A Fast, Model-Independent Method for Cerebral Cortical Thickness Estimation using MRI

M.L.J. Scott, P.A. Bromiley, N.A. Thacker, C.E. Hutchinson and A. Jackson

Last updated
11 / 11 / 2008

WWW.TINA-VISION.NET

Imaging Science and Biomedical Engineering Division,
Medical School, University of Manchester,
Stopford Building, Oxford Road,
Manchester, M13 9PT.

A Fast, Model-Independent Method for Cerebral Cortical Thickness Estimation using MRI

M.L.J. Scott, P.A. Bromiley, N.A. Thacker, C.E. Hutchinson and A. Jackson
Imaging Science and Biomedical Engineering Division
Medical School, University of Manchester
Manchester, M13 9PT, UK
`paul.bromiley@manchester.ac.uk`

Abstract

Several algorithms for measuring the cortical thickness in the human brain from MR image volumes have been described in the literature, the majority of which rely on fitting deformable models to the inner and outer cortical surfaces. However, the constraints applied during the model fitting process in order to enforce spherical topology and to fit the outer cortical surface in narrow sulci, where the cerebro-spinal fluid (CSF) channel may be obscured by partial voluming, may introduce bias in some circumstances, and greatly increase the processor time required.

In this paper we describe an alternative, voxel based technique that measures the cortical thickness using inversion recovery anatomical MR images. Grey matter, white matter and CSF are identified through segmentation, and edge-detection is used to identify the boundaries between these tissues. The cortical thickness is then measured along the local 3D surface normal at every voxel on the inner cortical surface. The method was applied to 119 normal volunteers, and validated through extensive comparisons with published measurements of both cortical thickness and rate of thickness change with age. We conclude that the proposed technique is generally faster than deformable model based alternatives, and free from the possibility of model bias, but suffers no reduction in accuracy. In particular, it will be applicable in data sets showing severe cortical atrophy, where thinning of the gyri leads to points of high curvature, and so the fitting of deformable models is problematic.

1 Introduction

The human cerebral cortex makes up the largest part of the brain, and consists of a highly convoluted layer of neuronal cells with the topology of a 2D sheet, surrounding a core of white matter. Its thickness varies considerably, from approximately 2mm in the calcarine sulcus to approximately 4mm in the precentral gyrus, with an average of approximately 3mm [3, 10, 15, 16]. Changes in cortical volume have been observed to correlate with both normal brain maturation and a variety of cortical pathologies. For example, grey matter (GM) volume loss is seen throughout adulthood to old age [42, 6], and Alzheimer's disease subjects show a significant reduction in cortical volume relative to controls [8], particularly in the medial temporal lobe, with the extent of the decrease correlating with greater cognitive decline.

Although they are three-dimensional, measurements of cerebral volumes give little or no information about the dimensions of the structure under investigation, whereas cortical thickness measures present a method for continuous measurement across the cortical surface. They have shown considerable potential both in the study of normal brain growth and maturation [25, 44, 42, 35, 48] and in the diagnosis or measurement of the progression of a wide variety of cortical pathologies, including Huntington's disease [33, 39] Multiple Sclerosis [34], Schizophrenia [18], depression in elderly patients [2], cerebral microangiopathy [30], Alzheimer's disease [46], Williams syndrome [47], attention-deficit-hyperactivity disorder [48], and fetal alcohol syndrome [48].

All cortical thickness measurement techniques require two core components: a method for locating the inner and outer cortical surfaces, and a method to measure the distance between them. However, the thickness measure chosen forms the definition of cortical thickness: the accuracy with which the measurement is made is dictated by the method chosen to define the cortical surfaces. The techniques presented in the literature can be coarsely divided into two groups in this respect: model-based approaches that involve fitting deformable models to the inner and outer cortical surfaces i.e. the boundaries between white matter (WM) and grey matter (GM) and GM and pia matter (PM), and data-driven approaches that detect these interfaces using image intensities alone. Model-based approaches typically involve segmenting the WM, fitting the model to the WM/GM interface, and then expanding the model to the GM/PM interface. The inner cortical boundary is smoother, and so fitting this boundary first

