Measurement of Brain Change Over Time

FMRIB Technical Report TR00SMS1
(A related paper has been submitted to JCAT)

Stephen M. Smith[1]**, Nicola De Stefano**[2]**, Mark Jenkinson**[1] **and Paul M. Matthews**[1]

1: Oxford Centre for Functional Magnetic Resonance Imaging of the Brain (FMRIB), Department of Clinical Neurology, University of Oxford, John Radcliffe Hospital, Headley Way, Headington, Oxford, UK
2: Neurometabolic Unit & NMR Centre, University of Siena, Italy

Corresponding author - S. Smith - steve@fmrib.ox.ac.uk

Abstract

Quantitative measurement of change in brain size and shape (for example, in order to estimate atrophy) is an important current area of research. New methods of change analysis attempt to improve robustness, accuracy and extent of automation. A fully automated method is presented here, which is accurate (around 0.2% brain volume change error) and achieves high robustness (no failures in several hundred analyses over a range of different data sets).

Keywords: Atrophy measurement, normalised registration

1 Introduction

Various methods have been proposed and implemented for cross-sectional (single time point) or longitudinal (multiple time points) analysis of brain atrophy or more general changes in brain size and shape using magnetic resonance imaging (MRI). A major potential application of atrophy measurements is as a surrogate marker for the progression of neuro-degenerative diseases such as Alzheimer's disease, or of diseases with secondary neuronal or axonal injury, such as multiple sclerosis.

Cross-sectional methods (e.g., [1]) work by measuring brain tissue volume - normally white plus grey matter - and comparing this against a normalisation volume - normally either brain tissue plus cerebrospinal fluid (CSF) volume, or intra-cranial volume. Longitudinal methods (e.g., [2, 7]) typically register (align) two scans separated in time and find areas of change. In general, cross-sectional analysis tends

to incur higher measurement error than longitudinal analysis (related mainly to the practical difference between indirect - cross-sectional - and direct - longitudinal - measurement of change). Additionally, most data sets do contain multiple time point measurements.

This paper presents a completely automated longitudinal measurement method named SIENA (Structural Image Evaluation, using Normalisation, of Atrophy). It performs segmentation of brain from non-brain tissue in the head, estimates the outer skull surface, and uses these results to register the two scans, while correcting (normalising) for imaging geometry changes. Then the registered segmented brain images are used to find local atrophy, measured on the basis of the movement of image edges, found to sub-pixel accuracy.

SIENA overcomes some important limitations of current methods. First, all stages are fully automated. Second, it is robust, allowing analysis of less than ideal MR images. Third, it can be applied to data acquired with different pulse sequences. The method appears to be relatively insensitive to slice thickness. The accuracy in the measurement of brain volume change (BVC) has been shown to be around 0.2%BVC with reasonable quality MR images.

2 Method

2.1 Brain Extraction

The first processing stage is the extraction of the brain from each of the input images, that is, the segmentation of brain from non-brain tissue. The method used is known as BET - Brain Extraction Tool. BET uses a tessellated mesh to model the surface; this model is allowed to deform according to various dynamic local controlling terms until it optimally fits the brain surface [14]. Results are extremely good, even around the eyes, one of the most difficult areas to segment from the brain.

BET provides a binary brain mask, the segmented brain image and an external skull surface image as output. For an example brain surface, see Figure 1. The cerebellum is included in the segmented brain, as is the upper part of the brain stem - the stem is automatically cut according to a surface interpolated sagittally across the ventral cerebellum, pons and temporal lobes.

Figure 1: Example brain surface found by BET.

2.2 Skull Extraction

Measurement of changes in brain size independent of that of the skull (which is of fixed size in the adult) benefits from the estimation of the skull as a normalising factor in both cross-sectional and longitudinal measurements. The importance of this in the latter case will now be explained in more detail.

Before brain change can be measured, the two images of the brain have to be registered (aligned). Clearly this registration cannot allow rescaling, otherwise the overall atrophy will be underestimated. However, because of possible changes in imaging geometry over time (due to gradient calibration drift, local field distortions or varying head placement in the scanner), it is necessary to hold the scale constant (see also [5] for previous work on this problem; note that some longitudinal methods have failed to take account of this problem, although methods based primarily on cross-sectional measurements tend to normalise against it). With the method described here, this can be achieved by using the exterior skull surface, which is assumed to be constant in size and shape, as a scaling constraint in the registration.

