NeuroImage **11,** 805–821 (2000)
doi:10.1006/nimg.2000.0582, available online at http://www.idealibrary.com on IDEAL®

Voxel-Based Morphometry—The Methods

John Ashburner and Karl J. Friston

The Wellcome Department of Cognitive Neurology, Institute of Neurology, Queen Square, London WC1N 3BG, United Kingdom

Received October 22, 1999

At its simplest, voxel-based morphometry (VBM) involves a voxel-wise comparison of the local concentration of gray matter between two groups of subjects. The procedure is relatively straightforward and involves spatially normalizing high-resolution images from all the subjects in the study into the same stereotactic space. This is followed by segmenting the gray matter from the spatially normalized images and smoothing the gray-matter segments. Voxel-wise parametric statistical tests which compare the smoothed gray-matter images from the two groups are performed. Corrections for multiple comparisons are made using the theory of Gaussian random fields. This paper describes the steps involved in VBM, with particular emphasis on segmenting gray matter from MR images with nonuniformity artifact. We provide evaluations of the assumptions that underpin the method, including the accuracy of the segmentation and the assumptions made about the statistical distribution of the data.

INTRODUCTION

A number of studies have already demonstrated structural brain differences among different patient populations using the technique of *voxel-based morphometry* (VBM) (Wright *et al.*, 1995, 1999; Vargha-Khadem *et al.*, 1998; Shah *et al.*, 1998; Krams *et al.*, 1999; Abell *et al.*, 1999; Woermann *et al.*, 1999; Sowell *et al.*, 1999; May *et al.*, 1999). This paper summarizes, and introduces some advances to, existing methods and provides evaluations of its components.

Studies of brain morphometry have been carried out by many researchers on a number of different populations, including patients with schizophrenia, autism, dyslexia, and Turner's syndrome. Often, the morphometric measurements used in these studies have been obtained from brain regions that can be clearly defined, resulting in a wealth of findings pertaining to these particular measurements. These measures are typically volumes of unambiguous structures such as the hippocampi or the ventricles. However, there are a number of morphometric features that may be more difficult to quantify by inspection, meaning that many structural differences may be overlooked. The importance of the VBM approach is that it is not biased to one particular structure and gives an even-handed and comprehensive assessment of anatomical differences throughout the brain.

Computational Neuroanatomy

With the increasing resolution of anatomical scans of the human brain and the sophistication of image processing techniques there have emerged, recently, a large number of approaches to characterizing differences in the shape and neuroanatomical configuration of different brains. One way to classify these approaches is to broadly divide them into those that deal with differences in brain shape and those that deal with differences in the local composition of brain tissue after macroscopic differences in shape have been discounted. The former use the deformation fields that map any individual brain onto some standard reference as the characterization of neuroanatomy, whereas the latter compare images on a voxel basis after the deformation fields have been used to spatially normalize the images. In short, computational neuroanatomic techniques can either use the deformation fields themselves or use these fields to normalize images that are then entered into an analysis of regionally specific differences. In this way, information about overall shape (deformation fields) and residual anatomic differences inherent in the data (normalized images) can be partitioned.

Deformation-Based and Tensor-Based Morphometry

We will use deformation-based and tensor-based morphometry in reference to methods for studying brain shapes that are based on deformation fields obtained by nonlinear registration of brain images. When comparing groups, deformation-based morphometry (DBM) uses deformation fields to identify differences in the relative positions of structures within the subjects' brains, whereas we use the term tensor-based morphometry to refer to those methods that localize differences in the local shape of brain structures (see Fig. 1).

1053-8119/00 $35.00

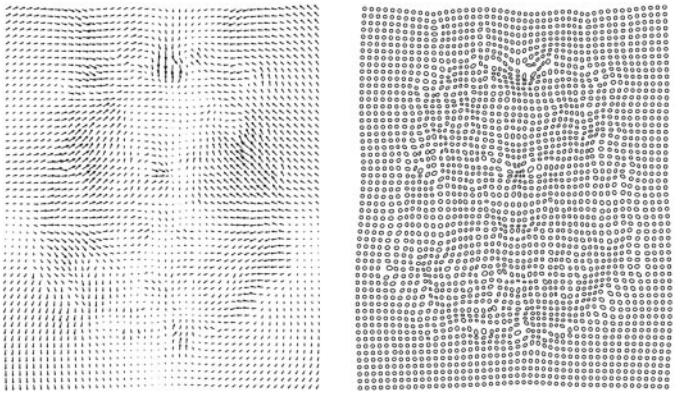

FIG. 1. We refer to deformation-based morphometry to describe methods of studying the positions of structures within the brain (left), whereas we use the term tensor-based morphometry for looking at local shapes (right). Currently, the main application of tensor-based morphometry involves using the Jacobian determinants to examine the relative volumes of different structures. However, there are other features of the Jacobian matrices that could be used, such as those representing elongation and contraction in different directions. The arrows in the image on the left show absolute displacements after making a global correction for rotations and translations, whereas the ellipses on the right show how the same circles would be distorted in different parts of the brain.

Characterization using DBM can be global, pertaining to the entire field as a single observation, or can proceed on a voxel-by-voxel basis to make inferences about regionally specific positional differences. This simple approach to the analysis of deformation fields involves treating them as vector fields representing absolute displacements. However, in this form, in addition to the shape information that is of interest, the vector fields also contain information on position and size that is likely to confound the analysis. Much of the confounding information can be removed by global rotations, translations, and a zoom of the fields in order to analyze the Procrustes shape (Bookstein, 1997a) of the brain.

DBM can be applied on a coarse (global) scale to simply identify whether there is a significant difference in the global shapes (based on a small number of parameters) among the brains of different populations. Generally, a single multivariate test is performed using the parameters describing the deformations—usually after parameter reduction using singular value decomposition. The Hotelling's T^2 statistic can be used for simple comparisons between two groups of subjects (Bookstein, 1997a, 1999), but for more complex experimental designs, a multivariate analysis of covariance can be used to identify differences via the Wilk's λ statistic (Ashburner *et al.*, 1998).

The alternative approach to DBM involves producing a statistical parametric map that locates any regions of significant positional differences among the groups of subjects. An example of this approach involves using a voxel-wise Hotelling's T^2 test on the vector field describing the displacements (Thompson and Toga, 1999; Gaser *et al.*, 1999) at each and every voxel. The significance of any observed differences can be assessed by assuming that the statistic field can then be approximated by a T^2 random field (Cao and Worsley, 1999). Note that this approach does not directly localize brain regions with different shapes, but rather identifies those brain structures that are in relatively different positions.

In order to localize structures whose shapes differ between groups, some form of tensor-based morphometry (TBM) is required to produce statistical parametric maps of regional shape differences. A deformation field that maps one image to another can be considered a discrete vector field. By taking the gradients at each element of the field, a Jacobian matrix field is obtained, in which each element is a tensor describing the relative positions of the neighboring elements. Morphometric measures derived from this tensor field can be used to locate regions with different shapes. The field obtained by taking the determinants at each point gives a map of the structure volumes relative to those of a reference image (Freeborough and Fox, 1998; Gee and Bajcsy, 1999). Statistical parametric maps of these determinant fields (or possibly their logs) can then be used to compare the anatomy of groups of subjects. Other measures derived from the tensor fields have also been used by other researchers, and these are described by Thompson and Toga (1999).

Voxel-Based Morphometry

The second class of techniques, which are applied to some scalar function of the normalized image, are referred to as voxel-based morphometry. The most prevalent example of this sort of approach, described in this paper, is the simple statistical comparison of gray matter partitions following segmentation. Other variants will be discussed later. Currently, the computational expense of computing very high resolution deformation fields (required for TBM at small scales) makes voxel-based morphometry a simple and pragmatic approach to addressing small-scale differences that is within the capabilities of most research units.

Overview

This paper describes the steps involved in voxel-based morphometry using the *SPM99* package (available from http://www.fil.ion.ucl.ac.uk). Following this we provide evaluations of the assumptions that underpin the method. This includes the accuracy of the segmentation and the assumptions made about the normality of the data. The paper ends with a discussion about the limitations of the method and some possible future directions.

VOXEL-BASED MORPHOMETRY

Voxel-based morphometry of MRI data involves spatially normalizing all the images to the same stereotactic space, extracting the gray matter from the normalized images, smoothing, and finally performing a statistical analysis to localize, and make inferences about, group differences. The output from the method is a statistical parametric map showing regions where gray matter concentration differs significantly between groups.

Spatial Normalization

Spatial normalization involves transforming all the subjects' data to the same stereotactic space. This is achieved by registering each of the images to the same template image, by minimizing the residual sum of squared differences between them. In our implementation, the first step in spatially normalizing each image involves matching the image by estimating the optimum 12-parameter affine transformation (Ashburner *et al.*, 1997). A Bayesian framework is used, whereby the maximum *a posteriori* estimate of the spatial transformation is made using prior knowledge of the normal variability of brain size. The second step accounts for global nonlinear shape differences, which are modeled by a linear combination of smooth spatial basis functions (Ashburner and Friston, 1999). The nonlinear registration involves estimating the coefficients of the basis functions that minimize the residual squared difference between the image and the template, while simultaneously maximizing the smoothness of the deformations.

It should be noted that this method of spatial normalization does not attempt to match every cortical feature exactly, but merely corrects for global brain shape differences. If the spatial normalization was perfectly exact, then all the segmented images would appear identical and no significant differences would be detected: VBM tries to detect differences in the regional concentration of gray matter at a local scale having discounted global shape differences.

It is important that the quality of the registration is as high as possible and that the choice of the template image does not bias the final solution. An ideal template would consist of the average of a large number of MR images that have been registered to within the accuracy of the spatial normalization technique. The spatially normalized images should have a relatively high resolution (1 or 1.5 mm isotropic voxels), so that the gray matter extraction method (described next) is not excessively confounded by partial volume effects, in which voxels contain a mixture of different tissue types.

