Chapter 6

Morphometry

Contents

6.1 Introduction

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. The 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 observable structural differences may be overlooked. This chapter describes morphometric approaches that are not biased to one particular structure or tissue and give even-handed and comprehensive assessments of anatomical differences throughout the brain.

Morphometric methods have a number of different aims. They can be used for localising significant structural differences among populations, or for showing that overall brain structure is related to some effect of interest. When testing the overall brain structure, multivariate statistical methods are used to analyse groups of parameters for the whole brain (e.g., the deformation-based morphometry described in this chapter). The result of the forms of morphometry that localise structural differences would typically be a statistical parametric map (Friston *et al.*, 1995d) of regional differences. Statistical parametric maps (SPMs) can be derived from uni-variate data where there is a single variable at each voxel (e.g., the voxel-based morphometry method described here), or from multi-variate data, where there are several different variables at each voxel (e.g., tensor-based morphometry).

Another use for morphometric methods is for characterising essential differences, or for producing some form of classification. Linear methods such as canonical correlation analysis, or nonlinear classification methods can be used for these purposes. This chapter will be restricted to simple linear classification methods, which model data as multivariate normal distributions. Nonlinear classification methods can assume more complex distributions for the data, but they tend to be much more computationally expensive.

Statistical tests generally involve disproving a null hypothesis with a particular level of confidence. In morphometry, the null hypothesis is usually that there are no significant structural differences among a number of populations, or due to particular covariates, such as age. The objective of the tests is to demonstrate improbability of the null hypothesis. For example, if p values of less than 0.01 are deemed to be significant, then false positive results would be expected only about once out of a hundred tests.

SPMs of univariate statistical measures often allow relatively simple questions to be addressed, such as where is there significantly more of a particular measure, that happens to correlate with a particular effect of interest. Standard parametric statistical procedures (t-tests and F-tests) can be used to test the hypotheses within the framework of the general linear model (LM), whereby a vector of observations is modelled by a linear combination of user specified regressors (Friston *et al.*, 1995d). The LM is a flexible framework that allows many different tests to be applied, ranging from group comparisons and identifying differences that are related to specified covariates such as disease severity or age, to complex interactions between different effects of interest.

Performing comparisons at each voxel results in many statistical tests being performed. Without any correction, the number of false-positive results would be proportional to the number of independent tests. A Bonferroni correction would be applied if the tests were independent, but this is not normally the case because of the inherent spatial smoothness of the data. In practice, the effective number of independent statistical tests is determined using Gaussian Random Field (GRF) theory (Friston *et al.*, 1995b; Worsley *et al.*, 1996). By using GRF theory, a correction for multiple dependent comparisons can be made to produce the appropriate rate of false-positive results.

SPMs can also be obtained from the results of voxel-wise multi-variate tests. Instead of one image per subject, multivariate tests effectively involve two or more images, and use a statistic such as Wilk's Λ. Following the tests, similar corrections based on GRF theory can be applied as in the univariate case. GRF theory has not yet been worked out for Wilk's Λ fields, so approximations are made that involve transforming the resulting Wilk's Λ fields to random fields

of other statistics, such as χ^2 or F fields. Subsequent processing assumes that the transformed fields have the same properties as true χ^2 or F fields.

There now follows a description of how multivariate statistical tests can be performed using Multivariate Analysis of Covariance. These can be used to produce SPMs that localise structural differences, or can be applied to sets of parameters describing whole brains. This is followed by a brief description of how differences between groups can be characterised by multivariate methods. The remainder of the chapter will describe three different morphometric methods: voxel-based, deformation-based and tensor-based morphometry.

6.1.1 Multivariate Analysis of Covariance

A multivariate analysis of covariance (MANCOVA) assumes that there are several dependent variables for each observation, where an observation refers to a collection of data for a subject. This data can be represented by an $M \times I$ matrix $\mathbf{X}$, where M is the number of subjects in the analysis, and I is the number of variables for each subject.

The columns of $\mathbf{X}$ are modelled by linear combination of basis functions. Some of these basis functions represent effects that are not considered interesting in the study, but may still be significant. For example, linear age effects may confound a study of handedness. If a left-handed group is not perfectly age matched with a right-handed group, then differences due to age could be attributed to handedness differences. In addition, the inclusion of confounding effects (such as age) in a model can result in a better fit to the data, and possibly make the test more sensitive to the effects of interest.

Confounding effects are modelled by an $M \times K$ design matrix $\mathbf{G}$. Each column of $\mathbf{G}$ can be a vector of covariates (e.g., the age of each subject), or alternatively can be arranged in blocks (e.g., there may be a column containing ones for left-handed subjects, and zeros for right-handed). In almost all cases, a column of ones is included in order to model a mean effect. First of all, any variance in the data that could be attributed to the confounds is removed by:

$$\mathbf{X_a} = \mathbf{X} - \mathbf{G}\left(\mathbf{G}^T\mathbf{G}\right)^{-1}\mathbf{G}^T\mathbf{X} \tag{6.1}$$

Similarly, the effects of interest are modelled by an $M \times J$ design matrix $\mathbf{Y}$. The columns of $\mathbf{Y}$ can represent group memberships, where elements contain a one for subjects who are in a particular group, or a zero if they do not. Alternatively, the columns may contain covariates of interest such as disease severity for each subject. The columns in this design matrix are orthogonalised with respect to matrix $\mathbf{G}$:

$$\mathbf{Y_a} = \mathbf{Y} - \mathbf{G}\left(\mathbf{G}^T\mathbf{G}\right)^{-1}\mathbf{G}^T\mathbf{Y} \tag{6.2}$$

A MANCOVA involves assessing how the predictability of the observations change when the effects of interest are discounted. This is based on the distributions of the residuals, which are assumed to be multinormal. The statistic is related to the determinants of the covariance matrices describing these distributions. In practice, the residual sum of squares and products (SSP) matrix ($\mathbf{W}$), is compared to the SSP matrix of the fitted effects ($\mathbf{B}$). These matrices are obtained by:

$$\mathbf{T} = \quad \mathbf{Y_a}(\mathbf{Y_a}^T\mathbf{Y_a})^{-1}\mathbf{Y_a}^T\mathbf{X_a}$$

$$\begin{aligned} \mathbf{B} &= \mathbf{T}^T\mathbf{T} \\ \mathbf{W} &= (\mathbf{X_a} - \mathbf{T})^T (\mathbf{X_a} - \mathbf{T}) \end{aligned} \tag{6.3}$$

The statistic is called *Wilk's Lambda* (Λ), and is based upon the ratios of the determinants:

$$\Lambda = \frac{|\mathbf{W}|}{|\mathbf{B} + \mathbf{W}|} \tag{6.4}$$

The Wilk's Lambda statistic can range between zero and one, where a value of one suggests no relationship between the effects of interest and the data, and a value of zero indicates a perfect relationship. This statistic is transformed to a χ^2 statistic (with IJ degrees of freedom under the null hypothesis) using the approximation of Bartlett [1]:

$$\chi^2 \approx -(M - K - (I + J + 1)/2) log_e(\Lambda) \tag{6.5}$$

Finally, the cumulative χ^2 distribution function is used to make inferences about whether the null hypothesis (that there is no difference between the distributions) can be rejected.

This multivariate approach fails when the number of variables approaches the number of subjects ($M - K - (I + J + 1)/2$ approaches zero, or becomes negative). In many situations, it is necessary to regularise the problem by reducing the number of variables with respect to the number of subjects. One way of doing this involves using singular value decomposition to decompose the original data matrix $\mathbf{X}$ into unitary matrices $\mathbf{U}$ and $\mathbf{V}$, and diagonal matrix $\mathbf{S}$, such that $\mathbf{X} = \mathbf{USV}^T$. The diagonal elements of $\mathbf{S}$ are arranged in decreasing importance, so it is possible to reconstruct an approximation of $\mathbf{X}$ using only the first L diagonal elements of $\mathbf{S}$ and the first L columns of $\mathbf{U}$ and $\mathbf{V}$, such that $\mathbf{X} \simeq \mathbf{U}^*\mathbf{S}^*\mathbf{V}^{*T}$. The MANCOVA would be performed using $\mathbf{U}^*$ (the first L columns of $\mathbf{U}$).

6.1.2 Canonical Correlation Analysis

Canonical correlation analysis (CCA) is used to measure the strength of association between two sets of variables. In this case, the variables are $\mathbf{X_a}$ and $\mathbf{Y_a}$, which are the data and design matrix from Section 6.1.1 after having been orthogonalised with respect to a set of confounding effects. The first canonical variate pair is the linear combination of columns of $\mathbf{X_a}$ and the linear combination of columns of $\mathbf{Y_a}$, that has the maximum correlation. The second canonical variate pair consists of linear combinations that maximise the correlation subject to the constraint that they are orthogonal to the first pair of canonical variables. Similarly, all subsequent pairs maximise the correlations and are orthogonal to all the previous pairs.

The weights used to determine the linear combinations are derived from the unitary matrices ($\mathbf{U}$ and $\mathbf{V}$) obtained by singular value decomposition:

$$\mathbf{USV}^T = \left(\mathbf{X_a}^T\mathbf{X_a}\right)^{-\frac{1}{2}} \left(\mathbf{X_a}^T\mathbf{Y_a}\right) \left(\mathbf{Y_a}^T\mathbf{Y_a}\right)^{-\frac{1}{2}} \tag{6.6}$$

Then the weights ($\mathbf{A}$ and $\mathbf{B}$) are derived by:

$$\mathbf{A} = \left(\mathbf{X_a}^T\mathbf{X_a}\right)^{-\frac{1}{2}} \mathbf{U} \text{ and } \mathbf{B} = \left(\mathbf{Y_a}^T\mathbf{Y_a}\right)^{-\frac{1}{2}} \mathbf{V} \tag{6.7}$$

The canonical variate pairs are obtained from the columns of $\mathbf{X_aA}$ and $\mathbf{Y_aB}$.

[1] This is true providing the matrices do not contain any redundant columns, otherwise matrix pseudo-inverses are required in the computations, and the numbers of columns replaced by the matrix ranks.

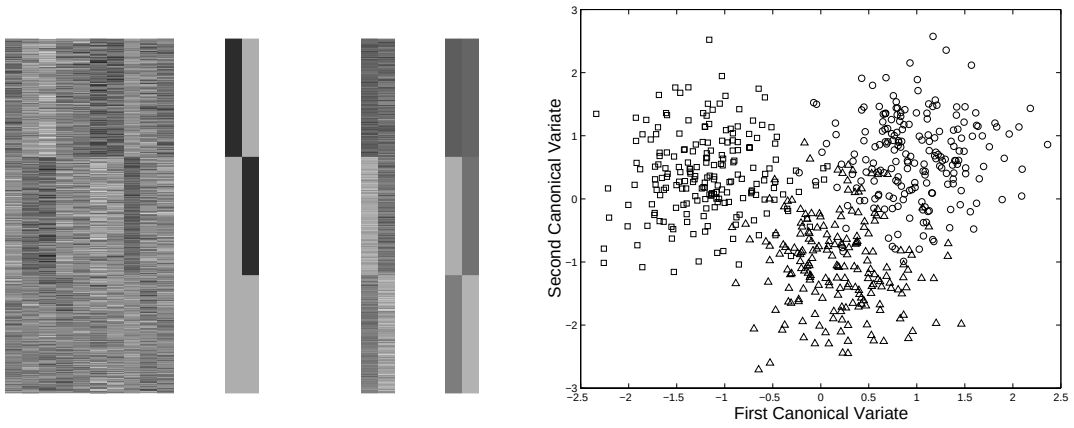

Figure 6.1: This figure illustrates canonical correlation analysis using simulated data. The two matrices on the left are a 600×10 data matrix ($\mathbf{X_a}$) and a 600×2 design matrix ($\mathbf{Y_a}$) after centering by orthogonalising with respect to a column of ones as described in Equations 6.1 and 6.2. The design matrix represents a partitioning of the data into three groups. The canonical variates for $\mathbf{X_a}$ and $\mathbf{Y_a}$ are shown in the third and fourth matrices ($\mathbf{X_aA}$ and $\mathbf{Y_aB}$). The left-hand columns of these two matrices contain the first canonical variate pair, whereas the second pair are in the right-hand columns. The graph on the right shows the two columns of $\mathbf{X_aA}$ plotted against each other, where the different symbols used represent memberships of the three different groups.

When the number of variables approaches the number of observations, the problem needs to be regularised. This can be done by computing the canonical variates from data that has been compacted using singular value decomposition as in Section 6.1.1. If $\mathbf{X}$ has been decomposed such that it can be approximated by $\mathbf{X} \simeq \mathbf{U^*S^*V^{*T}}$, and canonical correlation analysis performed on $\mathbf{U^*}$ and $\mathbf{Y}$ to give weight matrices $\mathbf{A^*}$ and $\mathbf{B^*}$, then the weights to be applied to the original data ($\mathbf{X}$) can be reconstructed by $\mathbf{V^*S^{*-1}A^*}$.