helps to guide the model into the more complex folding of the outer boundary. The model itself typically consists of a set of vertices that tessellate the cortical surface with triangular elements. The Freesurfer algorithm ([7, 11, 33], surfer.nmr.mgh.harvard.edu) fits the models using an objective function consisting of an intensity term to identify the boundary and a pair of constraints that impose local smoothness and regularise the tesselation. However, it is also desirable to prevent self-intersections in the models, ensuring that they have a simple, spherical topology, which in turn allows the surface and subsequent thickness measurements to be mapped onto a plane or sphere using standard projections. This can be advantageous both in terms of displaying the results and in the definitions of the thickness measures used. Self-intersections can be prevented either by explicitly removing them when they occur (e.g. [10]) or by introducing additional terms into the objective function. The latter approach is used in the ASP (Anatomic Segmentation using Proximities) algorithm [24], which introduces both a surface self-proximity term and a term based on the distance between corresponding vertices on the inner and outer surface models. Such terms may also be required in order to solve the notoriously difficult problem of fitting the model in tightly folded gyri [17, 47], where the sulcal banks oppose so closely that there is no clear cerebrospinal fluid (CSF) channel in the sulcus at the resolution of typical MR images. In such regions the outer surface model may fail to fit the pial surface within the sulcal fundus.

The deformable model based approach has several drawbacks. First, such algorithms require considerable computational resources, largely due to the topological constraints. The algorithms presented in [24] and [10] required 30 hours on a 180 MHz Silicon Graphics R10000 and 5 hours on a 500 MHz Pentium 3 respectively to process each image volume. Performance figures for modern hardware are available from the Freesurfer website, indicating processing times of the order of hours per image volume. Second, the use of topological constraints may bias the measured thickness [24, 17]. In particular, the constraint on the distance between corresponding vertices of the inner and outer surface models used in ASP penalises corresponding vertices as they deviate from an ideal distance, biasing the algorithm towards a fixed separation between the inner and outer cortical surface models and thus reducing the statistical power of the technique to identify abnormal cortical thicknesses. Smoothness constraints may be problematic at points of high curvature such as the gyral crowns and sulcal fundi [21], although imposing only second-order smoothness can alleviate this problem [10]. Kim et al. [17] have proposed an updated version of ASP (CLASP, Constrained Laplacian ASP) in which several of these constraints, most importantly that on the inter-surface distance, were eliminated. However, a third issue of selecting weights for each constraint term remains: there is no guarantee that weights obtained by optimising the performance on images from one group of subjects, for example normals, will be applicable to images from groups with significantly different cortical thicknesses resulting from pathological processes or even normal ageing in groups with mean ages significantly different from that used to optimise the weights.

These problems have led to an interest in data-driven techniques, in which the inner and outer cortical surfaces are determined using only image intensities (e.g. [21, 47, 22]). Such approaches also typically begin with segmentation of the WM, GM and CSF. Thickness measurements are then performed at each point on the inner cortical surface, by propagating away from the surface according to some thickness measure until the outer cortical surface is reached. In regions of tightly folded gyri as described above, if the pial surface is missed and another point on the inner cortical surface is reached, the thickness measurement can be halved to produce a value that is approximate but still entirely data driven. It has been shown that this approach has little effect on final, regional thickness measurements if local smoothing is applied [47, 23, 22]. Some authors have also combined the two approaches by fitting deformable models in order to define homologous points on the cortical surfaces across different subjects and using these in registration, for defining the points at which thickness measurements are made, and to define a surface on which the results may be displayed, whilst still using only image intensities in the thickness measurement itself [43, 22].

