we would use a robust method that fits a long-tailed distribution for ϵ_{ij}, unconstrained terms α_j for each class, and a 'smooth' function s. We cope with unknown class in two ways. In the early stages of the process we only include data points whose classification is nearly certain, and later we use

$$\log Y_{ij} = \mu + \sum_{\text{class } c} \beta_c \, p(c\,|Y_{ij}) + s(i,j) + \epsilon_{ij} \tag{2}$$

that is we average the class term over the posterior probabilities for the current classification.

For the smooth term s we initially fitted a linear trend plus a spline model in the distance from the central axis of the magnet, but this did not work well, so we switched to *loess* [4]. Loess is based on fitting a linear surface locally plus approximation techniques to avoid doing for the order of 27 000 fits. The results are shown in Figure 3; the bias fields for these images are not large and change intensity by 5–10% over the brain.

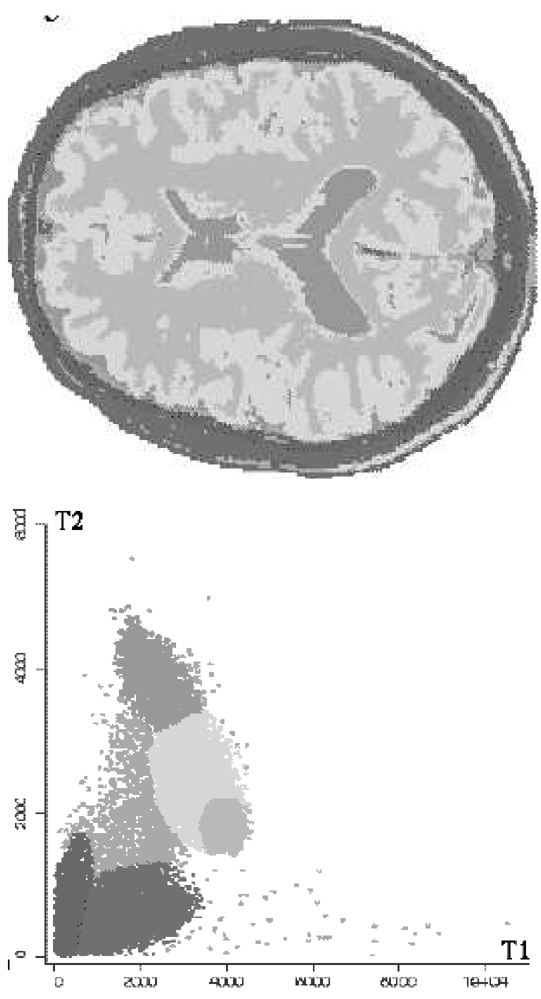

Figure 4: Classified image and T1–T2 plot showing the classification with normally-distributed clusters. The six populations are show in different grey levels and can be identified by reference to Figure 1.

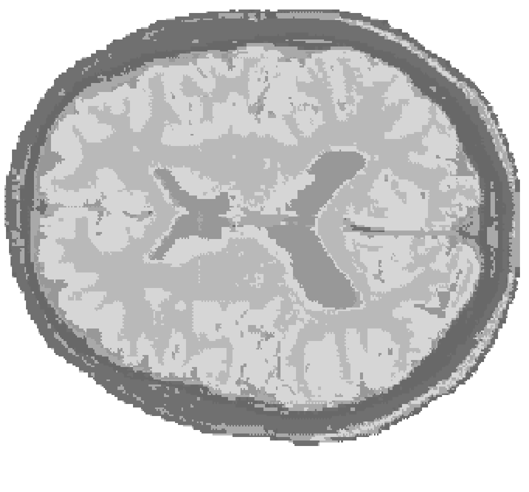

Figure 5: As Figure 4 after removing the bias field.

3 Classification

Our model is that for the bias-corrected $(Y_{ij}/\hat{s}(i,j))$ T1 and T2 measurements jointly follow a mixture of six distributions. These should correspond to white matter, grey matter, CSF, two types of skull (the distribution of skull in Figure 2 seems to split into two ellipses) and 'outlier', a diffuse distribution picking up the isolated points seen in Figure 2. The bivariate distributions for the six classes will have elliptical contours. We have experimented with normal and bivariate t distributions: the latter are often desirable as the normal density decays incredibly rapidly.