Figure 6.1 shows CCA as it would be used to graphically describe the differential features of three groups. However, it can also be used to aid classification. Once derived, the same weighting matrix ($\mathbf{A}$) can be applied to new data-sets that were not involved in its derivation. If there are only two groups involved in a study, then CCA can be used directly to assign group memberships to new observations[2]. With more groups, CCA serves as a graphical aid for assigning new observations to the groups.

6.2 Voxel-Based Morphometry

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), or a related method that involves parcelling the images into different regions (Goldszal *et al.*, 1998). This section summarises the procedures involved in VBM, and provides evaluations of its statistical component.

[2]Providing that the signs of the canonical variates are adjusted accordingly.

6.2.1 Methods

Voxel-based morphometry of MRI data involves spatially normalising all the the subjects' images to the same stereotactic space, extracting the grey matter from the normalised images, smoothing, and finally performing a statistical analysis to localise, and make inferences about, group differences. The output from the method is an SPM showing regions where grey matter concentration differs significantly among the groups.

Spatial Normalisation

Spatial normalisation involves transforming all subjects' data to the same stereotactic space. This is achieved by registering each image to the same template image by minimising the residual sum of squared differences between them according to the methods described in Chapter 3. It should be noted that this method of spatial normalisation does not attempt to match every cortical feature exactly, but merely accommodates differences in global brain shape. If the spatial normalisation was perfectly exact, then all the segmented images would appear identical and no significant differences could be detected: VBM tries to detect differences in the regional concentration of grey 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 template image does not bias the final solution. An ideal template consists of the average of a large number of MR images that have been registered to within the accuracy of the spatial normalisation technique. The spatially normalised images should have a relatively high resolution (1mm or 1.5mm isotropic voxels), so that the grey matter extraction method is not excessively confounded by partial volume effects, where voxels contain a mixture of different tissue types.

Image Partitioning

The spatially normalised images are then partitioned into grey matter (GM), white matter (WM), cerebro-spinal fluid (CSF), and three other background classes, using the tissue classification technique described in Chapter 5.

Smoothing the Grey Matter Segments

The grey 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 grey 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 "grey matter density", but should not be confused with cell packing density measured cytoarchitectonically. It will be referred to as "concentration" to avoid confusion. By the central limit theorem, smoothing also has the effect of rendering the data more normally distributed, thus 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 normalisation.

Logit Transform (optional)

In effect, each voxel in the smoothed image segments represents the local GM concentration (between zero and one). 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:

$$logit(p) = \log_e \left(\frac{p}{1-p} \right) \tag{6.8}$$

For concentrations very close to either one or zero, 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 other extreme. An alternative method of analysis involves using logistic regression within the framework of the general*ised* linear model (GLM) (Taylor *et al.*, 1998), but this is beyond the scope of this chapter as it requires iterative re-weighted least-squares methods. Whether or not the logit transform is a necessary processing step for VBM will be addressed later.

Statistical Analysis

Statistical analysis using the general linear model (LM) is used to identify regions of grey matter concentration that are significantly related to the particular effects under study (Friston *et al.*, 1995d). The LM is a flexible framework that allows many different tests to be applied, ranging from group comparisons and identification of regions of grey 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 zero and one, where most of the values are very close to either of the extremes. Only by smoothing the segmented images does the behaviour of the residuals become more normally distributed.

Following application of the LM, the significance of any differences is ascertained using GRF theory (Friston *et al.*, 1995b; Worsley *et al.*, 1996). A voxel-wise statistical parametric map (SPM) comprises the result of many statistical tests, so it is necessary to correct for multiple dependent comparisons.

6.2.2 Evaluations

A number of assumptions need to hold in order for VBM to be useful. Confounding effects must be eliminated or modelled as far as possible. For example, it is not appropriate to compare two different groups if the images are acquired on 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. An issue of validity concerns the assumptions required by the statistical tests. For parametric tests, it is

often important that the data are normally distributed. If the data are not well behaved, then it is useful to know what effects there may be on the statistical tests. If there is doubt about the validity of the assumptions, it is better to use a non-parametric statistical analysis (Holmes *et al.*, 1996; Bullmore *et al.*, 1999).

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 prove that data are normally distributed – they 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 in 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 applying the appropriate corrections derived from residual forming matrix) into increasing order $(x_1, x_2, \ldots x_J)$. The inverse cumulative distribution of each of the J elements is then computed as:

$$q_j = \sqrt{2} erf^{-1}\left(2\frac{j-\frac{3}{8}}{J+\frac{1}{4}} - 1\right) \tag{6.9}$$

where erf^{-1} 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 identify 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 work, the correlation coefficient is simply used as a "normality statistic", and its distribution is examined over voxels.

The data used to test the assumptions were T1 weighted MRI scans of 50 normal male right handed subjects aged between 17 and 62 (median 26, mean 29). The structural scans had been acquired as part of an ongoing programme of functional imaging research. The scans were performed on a Siemens MAGNETOM Vision scanner operating at 2 Tesla. An MPRAGE sequence was used with a 12^o tip angle, 9.7ms repeat time, 4ms echo time and 0.6ms inversion time, to generate sagittal images of the whole brain with voxel sizes of $1\times1\times1.5$mm. The images were spatially normalised, segmented and smoothed using a Gaussian kernel of 12mm full width at half maximum (FWHM).

Voxel-by-voxel correlation coefficients of the Q-Q plots were computed over all voxels where 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 a VBM analysis. This is because it is known that these low intensity voxels are most likely to deviate 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 modelled the confounding effect of the total amount of grey 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 (717191 voxels), along with histograms generated from simulated Gaussian noise. These are plotted in Figure 6.2, and show

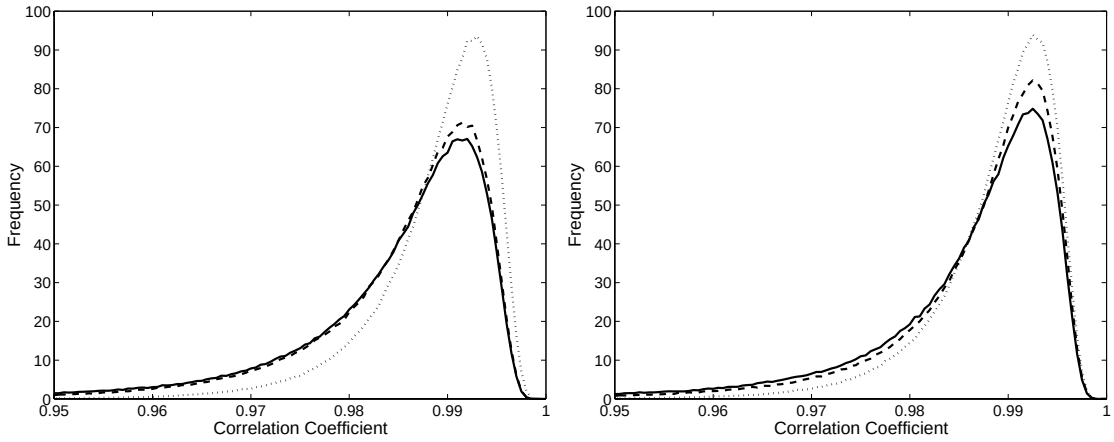

Figure 6.2: Histogram of correlation coefficients taken over the whole image volumes (using a total of 717191 voxels where the mean intensity over all images was greater than 0.05). The dotted line is the histogram that would be expected if the data were perfectly normally distributed. The solid line shows the histogram of the data without the logit transform, and the dashed line shows the histogram obtained using the logit transformed data. The plot on the left is based on the model that does not include global grey matter as a confound, whereas that on the right does model this confounding effect.

that the data does deviate slightly from a normal distribution. The logit transform appeared to make the residuals slightly more normally distributed. The normality of the residuals was also improved by modelling the total amount of grey matter as a confounding effect.

Testing the Rate of False Positives using Randomisation

The previous section showed that the data are not quite normally distributed, but it does not show how the non-normality influences any subsequent statistics. Ultimately, we wish to protect against false-positive results, and in this section, the frequency with which they arise is tested (Bullmore *et al.*, 1999). The statistics were evaluated using the same pre-processed 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 SPM99 (Wellcome Department of Cognitive Neurology, London, UK) to compare the groups. The particular 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 grey matter concentration of the smaller group. The end result is a series of 100 statistical parametric parametric maps of the t-statistic.

Within each of these SPMs, the local maxima of the t-statistic fields were corrected for the number of dependent tests performed, and a p value assigned to each (Friston *et al.*, 1995b; Worsley *et al.*, 1996). Using a corrected threshold[3] of p=0.05, one would expect about five

[3] A sufficiently smooth random field can be thresholded to reveal a number of distinct blobs, (more technically known as connected components of the excursion set, and often referred to as clusters). The number of clusters in the field minus the number of holes is the Euler characteristic of the excursion set. For high thresholds, the

clusters containing p values below this threshold by chance alone. Over the 100 SPMs, there were six clusters containing 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 modelling the total amount of grey matter as a confounding effect. The results of this analysis produced four significant clusters 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 SPM99 is based on the number of connected voxels in a cluster defined by a pre-specified threshold (extent statistic) (Friston *et al.*, 1995b). 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 neuro-anatomy. As noted by Worsley (1999), this nonstationary smoothness leads to inexact p values:

> "The reason is simple: 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 five significant clusters would be expected from the 100 SPMs if the smoothness was stationary. Eighteen significant clusters were found when the total amount of grey matter was not modelled as a confound, and fourteen significant clusters were obtained when it was. Assuming Poisson statistics, the probabilities of obtaining this many significant clusters, under the null hypothesis that the extent statistic is valid, are 0.000001 and 0.000226 respectively. 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's t distribution. This was verified using the computed t-fields, where each t-field contains 717191 voxels. Plots of the resulting histograms are shown in Figure 6.3. The top row presents distributions when global differences in grey matter are 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 pre-processed brain images, but used the total count of grey matter in the brains as a confound. Each subject was randomly assigning 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 10000 times, and Figure 6.4 shows an image of the number of false positives occurring

number of holes becomes negligible, so the Euler characteristic is approximately equal to the number of clusters. Therefore, it is possible to control the expected number of false-positive clusters arising in an SPM by choosing an appropriate threshold. For example, a particular threshold on the random field could predict five false positive clusters in every 100 fields. The threshold can therefore be equated with a particular "corrected probability value" of 0.05. Similarly, any reasonably high value in the field can have a corrected p value associated with it.

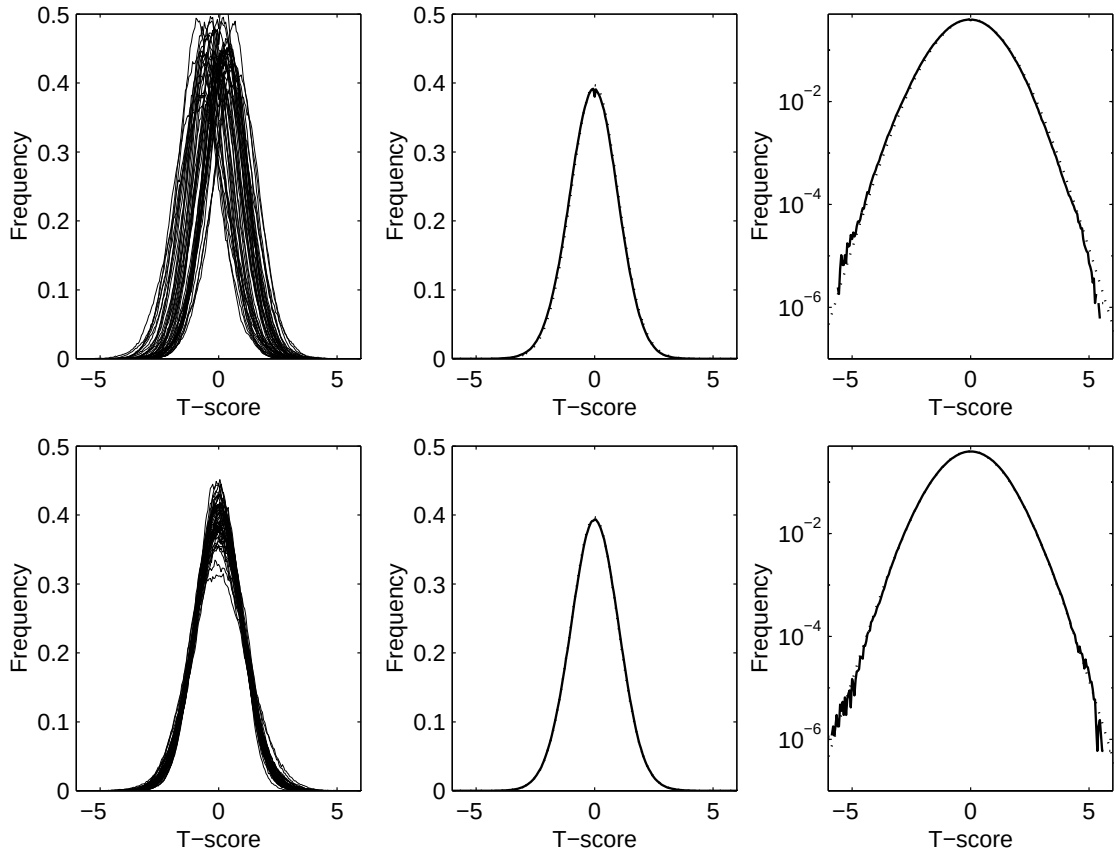

Figure 6.3: Histograms of t-scores from randomly generated tests. Above: Not modelling mean effect (48 degrees of freedom). Below: Modeling a mean effect as a confound (47 degrees of freedom). Left: 50 histograms of t-scores testing randomly generated effects of interest. Centre: the mean (i.e., cumulative distribution over all voxels and volumes) of the 50 histograms is plotted as a solid line, and the probability density function of the Students t distribution for 47/48 degrees of freedom is shown by the dotted line. Right: The same as centre, except plotted on a logarithmic scale.