The second requirement for cortical thickness measurement is a definition of thickness. The distance between the corresponding points on the two surface models can be used in model-based approaches. Alternatively, the minimum distance between a given point on one surface and the closest point on the opposing surface or the distance along the normal to the inner surface can be used [24]. Lerch et al. [20] showed using a population simulation that the former method is the most sensitive in recovering true change in thickness and also minimised the variability across a population, and recommended this measure (out of 6) as the one to use. However, more complex approaches have also been attempted. For example, the inner and outer cortical surfaces can be treated as charged conductors and the Laplace equation applied to define hypothetical electric field lines between them [15]. Both cortical surfaces must be closed (i.e. have spherical topology) in order to apply this technique, and so a method to cap the hole introduced by the brainstem (such as described by [7] and [49]) is required. The detection of the surfaces must also be complete, which can be problematic in data with partial volume effects. However, the field lines are guaranteed to be diffeomorphic i.e. they do not cross. The thickness of the cortex at any given point then becomes the sum of the lengths of the surface normals at each field line, where the start point at each field line is the end of the surface normal from the previous field line. This ensures that every point has a unique

and reversible point on the opposing surface. The assumed model also has some analogy with reality [1], since the cortical mantle is divided into 6 layers. Yezzi et al. [50] present a generalised version of the Laplacian approach, in the form of a pair of first-order, linear partial differential equations, together with an efficient numerical solution allowing cortical thicknesses to be measured in 3 minutes on a 800MHz Pentium 3 processor. Other algorithms such as marching cubes or the Eikonal fire equation [36] have also been applied [43, 22]. Finally, [26] proposed a rather different approach that dispenses with the need for an explicit distance metric; a Bayesian segmentation was used to identify GM, WM and CSF, and a tesselation of the GM/WM surface was generated. The distance between each image voxel and this surface was then found, and the segmentation result used to generate the probability that a voxel would be a member of each tissue class as a function of distance from the GM/WM surface.

The aim of the work described here was to produce a fully automated cortical thickness measurement algorithm that does not depend on the use of deformable models of the inner and outer cortical surfaces. The inner cortical surface is located using edge detection, and the local 3D surface normal at each voxel obtained. A search is then performed along this vector in order to locate the outer cortical surface, including measures to deal with situations in which the outer surface is obscured by partial voluming, thus measuring the cortical thickness. The removal of the requirement to model the cortical surfaces has several advantages. It reduces the processor time required by an order of magnitude; typical execution times for model-based algorithms are, as described above, of the order of hours per image volume, whereas the algorithm described here processed the image volumes used in this study in an average of 8 minutes per volume. It also removes the need to empirically estimate weights for terms in the model fitting objective function. The algorithm is therefore more suited to widespread application of morphometric techniques in clinical imaging. In addition, it avoids the possibility of biases due to model constraints; this may increase the statistical power of the method to detect changes in cortical thickness across groups. The approach has some similarities with that proposed by [21] and used by [30] and [29]. However, those authors used a Euclidean distance transform to measure thickness, and thus obtained the shortest distance between the inner and outer surface, as opposed to the normal distance measure used here.

Adopting the normal distance may have several advantages in data-driven approaches to cortical thickness measurement. First, calculation of the shortest distance requires accurate delineation of both the inner and outer surfaces, which may introduce problems at points where the outer cortical surface is obscured by partial voluming or noise. Calculation of the normal distance requires only delineation of the inner surface, which is in general an easier task than location of the outer surface, thus removing a potential source of error. The search along the normal direction can then take account of multiple possible scenarios, as described in more detail below, for example points where partial voluming obscures the outer surface and leaves only a dip in intensity rather than a detectable edge. Furthermore, in regions of the cortex where there is a maximum in thickness, such as at gyral crowns, use of the shortest distance may produce a systematic under-estimation of the thickness by preferentially locating the shortest possible vectors along which to make the measurement. Such an effect would manifest as a bias towards lower thickness values in the final result, potentially reducing the ability of the algorithm to monitor pathological or age-related thickness changes. Finally, [24] found the normal distance to be more similar to the vertex-vertex distance commonly used in deformable model based approaches than the shortest distance, and so the proposed method presents an advantage in terms of facilitating comparisons between model-based and data-driven methods.

2 Method

Fig. 1 gives an overview of the algorithm. The processing of the data can be divided into two stages, the actual cortical thickness estimation, and the pre-processing to convert the original volume of data into the required form. Each stage of the algorithm is described in detail below.