Image Partitioning with Correction for Smooth Intensity Variations

The spatially normalized images are next partitioned into gray matter (GM), white matter (WM), cerebrospinal fluid (CSF), and three other background classes, using a modified mixture model cluster analysis technique. We have extended a previously described tissue classification method (Ashburner and Friston, 1997) so that it includes a correction for image intensity nonuniformity that arises for many reasons in MR imaging. Because the tissue classification is based on voxel intensities, the partitions derived using the older method can be confounded by these smooth intensity variations. Details of the improved segmentation method are provided in the Appendix.

Preprocessing of Gray Matter Segments

The gray matter images are now smoothed by convolving with an isotropic Gaussian kernel. This makes the subsequent voxel-by-voxel analysis comparable to a region of interest approach, because each voxel in the smoothed images contains the average concentration of gray matter from around the voxel (where the region around the voxel is defined by the form of the smoothing kernel). This is often referred to as "gray matter density," but should not be confused with cell packing density measured cytoarchitectonically. We will refer to "concentration" to avoid confusion. By the central limit theorem, smoothing also has the effect of rendering the data more normally distributed, increasing the validity of parametric statistical tests. Whenever possible, the size of the smoothing kernel should be comparable to the size of the expected regional differences between the groups of brains. The smoothing step also helps to compensate for the inexact nature of the spatial normalization.

Logit Transform

In effect, each voxel in the smoothed image segments represents the local concentration of the tissue (between 0 and 1). Often, prior to performing statistical tests on measures of concentration, the data are transformed using the *logit* transformation in order to render them more normally distributed. The logit transformation of a concentration p is given by

$$\text{logit}(p) = \frac{1}{2}\log_e\left(\frac{p}{1-p}\right).$$

For concentrations very close to either 1 or 0, it can be seen that the logit transform rapidly approaches infinite values. Because of this instability, it is advisable to exclude voxels from subsequent analyses that are too close to one or the other extreme. An improved

model for the data can be estimated using logistic regression (Taylor *et al.*, 1998), but this is beyond the scope of this paper as it requires iterative reweighted least-squares methods. Whether the logit transform is a necessary processing step for voxel-based morphometry will be addressed later.

Statistical Analysis

Statistical analysis using the general linear model (GLM) is used to identify regions of gray matter concentration that are significantly related to the particular effects under study (Friston *et al.*, 1995b). The GLM is a flexible framework that allows many different tests to be applied, ranging from group comparisons and identifying regions of gray matter concentration that are related to specified covariates such as disease severity or age to complex interactions between different effects of interest. Standard parametric statistical procedures (*t* tests and *F* tests) are used to test the hypotheses, so they are valid providing the residuals, after fitting the model, are independent and normally distributed. If the statistical model is appropriate there is no reason why the residuals should not be independent, but there are reasons why they may not be normally distributed. The original segmented images contain values between 0 and 1, of which most of the values are very close to either of the extremes. Only by smoothing the segmented images does the behavior of the residuals become more normally distributed.

Following the application of the GLM, the significance of any differences is ascertained using the theory of Gaussian random fields (Worsley *et al.*, 1996; Friston *et al.*, 1995a). A voxel-wise statistical parametric map (SPM) comprises the result of many statistical tests, and it is necessary to correct for these multiple dependent comparisons.

EVALUATIONS

A number of assumptions need to hold in order for VBM to be valid. First of all, we must be measuring the right thing. In other words, the segmentation must correctly identify gray and white matter, and consequently we have included an evaluation of the segmentation method. Also, confounding effects must be eliminated or modeled as far as possible. For example, it is not valid to compare two different groups if the images were acquired on two different scanners or with different MR sequences. In cases such as this, any group differences may be attributable to scanner differences rather than to the subjects themselves. Subtle but systematic differences in image contrast or noise can easily become statistically significant when a large number of subjects are entered in a study. A third issue of validity concerns the assumptions required by the statistical tests. For parametric tests, it is important that the data are normally distributed. If the data are not well behaved, then it is important to know what the effects are on the statistical tests. If there is doubt about the validity of the assumptions, it is better to use a nonparametric statistical analysis (Holmes *et al.*, 1996).

Evaluation of Segmentation

In order to provide a qualitative example of the segmentation, Fig. 2 shows a single sagittal slice through six randomly chosen T1-weighted images. The initial registration to the prior probability images was via an automatically estimated 12-parameter affine transformation (Ashburner *et al.*, 1997). The images were automatically segmented using the method described here, and contours of extracted gray and white matter are shown superimposed on the images.

In order to function properly, the segmentation method requires good contrast between the different tissue types. However, many central gray matter structures have image intensities that are almost indistinguishable from that of white matter, so the tissue classification is not very accurate in these regions. Another problem is that of partial volume. Because the model assumes that all voxels contain only one tissue type, the voxels that contain a mixture of tissues may not be modeled correctly. In particular, those voxels at the interface between white matter and ventricles will often appear as gray matter. This can be seen to a small extent in Figs. 2 and 3.

A Comparison of the Segmentation—With and without Nonuniform Sensitivity Correction

Segmentation was evaluated using a number of simulated images ($181 \times 217 \times 181$ voxels of $1 \times 1 \times 1$ mm) of the same brain generated by the BrainWeb simulator (Cocosco *et al.*, 1997; Kwan *et al.*, 1996; Collins *et al.*, 1998) with 3% noise (relative to the brightest tissue in the images). The contrasts of the images simulated T1-weighted, T2-weighted, and proton density (PD) images (all with 1.5-T field strength), and they were segmented individually and in a multispectral manner.[1] The T1-weighted image was simulated as a spoiled FLASH sequence, with a 30° flip angle, 18-ms repeat time, 10-ms echo time. The T2 and PD images were simulated by a dual echo spin echo technique, with 90° flip angle, 3300-ms repeat time, and echo times of 35 and 120 ms. Three different levels of image nonuniformity were used: 0%RF—which assumes that there is no intensity variation artifact, 40%RF—which assumes a fairly typical amount of nonuniformity, and 100%RF—which is more nonuniformity than would

[1] Note that different modulation fields that account for nonuniformity (see Appendix) were assumed for each image of the multispectral data sets.

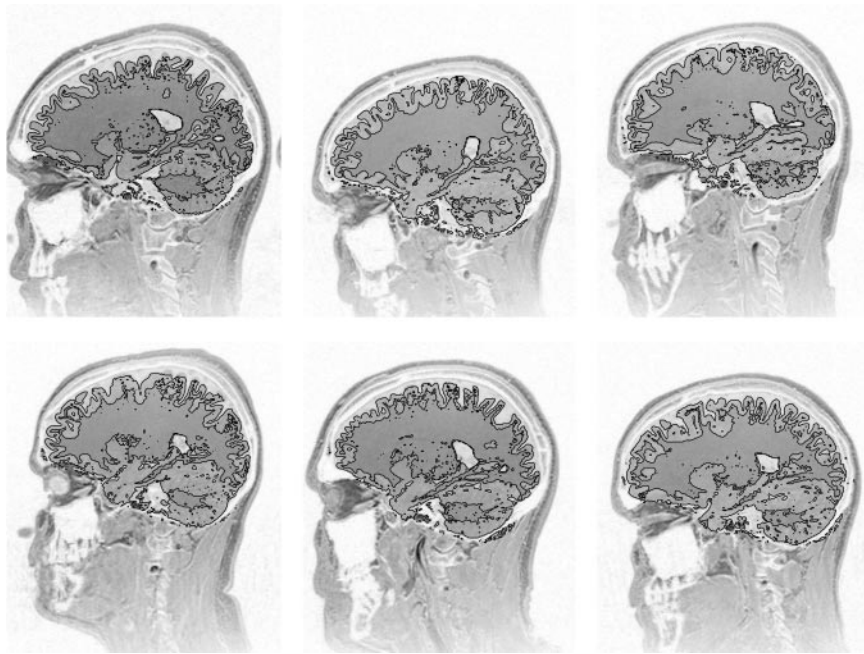

FIG. 2. A single sagittal slice through six T1-weighted images (2-T scanner, with an MPRAGE sequence, 12° tip angle, 9.7 ms repeat time, 4 ms echo time, and 0.6 ms inversion time). Contours of extracted gray and white matter are shown superimposed on the images.

normally be expected. The simulated images were segmented, both with and without sensitivity correction (see Appendix for further details). Three partitions were considered in the evaluation: gray matter, white matter, and other (not gray or white), and each voxel was assigned to the most likely partition. Because the data from which the simulated images were derived were available, it was possible to compare the segmented images with images of "true" gray and white matter using the κ statistic (a measure of interrater agreement),

$$\kappa = \frac{p_0 - p_e}{1 - p_e},$$

where p_0 is the observed proportion of agreement and p_e is the expected proportion of agreements by chance. If there are N observations in K categories, the observed proportional agreement is

$$p_0 = \sum_{k=1}^{K} f_{kk}/N,$$

where f_{kk} is the number of agreements for the kth category. The expected proportion of agreements is given by

$$p_e = \sum_{k=1}^{K} r_k c_k / N^2,$$

where r_k and c_k are the total number of voxels in the kth class for both the "true" and the estimated partitions.

The classification of a single plane of the simulated T1-weighted BrainWeb image with the nonuniformity is illustrated in Fig. 3. It should be noted that no preprocessing to remove scalp or other nonbrain tissue was performed on the image. In theory, the segmentation method should produce slightly better results of this nonbrain tissue is excluded from the computations. As the algorithm stands, a small amount of nonbrain tissue remains in the gray matter segment, which has arisen from voxels that lie close to gray matter and have similar intensities.

The resulting κ statistics from segmenting the different simulated images are shown in Table 1. These results show that the nonuniformity correction made little difference to the tissue classification of the images without any nonuniformity artifact. For images containing nonuniformity artifact, the segmentations using the correction were of about the same quality as the segmentations without the artifact and very much better than the segmentations without the correction.