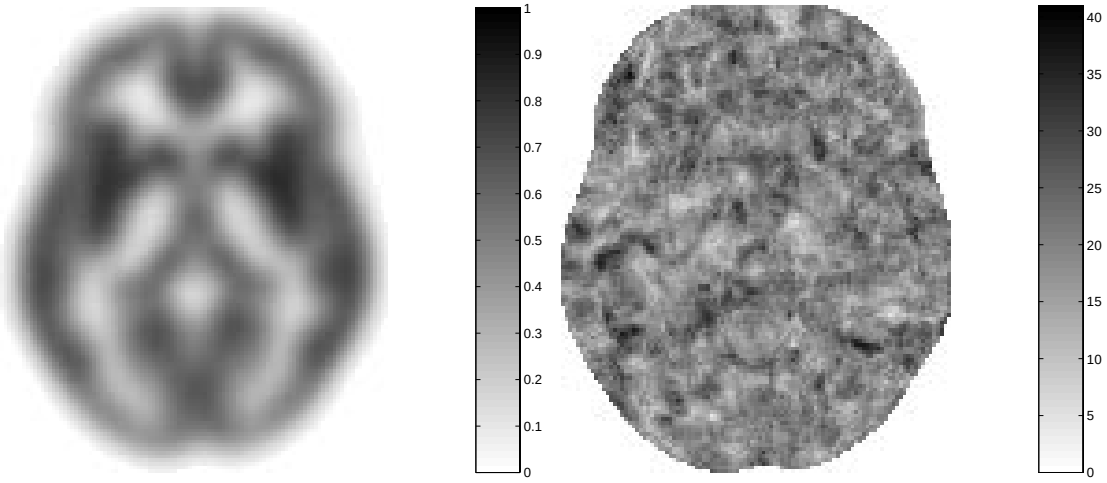

Figure 6.4: Left: Mean of 50 subjects pre-processed brain images. Right: Number of false positives occurring at each voxel at the uncorrected 0.002 level, after 10000 randomisations.

at each of the 10693 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 (10000×0.002). An average of 20.171 false positives were found, showing that the validity of statistical tests based on uncorrected t statistics are not severely compromised.

The results from randomisation have involved comparisons between two groups of subjects. Further evaluations are required in order to assess the rate of false positive results arising from comparisons between single subjects and whole groups. It is expected that non-normally distributed errors may have a much more detrimentel effect on these types of parametric tests (see Section 6.4).

6.3 Deformation Based Morphometry

This section introduces a different technique for characterising differences among structural or anatomical brain images. The anatomical differences between any two brains can be expressed at a microscopic scale (e.g., differences in cytoarchitectonics or myeloarchitectonics), at a mesoscopic scale (e.g., cortical dysplasia) or at a macroscopic level (e.g., ventricular enlargement or abnormal temporal lobe asymmetry). From the perspective of neuroimaging, differences at a mesoscopic and macroscopic level are amenable to measurement. The technique described here characterises global differences in macroscopic anatomy that complements voxel-based morphometry, allowing one to examine differences at both mesoscopic and macroscopic scales.

Deformation-based morphometry (DBM) is a characterisation of the differences in the vector fields that describe global or gross differences in brain shape. These vector fields are the deformation fields used to effect nonlinear variants of spatial normalisation, when one of the images is a template that conforms to some standard anatomical space. In what follows, the deformation fields that map a series of images onto the same template are compared to see if there are any systematic differences. Because the deformation fields are multivariate, standard multivariate statistical techniques are employed to estimate the nature of the differences and to make infer-

ences about them. The endpoint of DBM is a p value pertaining to the significance of the effect and one or more canonical vectors, or deformations, that characterise their nature. These results are obtained using multivariate analysis of covariance (MANCOVA) and canonical correlation analysis (CCA) respectively.

In order to demonstrate the technique, a study of dimorphism in relation to handedness and sex has been chosen. This should be seen as a vehicle to explain and illustrate how to do these analyses.

6.3.1 Methods

High resolution structural T1 MR images were collected from 61 normal healthy volunteers. These were all acquired on the same 2 Tesla Siemens Magnetom Vision scanner, using an MPRAGE sequence. The resolution of the images was $1mm \times 1mm \times 1.5mm$. The data consisted of 15 female right-handed subjects, 5 female left-handed subjects, 30 male right-handed subjects and 11 male left-handed subjects, all between the ages of 20 and 37. The scans were all acquired as part of ongoing functional imaging projects within the department, and all subjects had no neurological or psychiatric history.

The data were spatially normalised to the same template, and their shapes extracted from the estimated deformations. The shapes were compacted into their most significant "eigen-warps", before multi-variate statistical analysis was performed.

Spatial Normalisation

The images were spatially normalised using the methods described in Chapter 3. The template image consisted of an average of twelve 12-parameter affine registered T1 MR images of the head, and was rendered symmetric (so it could be used to examine brain asymmetry) by averaging with itself reflected across the sagittal midplane. The MRI sequence used to generate the images that constituted the template, was identical to that used for all the other images, thus ensuring that more accurate registrations could be achieved.

The first step of the normalisation was to determine the optimum 12-parameter affine transformation. Initially, the registration was performed by matching the whole of the head (including the scalp) to the template. Following this, the registration proceeded by only matching the brains together, by appropriate weighting of the template voxels (see Figure 6.5). This is a completely automated procedure that largely discounts the confounding effects of skull and scalp differences.

The affine registration was followed by estimating nonlinear deformations. Each of the deformation fields was described by 1176 parameters, where these represent the coefficients of the deformations in three orthogonal directions. The mean of the spatially normalised images is shown for each group in Figure 6.6. It can be seen that in terms of gross anatomy, following normalisation, they are virtually indistinguishable.

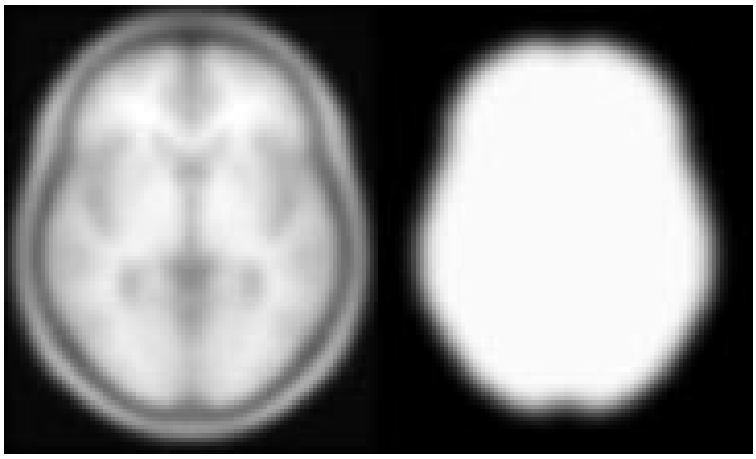

Figure 6.5: The template (left) and weighting image (right) used by the registration. Note that the images have been smoothed using an 8mm FWHM isotropic Gaussian kernel in order to facilitate the registration.

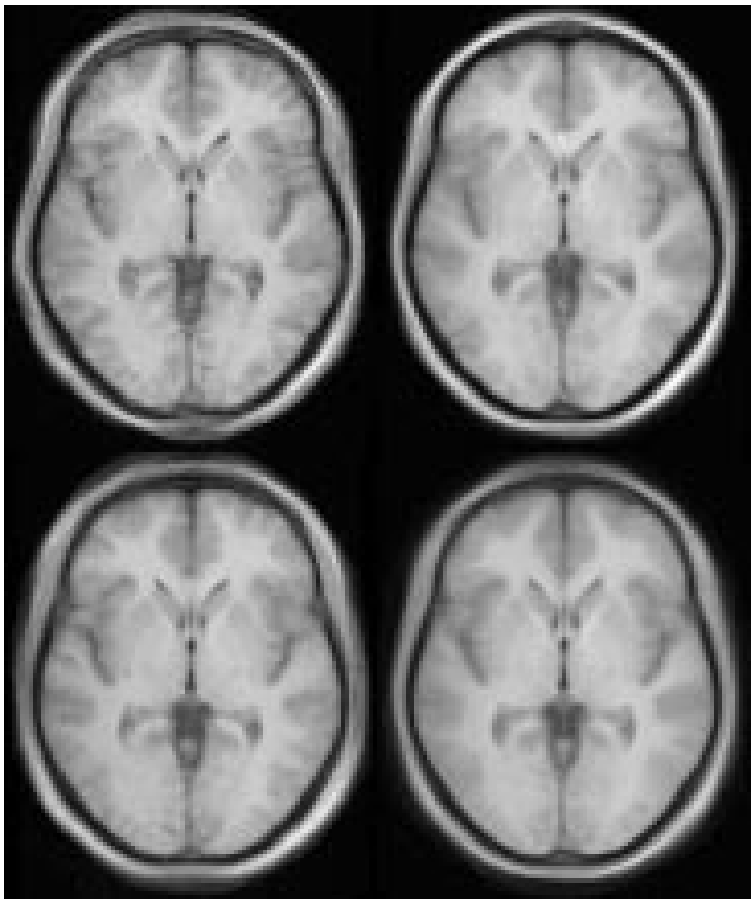

Figure 6.6: The mean of the spatially normalised images for each group: left handed females (above left), right handed females (above right), left handed males (below left) and right handed males (below right).

Extracting Shape Information

Each set of spatial normalisation parameters (affine and nonlinear) encodes a deformation field relating to the position, size and shape of the brain. For the analysis presented here, only the information relating to the shapes of the brains was used, by removing the effects of size and position.

The deformation fields are defined by both nonlinear and affine components. In order to proceed, it is necessary to decompose the transformation into components relating to position and size (uninteresting components), and shape (the components of interest). In order to effect this decomposition, each deformation field was reconstructed from its parameters. Each field provides a mapping from points in the template to points in the source image, allowing the landmark methods described by Bookstein (1997b) to be used to extract the size and positional information. The extracted measures were such that the remaining shapes minimised the squared Procrustes distance between the template and images. Rather than basing the procedure on a few landmarks, all the elements of the deformation field corresponding to voxels within the brain were considered (by weighting with the image shown in Figure 6.5). This involved first determining translations by computing centres of mass:

$$\begin{aligned}\bar{\mathbf{x}} &= \frac{\sum_{i=1}^{I} \mathbf{x}_i w_i}{\sum_{i=1}^{I} w_i} \\ \bar{\mathbf{y}} &= \frac{\sum_{i=1}^{I} \mathbf{y}_i w_i}{\sum_{i=1}^{I} w_i}\end{aligned} \tag{6.10}$$

where $\mathbf{x}_i$ is the co-ordinate of the ith voxel of the template, $\mathbf{y}_i$ is the location that it maps to, and w_i is the weighting for that element. The rotations were computed from the cross-covariance matrix ($\mathbf{C}$) between the elements and deformed elements (after removing the effects of position):

$$c_{jk} \propto \sum_{i=1}^{I} w_i (x_{ij} - \bar{x}_j)(y_{ik} - \bar{y}_k) \tag{6.11}$$

The 3×3 matrix $\mathbf{C}$ was decomposed using singular value decomposition to give three matrices, $\mathbf{U}$, $\mathbf{S}$ and $\mathbf{V}$ (such that $\mathbf{C} = \mathbf{USV}^T$, where $\mathbf{U}$ and $\mathbf{V}$ are unitary, and $\mathbf{S}$ is a diagonal matrix). The rotation matrix ($\mathbf{R}$) could then be reconstituted from these matrices by $\mathbf{R} = \mathbf{UV}^T$. Finally, moments around the centres were used to correct for relative size differences (z):

$$z = \sqrt{\frac{\sum_{j=1}^{3} \sum_{i=1}^{I} (x_{ij} - \bar{x}_j)^2 w_i}{\sum_{j=1}^{3} \sum_{i=1}^{I} (y_{ij} - \bar{y}_j)^2 w_i}} \tag{6.12}$$

After removing the effects of translation, rotation and scaling from the deformation fields, they were then reparameterised by the lowest frequency coefficients of their three dimensional DCTs.