2.1 Subjects and Scan Parameters

The subject group used in the validation of the algorithm was composed of groups from three different studies, scanned on different occasions on the same scanner (1.5 Tesla whole-body scanner Philips ACS PT 6000 NT, Philips Medical Systems, Best, The Netherlands). Scan parameters and demographics are given in Table 1. The acquisition used was a T1 inversion recovery sequence. All subjects were normal volunteers and gave informed consent at least 24 hours before scanning. The Central Manchester Local Research Ethics Committee approved all three studies. Group 1 were part of a perfusion reproducibility study, Group 2 were normal volunteers in an ME study and Group 3 were part of a Wellcome study on amnestic cognitive impairment. The principle reason for using a range of age groups was to investigate the effect of age on cortical thickness. Various calibration stages of the algorithm were performed on a subset (18 subjects) of the data, namely the entirety of Groups 1 and 2, and five representative subjects from Group 3 spanning the age range of the dataset. Although the scans were acquired
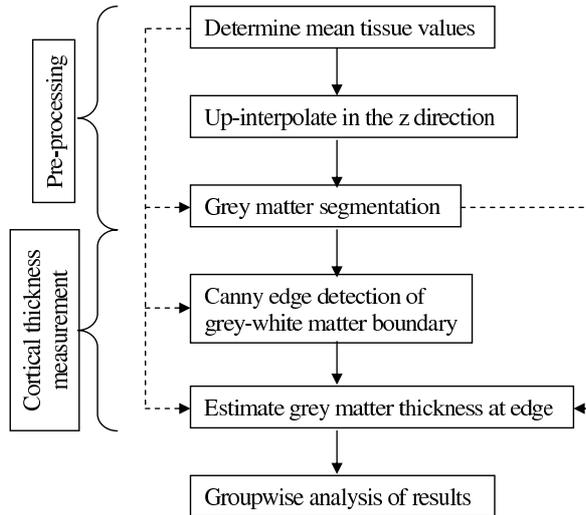
Figure 1: Overview of the algorithm, including pre-processing stages. Solid-line arrows illustrate the flow of operations performed, dashed-line arrows illustrate where the result of one stage of the processing is used as an input to a later step.

for the different groups at different times, [13] have demonstrated using a Siemens scanner that scanner upgrades do not increase variability or introduce bias.

The imaging protocol was specifically selected in order to maximise signal to noise for grey and white matter separation and to minimise the effects of field inhomogeneity. The segmentation and edge-detection stages of the proposed algorithm both make specific assumptions about the absolute image intensities, and so would be sensitive to image inhomogeneity. However, both are applied only to individual slices, rather than to entire image volumes simultaneously, and so inter-slice inhomogeneity can be ignored for the purposes of the algorithm. Intra-slice inhomogeneity was evaluated using the N3 algorithm [41]: application of this algorithm to five image volumes randomly selected from those described above indicated mean intra-slice inhomogeneities of approximately 2%, which was insignificant compared to the image noise. This removed the need for the application of an inhomogeneity correction algorithm prior to data processing and thus simplified the analysis of the results, since no consideration of the inhomogeneity correction was needed in the analysis of the performance of the algorithm. However, several authors (e.g. [24]) have presented cortical thickness measurements performed on inhomogeneity corrected data. This implies that the algorithm presented here will be applicable to other MR image types following the application of a suitable correction algorithm. The scanning protocols used are typical of scans acquired clinically, in that the in-plane resolution is sub-millimetre, whereas the through-plane resolution is of the order of a few millimetres, in order to increase the signal to noise ratio of the data. From a purely geometrical perspective, if it is possible to show that the method presented below produces reliable results when applied to such data, then there is no reason to expect that it will not work on more research-oriented protocols that may have isotropic voxels: the reverse may not be the case, hence the use here of the clinical scans. Reducing the slice thickness will inevitably reduce the signal-to-noise ratio, but this will manifest itself as an increase in the random errors on the results, rather than the introduction of a systematic error.