In most MR images, the skull appears very dark. In T1-weighted images, the internal surface of the skull is largely indistinguishable from the CSF, which is also dark. Thus the exterior surface is searched for. This also can be difficult to identify, even for human experts, but is the most realistic surface to aim to find. The exterior skull surface is found automatically as the final stage of brain extraction, using BET. Starting with the estimated brain surface, each surface point is taken as the start of a search outwards for the optimal skull position. The most distant (from the brain) point of low intensity (before the bright scalp) is found, and the first peak in gradient outside of this is then defined, as the exact position of the exterior of the skull surface. This method is quite successful, even in regions of overlying (dark) muscle or where there is significant (bright) marrow within the bone.

Thus a skull image is generated for each input image, to be used in registration.

For example, see Figure 2. Here the estimate of the skull is poor at the very top

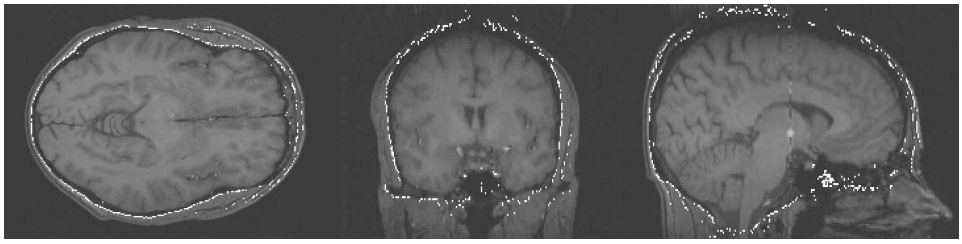

Figure 2: Example exterior skull surface found by BET.

of the head, as intensity inhomogeneity causes the entire signal to fade here. This is fairly typical, and this amount of "missing skull" is not in general a problem in terms of the quality of following registration.

2.3 Registration

As already stated, before the differences between two images can be found, the brains in the two images must be aligned, using a registration procedure. The registration carried out uses a robust and accurate automated linear registration tool, FLIRT (FMRIB's Linear Image Registration Tool) [9]. This uses the correlation ratio cost function [12] and multi-scale search followed by minimisation.

The use of FLIRT in this application is more complex than in the more normal case of registration of two single images. Instead, we are attempting to register two brains, whilst using the two skull images to constrain the scale and skew. In the first stage, the second brain is registered to the first, with a full affine (linear) transformation. Thus translation, rotation, scaling and skew are all allowed to vary (12 parameters). The brain images are used in this first step in preference to the head or the skull images; the head image contains a lot of highly variable components which detract from the registration, such as different ear or scalp appearance, whilst the skull images are far less rich than the brain images, and would therefore result in reduced registration robustness. The resulting transformation is then applied to the second skull image.

The transformed second skull image is then registered to the first skull image, allowing only the scaling and skew to vary. This forces the scaling and skew constraint on the whole procedure, correcting for changes in image geometry. The resulting transform is applied to the registered second brain, thus applying the con-

straint to the brain images.

Next, the resulting second brain image is re-registered to the first brain, allowing only translation and rotation to vary, in order to optimise the final registration between them, the scaling and skew constraint now having been applied.

One could stop here and apply change analysis to the registered second brain and the original first brain. However, this is not optimal, as the second brain image has been through a processing step that the first brain image has not, namely a spatial transformation (involving interpolation of its values). (Note, this is only one transformation, not three, as the three registrations' transformations can be mathematically combined into one before being applied to the initial second brain image.) The images will therefore look slightly different; the transformed second brain image will be slightly more blurred than the first brain image.

To ensure that the images being compared undergo equivalent processing steps, both input images are transformed to a position which is halfway between the two. In this way both images are subjected to an similar degree of interpolation-related blurring. Thus the total transformation that took the initial second brain image to its final position is decomposed into two. Firstly, the "square root" of the matrix is found, that takes the initial second brain image halfway to its final position. (In fact, not all affine transformations have exact matrix square roots which are also affine transformations, so the transformation is actually decomposed into the affine components - translation, rotation, scale and skew - and each component's effect is halved. The components are then reformed into an affine matrix, which is close to the square root of the original matrix, but is guaranteed to be affine.) This matrix then takes the initial second brain image halfway towards the first brain image. Now the transformation is found which takes the first brain image also to the midway point. Because the "square root" matrix already found may not be exact, it is not correct to simply invert this - the images would not then be in perfect registration. Instead, the inverse of the complete registration is multiplied by the "square root" matrix to give the exact transformation that will take the first brain image to the midway position.

The result of all of the above calculations, therefore, is two linear transformations. One is to be applied to the first input image and the other is applied to the second. The typical quality of this brain registration is illustrated in Figure 3, an example subtraction of a registered pair of head images, which shows only appreciable motion outside of the brain.