Statistical Analysis

Following this, a data matrix was generated, where each row contained the coefficients of the nonlinear basis functions describing the difference in shape between the template and each image. For the multivariate analysis that followed, it was necessary to reduce the number of these coefficients relative to the number of images. Singular value decomposition was used to compact this information, such that about 96% of the variance of the nonlinear deformations was represented by 20 parameters for each image.

MANCOVA was used to make inferences about the effects of interest (i.e. provide p values). In the simplest case of comparing two groups, the MANCOVA becomes the special case of Hotelling's T^2 test. MANCOVA does not simply tell one what the difference is. To characterise these differences, one usually uses CCA based upon the parameters estimated by the MANCOVA. CCA is a device that finds the linear combination of the dependent variables (in this case the deformations) that is maximally correlated with the explanatory variables (e.g., male vs. female). In the simple case of one categorical explanatory variable (e.g., sex) this will be the deformation field that best discriminates between males and females. Note that this is not the same as simply subtracting the deformation fields of two groups. This is because (i) the MANCOVA includes the effects of confounds that are removed and (ii) some aspects of the deformations may be less reliable than others (CCA gives deformations that explicitly discount error in relation to predicted differences). The canonical deformations can either be displayed directly as deformation fields, or can be applied to some image to "caricature" the effect detected. In this section, both are combined in order to illustrate the deformations more clearly.

6.3.2 Results

Tests for significant differences between groups of subjects were performed using a MANCOVA on the deformation parameters. A number of tests were performed, including tests relating to the handedness of the subjects, of sexual dimorphism, looking at brain asymmetry, and interactions among these factors.

Handedness, Sex and the Interaction Between Them

A MANCOVA testing for the effects of both handedness and sex simultaneously, suggested extremely significant effects ($p = 2.1 \times 10^{-7}$). Because there were two effects of interest, CCA could be used to generate a scatter-plot representing the optimum separation of the groups (see Figure 6.7) in terms of the two corresponding canonical variates. It can be seen that the first canonical variate is mainly sensitive to the differences between men and women, whereas the second discriminates between handedness.

The effects of sex and handedness were then tested individually, both showing significant differences ($p = 0.00014$ and $p = 0.00020$ respectively). The test for sex differences used handedness as a confound and that for handedness used sex as a confound. A further test failed to show any interaction between handedness and sex ($p = 0.90$). Differences between the groups were characterised using CCA, the results of which are illustrated in Figure 6.8. These can be compared to the difference between the brain shapes (after removing confounding effects) as shown in Figure 6.9.

The effect of sex can be most clearly seen in the sagittal view and suggests that men have a more protruding occipital pole, whereas women have more prominent frontal poles. The effect of handedness involves more asymmetric differences affecting predominantly the right frontal lobe (transverse section, middle row of Figure 6.8).

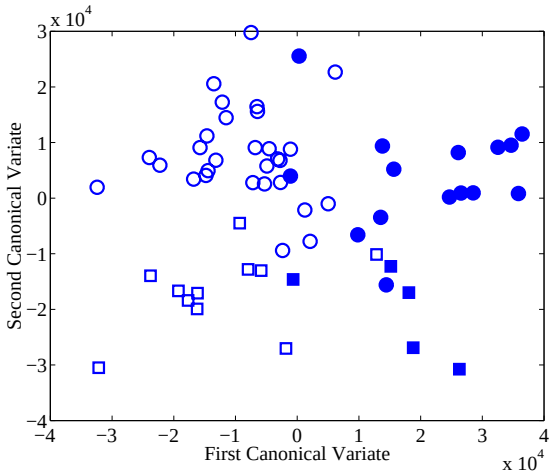

Figure 6.7: Separation of subjects using canonical correlation analysis. Right handed females (filled circles), left handed females (filled squares), right handed males (empty circles) and left handed males (empty squares).

Brain Asymmetry

Because the template used by the spatial normalisation was symmetric, it was possible to look at left/right brain asymmetry. The coefficients of the DCT can be divided into those that account for nonlinear deformations that are symmetric, and those that relate to anti-symmetric deformations. Very significant brain asymmetries were detected ($p < 0.001$) by testing that the coefficients of the anti-symmetric warps differed from zero. Geschwind & Galaburda (1987) discuss many of the asymmetries found by a number of researchers. These include the fact that the left occipital lobe is broader and longer than that on the right, which is confirmed in Figure 6.9. However, because of the large amount of variability in the occipital lobes, this is not a feature of brain asymmetry that is strongly characterised by CCA (see Figure 6.8). Another asymmetry (that was not really confirmed in Geschwind & Galaburda) is that the right frontal lobe is usually larger than that on the left. However, the results obtained here contradict this finding, in that the left frontal lobe appears to be the larger of the two. From Figure 6.9 one can see that the magnitude of the difference is relatively small, but it is still a feature that is strongly characterised by CCA. Differences in asymmetry between males and females and between left and right handed subjects were both found to be significant ($p = 0.026$ and $p = 0.0076$ respectively).

In short, reliable features of asymmetry and dimorphism may not necessarily be the biggest or most obvious. Furthermore, the approach presented here gives estimates of dimorphism that explicitly discount differences due to other factors.

6.4 Tensor-Based Morphometry

The objective of *tensor-based morphometry* (TBM) is to localise regions of shape differences among groups of brains, based on deformation fields that map points in a template (x_1, x_2, x_3) to equivalent points in individual source images (y_1, y_2, y_3). In principle, the Jacobian matrices of the deformations (a 2nd order tensor field relating to the spatial derivatives of the transformation)

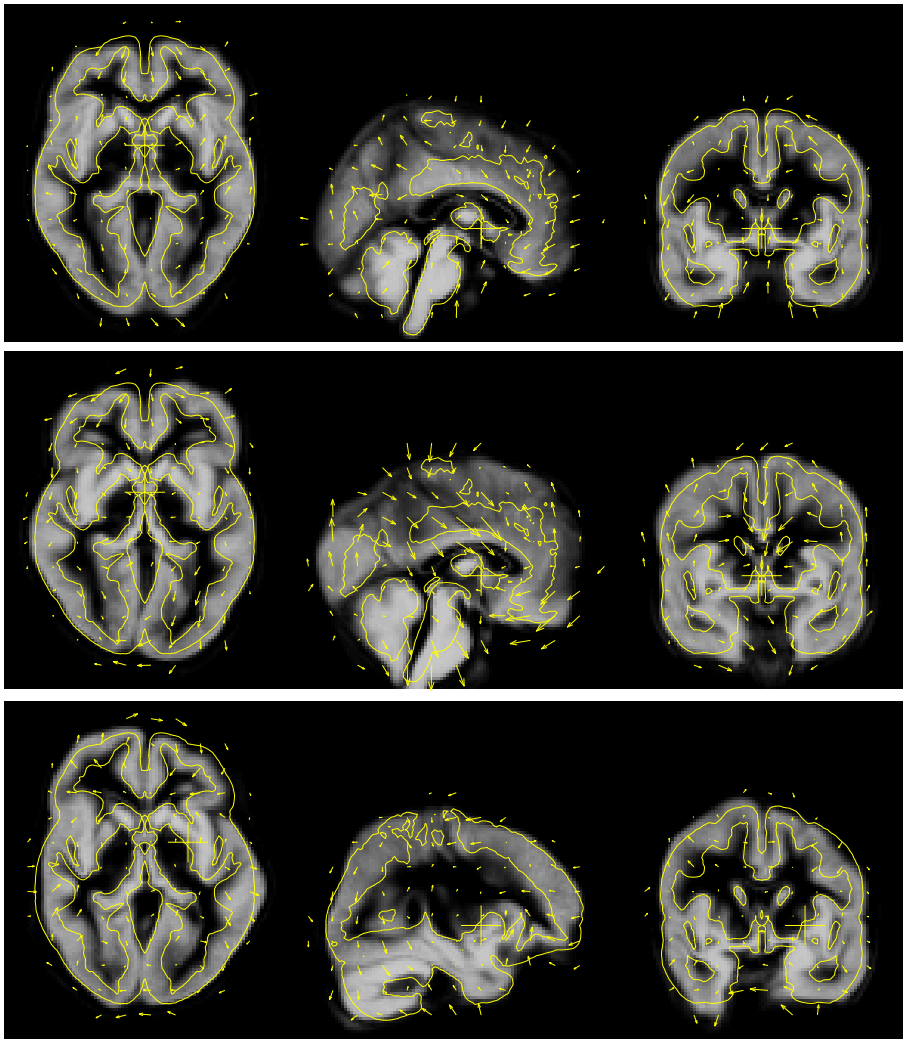

Figure 6.8: Nonlinear warps that caricature a male brain (above), a right-handed brain (centre), and natural brain asymmetry (below). These have been characterised by canonical correlation analysis. The images of grey matter show the caricature of the deformations. Superimposed on this is a contour from the undeformed image. The arrows show the direction of the nonlinear warps characterised by CCA (from undeformed to deformed). In this figure, the deformations have been arbitrarily scaled for better visualisation. These are not the mean differences between the brain shapes, but rather the differences that most clearly distinguish them. In it's most general form, CCA produces a set of vectors that best partition the data according to the design matrix. If there are multiple effects of interest, then there is no simple relationship between these effects and the canonical variates, but with only one factor of interest (as in these examples), the canonical variates can be directly related to the factor. In the transverse and coronal sections, the left side of the brain is on the left side of the figure.

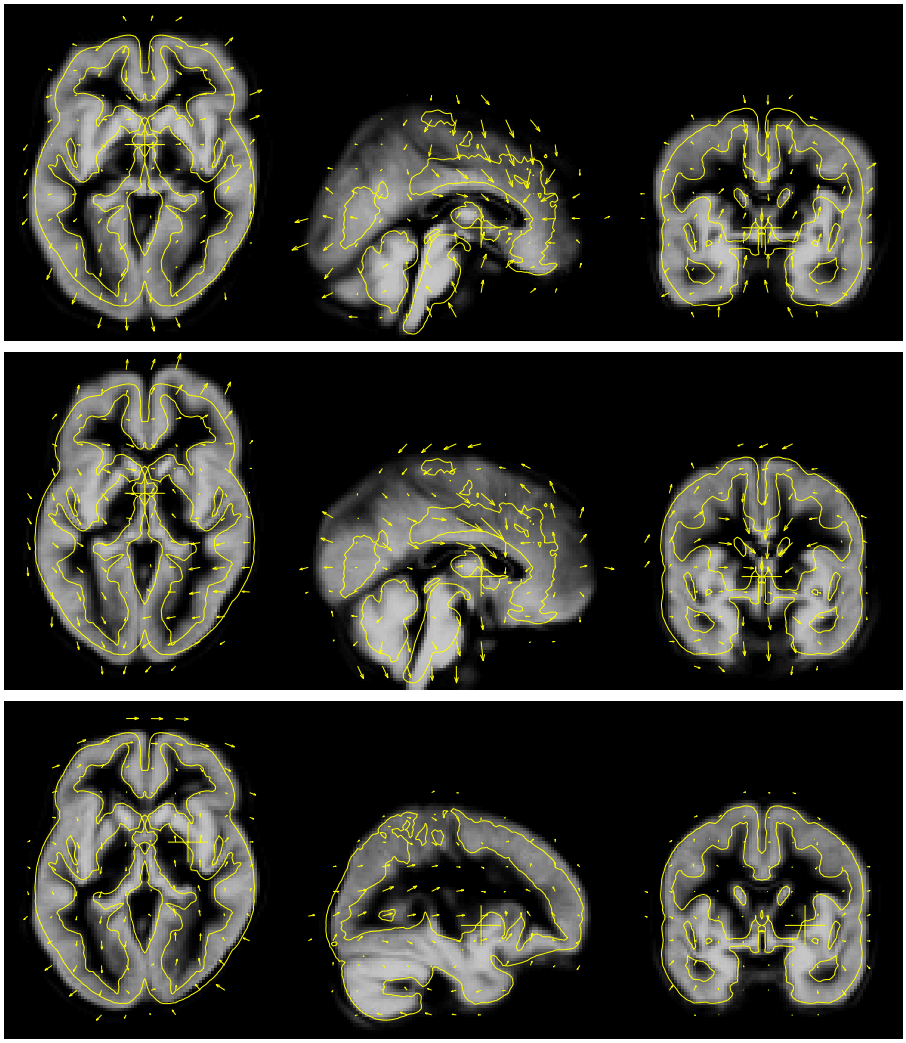

Figure 6.9: The deformation required to map from a female to male brain (above), left-handed to right-handed brain (centre), and antisymmetric deformations from a symmetric template to an asymmetric brain (below), all multiplied by a factor of 5. The deformations were computed after first removing the effects of confounds, and are a direct characterisation of the parameter estimates without referring to the errors or reliability of the differences (c.f., CCA).

should be more reliable indicators of local brain shape than absolute deformations. Absolute deformations represent positions of brain structures, rather than local shape, and need to be quantified relative to some arbitrary reference position.