2.2 Data Pre-Processing

The first of the data preprocessing stages involved application of the partial volume segmentation algorithm described by [28]. The algorithm fitted a model consisting of Gaussian distributions for each pure tissue and uniform distributions, convolved with Gaussian distributions representing the noise process, for each partial volume class, to the intensity histogram of the image volume. This resulted in an estimate of the mean and standard deviation of each pure tissue for use in later stages of the cortical thickness measurement algorithm. The histogram was constructed using voxel intensities within a block of the brain containing only WM, GM and CSF, manually defined in the Talairach coordinates and then transformed into the coordinate system of each image volume using the result of an affine registration. The position of the block is illustrated in Fig. 2. The positioning of the volume in the frontal lobe was selected in order to obtain a sufficient number of voxels of the three pure tissue types. In elderly brains, the differentiation between all three tissue types becomes less distinct for several reasons. The boundary between GM and WM appears much more diffuse than in younger brains (the gradient between the two becomes

Parameter	Group 1	Group 2	Group 3
Number (males)	4 (4)	9 (5)	106 (43)
Mean Age (range)	40.2 (34 - 46)	35.4 (19 - 53)	74.4 (60 - 86)
TI (ms)	300	300	300
TE(ms)	18	18	18
TR (ms)	6850	6850	6850
Flip Angle (degrees)	90	90	90
Echo Train Length	9	9	9
Slice Orientation	Axial	Coronal	Axial
In-plane Resolution (mm)	0.9x0.9	0.9x0.9	0.9x0.9
Slice Thickness (mm)	3.0	4.0	3.0

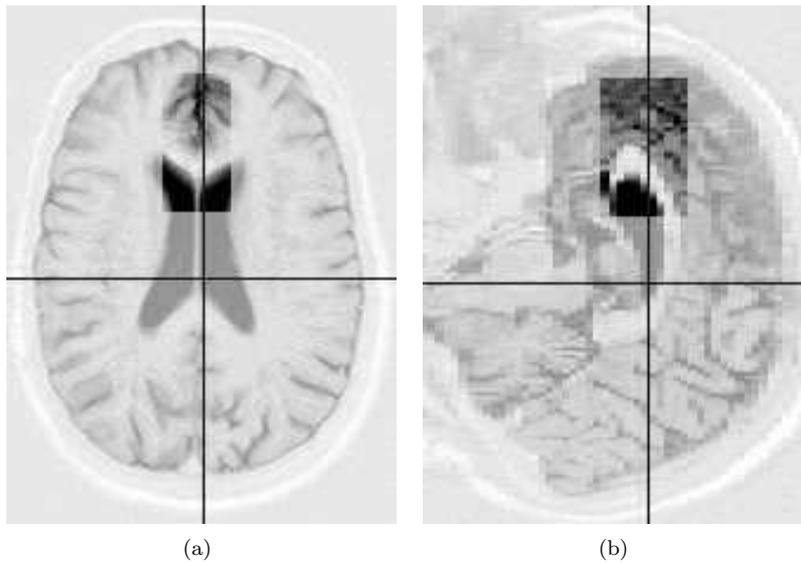Table 1: Subject demographics and scan parameters.

(a) (b)

Figure 2: Sections illustrating the approximate position of the cuboid, here shown in greater contrast than the surrounding tissue, from which the three-tissue histogram is constructed: (a) axial view, (b) sagittal view.
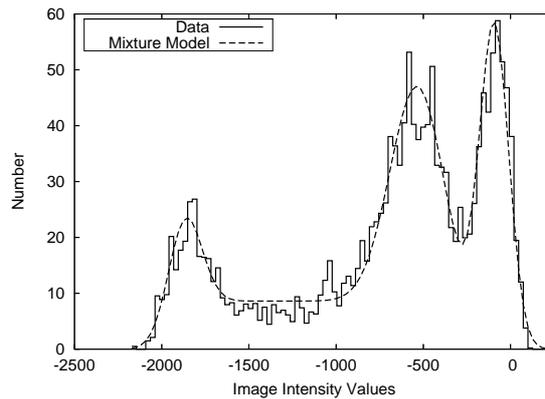

Figure 3: Histogram of voxel intensity values (solid line) and fit (dashed line) of the Gaussian mixture model to the histogram. There are clear peaks for the cerebro-spinal fluid, grey matter and white matter at approximately -1800, -500 and -100 grey levels, respectively.