All of the brain and skull images are now discarded; only the original unsegmented

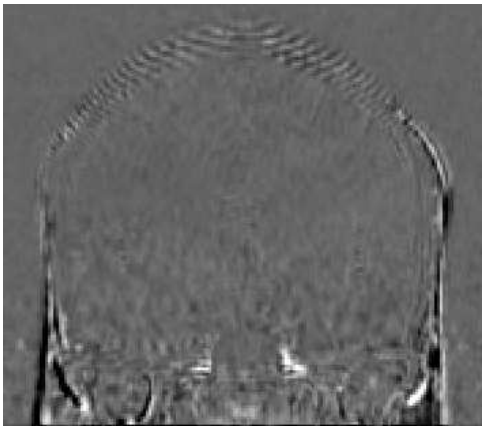

Figure 3: Example subtraction of a registered pair of head images showing only appreciable motion outside of the brain.

images and the brain mask images are kept. The transformations are applied to these images so that two registered head images and two registered brain mask images result. These four images are passed on to the next stage.

2.4 Masking

The registered binary brain masks are now combined into a single mask which will be applied to the registered head images to produce two new registered brain images. The reason for this (rather than keeping the original registered brain images) is that even slight differences in the original brain segmentations (i.e., the production of the brain masks) would cause the artefactual appearance of brain change. Thus the two masks are "binary ORed" - i.e., if either is 1 at a particular voxel, the output is 1. (They cannot be "ANDed" as the brain from the second time point would cause incorrectly reduced masking of the first time point image in the case of atrophy.)

The resulting combined mask is then applied to the registered head images to produce to registered brain images. These two images are passed to the final stage for the analysis of change.

2.5 Change Analysis

The final stage in the analysis is the change estimation itself. There is great variety in how this is achieved amongst published methods. Some researchers (e.g., [8, 7, 10]) use normalized subtraction of the images, assuming that resulting areas of significant deviation from zero correspond to areas of interesting brain change. This relies on the assumption that the images will appear exactly the same (apart from valid change); various procedures such as histogram-matching and relative bias field correction have been suggested [10], in order to attempt to make the images look as similar as possible. Others look more directly for changes around tissue boundaries. For example, [3, 2, 5, 4] use the "boundary shift integral" (the area under the intensity profile across a boundary in image 1 is subtracted from that for image 2, and normalised by the boundary height, resulting in an accurate measure of lateral motion), which gives the motion of each edge, even if blurred, but only if image contrasts in general are well matched between scans. Methods that are principally cross-sectional in nature, such as that of Fisher [1, 13], Ge [6] and Reddick [11] avoid the need to address the issue of change analysis.

The system presented here finds all brain surface edge points (including internal brain-CSF edge points) to sub-voxel accuracy in both images and estimates the apparent motion of each brain edge point perpendicular to the local edge. By finding edge points the method is relatively insensitive to changes in intensity of tissues from one scan to the next - boundaries are fairly objective, even if general tissue intensities change. Thus this method requires no (intensity) normalisation of the images, and is not sensitive to problems arising from intensity inhomogeneities across the images.

Edge points are found using a simple gradient-based edge detector. At each voxel the intensity gradient in x, y and z is found, and the sum-of-squares gradient is taken as the total edge strength. Next, at each voxel whose edge strength exceeds a threshold (automatically set on the basis of image contrasts measured and output by BET), the direction of maximum absolute gradient strength is found. The two immediate neighbours in this direction are checked to see if their edge strength value is higher. If either is, then the current voxel is discarded. This step is known as non-maximum suppression. The resulting set of voxels are assumed to be brain surface points. The position of the surface within these voxels is measured, i.e., the edge is found to sub-voxel accuracy. This is achieved by fitting a quadratic through the edge strength values of the edge voxel and its two neighbours in the direction perpendicular to the edge; the peak of the quadratic is assumed to be a better estimate of the position of the edge than the centre of the edge voxel.

Edge detection is carried out on both registered brain images. Now, for every edge point in image 1, voxels along a line perpendicular to the edge at that point are searched in image 2, in an attempt to find the closest matching edge point. Both edge points must have intensity gradients in the same direction to be allowed to match. Once an edge point match is found, the subvoxel positions are taken into account, resulting in a single number which describes how far and in which direction the edge point from image 1 has moved. Then, according to the position in the image and the combination of the gradient direction and the edge point movement direction, the movement is marked as positive ("growth") or negative ("atrophy") in a new image, at the position of the edge point in image 1. This new image contains mostly zeros, except at the brain edge points, where it contains small positive or negative values encoding whether the brain surface at each edge point has grown or shrunk between the first and second scans, (for example, see Figure 4).