A by-product of the segmentation is the estimation of an intensity nonuniformity field. Figure 4 shows a comparison of the intensity nonuniformity present in a simulated T1 image with 100% nonuniformity (created by dividing noiseless simulated images with 100% nonuniformity and no nonuniformity) with that recovered by the segmentation method. A scatterplot of "true"

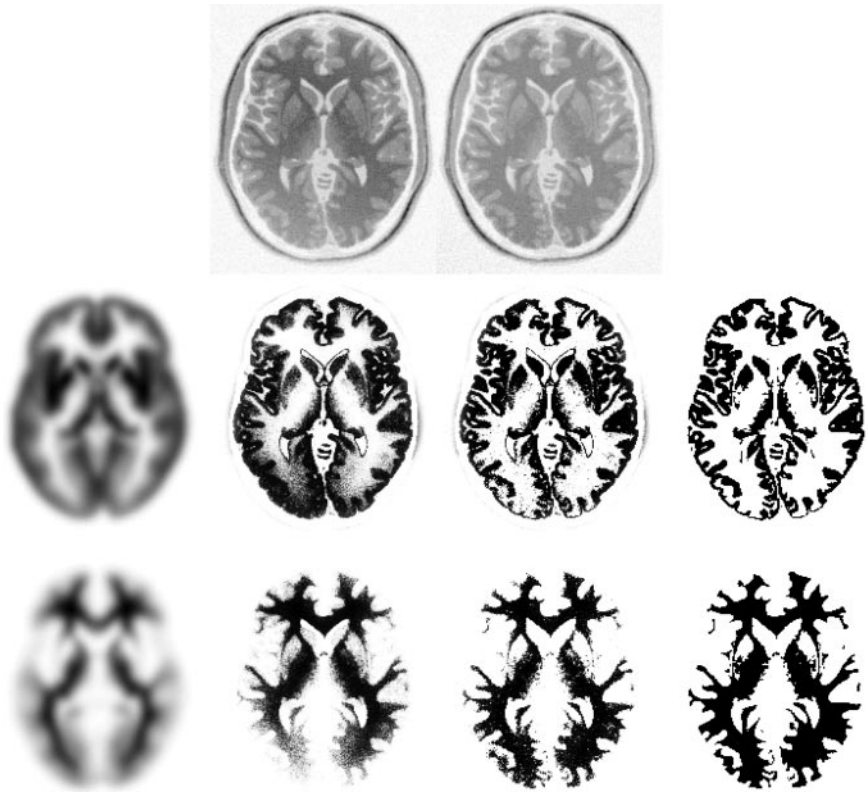

FIG. 3. The classification of the simulated BrainWeb image. The top row shows the original simulated T1-weighted MR image with 100% nonuniformity and the nonuniformity corrected version. From left to right, the middle row shows the *a priori* spatial distribution of gray matter used for the segmentation, gray matter segmented without nonuniformity correction, gray matter segmented with nonuniformity correction, and the "true" distribution of gray matter (from which the simulated images were derived). The bottom row is the same as the middle, except that it shows white matter rather than gray. Without nonuniformity correction, the intensity variation causes some of the white matter in posterior areas to be classified as gray. This was also very apparent in the cerebellum because of the intensity variation in the inferior–superior direction.

versus recovered nonuniformity shows a straight line, suggesting that the accuracy of the estimated nonuniformity is very good.

Stability with Respect to Misregistration with the a Priori Images

In order for the Bayesian segmentation to work properly, the image volume must be in register with a set of *a priori* probability images used to instate the priors. Here we examine the effects of misregistration on the accuracy of the segmentation, by artificially translating (in the left–right direction) the prior probability images by different distances prior to segmenting the whole simulated volume. The 1-mm slice thickness, 40% nonuniformity, and 3% noise simulated T1-weighted image (described above) was used for the segmentation, which included the nonuniformity correction. The κ statistic was computed with respect to

TABLE 1

	Single image			Multispectral			
	T1	T2	PD	T2/PD	T1/T2	T1/PD	T1/T2/PD
0%RF—uncorrected	0.95	0.90	0.90	0.93	0.94	0.96	0.94
0%RF—corrected	0.95	0.90	0.90	0.93	0.94	0.96	0.95
40%RF—uncorrected	0.92	0.88	0.79	0.90	0.93	0.95	0.94
40%RF—corrected	0.95	0.90	0.90	0.93	0.94	0.96	0.94
100%RF—uncorrected	0.85	0.85	0.67	0.87	0.92	0.94	0.93
100%RF—corrected	0.94	0.90	0.88	0.92	0.93	0.95	0.94

Note. The different κ statistics that were computed after segmenting the simulated images are shown.

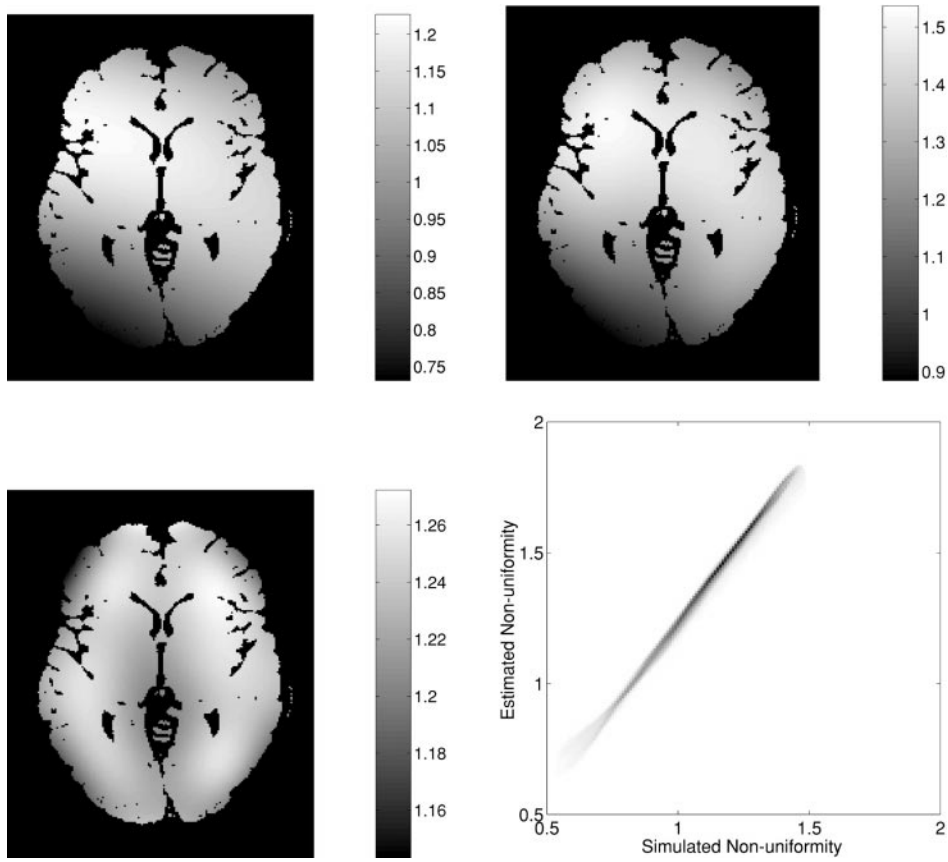

FIG. 4. Top left: The true intensity nonuniformity field of the simulated T1 image. Top right: The nonuniformity recovered by the segmentation algorithm. Bottom left: The recovered divided by the true nonuniformity. Bottom right: A scatterplot of true intensity nonuniformity versus recovered nonuniformity, derived from voxels throughout the whole volume classified as either white or gray matter. Note that the plot is a straight line, but that its gradient is not because it is not possible to recover the absolute scaling of the field.

the true gray and white matter for the different translations, and the results are plotted in Fig. 5.

In addition to illustrating the effect of misregistration, this also gives an indication of how far a brain can deviate from the normal population of brains (that constitute the prior probability images) in order for it to be segmented adequately. Clearly, if the brain cannot be adequately registered with the probability images, then the segmentation will not be as accurate. This also has implications for severely abnormal brains, as these are more difficult to register with the images that represent the prior probabilities of voxels belonging to different classes. Segmenting these abnormal brains can be a problem for the algorithm, as the prior probability images are based on normal healthy brains. Clearly the profile in Fig. 5 depends on the smoothness or resolution of the *a priori* images. By not smoothing the *a priori* images, the segmentation would be optimal for normal, young, and healthy brains. However, the prior probability images may need to be smoother in order to encompass more variability when patient data are to be analyzed.

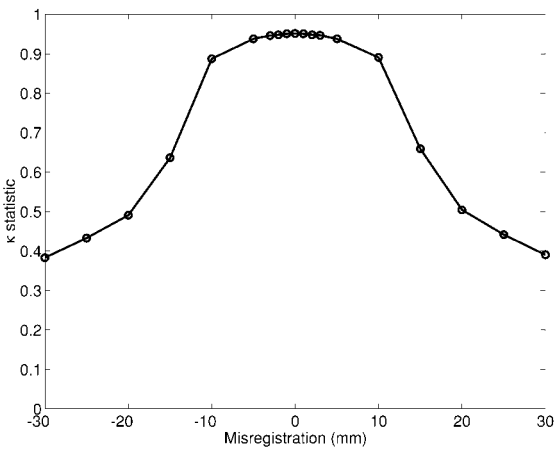

FIG. 5. Segmentation accuracy with respect to misregistration with the *a priori* images.