6.4.1 Theory

A Jacobian matrix contains information about the local stretching, shearing and rotation involved in the deformation, and is defined at each point by:

$$\mathbf{J} = \begin{bmatrix} \partial y_1/\partial x_1 & \partial y_1/\partial x_2 & \partial y_1/\partial x_3 \\ \partial y_2/\partial x_1 & \partial y_2/\partial x_2 & \partial y_2/\partial x_3 \\ \partial y_3/\partial x_1 & \partial y_3/\partial x_2 & \partial y_3/\partial x_3 \end{bmatrix} \tag{6.13}$$

A simple form of TBM involves comparing relative volumes of different brain structures, where the volumes are derived by taking Jacobian determinants at each point (see Figures 6.10 and 6.11). Simple uni-variate statistics (t- or F- tests) can then be used to make inferences about regional volume differences among populations. This type of morphometry is useful for studies that have specific questions about whether growth or volume loss has occurred.

When many subjects are included in a study, a potentially more powerful form of TBM can be attained using multi-variate statistics on other shape measures derived from the Jacobian matrices. This use of multivariate statistics does not explicitly test for growth or shrinkage, but indicates whether there is any shape difference. It is therefore useful when there is no clear hypothesis about the nature of the differences, as may be the case when studying the effects of maturation on the human brain. This form of morphometry should be able to identify shape differences even when volumes are the same.

Because a Jacobian matrix encodes both local shape (zooms and shears) and orientation, it is necessary to remove the latter before making inferences about shape. According to the polar decomposition theorem (Ogden, 1984), a non-singular Jacobian matrix can be decomposed into a rotation matrix ($\mathbf{R}$) and a symmetric positive definite matrix ($\mathbf{U}$ or $\mathbf{V}$), such that $\mathbf{J} = \mathbf{RU} = \mathbf{VR}$. Matrices $\mathbf{U}$ and $\mathbf{V}$ (called the *right* and *left stretch* tensors respectively) are derived by $\mathbf{U} = \left(\mathbf{J}^T\mathbf{J}\right)^{1/2}$ and $\mathbf{V} = \left(\mathbf{J}\mathbf{J}^T\right)^{1/2}$. Matrix $\mathbf{R}$ is then given by $\mathbf{R} = \mathbf{J}\mathbf{U}^{-1}$ or $\mathbf{R} = \mathbf{V}^{-1}\mathbf{J}$.

For a purely rigid body transformation, $\mathbf{U} = \mathbf{V} = \mathbf{I}$ (the identity matrix). Similarly, if $\mathbf{R} = \mathbf{I}$, then the deformation can be considered as *pure strain*. Deviations of $\mathbf{U}$ or $\mathbf{V}$ away from $\mathbf{I}$ indicate a shape change, which can be represented by a strain tensor $\mathbf{E}$. Depending upon the reference co-ordinate system, the strain tensor is referred to as either a *Lagrangean* or an *Eulerian* strain tensor. When the strain tensor is derived from $\mathbf{U}$, it is referred to as a Lagrangean strain tensor, whereas when it is derived from $\mathbf{V}$, it is Eulerian. Spatial normalisation of a series of source images involves determining a mapping from each point in the template image, to corresponding points in the source images. In order to compare image shapes, it is necessary to derive measures of shape within the co-ordinate system of the template image, rather than within the different co-ordinate systems of the individual source images. Therefore, the Lagrangean framework should be used.

For any deformation, there is a whole continuum of ways of defining strain tensors, based on a parameter m. When m is non-zero, the family of Lagrangean strain tensors $\mathbf{E}$ are given by $\mathbf{E}^{(m)} = m^{-1}\left(\mathbf{U}^m - \mathbf{I}\right)$. For the special case when m is zero, the strain tensor is given by

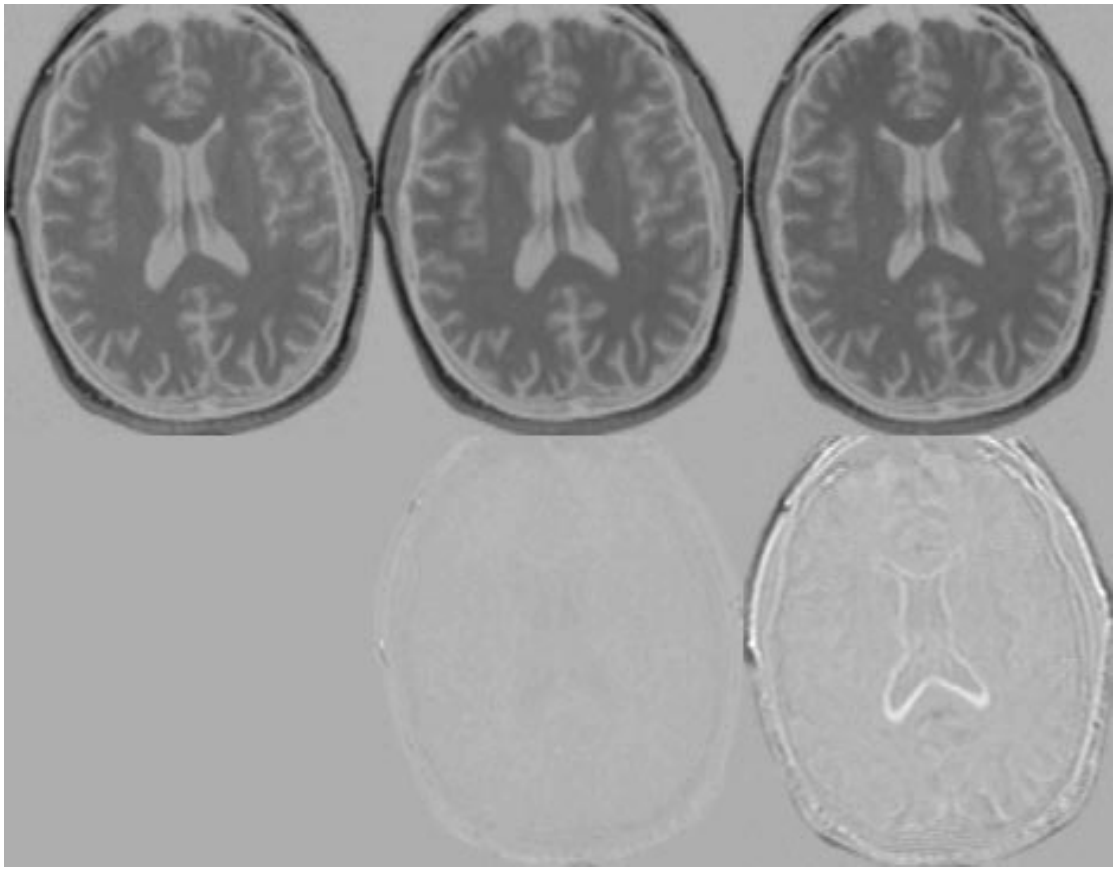

Figure 6.10: This figure illustrates warping together a pair of serial scans of the same Alzheimers subject, using the three dimensional method presented in Chapter 4. The data were provided by the Dementia Research Group, National Hospital for Neurology and Neurosurgery, London, UK. From left to right, the top row shows a slice from the late scan (used as the template), the early scan after rigid body registration followed by warping, and the early scan after rigid registration only. The bottom row shows the difference between the late and early scan, with and without warping. All images are displayed with the same intensity scaling. With only rigid registration, the difference between the early and late scan exhibits a difference at the edges of the ventricles (among other regions). This difference is greatly reduced by including warping in the spatial transformation. The information contained in the difference image is transferred to the deformation fields. One informative feature of the deformation fields are the Jacobian determinants, which reflect volume changes. These are shown in Figure 6.11.

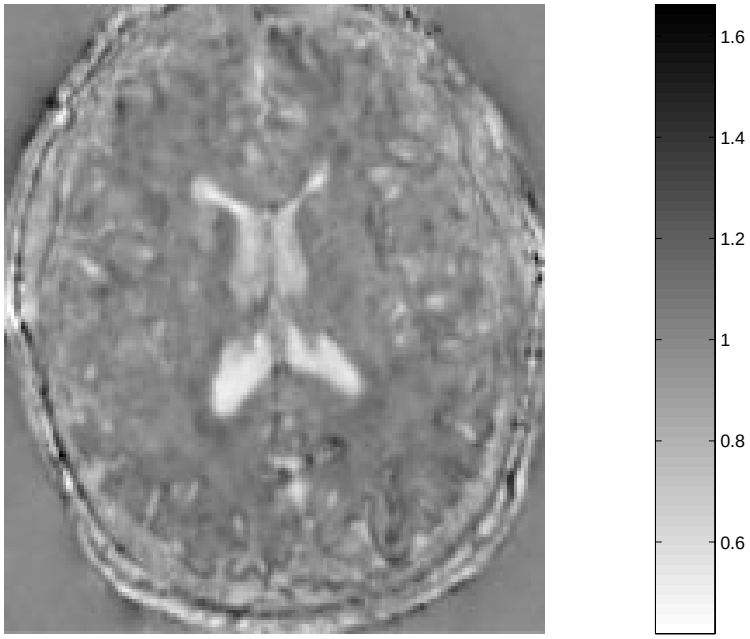

Figure 6.11: This figure illustrates the volume changes estimated by warping together the images shown in Figure 6.10. The relative volumes are the Jacobian determinants of the deformation field. Smaller determinants are obtained when a region of the template maps to a smaller region in the source image. In this example, they represent regions that have expanded between the early and late scans. Regions where there are no measurable volume changes have Jacobian determinants with a value of one.

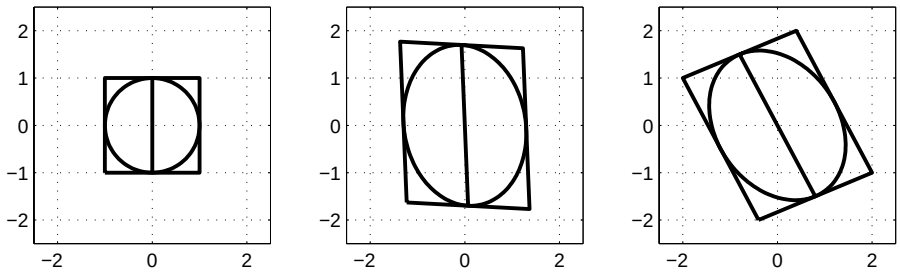

Figure 6.12: This figure illustrates polar decomposition, whereby the regular shape in the left-hand panel is transformed via shears and zooms to give the shape in the centre, before being rotated to give the right-hand shape.

$\mathbf{E}^{(0)} = ln\,(\mathbf{U})$, where the ln refers to a matrix logarithm. When m assumes values of -2, 0, 1 or 2, then the tensors are called the *Almansi*, *Hencky*, *Biot* or *Green* strain tensors respectively.

For deformations derived using the methods in Chapter 4[4], the Hencky tensor may be the most appropriate measure of local shape, providing the priors exert enough influence on the resulting deformation fields. One advantage of basing measures on the Hencky tensor is that it allows relative changes in length, volume and area to be modelled using similar log-normal distributions. This would not be possible if the measures were modelled by normal distributions, since if a variable l is normally distributed, then l^3 is not. Also, it would not make sense to assume distributions that allow negative lengths or volumes, if these are explicitly prohibited by the warping algorithm.

In three dimensions, a strain tensor has six unique elements (as it is a symmetric 3×3 matrix). Multivariate statistics on these elements can model their means, and also their distribution about the mean via a variance-covariance matrix. Such a covariance matrix is able to describe many properties of the warps. The simplest of these are modelled by the diagonal elements of the variance-covariance matrix, which are able to characterise different amounts of warping in different directions. The off diagonal elements show correlations between the strains in different directions, allowing more complex material properties to be modelled. For example, a covariance matrix that enforces the trace of the Hencky tensor to equal zero would model transformations where volumes are preserved (isochoric transformations). In two dimensions, such a covariance matrix could take the following form:

$$\mathbf{C} = \begin{bmatrix} t_1 & -t_1 & -t_3 \\ -t_1 & t_1 & t_3 \\ -t_3 & t_3 & t_2 \end{bmatrix} \tag{6.14}$$

where t_1, t_2 and t_3 are arbitrary values, and the Hencky tensor has been parameterised by $\mathbf{p}$ which has been drawn from a multi-normal distribution of zero mean and variance-covariance $\mathbf{C}$:

$$\mathbf{E}^{(0)} = \begin{bmatrix} p_1 & p_3 \\ p_3 & p_2 \end{bmatrix} \tag{6.15}$$

6.4.2 Data for Evaluations

The data used for the evaluation were MR scans of 58 male right-handed subjects. Sagittal images of the whole brain with voxel sizes of $1\times1\times1.5$mm were acquired on a Siemens MAGNETOM Vision scanner operating at 2 Tesla using an MPRAGE sequence with a 12^o tip angle, 9.7ms repeat time, 4ms echo time and 0.6ms inversion time.

