

# Accurate, Robust, and Automated Longitudinal and Cross-Sectional Brain Change Analysis

Stephen M. Smith,\* Yongyue Zhang,\* Mark Jenkinson,\* Jacqueline Chen,\*† P. M. Matthews,\*  
Antonio Federico,‡ and Nicola De Stefano‡

\*Oxford Centre for Functional Magnetic Resonance Imaging of the Brain, Department of Clinical Neurology, FMRIB, University of Oxford, John Radcliffe Hospital, Headley Way, Headington, Oxford, United Kingdom; †Neurometabolic Unit & NMR Centre, University of Siena, Italy; and ‡MRS Lab, McConnell Brain Imaging Centre, MNI, Canada

Received April 12, 2001

**Quantitative measurement of brain size, shape, and temporal change (for example, in order to estimate atrophy) is increasingly important in biomedical image analysis applications. New methods of structural analysis attempt to improve robustness, accuracy, and extent of automation. A fully automated method of longitudinal (temporal change) analysis, SIENA, was presented previously. In this paper, improvements to this method are described, and also an extension of SIENA to a new method for cross-sectional (single time point) analysis. The methods are fully automated, robust, and accurate: 0.15% brain volume change error (longitudinal); 0.5–1% brain volume accuracy for single-time point (cross-sectional). A particular advantage is the relative insensitivity to differences in scanning parameters. The methods provide easy manual review of their output by the automatic production of summary images which show the results of the brain extraction, registration, tissue segmentation, and final atrophy estimation.** © 2002 Elsevier Science (USA)

**Key Words:** structural brain analysis; atrophy measurement; normalized 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 measurement is as a surrogate marker for the progression of neurodegenerative diseases such as Alzheimer's disease, or of diseases with secondary neuronal or axonal injury, such as multiple sclerosis.

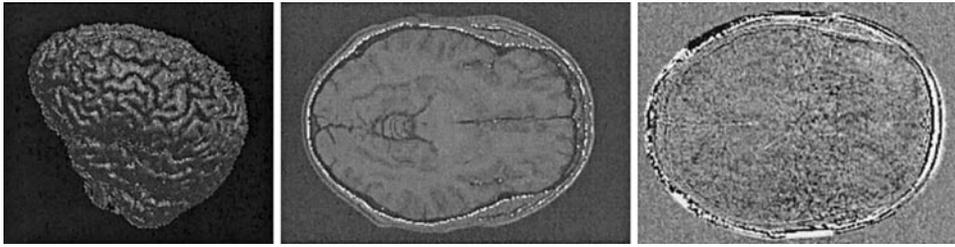
Cross-sectional methods (e.g., Fisher *et al.*, 1997) work by measuring brain tissue volume—normally white plus grey matter—and comparing this against a normalization volume—normally either brain tissue

plus cerebrospinal fluid (CSF) volume, or intracranial volume. Longitudinal methods (e.g., Fox and Freeborough, 1997; Hajnal *et al.*, 1995) typically register (align) two scans separated in time and find regions of change. In general, cross-sectional analysis tends to incur higher absolute measurement error than longitudinal analysis. This is related mainly to the practical difference between integrated (cross-sectional) and differential (longitudinal) measurement of change. Although many data sets do contain multiple time point measurements, there are also situations where only single time points are available, or where the question of interest relates to “absolute” atrophy rather than its rate. Thus cross-sectional and longitudinal studies/analyses contain many common features and purposes, but are also to some extent complementary.

This paper presents completely automated longitudinal and cross-sectional measurement methods named respectively SIENA (Structural Image Evaluation, using Normalization, of Atrophy) and SIENAX (an adaptation of SIENA for cross-sectional measurement).

SIENA performs segmentation of brain from non-brain tissue in the head and estimates the outer skull surface (for both time-points), and uses these results to register the two images, while correcting (normalizing) 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.

SIENAX also performs segmentation of brain from non-brain tissue in the head and estimates the outer skull surface, with data from a single time-point. The brain and skull images are then registered to a standard space brain and skull image pair. This step normalises for skull size, and means that it is not necessary to measure CSF volume (otherwise a problem in T1-weighted images as it is hard to accurately separate CSF and skull). Next a probabilistic brain mask derived in standard space is applied to make sure that certain structures such as eyes/optic nerve have not



**FIG. 1.** Left: Example brain surface found by BET (3D rendering using MEDx, Sensor Systems). Middle: Example exterior skull surface voxels found by BET. Right: example subtraction image (displayed intensity range covers approximately 5% of the images' original intensity ranges) after registration (without intensity normalization) of two images from a subject without atrophy, showing large changes outside the skull but very little change in the brain.

been included in the brain segmentation. Finally tissue-type segmentation is carried out (including partial volume estimation) and a (normalised) brain volume estimate is produced.

Note that most methods of atrophy analysis (whether cross-sectional or longitudinal) will not correctly interpret other types of brain change or pathology (such as developing tumour or lesion), unless they are explicitly additionally developed to cope with such cases.

## 2. SIENA—LONGITUDINAL METHOD

The SIENA method was originally described in Smith *et al.* (2001). The major subsequent improvement is in the final stage (the change analysis), so the initial stages are only described briefly.

### 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 nonbrain tissue. The method used is known as BET—Brain Extraction Tool (Smith, 2000a,b). 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. Results are normally in extremely good correspondence with manually segmented output, 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. 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.