There is an extensive literature [10, 1] and several books [14, 7, 8] on fitting mixtures of densities, and using them as a model for classification. There are two distinct approaches in the literature, and the one we have taken is 'Bayesian' rather than 'maximum likelihood' in character. That is, we estimate the parameters in a finite mixture model

$$p(T1, T2) = \sum_{i=1}^{6} \pi_i \, p_i(T1, T2; \phi_i) \qquad (3)$$

and then use the posterior probabilities for a 'soft' classification (rather than the 'hard' classification used by Banfield & Raftery [1] that simultaneously maximizes over the parameters *and* the classification of the observations.). Fitting the parameters (π_i) and (ϕ_i) is currently by maximum-likelihood, although we will investigate using Markov Chain Monte Carlo methods for a full Bayesian scheme with linked hyperpriors.

We need to specify more precisely the class-conditional distributions used in (3). In all our experiments these are elliptically symmetric distributions, either normal distributions or t distributions with a fixed number of 'degrees of freedom' (that controls the rate of decay of the tails). Thus the parameter ϕ_i has to encompass the location (two parameters) and the scatter matrix (three parameters, for example the spread along the major and minor axes of the ellipse and the rotation angle of the major axis). So far we have used unconstrained maximum-likelihood fitting of these five parameters. It would certainly be useful to introduce prior information on the shape and size of the clusters, but with so many data points we found no difficulty in finding repeatedly the same plausible solution provided we used suitable starting values. This is an interesting technical statistical fitting problem: the likelihood is unbounded (so strictly there is no maximum likelihood estimate) and there are usually many local maxima of the log-likelihood. Yet the so-called 'EM algorithm' is very widely used despite evidence [10, 6, 13] of the superiority of conventional optimization algorithms.

This fitting scheme is unsupervised, but it is started by manually selecting and labelling some voxels. This lead to a fitted model for a 'normal' subject, and the parameter values fitted to this subject are used as the starting point for the fitting algorithm when applied to a future image (after some approximate matching of images in T1–T2 space).

Some results are shown in Figures 4 and 5. The removal of the bias field makes some small but noticeable differences. However, it makes very little difference to the cluster spread shown in the T1–T2 plot of Figure 4, and the spread cannot be accounted for by partial-volume effects as these would only 'smear' the clusters along the CSF–white matter alignment shown in Figure 1. (We checked that the spread in the other direction is not due to voxels near the skull). Interactive exploration of linked views of the two images shown in Figure 4 shows that there is both appreciable local variation in response, and a systematic spatial variation in response from CSF which is different from the spatial response variation for white and grey matter.

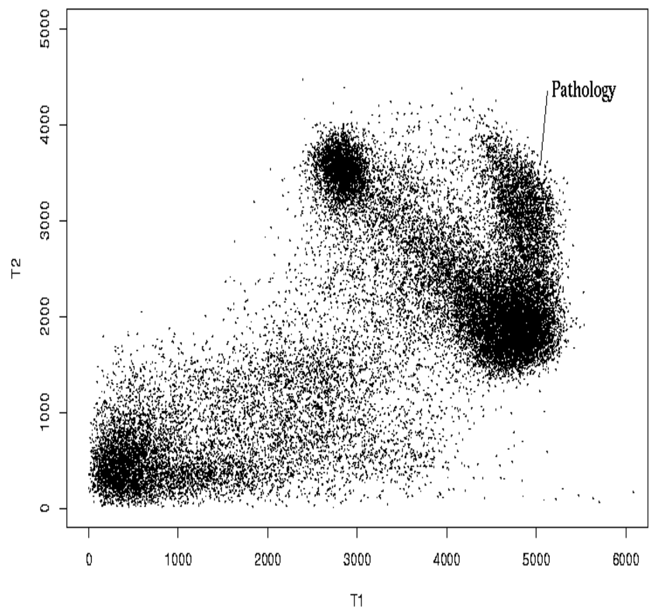

Figure 6: T1–T2 plot of a brain slice of a brain with a pathology (hydrocephalus).