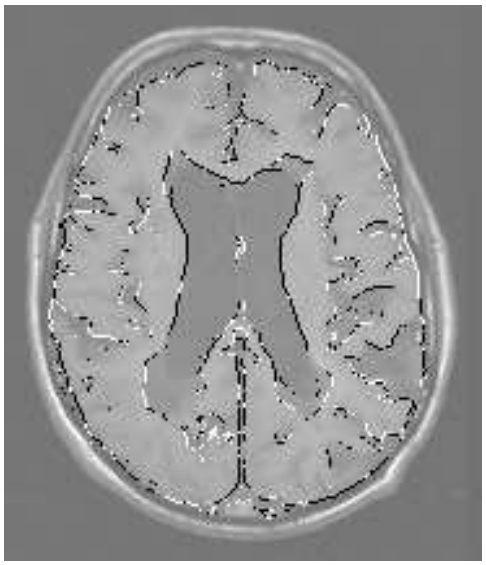

Figure 4: Example slice showing edge motion estimation (atrophy is dark, "growth" is light), overlaid on original image.

Brain atrophy is conventiently quantified by a single number such as the percentage volume change (%BVC). The initial value obtained from the change image is the sum of all edge point motions (linear voxel units), which, when multiplied by voxel volume, gives the total BVC. This is one possible measure, as would be a %BVC derived directly from this. However, a more invariant measure is obtained by dividing this volume by the number of edge points found times the voxel "area".

(Note, the final stages of SIENA are always carried out with cubic voxels, so there is no ambiguity in the definition of area here.) This measure is then the mean perpendicular brain surface motion. The reason why this is preferable to the total volume change is that it is not (to first order) dependent on the number of edge points found. As the number of edge points depends on slice thickness (see below - typically by a factor of two between 1mm slices and 6mm) and (to a lesser extent) other scanning details, it is a good idea to normalize for the number of points found. Finally, if it is required to convert the mean surface motion to a %BVC, the ratio of the brain volume to the brain surface area needs to be estimated.

In this formulation:

$$l = \frac{v \sum m}{aN}, \tag{1}$$

where l is the mean surface motion, $\sum m$ is the edge motion (voxels) summed over all edge points, v is voxel volume, N is the number of detected edge points and a is voxel cross-sectional area. Thus,

$$\text{\% brain volume change} = \frac{100\,lA}{V} = \frac{100\,lfV}{V} = 100\,lf, \tag{2}$$

where A is the brain surface area (actual, i.e., not aN), V is the actual brain volume, f is the ratio of actual area to volume.

It is possible to find f directly for any given image without knowing A or V; if a single image is scaled by a known amount and then compared with the unscaled version using the above change analysis, the correct %BVC is known from the scaling that was applied, and the measurement of l then allows f to be found. It varies across scanners, slice thicknesses and pulse sequence, but normally lies between 0.1 and 0.2mm^{-1}. Applying this method (referred to as self-calibration) ensures that there is no bias (systematic error) in the reported estimates of %BVC.

The complete method is summarised in Figure 5.

3 Validation / Results

3.1 Investigation of Accuracy as a Function of Slice Thickness

To test the accuracy of SIENA, 16 normal volunteers were scanned in two separate sessions each. Each session consisted of 8 scans, at a range of slice thicknesses, to enable the dependence on slice thickness of the accuracy to be determined. The