For an example of extracted brain surface, see Fig. 1 (left).

### 2.2. Skull Extraction

Measurement of changes in brain size benefits from the estimation of the skull (which is of fairly unvarying size over time in an adult) as a normalizing 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 or variable local field distortions), it is necessary to hold the scale constant (see also Freeborough, 1996, 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 (assumed to be constant in size and shape for an individual) 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 Fig. 1 (middle).

### 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) (Jenkinson and Smith, 2001), which by default uses the correlation ratio cost function (Jenkinson and Smith, 2001) and a very robust multistart multiscale optimisation strategy.

The use of FLIRT in this application is more complex than in the more normal case of registration of two single images. A three-step procedure is used, where the brain images are used to optimise the initial registration and the final translation and rotation, whilst the skull images are used to optimise the scaling and skew.

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). 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 a similar degree of interpolation-related blurring.

The typical quality of this brain registration is illustrated in Fig. 1 (right), an example subtraction of a registered pair of head images, which shows only appreciable motion outside of the skull.

All of the brain and skull images are now discarded; only the original unsegmented images and the brain mask images are kept. The transformations are applied to these images so that two registered ("common-space") 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 two registered brain

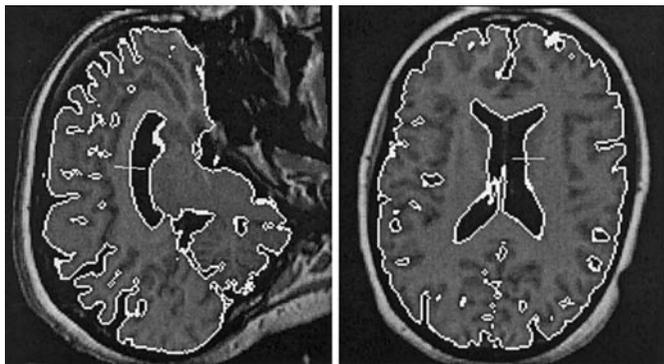
images. These two images are passed to the final stage for the analysis of change.

#### 2.5. Change Analysis

The next stage in the analysis is the change estimation itself. There is great variety in how this is achieved amongst published longitudinal atrophy methods. Some researchers (e.g., Hajnal *et al.*, 1995a,b; Lemieux *et al.*, 1998) 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 the change of interest); various procedures such as histogram-matching and relative bias field correction have been suggested (Lemieux *et al.*, 1998), in order to attempt to make the images look as similar as possible. Others look more directly for changes around tissue boundaries. For example, Fox and colleagues (Fox and Freeborough, 1997; Fox *et al.*, 1996; Freeborough and Fox, 1997; Freeborough *et al.*, 1996) 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 normalized 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 *et al.* (1997); Ge *et al.* (1999); Rudick *et al.* (1999), and Reddick *et al.* (1999) avoid the need to address the issue of change analysis.

The system presented here first attempts to find all brain surface edge points and then estimates the motion of these edge points from one time point to the next. This edge motion is found for the whole brain surface, enabling the total volume change to be estimated. The previously published version of SIENA found edges on the basis of edge strength, and then found edge motion by searching for matching edge points from one image to the next. This suffered slightly from relatively imprecise definition of edge points, i.e., discrimination was imperfect. The current version uses full tissue-type segmentation to find edge points, and thus is more correctly selective, and also enforces continuity of the estimated brain surface. Thus the system presented here finds all brain surface edge points (including internal brain-CSF edge points, such as those around the ventricles) and then finds the motion of these points, in a Bayesian framework, perpendicular to the local edge, to subvoxel accuracy.

In order to find all brain surface edge points, tissue segmentation is performed on the image from time point 1 after application of the joint brain mask (see previous section). The tool used (Zhang *et al.*, 2001) carries out tissue (Grey Matter, White Matter and



**FIG. 2.** Example slices through an image after edge point detection and example perpendicular image profiles.

CSF) segmentation and bias field correction. The method is based on a hidden Markov random field (segmentation labelling) model and an associated Expectation-Maximization algorithm for estimating tissue intensity parameters and bias field (spatial intensity inhomogeneity). The whole process is fully automatic (after being instructed as to whether the image is T1 or T2, and whether to attempt to segment grey and white matter as a single class or as separate classes), producing a tissue-labeled segmentation. It is robust and reliable, compared to the more common finite-mixture-model-based methods, which are sensitive to noise, particularly as they use no spatial neighbourhood information.

The tissue segmentation labels are used to find all brain edge points. First, gray and white voxels are combined into a single class, as are also CSF and background voxels. All boundary voxels between these two resulting classes are used for the next processing stage. Note that this method of finding brain edge voxels enforces a continuous surface (without breaks), although not necessarily a topologically simple one. Figure 2 shows example slices through an image after edge point detection (and also example perpendicular image profiles as described below).

Next, the common-space registered image from time point one is processed at each brain edge point. First the image gradient direction (in 3-D) is found, using a simple  $3 \times 3 \times 3$  Gaussian-weighted derivative operator. This is used to find the surface normal unit vector (and will always point from the darker side of the boundary to the lighter side—this information will later be used to tell the difference between atrophy and “growth”).