4 Outliers and anomalies

We have found our scheme to be quite robust to variation in imaging conditions and to different 'normal' subjects. The 'background' class helps greatly in achieving this robustness as it 'mops up' the observations which do not agree with the model.

However, outliers can be more extreme, as shown in Figure 6, which is from a subject with a different response class. This illustrates the dangers [13, pp. 24–6] of classifying all the points. This is a particularly common mistake when neural networks are used for classification, and we have seen MRI brain scans classified by neural networks where common sense suggested an 'outlier' report was the appropriate one.

The procedure presented here almost entirely ignores the spatial nature of the image. For some purposes this would be a severe criticism, as *contextual* classification [12] would be appropriate. However, our (and our collaborators') interest in these images is not a pretty picture but is indeed in the anomalies, and for that we prefer to stay close to the raw data. The other interest is in producing summary measures that can be compared across time.

5 Future work

The study was partly to see what could be achieved with the relatively high resolution T1 and T2 images. Spatial resolution can be traded for further information per voxel: in functional MRI this is temporal information but in MR spectroscopy [9, Chapter 10] this is spectroscopic information on the chemical constitution of the (appreciably larger) voxel. On those images the segmentation issues are expected to be much harder, and issues of statistical significance of findings needs to be addressed.

Acknowledgments

JM is supported by an EPSRC research studentship. We are grateful to Professor Peter Styles (MRC Biochemical and Clinical Magnetic Resonance Spectroscopy Unit) and Dr Stephen M. Smith (Oxford Centre for Functional Magnetic Resonance Imaging of the Brain) for many discussions and providing both data and insight into the physical and physiological processes involved.

References

[1] Banfield, J. D. and Raftery, A. E. (1993) Model-based Gaussian and non-Gaussian clustering. *Biometrics* **49**, 803–821.

[2] Bezdek, J. C., Hall, L. O. and Clarke, L. P. (1993) Review of MR image segmentation techniques uing pattern recognition. *Medical Physics* **20**, 1033–1048.

[3] Clarke, L. P., Velthuzen, R. P., Camacho, M. A., Heine, J., M, V., Hall, L. O., Thatcher, R. W. and l. Silbiger, M. (1995) MRI segmentation: Methods and application. *Magnetic Resonance Imaging* **13**, 343–368.

[4] Cleveland, W. S., Grosse, E. and Shyu, W. M. (1992) Local regression models. In *Statistical Models in S*, eds J. M. Chambers and T. J. Hastie, Chapter 8. New York: Chapman and Hall.

[5] Guillemaud, R. and Brady, M. (1997) Enhancement of MR images. *IEEE Transactions on Medical Imaging* **16**, 238–251.

[6] Ingrassia, S. (1992) A comparison between the simulated annealing and the EM algorithms in normal mixture decompositions. *Statistics and Computing* **2**, 203–211.

[7] McLachlan, G. J. and Basford, K. E. (1988) *Mixture Models: Inference and Applications to Clustering*. New York: Marcel Dekker.

[8] McLachlan, G. J. and Krishnan, T. (1997) *The EM Algorithm and Extensions*. New York: Wiley.

[9] Orrison, Jr., W. W., Lewin, J. D., Sanders, J. A. and Hartshorne, M. F. (1995) *Functional Brain Imaging*. St. Louis: Mosby.

[10] Redner, R. A. and Walker, H. F. (1984) Mixture densities, maximum likelihood and the EM algorithm. *SIAM Review* **26**, 195–239.

[11] Ripley, B. D. (1981) *Spatial Statistics*. New York: John Wiley and Sons.

[12] Ripley, B. D. (1988) *Statistical Inference for Spatial Processes*. Cambridge: Cambridge University Press.

[13] Ripley, B. D. (1996) *Pattern Recognition and Neural Networks*. Cambridge: Cambridge University Press.

[14] Titterington, D. M., Smith, A. F. M. and Makov, U. E. (1985) *Statistical Analysis of Finite Mixture Distributions*. Chichester: John Wiley and Sons.

[15] Wells, III, W. M., Grimson, W. E. L., Kikinis, R. and Jolesz, F. A. (1996) Adaptive segmentation of MRI data. *IEEE Transactions on Medical Imaging* **15**, 429–442.