Evaluation of the Assumptions about Normally Distributed Data

The statistics used to identify structural differences make the assumption that the residuals after fitting the model are normally distributed. Statistics cannot

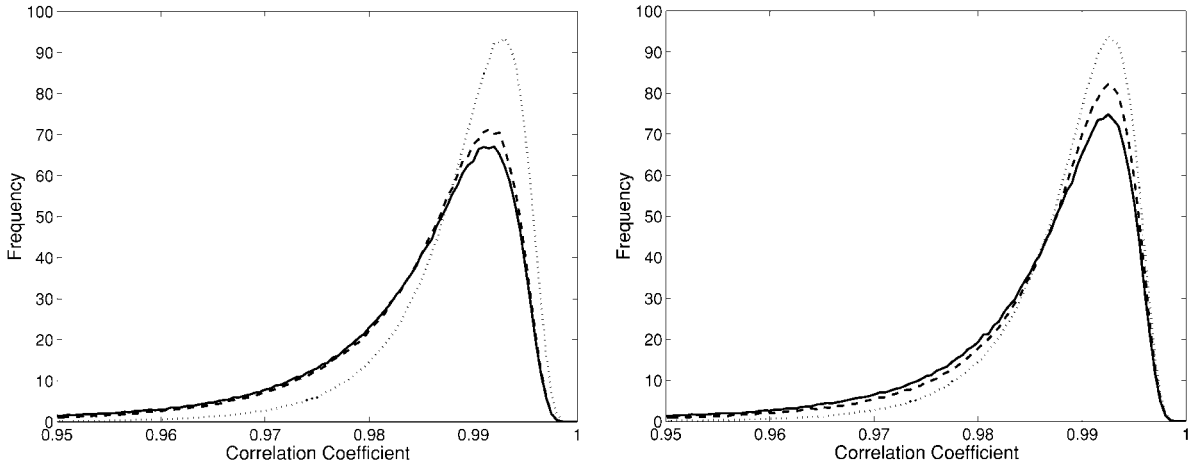

FIG. 6. Histograms of correlation coefficients taken over the whole-image volumes (using a total of 717,191 voxels for which the mean intensity over all images was greater than 0.05). The dotted lines are the histograms that would be expected if the data were perfectly normally distributed. The solid lines show the histograms of the data without the logit transform, and the dashed lines show the histograms obtained using the logit transformed data. The plot on the left is based on the model that does not include global gray matter as a confound, whereas that on the right does model this confounding effect.

prove that data are normally distributed—it can only be used to disprove the hypothesis that they are normal. For normally distributed data, a *Q-Q plot* of the data should be a straight line. A significant deviation from a straight line can be identified by computing the correlation coefficient of the plot as described by Johnson and Wichern (1998).

A Q-Q plot is a plot of the sample quantile versus the sample quantile that would be expected if the residuals were normally distributed. Computing the sample quantile involves first sorting the J residuals (after dividing by the square root of the diagonal elements of the residual forming matrix) into increasing order (x_1, x_2, . . . , x_J). The inverse cumulative distribution of each of the J elements is then computed as

$$q_j = \sqrt{2}\ \mathrm{erfinv}\left(2\,\frac{j-\frac{3}{8}}{J+\frac{1}{4}} - 1\right),$$

where erfinv is the inverse error function. A Q-Q plot is simply a plot of $\mathbf{q}$ versus $\mathbf{x}$ and should be a straight line if the data in $\mathbf{x}$ are normally distributed. To test normality, the correlation coefficient for the Q-Q plot is used to test for any significant deviation from a straight line. A lookup table is used to reject the null hypothesis if the correlation coefficient falls below a particular value, given a certain sample size. However, in this paper we simply use the correlation coefficient as a "normality statistic" and examine its distribution over voxels.

The data used to test the assumptions were T1-weighted MRI scans of 50 normal male right-handed subjects ages between 17 and 62 (median 26, mean 29), whose structural scans had been acquired as part of an ongoing program of functional imaging research. The scans were performed on a Siemens Magnetom Vision scanner operating at 2 T. An MPRAGE sequence was used with a 12° tip angle, 9.7-ms repeat time, 4-ms echo time, and 0.6-ms inversion time, to generate sagittal images of the whole brain with voxel sizes of 1 × 1 × 1.5 mm. The images were spatially normalized, segmented, and smoothed using a Gaussian kernel of 12 mm full width at half-maximum (FWHM).

Voxel-by-voxel correlation coefficients of the Q-Q plots were computed over all voxels of the data for which the mean intensity over all images was greater than 0.05. Voxels of low mean intensity were excluded from the computations, because they would not be included in the VBM analysis. This is because we know that these low-intensity voxels are most likely to deviate most strongly from the assumptions about normality. Q-Q plots were computed using two different linear models. The first model involved looking at the residuals after fitting the mean, whereas the second was more complex, in that it also modeled the confounding effect of the total amount of gray matter in each volume. Q-Q plots were computed both with and without the logit transform. Histograms of the correlation coefficients were computed over the whole-image volumes (717,191 voxels), along with histograms generated from simulated Gaussian noise. These are plotted in Fig. 6 and show that the data do deviate slightly from normally distributed. The logit transform appeared to make the residuals slightly more normally distributed. The normality of the residuals was also improved by modeling the total amount of gray matter as a confounding effect.

Testing the Rate of False Positives Using Randomization

The previous section showed that the data are not quite normally distributed, but it does not show how the nonnormality influences any subsequent statistics.

Ultimately, we wish to protect against false-positive results, and in this part of the paper, we test how frequently they arise. The statistics were evaluated using the same preprocessed structural brain images of 50 subjects as were used in the previous section. The subjects were randomly assigned, with replacement, to two groups of 12 and 38, and statistical tests performed using SPM99b (Wellcome Department of Cognitive Neurology, London, UK) to compare the groups. The numbers in the groups were chosen as many studies typically involve comparing about a dozen patients with a larger group of control subjects. This was repeated a total of 50 times, looking for both significant increases and decreases in the gray matter concentration of the smaller group. The end result is a series of 100 parametric maps of the t statistic. Within each of these SPMs, the local maxima of the t statistic field were corrected for the number of dependent tests performed, and a P value was assigned to each (Friston *et al.*, 1995a). Using a corrected threshold of $P = 0.05$, we would expect about five local maxima with P values below this threshold by chance alone. Over the 100 SPMs, there were six local maxima with corrected P values below 0.05. The same 50 subjects were randomly assigned to either of the two groups and the statistics performed a further 50 times, but this time modeling the total amount of gray matter as a confounding effect. The results of this analysis produced four significant local maxima with corrected P values below 0.05. These results suggest that the inference procedures employed are robust to the mild deviations from normality incurred by using smooth image partitions.

Another test available within SPM is based on the number of connected voxels in a cluster defined by a prespecified threshold (extent statistic). In order to be valid, this test requires the smoothness of the residuals to be spatially invariant, but this is known not to be the case by virtue of the highly nonstationary nature of the underlying neuroanatomy. As noted by Worsley *et al.* (1999), this nonstationary smoothness leads to inexact P values.

> The reason is simply this: by chance alone, large size clusters will occur in regions where the images are very smooth, and small size clusters will occur in regions where the image is very rough. The distribution of cluster sizes will therefore be considerably biased towards more extreme cluster sizes, resulting in more false positive clusters in smooth regions. Moreover, true positive clusters in rough regions could be overlooked because their sizes are not large enough to exceed the critical size for the whole region.

Corrected probability values were assigned to each cluster based on the number of connected voxels exceeding a t value of 3.27 (spatial extent test). Approximately 5 significant clusters would be expected from the 100 SPMs if the smoothness was stationary. Eighteen significant clusters were found when the total amount of gray matter was not modeled as a confound, and 14 significant clusters were obtained when it was. These tests confirmed that the voxel-based extent statistic should not be used in VBM.

Under the null hypothesis, repeatedly computed t statistics should assume the probability density function of the Student t distribution. This was verified using the computed t fields, of which each t field contains 717,191 voxels. Plots of the resulting histograms are shown in Fig. 7. The top row presents distributions when global differences in gray matter were not removed as a confound. Note that global variance biases the distributions of t values from any particular comparison.

Further experiments were performed to test whether false positives occurred evenly throughout the brain or were more specific to particular regions. The tests were done on a single slice through the same 50 subjects' preprocessed brain images, but used the total count of gray matter in the brains as a confound. Each subject was randomly assigned to two groups of 12 and 38, pixel-by-pixel two-tailed t tests were done, and locations of t scores higher than 3.2729 or lower than -3.2729 were recorded (corresponding to an uncorrected probability of 0.002). This procedure was repeated 10,000 times, and Fig. 8 shows an image of the number of false positives occurring at each of the 10,693 pixels. Visually, the false positives appear to be uniformly distributed. According to the theory, the number of false positives occurring at each pixel should be 20 ($10{,}000 \times 0.002$). An average of 20.171 false positives was found, showing that the validity of statistical tests based on uncorrected t statistics is not severely compromised.

DISCUSSION

Possible Improvements to the Segmentation

One of the analytic components described in this paper is an improved method of segmentation that is able to correct for image nonuniformity that is smooth in all three dimensions. The method has been found to be robust and accurate for high-quality T1-weighted images, but is not beyond improvement. Currently, each voxel is assigned a probability of belonging to a particular tissue class based only on its intensity and information from the prior probability images. There is a great deal of other information that could be incorporated into the classification. For example, we know that if all a voxel's neighbors are gray matter, then there is a high probability that it should also be gray matter. Other researchers have successfully used

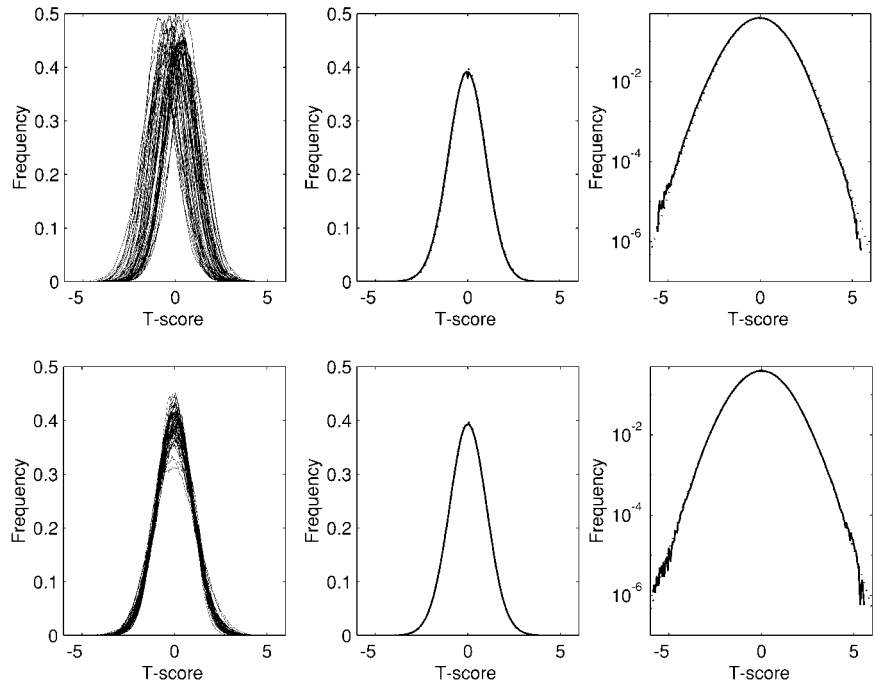

FIG. 7. Histograms of *t* scores from randomly generated tests. Top: Not modeling mean effect (48 degrees of freedom). Bottom: Modeling a mean effect as a confound (47 degrees of freedom). Left: 50 histograms of *t* scores testing randomly generated effects of interest. Center: The mean (i.e., cumulative distribution over all voxels and volumes) of the 50 histograms is plotted as a solid line, and the probability density function of the Student *t* distribution for 47/48 degrees of freedom is shown by the dotted line. Right: The same as center, except plotted on a logarithmic scale.