Next, a 1-D array (an intensity profile perpendicular to the edge) is filled with values from the image. These values are sampled at subvoxel positions (using trilinear interpolation) as the array’s elements will not in general fall exactly at voxel grid positions. The length of the array is preset to a fixed number of millimeters (typically  $\pm 3$ ); the extent will also be limited by the

presence of a second edge, for example, the far side of a sulcus, in order to prevent other nearby edges from confusing the motion estimation. A second 1-D array is filled with values from exactly the same image positions from the (common-space registered) image from time point two.

Edge motion is now estimated by finding the relative shift, between the arrays, which produces the maximum correlation (to sub-voxel accuracy using interpolation of the correlation scores). However, before the correlation, each array is preprocessed in two ways.

First each profile is convolved with a differentiating kernel, as it makes sense to correlate the derivatives (edge-enhancements) of the two 1-D image profiles rather than the raw image values; if there are intensity or contrast differences between the two images, the position of maximum correlation could be skewed, but this effect is much reduced if correlating edge-enhanced versions of the profiles. Thus this method requires no (intensity) normalization of the images and is not sensitive to problems arising from intensity inhomogeneities across the images.

The second process is the multiplication of each profile by a high-power exponential profile (smoothed sharp cutoff); this acts as a prior on the expected motion by weighting the correlation score, so that higher motions are less likely than small ones—this helps reduce the effect of large motion mismatches (which otherwise make a large contribution to error in the overall method). This can be viewed as a Bayesian prior:

$$P(\text{displacement}|\text{data}) \propto P(\text{data}|\text{displacement})P(\text{displacement}), \quad (1)$$

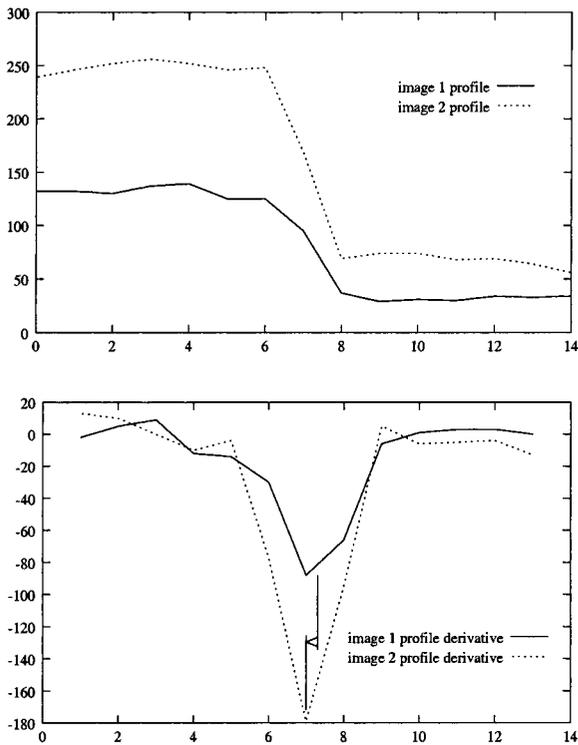
where the first term on the right can be thought of as the raw correlation score, and second term is the prior on the displacement between the profiles

$$P(\text{displacement}) \propto e^{\frac{-\text{displacement}^4}{2\sigma^4}}, \quad (2)$$

which has  $\sigma$  set to a suitable length such as 7 mm. Because the posterior on the displacement is simply used to find the maximum probability the constants of proportionality are unimportant.

Thus the optimal displacement is found for each edge point, and, as stated earlier, the direction of the edge normal determines whether atrophy or “growth”<sup>1</sup> is taking place at this point. The position of optimal displacement is estimated to subvoxel accuracy by fitting a quadratic through the correlation values at the peak

<sup>1</sup> “Growth” could either be real, e.g., oedema, or due to nonlinear motion between the two time points. In the latter case, it will tend to cancel out with atrophy measurements at other points in the image.



**FIG. 3.** Example profiles from one edge point with a slight shift between time points and the derivatives of these profiles.

and its two neighbors.<sup>2</sup> Figure 3 shows example profiles from one edge point with a slight shift between time points, and the derivatives of these profiles.

For example slices showing atrophy as dark edge points and “growth” as light, see Fig. 4.

### 2.6. Percentage Brain Volume Change Quantification

Brain atrophy is conveniently quantified by a single number such as the percentage brain volume change (PBVC). 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 PBVC 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 confusion about 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 1 mm slices and 6 mm) 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 PBVC, the ratio of the brain volume to the brain surface area needs to be estimated.

In this formulation:

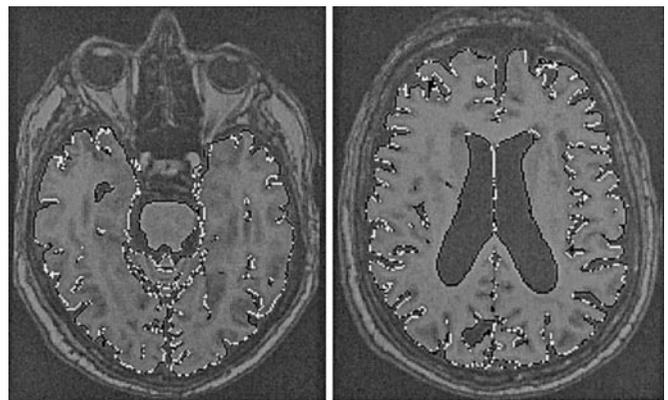
$$l = \frac{v \sum m}{aN}, \quad (3)$$

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,

$$\text{Percentage brain volume change} = \frac{100IA}{V} = \frac{100IfV}{V} = 100If, \quad (4)$$

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 PBVC 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 0.2  $\text{mm}^{-1}$ . Applying this method (referred to as self-calibration) helps reduce bias (systematic error) in the reported estimates of PBVC.