The images were all approximately warped to the same stereotactic space using the methods described in Chapter 3. Following this, a $41 \times 67 \times 59$mm region around the right hippocampus was identified, and the deformation fields refined for this region using a similar procedure to that described in Section 4.3.3. First of all, the approximately spatially normalised images were used to create an average image for this region. This average was used as a template for subsequent high-dimensional registrations, but was modified every four iterations to reflect the new warped

[4]It should be noted that the sum of squares of the Hencky tensor elements is equivalent to the sum of squares of the logs of the singular values of $\mathbf{J}$, meaning that the priors described in Chapter 4 are effectively minimising the squares of the Hencky tensor.

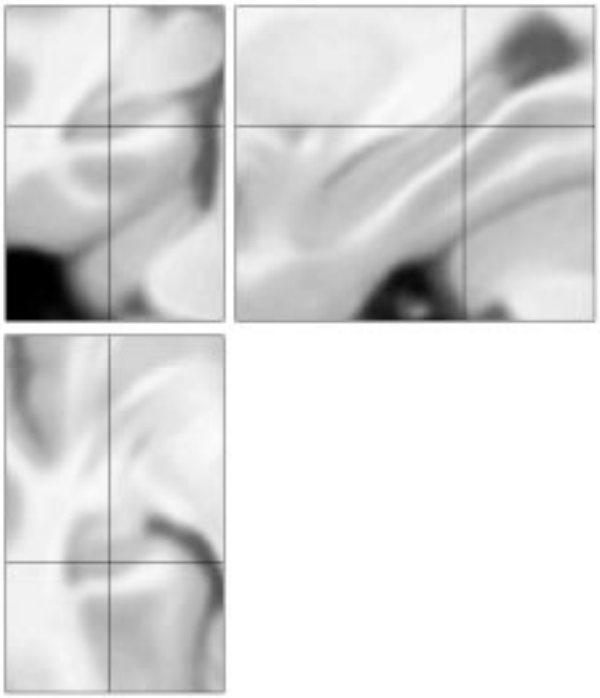

Figure 6.13: Orthogonal views through the mean of 58 warped hippocampus images.

average. A value of 4 was used for λ (the regularisation parameter), and a total of approximately 220 iterations were performed for each image. The resulting deformation fields had an isotropic resolution of 1mm, meaning that each of the 58 deformation fields was defined by 486219 parameters. Figure 6.13 shows orthogonal views through the mean of the normalised images. Six randomly chosen warped images from the dataset are shown in Figure 6.14.

A number of evaluations were performed using this dataset, both in the context of comparing Jacobian determinants between the groups, and also in the context of using multivariate methods to compare strain tensors. No evaluation of the validity of using GRF theory for correcting for multiple dependent comparisons was performed. This was partly because a solid theory for Wilk's Lambda fields has not yet been formulated. Also, some of the transforms applied to the fields involved smoothing, followed by taking logs, which would lead to violations of the GRF theories as they currently stand.

6.4.3 Morphometry on Jacobian Determinants

One form of TBM involves identifying structure volume differences among groups, based on the Jacobian determinants of the deformation fields. Generally, simple voxel-by-voxel t or F tests would be used to produce an SPM of significant volume differences among the groups.

The Jacobian determinant fields from the deformations described in Section 6.4.2 were written. Each element in the field was considered a node in a tetrahedral mesh. A Jacobian determinant was computed at each node from the average volume of the tetrahedra with which it formed a vertex. For half the elements, the value was the average of 8 tetrahedral volumes, whereas it was the average of 24 volumes for the other half (see Section 4.2.4). Images of six determinant fields

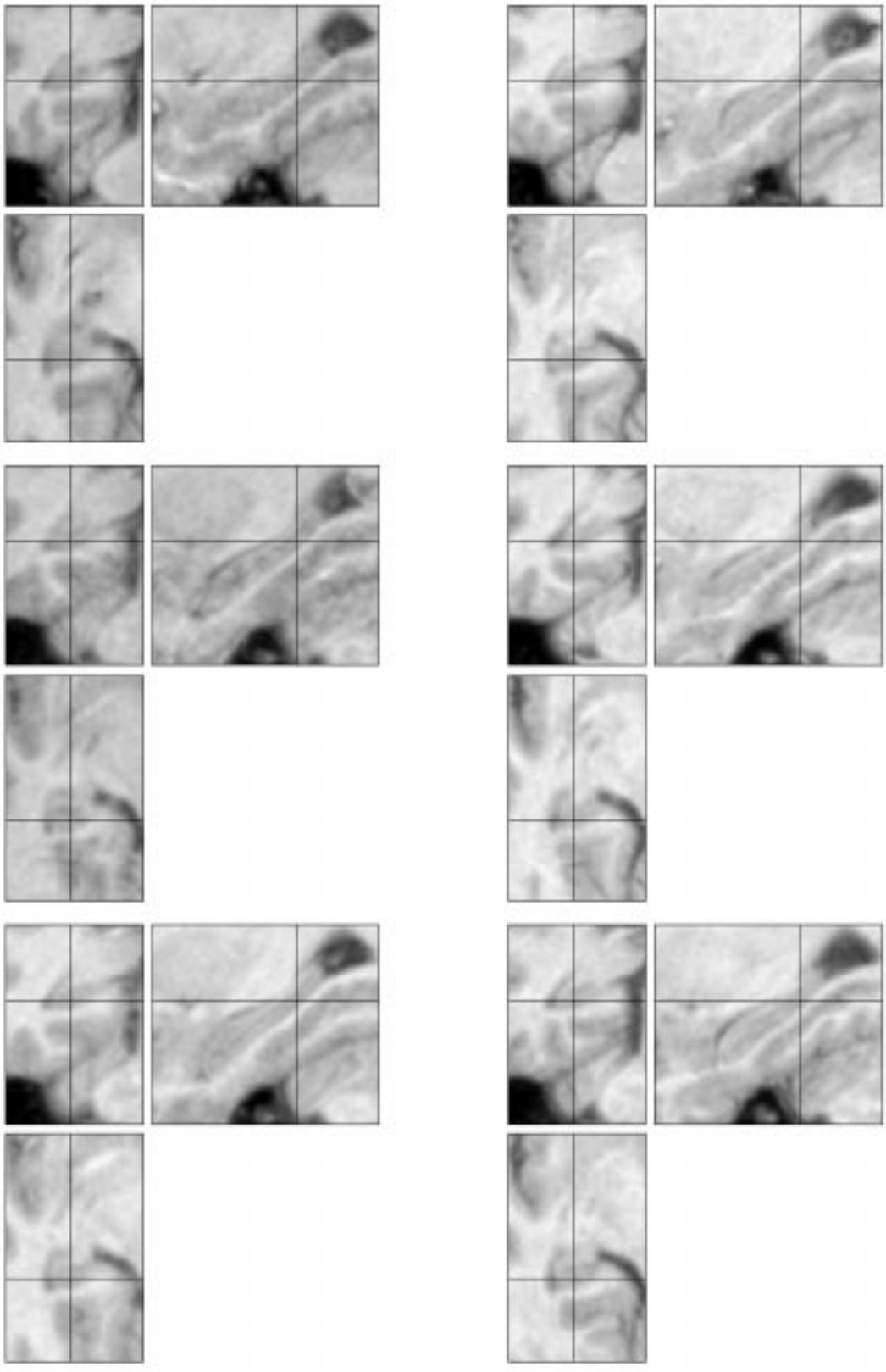

Figure 6.14: A selection of orthogonal views through randomly chosen warped images of hippocampi (see also Figure 6.15).

are shown in Figure 6.15.

Distributions of Jacobian Determinants

A number of assumptions are required in order to perform statistical tests, and when parametric tests are used within the framework of the general linear model, one of the requirement is that the residuals from the model are normally distributed. Section 6.2.2 described a method of testing normality based on correlation coefficients from Q-Q plots. A similar procedure is used here in order to examine the distributions of the residuals. A number of different transformations were also applied to the Jacobian determinants in order to assess their effects on the residuals. The transformations include spatial smoothing by an isotropic 8mm FWHM Gaussian kernel, and taking logs of the Jacobians both before and after smoothing. Histograms of the correlation coefficients over all voxels were computed as described in Section 6.2.2.

The data used in the analysis were derived from 42 of the subjects, the results of which are plotted in Figure 6.16. The first conclusion to be drawn from this figure is that when the warping scheme of Chapter 4 is used, the logs of the Jacobian determinants show more normally distributed residuals than the Jacobian determinants themselves. Smoothing also rendered the residuals more normally distributed, both with and without taking logs of the determinants. Whether logs were taken before or after smoothing appeared to make little difference to the behaviour of the residuals, but this still has implications for the analysis. When logs are taken before smoothing, any analysis is based on logs of the geometric mean of the relative volumes (integrated under the smoothing kernel), and therefore does not directly relate to arithmetic mean volumes. If logs are taken after smoothing, then any corrections for multiple comparisons made using GRF theory are likely to be compromised.

False Positive Rates from T-Tests

The previous section showed that the Jacobian determinants were not normally distributed, but that the normality can be improved by taking logs of the determinants, and also by smoothing. This section evaluates the rate of false positives that are likely to be encountered when performing voxelwise t tests on the logs of the smoothed Jacobian determinant field.

Voxelwise t tests were performed 1000 times over all 58 logged, 8mm smoothed determinant fields. The tests involved randomly assigning the fields to two groups, where one group contained 12 subjects and the other contained 46. The results are histograms of 162073000 t values, which are plotted in Figure 6.17. As hoped for, the resulting t values appear to be distributed reasonably well according to a t distribution with 56 degrees of freedom, implying that voxelwise t tests should produce the appropriate rate of false positive results when comparing groups containing several subjects.

Further tests were performed that involved comparing Jacobian determinants from each single subject with those of the remaining 57. The evaluations were done on the original determinant fields, on the fields after 8mm smoothing and after taking logs of the smoothed and unsmoothed data. Higher t-scores indicate larger Jacobian determinants in the deformation field of the single subject. The results (shown in Figure 6.18) suggest that parametric tests may be less valid for comparing single subjects with groups. Many of the effects of non-normally distributed data are

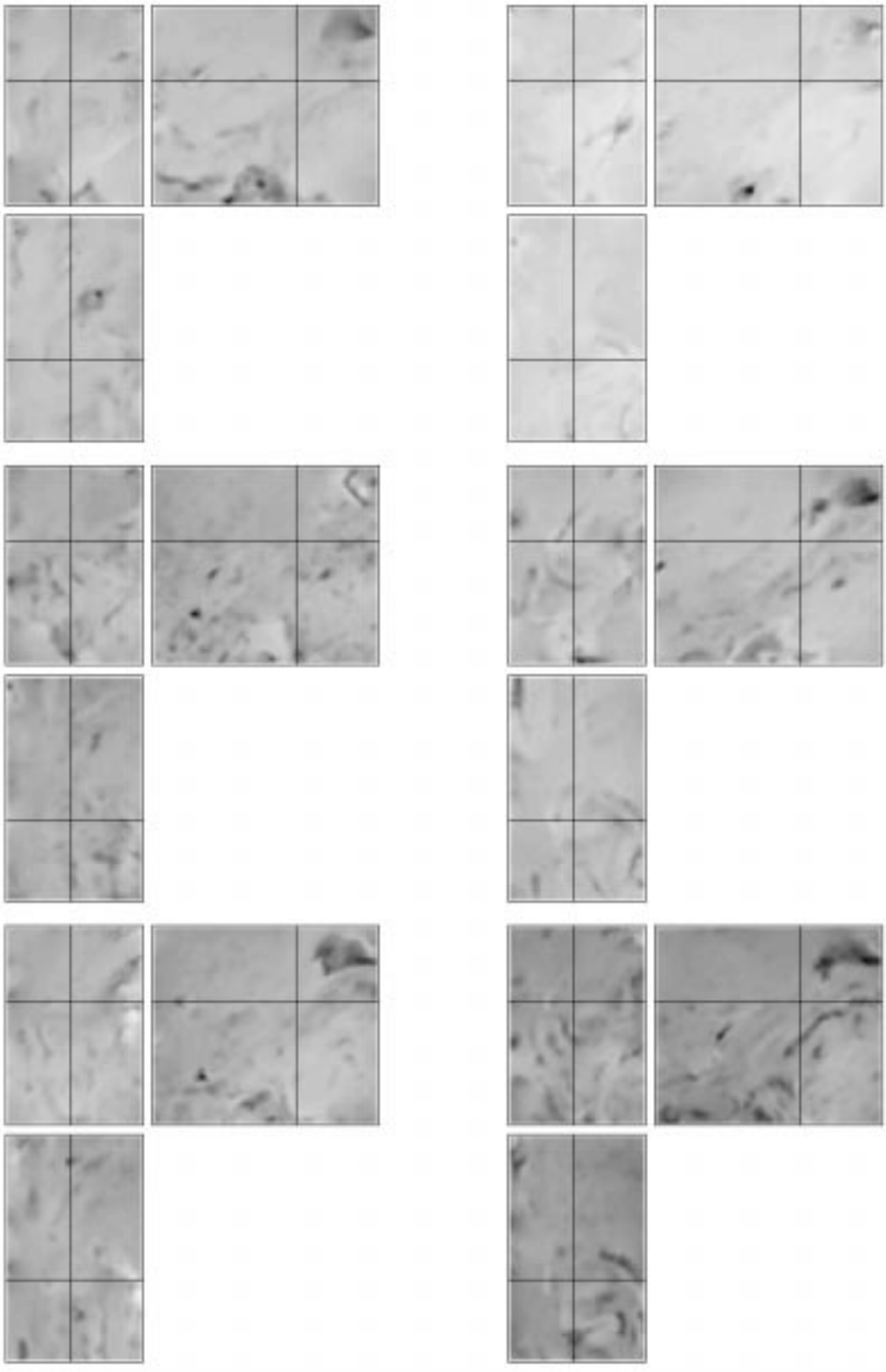

Figure 6.15: A selection of orthogonal views through the Jacobian determinant fields from the same images shown in Figure 6.14.