Markow random field models to include this information in the tissue classification model (Vandermeulen *et al.*, 1996; Van Leemput *et al.*, 1999b). A very simple prior, that can be incorporated, is the relative intensity of the different tissue types. For example, if we are segmenting a T1-weighted image, we know that the white matter should have a higher intensity than the gray matter, which in turn should be more intense than that of the CSF. When computing the means for each cluster, this prior information could sensibly be used to bias the estimates.

The Effect of Spatial Normalization

Because of the nonlinear spatial normalization, the volumes of certain brain regions will grow, whereas others will shrink. This has implications for the interpretation of what VBM is actually testing for. The

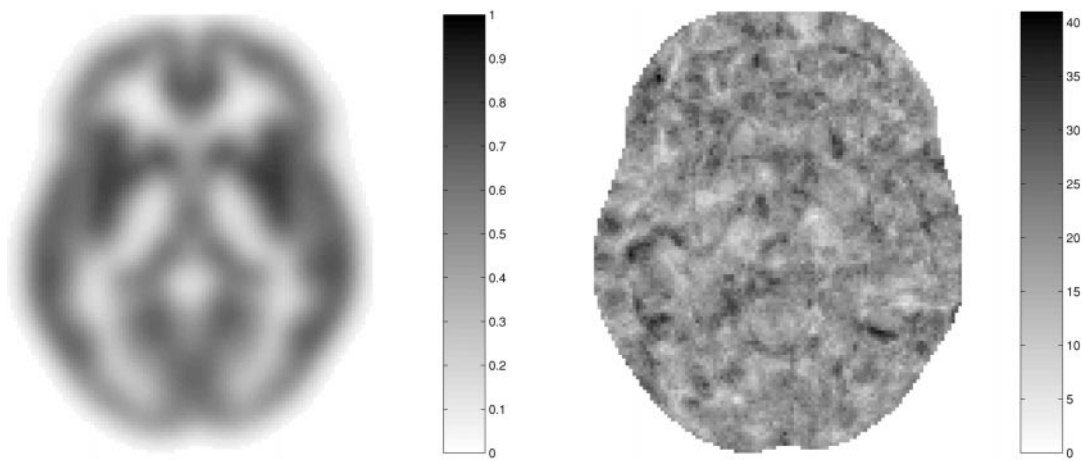

FIG. 8. Left: Mean of 50 subjects' preprocessed brain images. Right: Number of false positives occurring at each voxel at the uncorrected 0.002 level, after 10,000 randomizations.

objective of VBM is to identify regional differences in the concentration of a particular tissue (gray or white matter). In order to preserve the actual amounts of gray matter within each structure, a further processing step that multiplies the partitioned images by the relative voxel volumes can be incorporated. These relative volumes are simply the Jacobian determinants of the deformation field. This augmented VBM can therefore be considered a combination of VBM and TBM, in which the TBM employs the testing of the Jacobian determinants. VBM can be thought of as comparing the relative concentration of gray matter (i.e., the proportion of gray matter to other tissue types within a region). With the adjustment for volume change, VBM would be comparing the absolute amounts of gray matter in the different regions. As mentioned under "Spatial Normalization," if the spatial normalization was perfect, then no gray matter differences would be observed if a volume change adjustment was not applied. In this instance, all the information would be in the deformation fields and would be tested using TBM. However, if the spatial normalization is only removing global differences in brain shape, the results of VBM show relative gray matter concentration differences. As faster and more precise registration methods emerge, then a TBM volume change adjustment may become more important. It is envisaged that, by incorporating this correction, a continuum will arise with simple VBM (with low-resolution spatial normalization) at one end of the methodology spectrum and statistical tests based on Jacobian determinants at the other (with high-resolution spatial normalization).

Another perspective on what VBM is actually comparing can be obtained by considering how a similar analysis would be done using volumes of interest (VOI). To simplify the analogy, consider that the smoothing kernel is the shape of a sphere (values of 1 inside and 0 outside) rather than a 3D Gaussian point spread function. After an image is convolved with this kernel, each voxel in the smoothed image will contain a count of the gray matter voxels from the surrounding spherical VOI. Now consider the effects of the spatial normalization and where the voxels within each VOI come from in the original gray matter images. The spheres can be thought of as being projected onto the original anatomy, but in doing so, their shapes and sizes will be distorted. Without multiplying by the relative voxel sizes, what would be measured would be the proportion of gray matter within each projected VOI (relative to other tissue types). With the multiplication, the total amount of gray matter within the VOI is being measured.

Multivariate Voxel-Based Morphometry

Ideally, a procedure like VBM should be able to automatically identify any structural abnormalities in a single brain image. However, even with many hundreds of subjects in a database of controls, as it stands, the method may not be powerful enough to detect subtle abnormalities in individuals. A possibly more powerful procedure would be to use some form of voxel-wise multivariate approach. Within a multivariate framework, in addition to images of gray matter concentration, other image features would be included. The first obvious feature to be included would be white matter concentration. Other features could include local indices of gyrification such as the curvature of the gray matter segment, image gradients, and possibly information from the spatial normalization procedure. With a larger database of controls, more image features can be included without seriously impacting on the degrees of freedom of the model. The Hotelling's T^2 test could be used to perform simple comparisons between two groups. However, for more complex models, the more general multivariate analysis of covariance would be necessary. By doing this, VBM and tensor-based morphometric techniques can be combined in order to provide a more powerful method of localizing regional abnormalities.

CONCLUSIONS

This paper has considered the various components of voxel-based morphometry. We have described and evaluated an improved method of MR image segmentation, showing that the modifications do improve the segmentation of images with intensity nonuniformity artifact. In addition, we tested some of the assumptions necessary for the parametric statistical tests used by SPM99 to implement VBM. We demonstrated that the data used for these analyses are not exactly normally distributed. However, no evidence was found to suppose that (with 12-mm FWHM smoothed data) uncorrected statistical tests or corrected statistical inferences based on peak height are invalid. We found that the statistic based on cluster spatial extent is not valid for VBM analysis, suggesting a violation of the stationariness assumptions upon which this test is based. Until the spatial extent test has been modified to accommodate nonstationary smoothness, then VBM should not use cluster size to assess significance (the peak height test has already been modified).

APPENDIX

The Tissue Classification Method

Although we actually use a three-dimensional implementation of the tissue classification method, which can also be applied to multispectral images, we will simplify the explanation of the algorithm by describing its application to a single two-dimensional image.

The tissue classification model makes a number of assumptions. The first is that each of the $I \times J$ voxels

of the image (**F**) has been drawn from a known number (K) of distinct tissue classes (clusters). The distribution of the voxel intensities within each class is normal (or multinormal for multispectral images) and initially unknown. The distribution of voxel intensities within cluster k is described by the number of voxels within the cluster (h_k), the mean for that cluster (v_k), and the variance around that mean (c_k). Because the images are spatially normalized to a particular stereotactic space, prior probabilities of the voxels belonging to the GM, the WM, and the CSF classes are known. This information is in the form of probability images—provided by the Montréal Neurological Institute (Evans *et al.*, 1992, 1993, 1994)—which have been derived from the MR images of 152 subjects (66 female and 86 male; 129 right handed, 14 left handed, and 9 unknown handedness; ages between 18 and 44, with a mean age of 25 and median age of 24). The images were originally segmented using a neural network approach, and misclassified nonbrain tissue was removed by a masking procedure. To increase the stability of the segmentation with respect to small registration errors, the images are convolved with an 8-mm full width at half-maximum Gaussian smoothing kernel. The prior probability of voxel f_{ij} belonging to cluster k is denoted by b_{ijk}. The final assumption is that the intensity of the image has been modulated by multiplication with an unknown scalar field. Most of the algorithm for classifying the voxels has been described elsewhere (Ashburner and Friston, 1997), so this paper will emphasize the modification for correcting the modulation field.

There are many unknown parameters in the segmentation algorithm, and estimating any of these requires knowledge about the other parameters. Estimating the parameters that describe a cluster (h_k, v_k, and c_k) relies on knowing which voxels belong to the cluster and also the form of the intensity modulating function. Estimating which voxels should be assigned to each cluster requires the cluster parameters to be defined and also the modulation field. In turn, estimating the modulation field needs the cluster parameters and the belonging probabilities.