**FIG. 4.** Example slices showing atrophy as dark edge points and “growth” as light. Within the sulci, atrophy approximately balances “growth”—this is due to slight (nonlinear) motion of the brain between scans, and does not contribute significantly to the total atrophy measurement. However, on the periphery of the cortex, and particularly around the ventricles, there is strong atrophy apparent.

<sup>2</sup> Analysis of the resulting motion estimates, in particular histogram analysis, was made during the development of this part of the method, to ensure that the subvoxel estimation was functioning correctly, and adding value to the overall method.

### 3. SIENAX—CROSS-SECTIONAL METHOD

SIENAX is closely related to the SIENA longitudinal method, but, instead of using images from two different time points, SIENAX attempts to estimate normalized brain volume (NBV)<sup>3</sup> from a single image, using the skull to normalize spatially, with respect to a standard image. Cross-sectional studies of brain atrophy normally attempt to relate brain size at a given point to the size of the brain at maturity (Sgouros *et al.*, 1999). In a cross-sectional study the latter can only be estimated from skull size, but the close relationship between normal skull and brain growth makes this a reliable marker. Thus, the goal for determination of relative brain atrophy is to accurately define brain size with respect to skull size, normalized to a standard template. Another way of looking at the value of normalizing for head/skull size is that it reduces within-group variations, making cross-group comparisons more sensitive. An alternative to using head size in normalisation is to estimate CSF volume (to end up with a measure of intracranial volume). One advantage of the approach described here is that it is not restricted to running on images where CSF can be robustly found. For example, in T1-weighted images, it is hard to distinguish between CSF and skull voxels, reducing the accuracy of CSF-volume-based normalization; clearly this is not a problem for SIENAX.

Because SIENAX uses several of the techniques already described above, the description of the method, given below, is relatively brief.

#### 3.1. Brain Extraction and Normalization

SIENAX uses BET to find the brain and skull images from the single input head image. These are then used, in a similar manner to the registration process in SIENA, to register the image to standard space brain and skull images (derived from the MNI152 standard image (Evans *et al.*, 1996, 1997)). Next a standard space mask is used to make sure that no parts of the eyes are left from the brain extraction (because of the connection of the optic nerve, this can occasionally happen)<sup>4</sup> and also to provide a consistent (i.e., nonarbitrary) cutoff point for the brain stem.

<sup>3</sup> “NBV” will be used to refer to brain volume after normalization to standard space; “BV” will be used to refer to original brain volume, i.e., before this normalization. Note that BV is in mm<sup>3</sup>, and that the normalization factor is dimensionless, so that NBV is also in mm<sup>3</sup>.

<sup>4</sup> The removal of optic nerves and eyes is a difficult problem; researchers have found it hard to find solutions which work well over the wide range of image types which are produced in different studies. This is why the two approaches (of brain extraction and the use of a standard space mask) are combined to give the best possible results. However, even this combined approach is not always perfect, and the resulting possible “segmentation” error forms a part of the reported quantitative errors evaluated in the results section.

#### 3.2. Tissue Segmentation

Next, the tissue segmentation program described above (Zhang *et al.*, 2001) is used to segment the extracted brain image into grey, white, CSF, and background, giving a BV estimate. However, unlike the segmentation carried out for SIENA (where the exact positioning of the brain boundary, to subvoxel position, is not important, because it is the motion of the profiles around this position that matter and not the exact central point of the profiles), the exact volumes of the different tissues are now very important. Thus the segmentation method includes estimation of partial volume effects (at all voxels, and not just at “edge” voxels), giving higher volumetric accuracy than a “binarized” segmentation. This is achieved by modelling the distributions of the intensity in each tissue class, and using these models to estimate partial volume effect for any particular voxel, given its intensity and its neighborhood.

This segmentation is actually carried out on the original extracted brain image, not the normalised one. This is so that no interpolation has been applied to the image, which would slightly degrade the image and lead to slightly less accurate segmentation. (The eye and brain stem masking discussed above is actually carried out on this image, by applying the reverse normalisation registration transform to the standard space mask to bring it into register with the input image.)

Thus the segmentation gives a total volume for brain tissue (BV). This is multiplied by a volumetric scaling factor,<sup>5</sup> to give normalised brain volume (NBV). The NBV can optionally be split into grey and white volumes.

## 4. VALIDATION/RESULTS

#### 4.1. SIENA—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, with a range of slice thicknesses taken during each session, to enable the dependence on slice thickness of the accuracy to be determined (this data is also used, in Section 4.4, to test SIENAX). The 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 1 mm to 6 mm slice thickness, T1-weighted axial 2-D fast field echo, TE = 11 ms, TR = 35 ms, flip = 40°, NAqc = 1. The 1-mm scan lasted 18 min, and each successive scan took less time,

<sup>5</sup> This scaling factor is derived from the normalisation transform matrix—a simple formula relates the 12 affine registration parameters to a volumetric scaling. This is very close to simply multiplying the three linear scaling factors together.