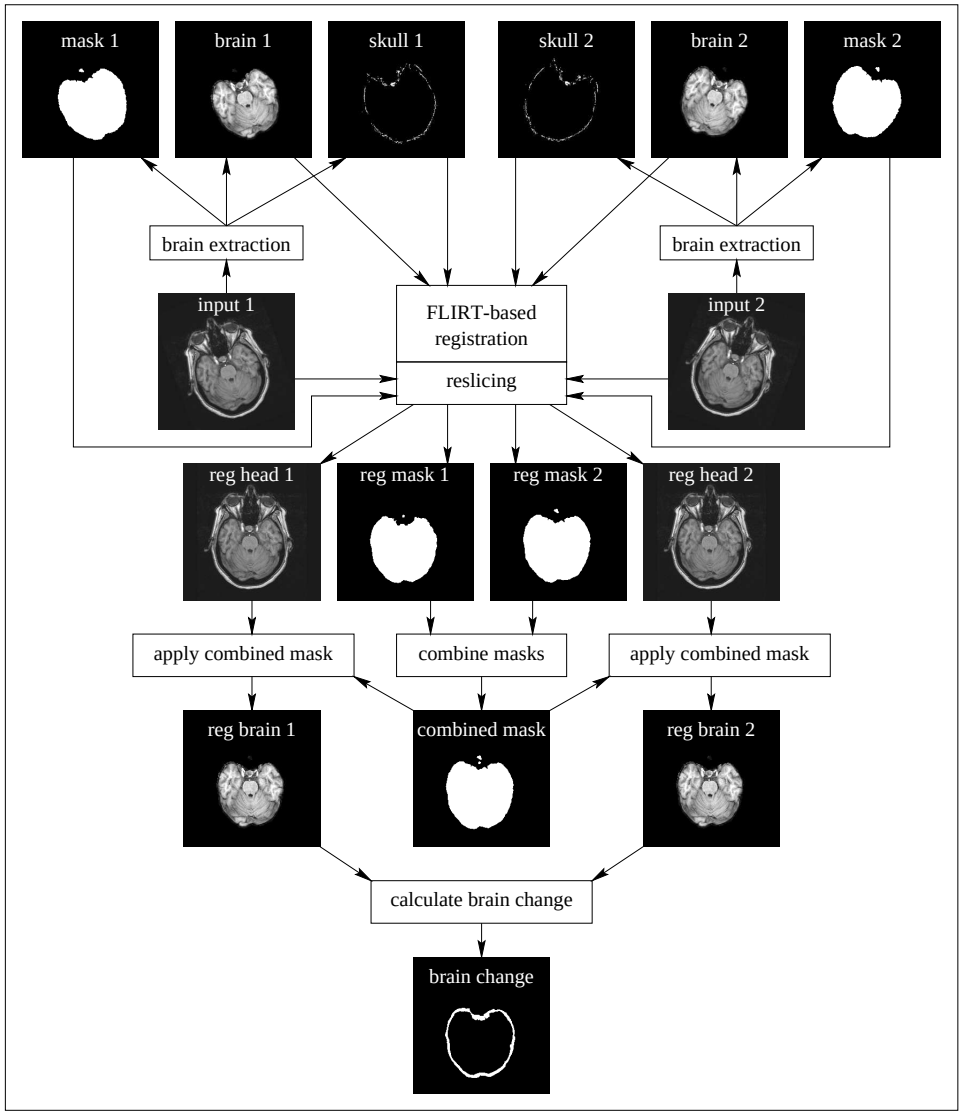

Figure 5: Overview of SIENA.

subjects' ages ranged from 26 to 44; half were female. The scanner was a Philips NT 1.5T operating at the NMR Center of the University of Siena. Scans 1 to 6 were 1mm to 6mm slice thickness, T1-weighted axial 2D fast field echo, TE=11ms, TR=35ms, flip=40^o, NAqc=1. The 1mm scan lasted 18 minutes, and each successive scan took less time, with the 6mm scan lasting 3 minutes. Scan 7 was a 3mm slice thickness axial volumetric fast field echo, TE=3ms, TR=20ms, flip=30^o, NAqc=1, and lasted 4 minutes. Scan 8 was the same as scan 7, but with coronal slices, lasting 4 minutes. For all scans the in-slice resolution was 1mm by 1mm, and enough slices were taken to include the top of the scalp and the bottom of the cerebellum. The inter-session interval was mostly between 1 and 7 days. Half of the subjects were scanned with the slice thickness range order reversed, to control for order effects.

The resulting 128 pairs of images were processed with the fully automated system, with no manual intervention. On average, each image pair took just under an hour to process on a state-of-the-art single processor PC running Linux. The registration results and BET segmentation were checked manually - no obviously incorrect segmentations were found for any of the 256 images and no obviously incorrect registrations were found in any of the 128 pairs. Self-calibration was used as described above to ensure that the %BVC estimates contained no systematic error.

The main results from the above experiment are shown in Figure 6. The %BVC is shown as a function of scan number for all 16 subjects. All results should ideally be zero, as the subjects should not be showing any atrophy. There are two clear results from this figure. Firstly, there is no clear slice-dependence to the errors. Secondly, the error size is small - the median absolute error over all results is 0.2%. The fact that 1mm slices do not generate significantly better results than thicker slices may at first seem surprising. However, one possible reason for this result is that the lower resolution scans are taken more quickly and therefore probably contain less image distortion due to subject motion during the scan. There was no significant difference between male and female subjects, and each subject in general did not show a mean error that was significantly different to zero, i.e., the error across slices thicknesses for each subject varied around zero.