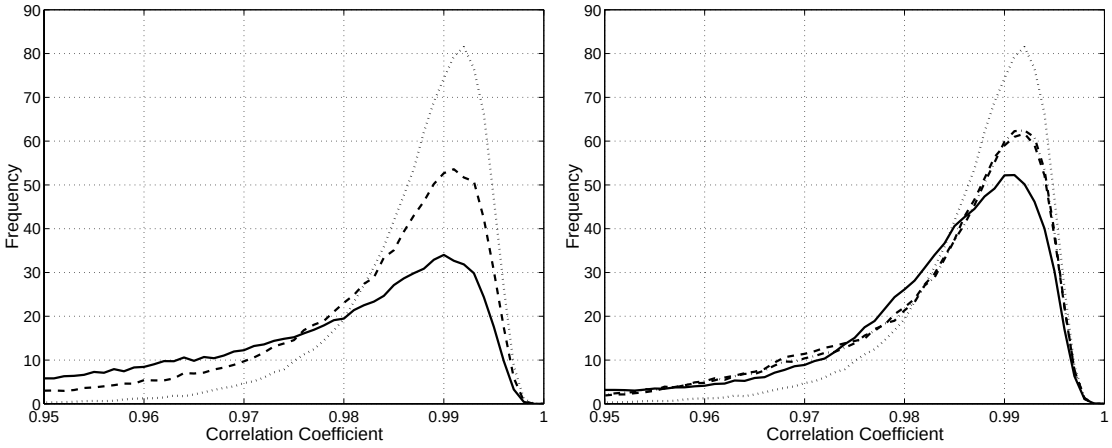

Figure 6.16: Histogram of correlation coefficients taken over the whole volume volumes (144495 voxels). The dotted line in both panels is a histogram obtained from simulated random data, and represents the histogram that would be expected if the Jacobian determinants were perfectly normally distributed. The left panel shows histograms of Q-Q plot correlation coefficients obtained from the unsmoothed Jacobian determinants, both without (solid line) and with taking logs. The panel on the right shows histograms of Q-Q plot correlation coefficients after taking 8mm FWHM Gaussian smoothing the Jacobian determinants. The solid line is for the smoothed determinants, the dot-dash line represents determinants that were smoothed before taking logs, and the dash-dash line was generated by taking logs and then smoothing.

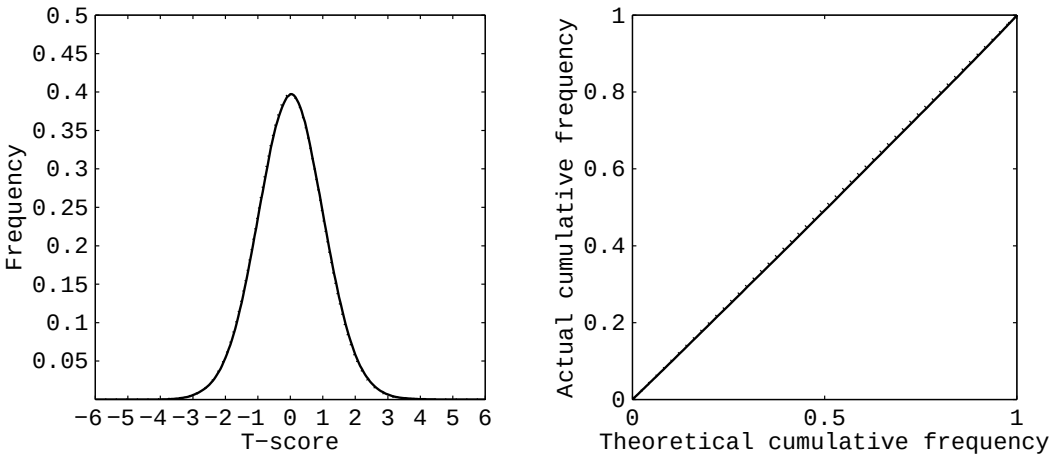

Figure 6.17: Histograms of t-scores from randomly generated tests involving comparing a group of 12 subjects with a group of 46. Left: the histogram resulting from 1000 random voxelwise t tests is plotted as a solid line, and the probability density function of the Students t distribution for 56 degrees of freedom is shown by the dotted line. For each test, a group of 12 subjects was compared with the remaining 46. Right: The same data plotted as a cumulative histogram against the theoretical cumulative histogram.

hidden when making group comparisons, because the group averages tend to be more normally distributed because of the central-limit theorem. This does not apply when comparing single subjects, so the validity of the results is strongly influenced by any non-normality of the data.

The least valid results were obtained from t-tests on the original determinant fields. Slight improvements in the validity of the tests were obtained after smoothing the fields or taking logs. However, closer examination of the cumulative histograms showed that much higher rates of false positive results than expected would occur, even after these transformations.

6.4.4 Morphometry on Strain Tensors

TBM based on Jacobian determinants has just been illustrated. Although this method identifies relative volume differences, there are some shape differences that it can not identify. A potentially more powerful approach may involve using multivariate statistics on strain tensors derived from the deformation fields.

Again, the deformation fields from Section 6.4.2 were used in the evaluations. Hencky and Biot strain tensors fields were derived from the deformations and written to disk. These tensor fields were computed from the deformations in much the same way as the Jacobian determinants used previously. Each element in a tensor field was computed from the average of the strain tensors of the neighbouring tetrahedra.

Distribution of Tensor Field Elements

Ideally, the tensor elements on which subsequent multivariate analysis is performed should be multi-normally distributed. The parameterisation of the distribution should be related to the registration scheme used to estimate the deformations. Because the schemes in Chapter 4 attempt to enforce log-normal distributions on the relative lengths, areas and volumes, then it should bias the elements of the Hencky tensor fields derived from the deformations to be multi-normally distributed.

This section examines the distributions of various tensor fields derived from the deformations, by looking at histograms of the correlation coefficients obtained from Q-Q plots. For the analysis, the tensor fields can be considered as consisting of six scalar fields. Univariate normality of each scalar field was assessed by the correlation coefficients of Q-Q plots, and the results combined into single histograms. First of all the the distributions of unsmoothed Biot and Hencky strain tensors were examined. This was followed by convolving the the Biot strain tensor fields with an 8mm FWHM Gaussian kernel, to identify any increases in normality that would be expected because of the central limit theorem. Finally, the smoothed Biot strain tensor fields were converted to Hencky tensors, and their distributions examined. All the resulting histograms are plotted in Figure 6.19.

As expected, smoothing made the data more normally distributed. Similarly, taking logs of the unsmoothed data also increased normality. However, one unexpected result is that the smoothed Hencky tensor data and the smoothed Biot tensor data transformed to Hencky tensors appears to be slightly less normally distributed than the smoothed Biot tensor data.

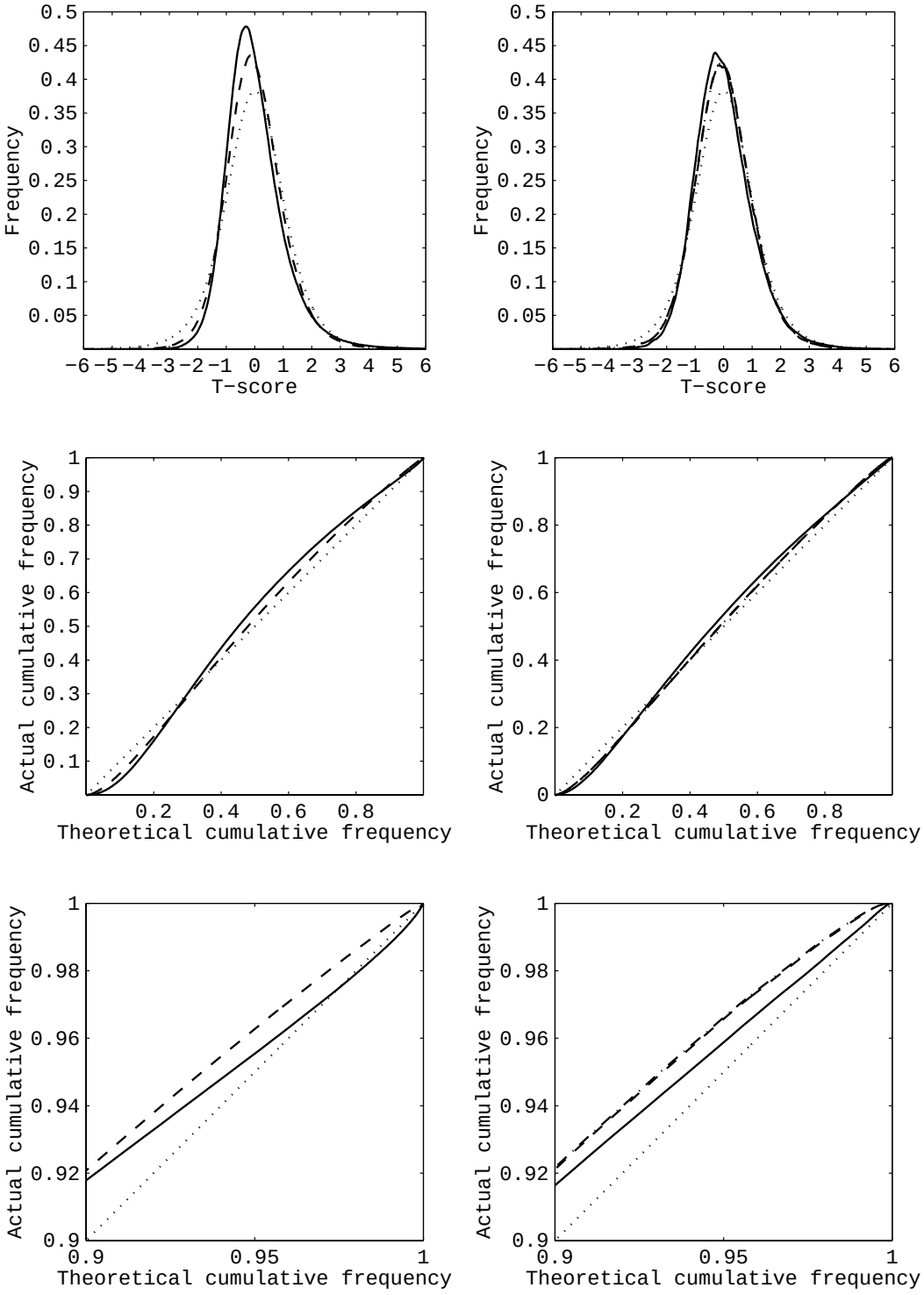

Figure 6.18: Histograms of t-scores from comparing single subjects with a group of 57. The top row shows histograms of the computed t-statistics, whereas the centre and bottom rows show plots of theoretical versus measured cumulative histograms. In all plots, the dotted line shows the theoretical t-distribution for 56 degrees of freedom. Left: Statistics derived from Jacobian determinants (solid line), and logged determinants (dashed line). Right: Statistics derived from 8mm smoothed Jacobian determinants (solid line), logged smoothed determinants (dash-dash line) and smoothed logged determinants (dot-dash line).