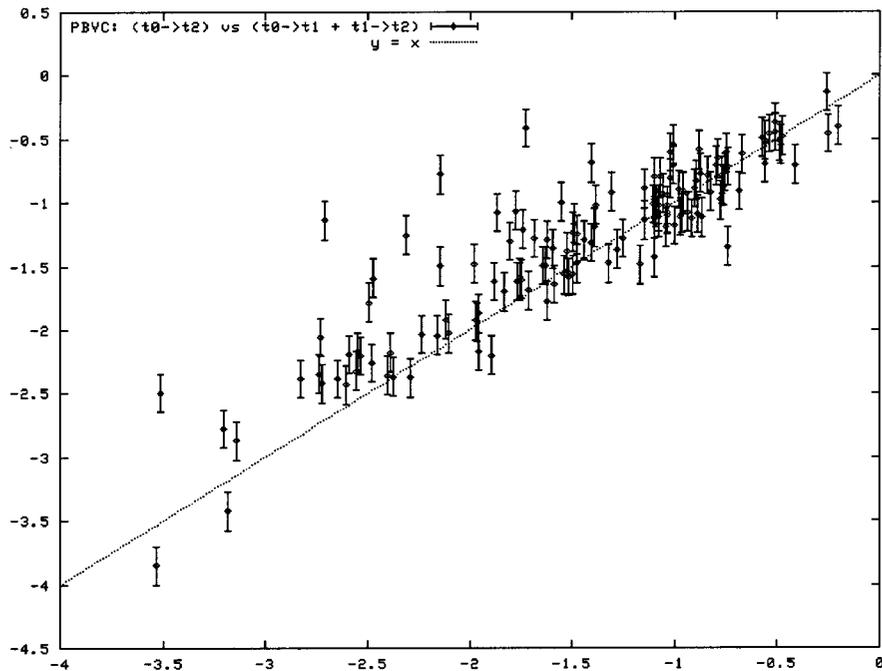


FIG. 5. Plot of (PBVC from  $t_0$  to  $t_2$ ) vs (PBVC from  $t_0$  to  $t_1$  plus that for  $t_1$  to  $t_2$ ).

with the 6-mm scan lasting 3 min. Scan 7 was a 3-mm slice thickness axial volumetric fast field echo, TE = 3 ms, TR = 20 ms, flip = 30°, NAqc = 1, and lasted 4 min. Scan 8 was the same as scan 7, but with coronal slices, lasting 4 min. For all scans the in-slice resolution was 1 by 1 mm, and enough slices were taken to include the top of the scalp and the bottom of the cerebellum. The intersession interval was mostly between 1 and 7 days. Half of the subjects were scanned with the slice thickness order reversed, to control for order effects.

The resulting 128 pairs of images were processed with SIENA, with no manual intervention. 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.

All PBVC measures should ideally be zero, as the subjects should not be showing any atrophy over such a short time interval. There are two clear results from the analysis. First, there is no clear slice-dependence to the errors. Second, the error in PBVC is small—the median absolute error over all results is less than 0.15% (a reduction of 0.05% from the median error of 0.2% found using the original version of SIENA, reported in Smith *et al.*, 2001). The fact that thinner 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.

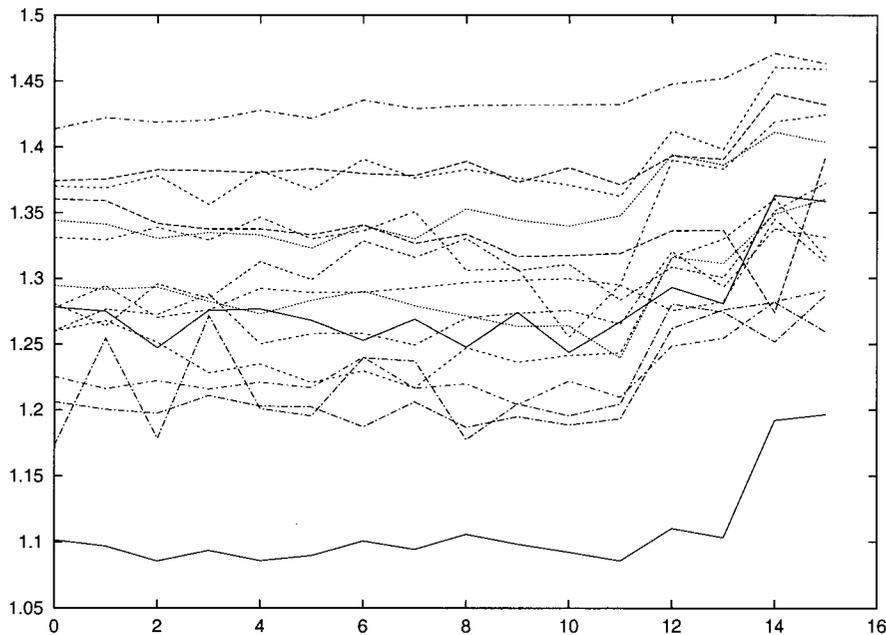
The contribution of the skull-based step in the registration was also investigated; the error introduced by this step was nearly as large as the total error, suggesting that the skull-based registration on average may contribute fairly highly to overall atrophy measurement error. However, this step is 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 0.4%, despite the differences between the images in each pairing.

The final outcome of these investigations, therefore, is that the error in measuring PBVC between images acquired using the same pulse sequence is around 0.15. This value is not strongly slice-thickness dependent.