The contribution of the skull-based step in the registration was investigated with the same data set. The results of the three-stage registration (before decomposition into two halfway transforms) were compared with a "control" registration, consisting of just stage one of the complete registration, i.e., a simple full linear transformation from brain image 2 to brain image 1. This can be thought of in the context of this data as a reasonable control, as we know that the subjects' brains will not

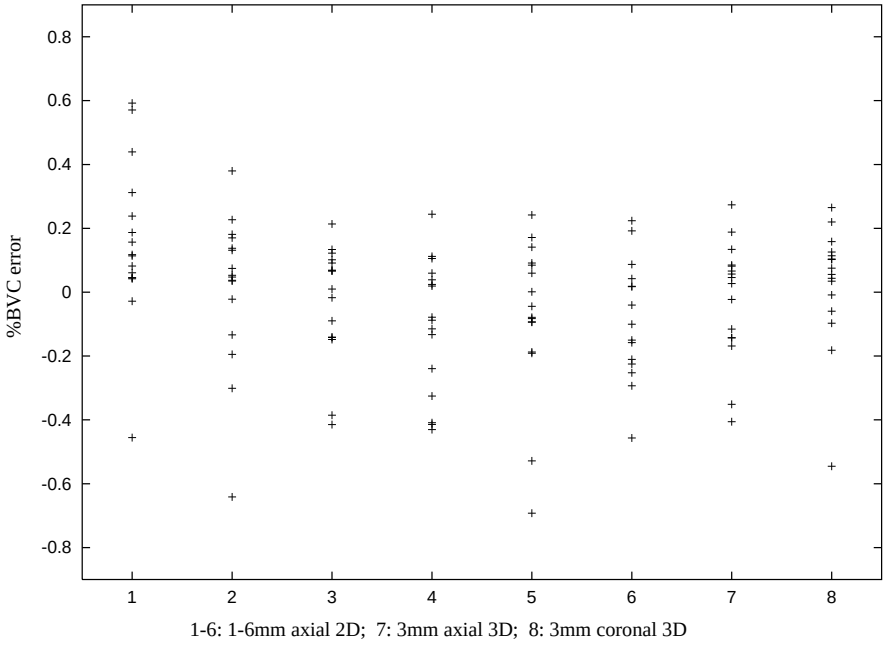

Figure 6: Plot of %BVC error vs different slice thicknesses.

have changed between scans, and the scanner itself is unlikely to have changed in characterisation either. Thus we attribute any difference between simple linear transformation and the three-stage skull-based registration to error in the latter. This error will be due to errors in skull surface estimation in each image, and error in the skull surface registration stage. Results are shown in Figures 7 and 8. The

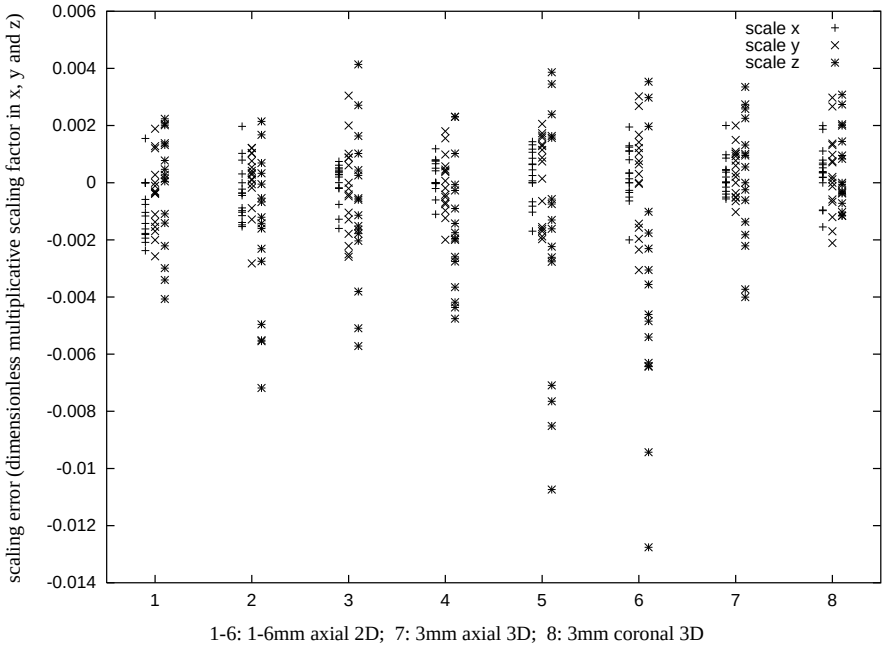

Figure 7: Scaling error vs different slice thicknesses.

median absolute values for scaling and skew (errors) are 0.0009 and 0.0010 respectively. Combining across scaling and skew and across x, y and z gives a %BVC error nearly as large as the complete error result, suggesting that the registration step on average may contribute fairly highly to overall atrophy measurement error. However, this step is important to include, given the common problem of imaging geometry drift.

As a further test of the SIENA method, one of the subject's data sets was tested across slice thicknesses - each image from time point 1 was tested against each image with a different slice thickness from time point 2. The median absolute error was only 0.42%, despite the substantial differences between the images in each pairing.