The problem requires an iterative algorithm (see Fig. 9). It begins with assigning starting estimates for the various parameters. The starting estimate for the modulation field is typically uniformly 1. Starting estimates for the belonging probabilities of the GM, WM, and CSF partitions are based on the prior probability images. Since we have no probability maps for background and nonbrain tissue clusters, we estimate them by subtracting the prior probabilities for GM, WM, and CSF from a map of all 1's and divide the result equally between the remaining clusters.[2]

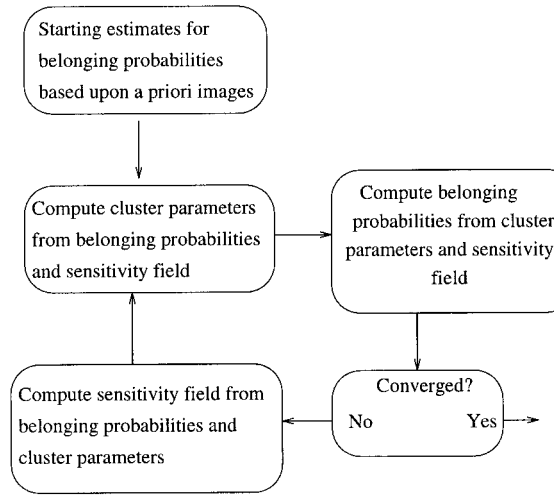

FIG. 9. A flow diagram for the tissue classification.

Each iteration of the algorithm involves estimating the cluster parameters from the nonuniformity corrected image, assigning belonging probabilities based on the cluster parameters, checking for convergence, and reestimating and applying the modulation function. This continues until a convergence criterion is satisfied. The final values for the belonging probabilities are in the range of 0 to 1, although most values tend to stabilize very close to one of the two extremes. The individual steps involved in each iteration will now be described in more detail.

Estimating the Cluster Parameters

This stage requires the original image to be intensity corrected according to the most recent estimate of the modulation function. Each voxel of the intensity-corrected image is denoted by g_{ij}. We also have the current estimate of the belonging probabilities for each voxel with respect to each cluster. The probability of voxel i, j belonging to class k is denoted by p_{ijk}.

The first step is to compute the number of voxels belonging to each of the K clusters (**h**) as

$$h_k = \sum_{i=1}^{I} \sum_{j=1}^{J} p_{ijk} \quad \text{over } k = 1..K.$$

Mean voxel intensities for each cluster (**v**) are computed. This step effectively produces a weighted mean of the image voxels, where the weights are the current belonging probability estimates:

[2] Where identical prior probability maps are used for more than one cluster, the affected cluster parameters need to be modified slightly. This is typically done after the first iteration, by assigning different values for the means uniformly spaced between 0 and the intensity of the white matter cluster.

$$v_k = \frac{\sum_{i=1}^{I} \sum_{j=1}^{J} p_{ijk} g_{ij}}{h_k} \quad \text{over } k = 1..K.$$

Then the variance of each cluster (**c**) is computed in a way similar to the mean:

$$c_k = \frac{\sum_{i=1}^{I} \sum_{j=1}^{J} p_{ijk} (g_{ij} - v_k)^2}{h_k} \quad \text{over } k = 1..K.$$

Assigning Belonging Probabilities

The next step is to recalculate the belonging probabilities. It uses the cluster parameters computed in the previous step, along with the prior probability images and the intensity modulated input image. Bayes' rule is used to assign the probability of each voxel belonging to each cluster,

$$p_{ijk} = \frac{r_{ijk} q_{ijk}}{\sum_{l=1}^{K} r_{ijl} q_{ijl}}$$

over $i = 1..I$, $j = 1..J$, and $k = 1..K$,

where p_{ijk} is the *a posteriori* probability that voxel i, j belongs to cluster k given its intensity of g_{ij}; r_{ijk} is the likelihood of a voxel in cluster k having an intensity of g_{ik}; and q_{ijk} is the *a priori* probability of voxel i, j belonging in cluster k.

The likelihood function is obtained by evaluating the probability density functions for the clusters at each of the voxels:

$$r_{ijk} = (2\pi c_k)^{-1/2} \exp\left(\frac{-(g_{ij} - v_k)^2}{2c_k}\right)$$

over $i = 1..I$, $j = 1..J$, and $k = 1..K$.

The prior (q_{ijk}) is based on two factors: the number of voxels currently belonging to each cluster (h_k) and the prior probability images derived from a number of images (b_{ijk}). With no knowledge of the *a priori* spatial distribution of the clusters or the intensity of a voxel, then the *a priori* probability of any voxel belonging to a particular cluster is proportional to the number of voxels currently included in that cluster. However, with the additional data from the prior probability images, we can obtain a better estimate of the priors:

$$q_{ijk} = \frac{h_k b_{ijk}}{\sum_{l=1}^{I} \sum_{m=1}^{J} b_{lmk}}$$

over $i = 1..I$, $j = 1..J$, and $k = 1..K$.

Convergence is ascertained by following the log-likelihood function:

$$\sum_{i=1}^{I} \sum_{j=1}^{J} \log\left(\sum_{k=1}^{K} r_{ijk} q_{ijk}\right).$$

The algorithm is terminated when the change in log-likelihood from the previous iteration becomes negligible.

Estimating and Applying the Modulation Function

Many groups have developed methods for correcting intensity nonuniformities in MR images, and the scheme we describe here shares common features. There are two basic models describing the noise properties of the images: multiplicative noise and additive noise. The multiplicative model describes images that have noise added before being modulated by the nonuniformity field (i.e., the standard deviation of the noise is multiplied by the modulating field), whereas the additive version models noise that is added after the modulation (standard deviation is constant). We have used a multiplicative noise model, which assumes that the errors originate from tissue variability rather than additive Gaussian noise from the scanner. Figure 10 illustrates the model used by the classification.

Nonuniformity correction methods all involve estimating a smooth function that modulates the image intensities. If the function is not forced to be smooth, then it will begin to fit the higher frequency intensity variations due to different tissue types, rather than the low-frequency intensity nonuniformity artifact. Thin-plate spline (Sled *et al.*, 1998) and polynomial (Van Leemput *et al.*, 1999a, b) basis functions are widely used for modeling the intensity variation. In these models, the higher frequency intensity variations are restricted by limiting the number of basis functions. In the current model, we assume that the modulation field (**U**) has been drawn from a population for which we know the *a priori* distribution. The distribution is assumed to be multinormal, with a mean that is uniformly 1 and a covariance matrix that models smoothly varying functions. In this way, a Bayesian scheme is used to penalize high-frequency intensity variations by introducing a cost function based on the "energy" of the modulating function. There are many possible forms for this energy function. Some widely used simple cost functions include the "membrane energy" and the "bending energy" (1997b), which (in three dimensions) have the forms $h = \sum_i \sum_{j=1}^{3} \lambda \left((\partial u(\mathbf{x}_i))/\partial x_{ji}\right)^2$ and $h = \sum_i \sum_{j=1}^{3} \sum_{k=1}^{3} \lambda((\partial^2 u(\mathbf{x}_i))/\partial x_{ji} \partial x_{ki})^2$, respectively. In these formulae, $\partial u(\mathbf{x}_i)/\partial x_{ji}$ is the gradient of the modulating function at the ith voxel in the jth orthogonal direction and λ is a user-assigned constant. However, for the purpose of modulating the images, we use a smoother cost function that is based on the squares of the third derivatives:

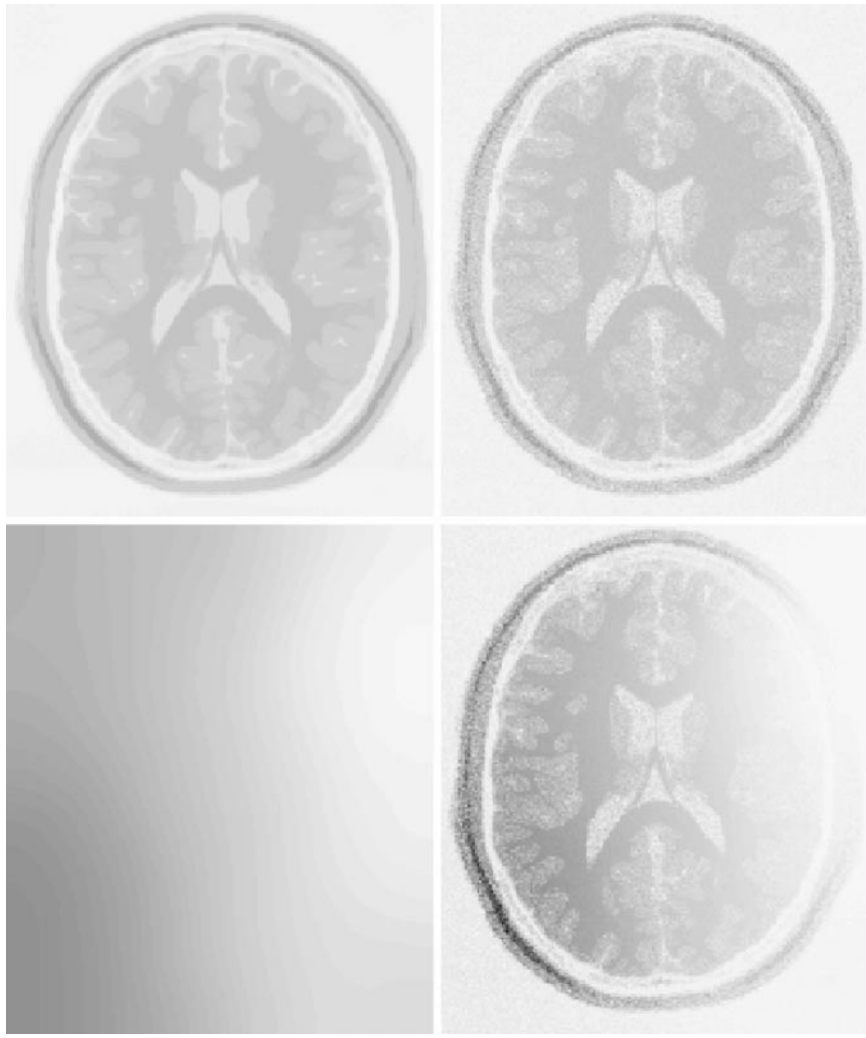

FIG. 10. The MR images are modeled as a number of distinct clusters (top left), with different levels of Gaussian random noise added to each cluster (top right). The intensity modulation is assumed to be smoothly varying (bottom left) and is applied as a straightforward multiplication of the modulation field with the image (bottom right).

$$h = \sum_i \sum_{j=1}^{3} \sum_{k=1}^{3} \sum_{l=1}^{3} \lambda \left(\frac{\partial^3 u(\mathbf{x}_i)}{\partial x_{ji} \partial x_{ki} \partial x_{li}} \right)^2 .$$

This model was chosen because it produces slowly varying modulation fields that can represent the variety of nonuniformity effects that we expect to encounter in MR images (see Fig. 11).