#### 4.2. SIENA—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 (Fox and Freeborough, 1997) can be carried out on such data; the atrophy measure from the first time point ( $t_0$ ) to the second ( $t_1$ ) is added to the measure from the second time point ( $t_1$ ) to the third ( $t_2$ ), and this sum compared with the direct measure from  $t_0$  to  $t_2$ . This will show up most sources of



**FIG. 6.** Plots (one plot for each subject) of volumetric scaling factor ( $y$ ) derived from normalization vs different slice thicknesses ( $x = 0, 1$  corresponds to the two sessions' 1-mm scans, 2,3 to 2-mm; 4,5 to 3-mm; 6,7 to 4-mm; 8,9 to 5-mm; 10,11 to 6-mm; 12, 13 to the 3-mm volumetric scan; and 14, 15 to the 3-mm volumetric coronal readout scan).

error in the atrophy estimation procedure 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  $t_0$  to  $t_1$  atrophy measure in the same way that it affected the  $t_0$  to  $t_2$  measure, this would not show in the three-time-point analysis.

Previously, MR images of brains of 39 multiple sclerosis patients were used in this manner to show that the error in PBVC estimation on patient data agrees with the estimate derived from the test-retest normals data (Smith *et al.*, 2001). A second data set has more recently been analysed using 141 ageing volunteers (mean age 75, standard deviation 6 years), courtesy of the Challenge/Optima project, Drs. Kevin Bradley and Marc Budge, Radcliffe Infirmary, Oxford. The results are shown in Fig. 5; points should ideally lie on the  $y = x$  line. The error bars show  $\pm 0.15\%$ , and are sufficient to explain the majority of the deviations from the line (the median distance from the line is 0.22%), demonstrating that the precision of the method as applied to "active atrophy" data sets is comparable to that with normal controls (other data sets have shown even closer agreement with the "Siena normals" error estimation). Note that there is a slight bias above the line for higher atrophy rates, suggesting that higher atrophy is slightly underestimated compared with lower rates. A possible explanation is in the Bayesian weighting against large edge motion (useful for reducing error

and thus increasing sensitivity)—this tradeoff is being investigated further.

#### 4.3. SIENA—Intersite Image Testing

As an extreme test of the robustness of SIENA to changes in imaging parameters, three images of the same subject, each using a different MR scanner, were used in an A-B-C-A atrophy estimation test. Clearly the atrophy results from these three tests should sum to zero. Two scans were taken at 1.5T with 3-mm slices (NMR Center, University of Siena and Montreal Neurological Institute) and one at 3T with 3.5-mm slices (FMRIB, Oxford). The scans span a period of 1 year. The mean absolute PBVC estimated is 1.2%, and the summed PBVC is 0.3%.

#### 4.4. SIENAX—Investigation of Error as a Function of Slice Thickness

To test the error of SIENAX, the data from the 16 normal volunteers described in Section 4.1 was used. Every pair of within-subject same-slice-thickness images was used in a test-retest of both BV and NBV. The mean error in BV (i.e., due to BET and segmentation only) was 0.4% (this error is expressed as a percentage of brain volume), and the mean error in NBV (i.e., for the whole process including normalization) was 1%. Thus, as expected, the normalization step increases test-retest error. However, the value of the normalisation step (by reducing cross-subject variability) can be

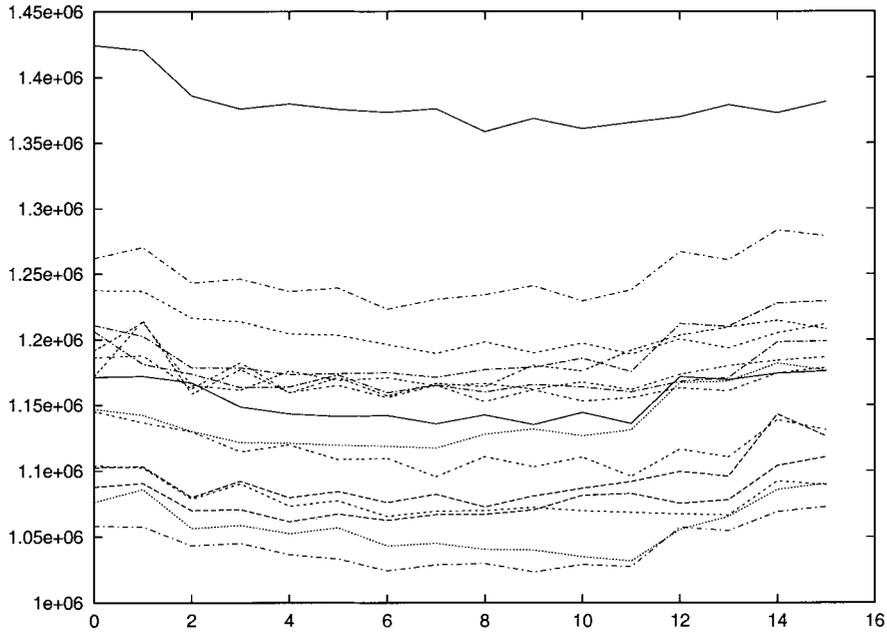


FIG. 7. Plots of BV vs different slice thicknesses.

seen when the standard error of the mean brain volume for the group is estimated with and without normalisation;  $SE(BV)$  for the group is 1.5% and  $SE(NBV)$  is 0.7%.