The final outcome of these investigations, therefore, is that the error in measuring

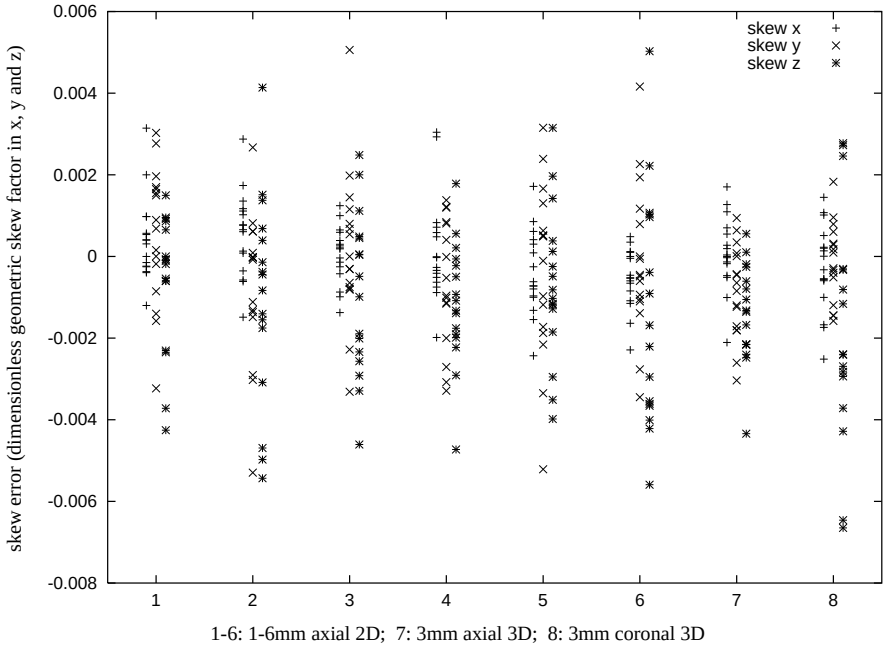

Figure 8: Skew error vs different slice thicknesses.

%BVC between images acquired using the same pulse sequence is around 0.2. This value is not strongly slice-thickness dependent.

3.2 Validation Using Patient Data from Three Time Points

Further investigations were carried out with data sets of three-time-point scans of patients. A sensitive method of error analysis [2] can be carried out on such data; the atrophy measure from t0 to t1 is added to the measure from t1 to t2, and this sum compared with the direct measure from t0 to t2. This will show up most sources of error in the atrophy estimation prodecure and is therefore a useful validation of the method. Sources of error not covered are those which affect both halves of the comparison equally; for example, if a scaling error was caused by inaccurate skull estimation at one time point, and affected the t0 to t1 atrophy measure in the same way that it affected the t0 to t2 measure, this would not show in the three-time-point analysis.

MR images of brains of 39 multiple sclerosis patients (courtesy of Valerie Stevenson and David Miller at the Institute of Neurology, London) were analysed. Each patient was scanned three times, with T1-weighted images, and slice thickness of 3mm. The results of atrophy analysis are shown in Figure 9. The points should ideally lie on the $y = x$ line. The error bars show $\pm 0.2\%$, and are clearly sufficient to explain deviation from the line, demonstrating that the precision of the method as applied to patient data sets is comparable to that with normal controls, despite differences in the pulse sequence for data acquisition.

4 Conclusions

A fully automated method is presented here, which is accurate (around 0.2% brain volume change error) and achieves high robustness. The method is robust to many of the problems typical of general morphometric analysis, and explicitly corrects for most of the remaining problems. A wide range of data types has been used to verify the robustness and accuracy of the method.