To reduce the number of parameters describing the field, it is modeled by a linear combination of low-frequency discrete cosine transform basis functions (chosen because there are no boundary constraints). A two (or three)-dimensional discrete cosine transform (DCT) is performed as a series of one-dimensional transforms, which are simply multiplications with the DCT matrix. The elements of a matrix ($\mathbf{D}$) for computing the first M coefficients of the DCT of a vector of length I are given by

$$d_{i1} = \frac{1}{\sqrt{I}}, \; i = 1..I,$$

$$d_{im} = \sqrt{\frac{2}{I}} \cos\left(\frac{\pi(2i-1)(m-1)}{2I} \right), \; i = 1..I, \; m = 2..M. \tag{1}$$

The matrix notation for computing the first $M \times M$ coefficients of the two-dimensional DCT of a modulation field $\mathbf{U}$ is $\mathbf{X} = \mathbf{D}_1^T \mathbf{U} \mathbf{D}_2$, where the dimensions of the DCT matrices $\mathbf{D}_1$ and $\mathbf{D}_2$ are $I \times M$ and $J \times M$, respectively, and $\mathbf{U}$ is an $I \times J$ matrix. The approximate inverse DCT is computed by $\mathbf{U} \simeq \mathbf{D}_1 \mathbf{X} \mathbf{D}_2^T$. An alternative representation of the two-dimensional DCT obtains by reshaping the $I \times J$ matrix $\mathbf{U}$ so that it is a vector ($\mathbf{u}$). Element $i + (j-1) \times I$ of the vector is then equal to element i, j of the matrix. The two-dimensional

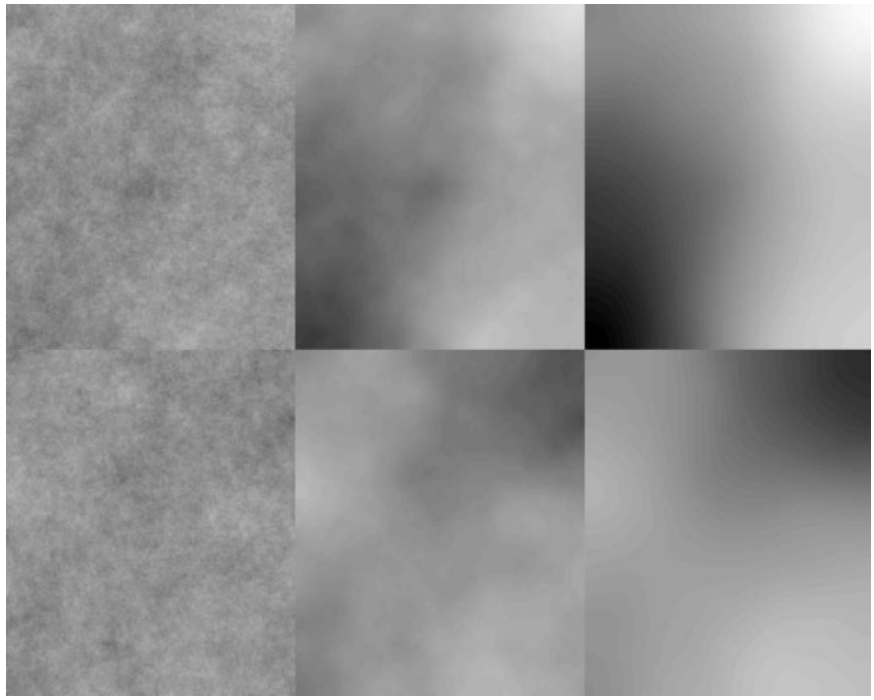

FIG. 11. Randomly generated modulation fields using the membrane energy cost function (left), the bending energy cost function (center), and the squares of the third derivatives (right).

DCT can then be represented by $\mathbf{x} = \mathbf{D}^T\mathbf{u}$, where $\mathbf{D} = \mathbf{D_2} \otimes \mathbf{D_1}$ (the Kronecker tensor product of $\mathbf{D_2}$ and $\mathbf{D_1}$) and $\mathbf{u} \simeq \mathbf{Dx}$.

The sensitivity correction field is computed by estimating the coefficients ($\mathbf{x}$) of the basis functions that minimize a weighted sum of squared differences between the data and the model and also the bending energy of the modulation field. This can be expressed using matrix terminology as a regularized weighted least-squares fitting,

$$\mathbf{x} = (\mathbf{A_1}^T\mathbf{A_1} + \mathbf{A_2}^T\mathbf{A_2} \cdots + \mathbf{C_0}^{-1})^{-1} \times (\mathbf{A_1}^T\mathbf{b_1} + \mathbf{A_2}^T\mathbf{b_2} \cdots + \mathbf{C_0}^{-1}\mathbf{x_0}),$$

where $\mathbf{x_0}$ and $\mathbf{C_0}$ are the means and covariance matrices describing the *a priori* distribution of the coefficients. Matrix $\mathbf{A_k}$ and column vector $\mathbf{b_k}$ are constructed for cluster k from

$$\mathbf{A_k} = \mathrm{diag}(\mathbf{p_k} c_k^{-1/2})\mathrm{diag}(\mathbf{f})\mathbf{D} \quad \text{and} \quad \mathbf{b_k} = \mathbf{p_k} c_k^{-1/2} v_k,$$

where $\mathbf{p_k}$ refers to the belonging probabilities for the kth cluster considered as a column vector. The objective is to find the smooth modulating function (described by its DCT coefficients) that will bring the voxel intensities of each cluster as close as possible (in the least-squares sense) to the cluster means, in which the vectors $\mathbf{p_k} c_k^{-1/2}$ are voxel-by-voxel weighting functions.

Computing $\mathbf{A_k}^T\mathbf{A_k}$ and $\mathbf{A_k}^T\mathbf{b_k}$ could be potentially very time consuming, especially when applied in three dimensions. However, this operation can be greatly speeded up using the properties of Kronecker tensor products (Ashburner and Friston, 1999). Figure 12 shows how this can be done in two dimensions using Matlab as a form of pseudo-code.

```
alpha_k = zeros(M*N,M*N);
beta_k  = zeros(M*N,1);
weight  = P_k*(c_k^(-0.5));
img1    = weight.*F;
img2    = weight*v_k;
for j = 1:J,
   tmp     = (img1(:,j)*ones(1,M)).*D1;
   alpha_k = alpha_k + kron(D2(j,:)'*D2(j,:), tmp'*tmp);
   beta_k  = beta_k  + kron(D2(j,:)', tmp'*img2(:,j));
end;
```

FIG. 12. The algorithm for computing $\mathbf{A_k}^T\mathbf{A_k}$ (*alpha_k*) and $\mathbf{A_k}^T\mathbf{b_k}$ (*beta_k*) in two dimensions using Matlab as a pseudo-code. The symbol "*" refers to matrix multiplication, whereas ".*" refers to element-by-element multiplication. "'" refers to a matrix transpose and "^" to a power. The jth row of matrix "*D2*" is denoted by "*D2*(j, :)", and the jth column of matrix "*img2*" is denoted by "*img2*(:, j)". The functions "*zeros*(a, b)" and "*ones*(a, b)" would produce matrices of size a × b of either all 0 or all 1. A Kronecker tensor product of two matrices is represented by the "*kron*" function. Matrix "*F*" is the $I \times J$ nonuniformity corrected image. Matrix "*P_k*" is the $I \times J$ current estimate of the probabilities of the voxels belonging to cluster k. Matrices "*D1*" and "*D2*" contain the DCT basis functions and have dimensions $I \times M$ and $J \times N$. "*v_k*" and "*c_k*" are scalers and refer to the mean and variance of the kth cluster.

The prior distribution of the coefficients is based on the cost function described above. For coefficients $\mathbf{x}$ this cost function is computed from $\mathbf{x}^T\mathbf{C}_0^{-1}\mathbf{x}$, where (in two dimensions),

$$\begin{aligned}\mathrm{C}_0^{-1} = {} & \lambda(\mathbf{D}''^T_2\mathbf{D}''_2) \otimes (\mathbf{D}_1^T\mathbf{D}_1) + 3\lambda(\mathbf{D}''^T_2\mathbf{D}''_2) \otimes (\mathbf{D}'^T_1\mathbf{D}'_1) \\ & + 3\lambda(\mathbf{D}'^T_2\mathbf{D}'_2 \otimes (\mathbf{D}''^T_1\mathbf{D}'_1) + \lambda(\mathbf{D}_2^T\mathbf{D}_2) \otimes (\mathbf{D}'''^T_1\mathbf{D}'''_1),\end{aligned}$$

where the notations $\mathbf{D}'_1$, $\mathbf{D}''_1$, and $\mathbf{D}'''_1$ refer to the first, second, and third derivatives (by differentiating Eq. (1) with respect to i) of $\mathbf{D}_1$, and λ is a regularization constant. The mean of the *a priori* distribution of the DCT coefficients is such that it would generate a field that is uniformly 1. For this, all the elements of the mean vector are set to 0, apart from the first element that is set to $\sqrt{MN}$.