A more detailed investigation of these results is now described. First, Fig. 6 shows separate plots for the different subjects of the volumetric scaling factor derived from the normalization step. The different slice

thicknesses are plotted in  $x$  in the same order as described in 4.1 there are two separate values for each slice thickness (one for each time point). The vertical spread of the different subjects' plots shows the variation in head size within this group. The relative constancy of each plot (particularly for the first 6 slice thicknesses) is encouraging. The slightly different results for the final two slice thicknesses ( $x$  values of 12

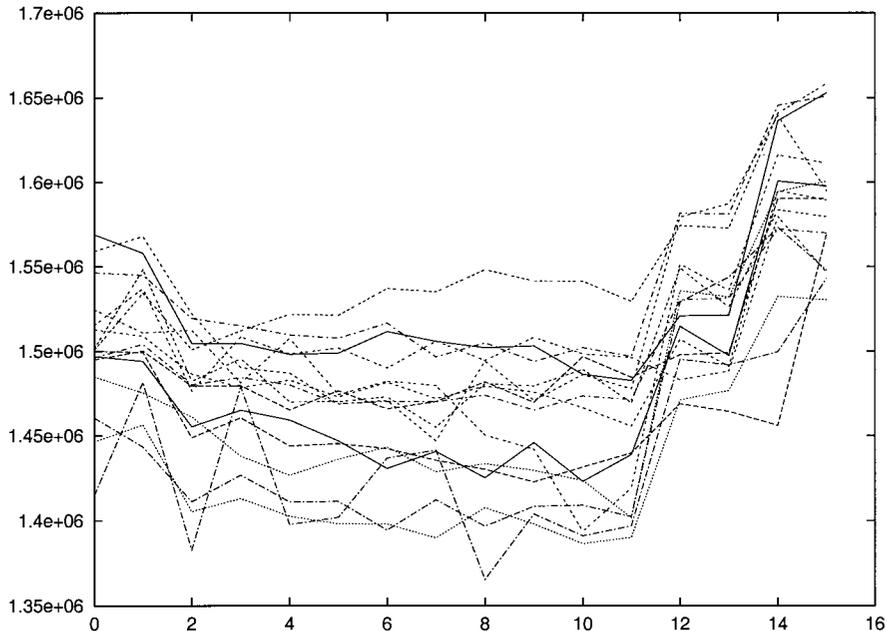
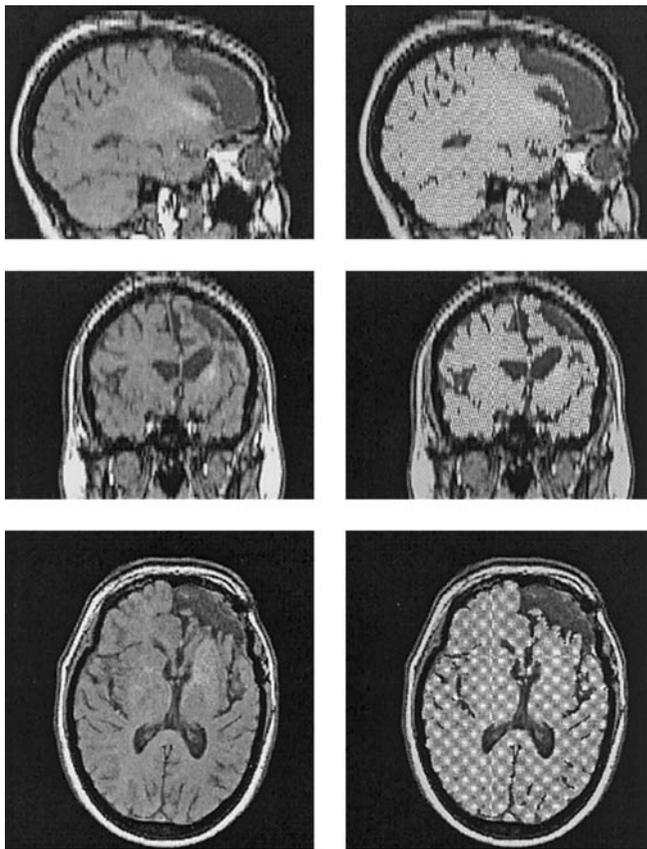


FIG. 8. Plots of NBV vs different slice thicknesses.



**FIG. 9.** SIENAX output for an epilepsy patient following frontal lobe surgery.

to 15) are still less than the spread across subjects, and are presumably due to the quite different sequences used for these images (including possibly slightly different imaging geometry calibration).

Figures 7 and 8 show plots of estimated BV and NBV. The BV estimates are more consistent (within subject) than the NBV estimates, as errors due to the normalization have not been added. However, the value of the normalisation can be seen by noting that the fractional spread across subjects is reduced by a factor of at least 2.

#### 4.5. SIENAX—Extreme Parenchyma Loss

Figure 9 shows the SIENAX output for an epilepsy patient following frontal lobe surgery. SIENAX has successfully coped with both the brain extraction and tissue segmentation. The NBV for this patient is estimated at  $1.26 \times 10^6 \text{ mm}^3$  (compared with  $1.45 \times 10^6 \text{ mm}^3$  in a relevant control group.)

## 5. CONCLUSION

This paper presents SIENA, a fully automated method of longitudinal (temporal) brain change analy-

sis, and an extension to a new method, SIENAX, for cross-sectional (single time point) analysis. SIENA is useful, for example, for longitudinal studies where maximal sensitivity to change over time is required. SIENAX is useful, for example, for differentiating two groups of subjects on the basis of single time point brain size measurement.