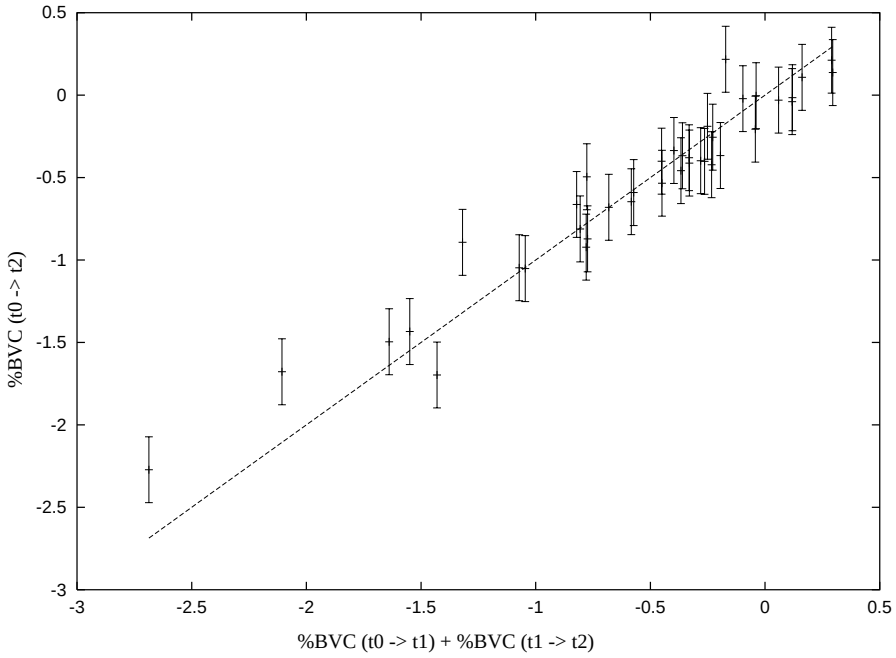

Figure 9: Plot of (%BVC from t0 to t2) vs (%BVC from t0 to t1 plus that for t1 to t2).

5 Acknowledgements

The authors thank Marzia Mortilla for carrying out most of the scanning for the set of normal subjects. SMS and PMM acknowledge funding from the MRC (UK) to support the FMRIB centre; MJ and NDS were funded by the EC Biomed II programme (MICRODAB).

References

[1] E. Fisher, R.M. Cothren, J.A. Tkach, T.J. Masaryk, and F. Cornhill. Knowledge-based 3D segmentation of the brain in MR images for quantitative multiple sclerosis lesion tracking. In *SPIE Proc. Medical Imaging: Image Processing*, pages 19–25, 1997.

[2] N.C. Fox and P.A. Freeborough. Brain atrophy progression measured from registered serial MRI: Validation and application to Alzheimer's disease. *Journal of Magnetic Resonance Imaging*, 7:1069–1075, 1997.

[3] N.C. Fox, P.A. Freeborough, and M.N. Rossor. Visualisation and quantification of rates of atrophy in Alzheimer's disease. *The Lancet*, 348:94–97, july 1996.

[4] P.A. Freeborough and N.C. Fox. The boundary shift integral: An accurate and robust measure of cerebral volume changes from registered repeat MRI. *IEEE Trans. on Medical Imaging*, 16(5):623–629, 1997.

[5] P.A. Freeborough, R.P. Woods, and N.C. Fox. Accurate registration of serial 3D MR brain images and its application to visualizing change in neurodegenerative disorders. *Journal of Computer Assisted Tomography*, 20(6):1012–1022, 1996.

[6] Y Ge, R J Grossman, J K Udupa, L Wei, L J Mannon, M Polansky, and D L Kolson. Longitudinal quantitative analysis of brain atrophy in relapsing-remitting and secondary-progressive multiple sclerosis. In *International Soc. of Magnetic Resonance in Medicine*, 1999.

[7] J.V. Hajnal, N. Saeed, A. Oatridge, E.J. Williams, I.R. Young, and G.M. Bydder. Detection of subtle brain changes using subvoxel registration and subtraction of serial MR images. *Journal of Computer Assisted Tomography*, 19(5):677–691, 1995.

[8] J.V. Hajnal, N. Saeed, E.J. Soar, A. Oatridge, I.R. Young, and G.M. Bydder. A registration and interpolation procedure for subvoxel matching of serially acquired MR images. *Journal of Computer Assisted Tomography*, 19(2):289–296, 1995.

[9] M. Jenkinson and S.M. Smith. A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, 5(2):143–156, June 2001.

[10] L. Lemieux, U.C. Wieshmann, N.F. Moran, D.R. Fish, and S.D Shorvon. The detection and significance of subtle changes in mixed-signal brain lesions by serial MRI scan matching and spatial normalization. *Medical Image Analysis*, 2(3):227–242, 1998.

[11] W E Reddick, J O Glass, and J W Langston. A non-invasive MRI measure of subtle longitudinal volume changes in brain. In *International Soc. of Magnetic Resonance in Medicine*, 1999.

[12] A. Roche, G. Malandain, X. Pennec, and N. Ayache. Multimodal image registration by maximization of the correlation ratio. Technical Report 3378, INRIA Sophia-Antipolis, 1998.

[13] R. A. Rudick, E. Fisher, J. C. Lee, J. Simon, and L. Jacobs. Use of the brain parenchymal fraction to measure whole brain atrophy in relapsing-remitting MS. *Neurology*, 53(8):1698–704, Nov 1999.

[14] S.M. Smith. Robust automated brain extraction. *NeuroImage*, 2000. submitted.