Finally, once the coefficients have been estimated, then the modulation field $\mathbf{u}$ can be computed from the estimated coefficients ($\mathbf{x}$) and the basis functions ($\mathbf{D}_1$ and $\mathbf{D}_2$):

$$u_{ij} = \sum_{n=1}^{N}\sum_{m=1}^{M} d_{2jn}x_{mn}d_{1im} \quad \text{over } i = 1..I \text{ and } j = 1..J.$$

The new estimates for the sensitivity-corrected images are then obtained by a simple element-by-element multiplication with the modulation field:

$$g_{ij} = f_{ij}u_{ij} \qquad \text{over } i = 1..I \text{ and } j = 1..J.$$

ACKNOWLEDGMENTS

Many thanks for discussions with John Sled and Alex Zijdenbos at McGill University who (back in 1996) provided the original inspiration for the image nonuniformity correction method described in the Appendix. The idea led on from work by Alex Zijdenbos on estimating nonuniformity from white matter in the brain images. Thanks also to Keith Worsley for further explaining the work of Jon Taylor, Chris Cocosco for providing information on the MRI simulator, Peter Neelin and Kate Watkins for information about the ICBM probability maps, and Tina Good and Ingrid Johnsrude for the data used in the evaluations. This work was supported by the Wellcome Trust. Most of the software for the methods described in this paper are available from the authors as part of the SPM99 package.

REFERENCES

Abell, F., Krams, M., Ashburner, J., Passingham, R. E., Friston, K. J., Frackowiak, R. S. J., Happé, F., Frith, C. D., and Frith, U. 1999. The neuroanatomy of autism: A voxel based whole brain analysis of structural scans. *NeuroReport* **10:**1647–1651.

Ashburner, J., and Friston, K. J. 1997. Multimodal image coregistration and partitioning—A unified framework. *NeuroImage* **6:**209–217.

Ashburner, J., and Friston, K. J. 1999. Nonlinear spatial normalization using basis functions. *Hum. Brain Mapp.* **7:**254–266.

Ashburner, J., Neelin, P., Collins, D. L., Evans, A. C., and Friston, K. J. 1997. Incorporating prior knowledge into image registration. *NeuroImage* **6:**344–352.

Ashburner, J., Hutton, C., Frackowiak, R. S. J., Johnsrude, I., Price, C., and Friston, K. J. 1998. Identifying global anatomical differences: Deformation-based morphometry. *Hum. Brain Mapp.* **6:**348–357.

Bookstein, F. L. 1997a. Landmark methods for forms without landmarks: Morphometrics of group differences in outline shape. *Med. Image Anal.* **1:**225–243.

Bookstein, F. L. 1997b. Quadratic variation of deformations. In *Information Processing in Medical Imaging* (J. Duncan and G. Gindi, Eds.), pp. 15–28. Springer-Verlag, Berlin/Heidelberg/New York.

Bookstein, F. L. 1999. *Brain Warping,* Chap. 10, pp. 157–182. Academic Press, San Diego.

Cao, J., and Worsley, K. J. 1999. The geometry of the Hotelling's T^2 random field with applications to the detection of shape changes. *Ann. Stat.* **27:**925–942.

Cocosco, C. A., Kollokian, V., Kwan, R. K.-S., and Evans, A. C. 1997. Brainweb: Online interface to a 3D MRI simulated brain database. *NeuroImage* **5:**S425.

Collins, D. L., Zijdenbos, A. P., Kollokian, V., Sled, J. G., Kabani, N. J., Holmes, C. J., and Evans, A. C. 1998. Design and construction of a realistic digital brain phantom. *IEEE Trans. Med. Imag.* **17:**463–468.

Evans, A. C., Collins, D. L., and Milner, B. 1992. An MRI-based stereotactic atlas from 250 young normal subjects. *Soc. Neurosci. Abstr.* **18:**408.

Evans, A. C., Collins, D. L., Mills, S. R., Brown, E. D., Kelly, R. L., and Peters, T. M. 1993. 3D statistical neuroanatomical models from 305 MRI volumes. In *Proc. IEEE—Nuclear Science Symposium and Medical Imaging Conference,* pp. 1813–1817.

Evans, A. C., Kamber, M., Collins, D. L., and Macdonald, D. 1994. An MRI-based probabilistic atlas of neuroanatomy. In *Magnetic Resonance Scanning and Epilepsy* (S. Shorvon, D. Fish, F. Andermann, G. M. Bydder, and H. Stefan, Eds.), NATO ASI Series A, Life Sciences, Vol. 264. pp. 263–274. Plenum, New York.

Freeborough, P. A., and Fox, N. C. 1998. Modelling brain deformations in Alzheimer disease by fluid registration of serial MR images. *J. Comput. Assisted Tomogr.* **22:**838–843.

Friston, K. J., Holmes, A. P., Poline, J.-B., Price, C. J., and Frith, C. D. 1995a. Detecting activations in PET and fMRI: Levels of inference and power. *NeuroImage* **4:**223–235.

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J.-B., Frith, C. D., and Frackowiak, R. S. J. 1995b. Statistical parametric maps in functional imaging: A general linear approach. *Hum. Brain Mapp.* **2:**189–210.

Gaser, C., Volz, H.-P., Kiebel, S., Riehemann, S., and Sauer, H. 1999. Detecting structural changes in whole brain based on nonlinear deformations—Application to schizophrenia research. *NeuroImage* **10:**107–113.

Gee, J. C., and Bajcsy, R. K. 1999. *Brain Warping,* Chap. 11, pp. 183–198. Academic Press, San Diego.

Holmes, A. P., Blair, R. C., Watson, J. D. G., and Ford, I. 1996. Non-parametric analysis of statistic images from functional mapping experiments. *J. Cereb. Blood Flow Metab.* **16:**7–22.

Johnson, R. A., and Wichern, D. W. 1998. *Applied Multivariate Statistical Analysis,* 4th ed. Prentice–Hall, Upper Saddle River, NJ.

Krams, M., Quinton, R., Ashburner, J., Friston, K. J., Frackowiak, R. S., Bouloux, P. M., and Passingham, R. E. 1999. Kallmann's syndrome: Mirror movements associated with bilateral corticospinal tract hypertrophy. *Neurology* **52:**816–822.

Kwan, R. K.-S., Evans, A. C., and Pike, G. B. 1996. An extensible MRI simulator for post-processing evaluation. In *Conference on Visualisation in Biomedical Computing*, pp. 135–140.

May, A., Ashburner, J., Büchel, C., McGonigle, D. J., Friston, K. J., Frackowiak, R. S. J., and Goadsby, P. J. 1999. Correlation between structural and functional changes in brain in an idiopathic headache syndrome. *Nat. Med.* **5:**836–838.

Shah, P. J., Ebmeier, K. P., Glabus, M. F., and Goodwin, G. 1998. Cortical grey matter reductions associated with treatment-resistant chronic unipolar depression. *Br. J. Psychiatry* **172:**527–532.

Sled, J. G., Zijdenbos, A. P., and Evans, A. C. 1998. A non-parametric method for automatic correction of intensity non-uniformity in MRI data. *IEEE Trans. Med. Imag.* **17:**87–97.

Sowell, E. R., Thompson, P. M., Holmes, C. J., Batth, R., Jernigan, T. L., and Toga, A. W. 1999. Localizing age-related changes in brain structure between childhood and adolescence using statistical parametric mapping. *NeuroImage* **9:**587–597.

Taylor, J., Worsley, K. J., Zijdenbos, A. P., Paus, T., and Evans, A. C. 1998. Detecting anatomical changes using logistic regression of structure masks. *NeuroImage* **7:**S753.

Thompson, P. M., and Toga, A. W. 1999. *Brain Warping*, Chap. 18, pp. 311–336. Academic Press, San Diego.

Van Leemput, K., Maes, F., Vandermeulen, D., and Suetens, P. 1999a. Automated model-based bias field correction of MR images of the brain. *IEEE Trans. Med. Imag.* **18:**885–896.

Van Leemput, K., Maes, F., Vandermeulen, D., and Suetens, P. 1999b. Automated model-based tissue classification of MR images of the brain. *IEEE Trans. Med. Imag.* **18:**897–908.

Vandermeulen, D., Descombes, X., Suetens, P., and Marchal, G. 1996. Unsupervised regularized classification of multi-spectral MRI. In *Proceedings of the International Conference on Visualization in Biomedical Computing*, pp. 229–234.

Vargha-Khadem, F., Watkins, K. E., Price, C. J., Ashburner, J., Alcock, K. J., Connelly, A., Frackowiak, R. S. J., Friston, K. J., Pembrey, M. E., Mishkin, M., Gadian, D. G., and Passingham, R. E. 1998. Neural basis of an inherited speech and language disorder. *Proc. Natl. Acad. Sci. USA* **95,** 12695–12700.

Woermann, F. G., Free, S. L., Koepp, M. J., Ashburner, J., and Duncan, J. D. 1999. Voxel-by-voxel comparison of automatically segmented cerebral grey matter—A rater-independent comparison of structural MRI in patients with epilepsy. *NeuroImage* **10:**373–384.

Worsley, K. J., Marrett, S., Neelin, P., Vandal, A. C., Friston, J. J., and Evans, A. C. 1996. A unified statistical approach for determining significant voxels in images of cerebral activation. *Hum. Brain Mapp.* **4:**58–73.

Worsley, K. J., Andermann, M., Koulis, T., MacDonald, D., and Evans, A. C. 1999. Detecting changes in non-isotropic images. *Hum. Brain Mapp.* **8:**98–101.

Wright, I. C., McGuire, P. K., Poline, J.-B., Travere, J. M., Murray, R. M., Frith, C. D., Frackowiak, R. S. J., and Friston, K. J. 1995. A voxel-based method for the statistical analysis of gray and white matter density applied to schizophrenia. *NeuroImage* **2:**244–252.

Wright, I. C., Ellison, Z. R., Sharma, T., Friston, K. J., Murray, R. M., and Mcguire, P. K. 1999. Mapping of grey matter changes in schizophrenia. *Schizophrenia Res.* **35:**1–14.