The methods are fully automated, robust and accurate: 0.15% brain volume change error (longitudinal) and 0.5–1% brain volume accuracy for single-time point (cross-sectional).

The SIENA and SIENAX software is freely available as part of the FMRIB Software Library (FSL) from the [www.fmrib.ox.ac.uk/fsl](http://www.fmrib.ox.ac.uk/fsl) website.

## ACKNOWLEDGMENTS

S.M.S. and P.M.M. acknowledge funding from the MRC (UK) to support the FMRIB centre; M.J. and N.D.S. were funded by the EC Biomed II programme (MICRODAB).

## REFERENCES

- Evans, A., Collins, L., Holmes, C., Paus, T., MacDonald, D., Zijdenbos, A., Toga, A., Fox, P., Lancaster, and J., Mazziotta, J. 1997. A 3D probabilistic atlas of normal human neuroanatomy. In *Third Int. Conf. on Functional Mapping of the Human Brain*, p. 349.
- Evans, A. C., Collins, D. L., and Holmes, C. J. 1996. Computational approaches to quantifying human neuroanatomical variability. In *Brain Mapping: The Methods* (J. C. Mazziotta and A. W. Toga, Eds.), pp. 343–361. Academic Press.
- Fisher, E., Cothren, R. M., Tkach, J. A., Masaryk, T. J., and Cornhill, F. 1997. Knowledge-based 3D segmentation of the brain in MR images for quantitative multiple sclerosis lesion tracking. In *SPIE Proc. Medical Imaging: Image Processing*, pp. 19–25.
- Fox, N. C., and Freeborough, P. A. 1997. Brain atrophy progression measured from registered serial MRI: Validation and application to Alzheimer's disease. *J. Magn. Reson. Imag.* **7**: 1069–1075.
- Fox, N. C., Freeborough, P. A., and Rossor, M. N. 1996. Visualisation and quantification of rates of atrophy in Alzheimer's disease. *Lancet* **348**: 94–97.
- Freeborough, P. A., and Fox, N. C. 1997. The boundary shift integral: An accurate and robust measure of cerebral volume changes from registered repeat MRI. *IEEE Trans. Med. Imag.* **16**(5): 623–629.
- Freeborough, P. A., Woods, R. P., and Fox, N. C. 1996. Accurate registration of serial 3D MR brain images and its application to visualizing change in neurodegenerative disorders. *J. Comput. Assist. Tomogr.* **20**(6): 1012–1022.
- Ge, Y., Grossman, R. J., Udupa, J. K., Wei, L., Mannon, L. J., Polansky, M., and Kolson, D. L. 1999. Longitudinal quantitative analysis of brain atrophy in relapsing-remitting and secondary-progressive multiple sclerosis. In *International Soc. of Magnetic Resonance in Medicine*.
- Hajnal, J. V., Saeed, N., Oatridge, A., Williams, E. J., Young, I. R., and Bydder, G. M. 1995a. Detection of subtle brain changes using subvoxel registration and subtraction of serial MR images. *J. Comput. Assist. Tomogr.* **19**(5): 677–691.
- Hajnal, J. V., Saeed, N., Soar, E. J., Oatridge, A., Young, I. R., and Bydder, G. M. 1995b. A registration and interpolation procedure for subvoxel matching of serially acquired MR images. *J. Comput. Assist. Tomogr.* **19**(2): 289–296.

- Jenkinson, M., and Smith, S. M. 2001. A global optimisation method for robust affine registration of brain images. *Med. Image Anal.* **5**(2): 143–156.
- Lemieux, L., Wieshmann, U. C., Moran, N. F., Fish, D. R., and Shorvon, S. D. 1998. The detection and significance of subtle changes in mixed-signal brain lesions by serial MRI scan matching and spatial normalization. *Med. Image Anal.* **2**(3): 227–242.
- Reddick, W. E., Glass, J. O., and Langston, J. W. 1999. A non-invasive MRI measure of subtle longitudinal volume changes in brain. In *International Soc. of Magnetic Resonance in Medicine*.
- Rudick, R. A., Fisher, E., Lee, J. C., Simon, J., and Jacobs, L. 1999. Use of the brain parenchymal fraction to measure whole brain atrophy in relapsing-remitting MS. *Neurology* **53**(8): 1698–704.
- Sgouros, S., Natarajan, K., Hockley, A. D., Goldin, J. H., and Wake, M. 1999. Skull base growth in childhood. *Pediatr. Neurosurg.* **31**(5): 259–68, 1999.
- Smith, S. M. 2000a. Robust automated brain extraction. In *Sixth Int. Conf. on Functional Mapping of the Human Brain*, p. 625.
- Smith, S. M. 2000b. Robust automated brain extraction. *Hum. Brain Mapp.*, in press.
- Smith, S. M., De Stefano, N., Jenkinson, M., and Matthews, P. M. 2001. Normalised accurate measurement of longitudinal brain change. *J. Comput. Assist. Tomogr.* **25**(3): 466–475.
- Zhang, Y., Brady, M., and Smith, S. 2001. Segmentation of brain MR images through a hidden Markov random field model and the expectation maximization algorithm. *IEEE Trans. Med. Imag.* **20**(1): 45–57.