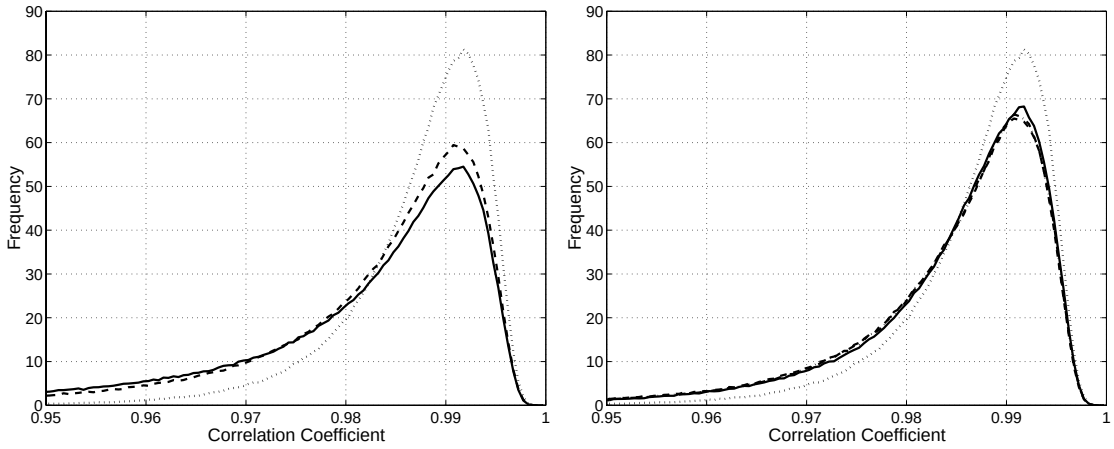

Figure 6.19: Histogram of Q-Q plot correlation coefficients taken over tensor fields. The dotted line in both panels is a histogram obtained from simulated random data, and represents the histogram that would be expected if the tensor elements were perfectly normally distributed. The left panel shows histograms of Q-Q plot correlation coefficients obtained from the all six unsmoothed sets of strain tensor elements. Histograms from Biot (solid line) and Hencky tensors (dashed line) are shown. The panel on the right shows histograms of Q-Q plot correlation coefficients of 8mm FWHM smoothed tensor fields. The solid line is derived from the Biot tensor field, the dot-dash line is from Hencky tensors that were derived from smoothed Biot tensors, and the dash-dash line was generated from 8mm smoothed Hencky tensor fields.

False Positive Rates from MANCOVA

This section assesses the rate of false positives that are likely to occur when multivariate methods are used to compare strain tensors among groups. The tests were for the worst case scenario, and involved comparing each of the 58 tensor fields with the remaining 57. A voxelwise Wilk's Λ statistic was computed at each point in the tensor fields. The same transformations of the tensor fields were used for the analysis as were used in Section 6.4.4, and the results are presented in Figure 6.20.

The plots of cumulative histograms show that a much higher rate of false positives than theoretically predicted would be obtained when comparing single subjects with groups. Smoothing the fields slightly reduces this rate of false positives, but smoothing by more than 8mm FWHM would be needed to bring the rate of false positive results to an acceptable rate.

6.5 Discussion

The morphometric methods developed here all attempt to automatically identify neuroanatomical differences, either from features of spatially normalised images, or from deformation fields that encode information about image shapes. Rather than focusing on particular structures, one important aspect of these methods is that the entire brain can be examined in a balanced way.

Two different types of results are obtained from the methods. One form localises structural differences among groups of brains (VBM and TBM) by producing SPMs of local shape differ-

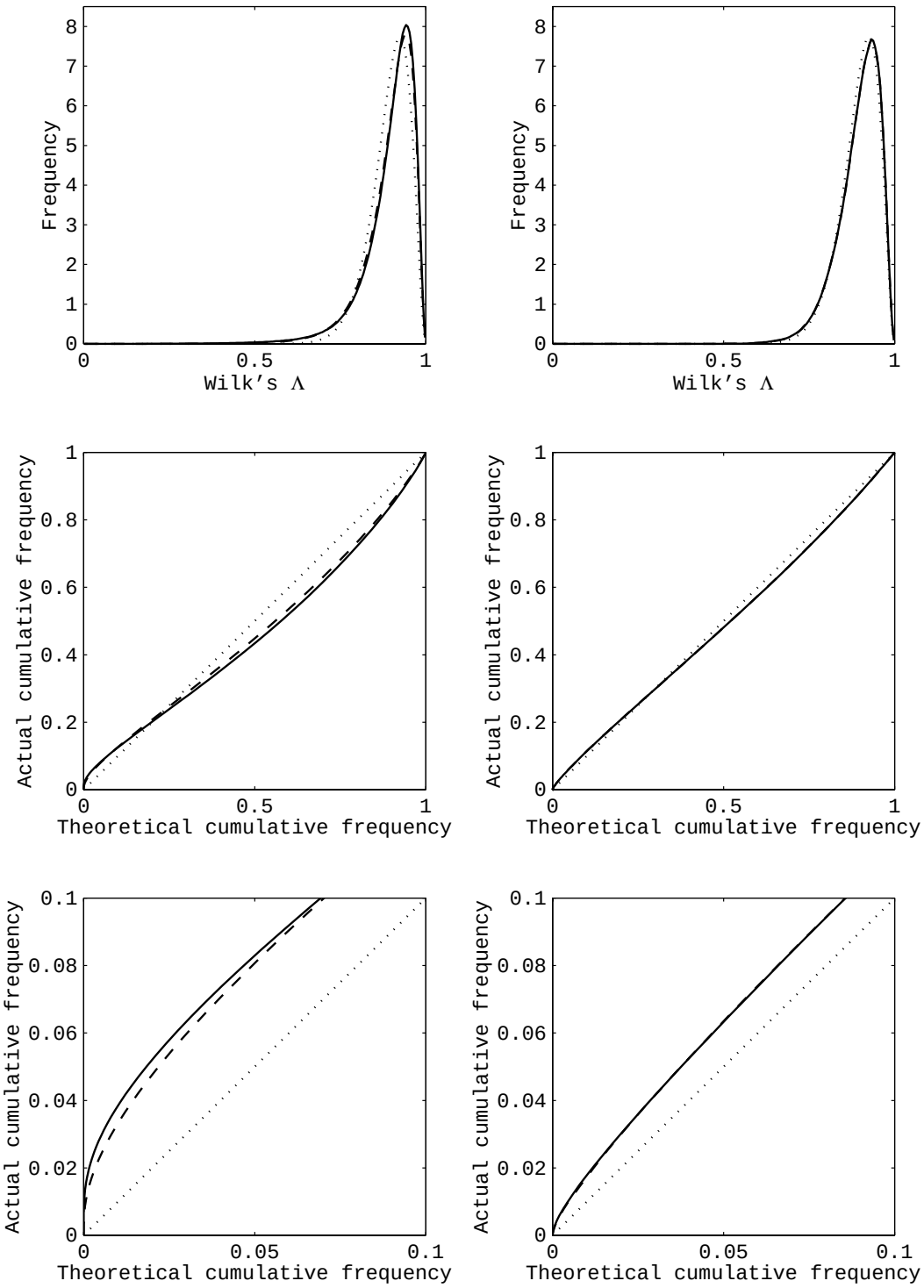

Figure 6.20: Histograms of Wilk's Λ statistics from comparing single subjects with a group of 57. The top row shows histograms of the computed Wilk's Λ statistics, whereas the centre and bottom rows show different parts of the cumulative histograms. In all plots, the dotted line shows the theoretical Wilk's Λ distributions. Left: Wilk's Λ statistics derived from Biot strain tensors (solid line), and Hencky strain tensors (dashed line). Right: Wilk's Λ statistics derived from 8mm smoothed Biot strain tensor field (solid line) and smoothed Biot strain tensor field converted to Hencky tensor field (dashed line).

ences. The other form performs a single, global multivariate test to compare the images (DBM). In principle, the difference between these approaches need not be quite so distinct. Rather than basing the tests either on information at each voxel, or on information from the whole brain, it is easy to see that forms of morphometry could be based on regional analyses. For example, DBM could be done region by region (after factoring out positional information), or TBM could be applied such that the tests include information from surrounding voxels. Both of these approaches should produce very similar results. The regions could be based on pre-defined structures on the template image. Alternatively, they could involve some weighted combination of information from the neighbourhood of each voxel, with weights based on a function such as a Gaussian kernel, or some other windowing function.

The different forms of morphometry have their own advantages and disadvantages, and the optimum approach may depend on the types of structural difference expected among the data. Where there are global patterns of difference, then the global approaches (that do not produce a SPM) may be more powerful, as they can model covariances between shapes of different structures. In contrast, the SPM approaches may be preferable where discrete focal differences are expected. The main disadvantages of global testing methods (as opposed to SPM methods), is that there are generally far more variables for each brain than there are brains in a study. This implies that covariances between all variables can not be computed, so some form of dimensionality reduction (or regularisation) is required, leading to inevitable data loss.

The most powerful forms of morphometry may ultimately depend on how well covariances can be modelled using as few parameters as possible. The current implementation of DBM effectively models covariance matrices by their most prominent eigenvectors, whereas the VBM implementation (and the univariate TBM) models a diagonal covariance matrix[5]. Voxelwise and regionwise multivariate approaches model banded covariance matrices.

A related issue is about how to normalise morphometric measures in order to produce the most useful results. This is particularly true for voxelwise approaches, because relationships between other, more global, factors are not modelled. For example, consider a group of subjects with smaller hippocampi, but also smaller brains. Should the reduced hippocampi be considered significantly different if they correlate with a smaller brain? This size difference could also relate to lengths or volumes of temporal lobe, or to a whole number of other measurements.

Another problem concerns visualising and communicating the results of morphometric tests. Three dimensional volumes are quite difficult to visualise, especially within the limited space of most journals. However, the results of morphometric tests are often vector or tensor fields. These are quite difficult to visualise in two dimensions, but in three dimensions the problem becomes much worse. The more useful results of global morphometric methods are the canonical variates that characterise the differences, and these are often some form of three dimensional vector or tensor field. In comparison, differences localised by voxelwise methods can be relatively easily presented as a SPM. Although the reasons may appear trivial, voxelwise approaches will probably come to dominate because their results can be explained and presented much more easily.

A potential future application for these approaches may be to automatically identify and locate unusual or abnormal brain regions in single subjects. This may involve comparing single subject images with images derived from a large group of controls. If the objective is to find anything

[5]With a few assumptions about some of the off-diagonal elements, which are modelled by Gaussian random field theory

unusual, then the most powerful methods may involve multivariate methods that combine VBM and TBM. Provided that the database of controls is sufficiently large, then the more variables that are included in a multivariate voxelwise test, then the more powerful the tests should be at localising differences.

When parametric statistical methods are used, comparisons between single subjects and whole groups may not be as reliable as comparisons between groups containing several subjects. It may be necessary to resort to non-parametric methods for such cases. The alternative may involve developing suitable transforms for the data that render them more normally distributed. The distributions of estimated deformation fields can be partially controlled by the amount of regularisation imposed on the warping model used. For example, approaches similar to those described in Section 7.3 would attempt to find the optimal amount of regularisation such that the distribution of the residuals is as close to normal as possible. This not only includes the residual difference between image and template, but also the distributions of the parameter estimates around their expected mean.

Other factors also influence the validity and sensitivity of the different morphometric approaches. In particular, the warping method, used to register the images to the same stereotactic space, has a significant influence on the results. For DBM and TBM, the statistical tests are based entirely on the deformation fields produced by the warping methods. The warping methods developed in this thesis are all nonlinear optimisation procedures, and therefore can be susceptible to reaching local minima, and hence non-optimal solutions. These local minima have negative consequences for subsequent statistical tests, as the estimated shapes do not reflect the true shapes of anatomical structures. Other problems occur when warping brains containing severe pathologies. For example, if a brain contains features that are not present in the template image, then an accurate match can not be achieved. The effects of this mismatch may also propagate to other brain regions because of the inherent smoothness of the deformation fields. Much more work is necessary in order to develop warping methods that can model the various forms of severe pathology that may be encountered.

In VBM, nonlinear spatial normalisation results in the volumes of certain brain regions increasing, whereas others decrease (Goldszal *et al.*, 1998). This has implications for the interpretation of what VBM actually tests. The objective of VBM is to identify regional differences in the concentration of a particular tissue (grey or white matter). In order to preserve the actual amounts of grey matter within each structure, a further processing step can be incorporated that multiplies the partitioned images by the relative voxel volumes. These relative volumes are simply the Jacobian determinants of the deformation field. VBM can be thought of as comparing the *relative concentration* of grey matter (i.e., the proportion of grey matter to other tissue types within a region). With the adjustment for volume change, VBM compares the *absolute amount* of grey matter in different regions. As mentioned in Section 6.2.1, if spatial normalisation was perfect, then no grey 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 normalisation is only removing global differences in brain shape, the results of VBM show relative grey 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 such a correction, a continuum will arise with simple VBM (with low resolution spatial normalisation) at one end of the methodolog-

ical spectrum, and statistical tests based on Jacobian determinants (with high resolution spatial normalisation) at the other.

Another perspective on what VBM compares can be obtained by considering how a similar analysis would be done using volumes of interest (VOIs). To simplify the analogy, consider that the smoothing kernel is the shape of a sphere (values of one inside, and zero outside) rather than a 3D Gaussian point spread function. After convolving an image with this kernel, each voxel in the smoothed image will contain a count of the grey matter voxels from the surrounding spherical VOI. Now consider the effects of spatial normalisation, and where the voxels within each VOI come from in the original grey matter images. The spheres can be thought of as projected on to 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 grey matter within each projected VOI (relative to other tissue types). With the multiplication, the total amount of grey matter within the VOI is measured.