much more shallow) presumably due to demyelination of the WM axons such that the WM appears more like the non-myelinated GM. In addition, the presence of fluid-filled lesions in WM may result in the apparent presence of GM in the WM, and the greater partial voluming of GM and CSF (due to increased CSF) results in the actual

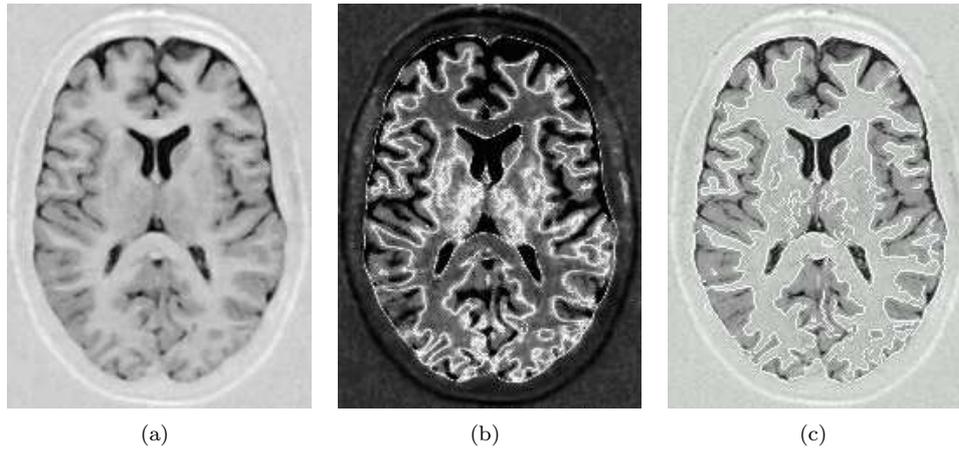|		
(a)	(b)	(c)

Figure 4: Example images from the edge-detection algorithm showing (a) the initial grey level image, (b) the likelihood enhancement image, (c) resulting edge detection for the same slice (overlaid in white). Note that, from (c) the edge detection appears to extend all the way to the back of the skull. The grey/white matter boundary is close to the skull here, and the likelihood image shows that both the skull/CSF boundary and the GM/WM boundary are highlighted. Due to their proximity and the fact that the edge detection algorithm as implemented here cannot cope with bifurcations, the detected edge follows the higher ridge through the image to the skull, rather than the slightly weaker GM/WM boundary.

GM appearing to have anomalous (lower) grey level values. In order to obtain realistic estimates of the pure tissue values, CSF values were taken from the ventricles, WM from the corpus callosum, where there did not appear to be much CSF/WM partial voluming, and GM from the cortex in front of the anterior cingulate, with relatively little CSF/GM partial voluming. This resulted in a histogram with well-defined peaks (see Fig. 3), which could be fitted in order to calibrate the mean tissue values.

Although not strictly necessary for algorithmic purposes, it was desirable to increase the degree of isotropy of the voxel dimensions in order to allow a single step size to be used in the search along the normal to the inner cortical surface, as described below, regardless of position within the image volume. Therefore, the data were up interpolated in the through-plane direction by a factor of two. The partial volume grey level analysis described above allowed estimation of the fractional volumetric contributions of pairs of tissues to partial volume voxels. This identifies a set of possible boundary locations within the voxel, differing in which surfaces of the voxel they intersect with. The most probable boundary location was then selected from amongst these possibilities according to the best agreement with the adjacent slices. Up-interpolation by a factor of two in the inter-slice direction was performed using this model.

The final pre-processing stage involved production of GM maps from the up-interpolated data. The intensity model fitted in the first stage of the algorithm was used to determine the most likely volumetric contribution of each tissue to each voxel. The intensity model described only GM, WM and CSF; however, since the other tissues present (e.g. fat, bone etc.) had mean intensities higher than that for white matter for the MR images used here, whilst GM and CSF had lower intensities, this did not introduce the possibility of confusing GM with these other tissues. The GM boundary was defined at voxels representing 50% GM by volume.

2.3   Cortical Thickness Measurement

Identification of the inner cortical surface (i.e. the GM/WM interface) was performed using an edge-detection algorithm, based on an edge-enhancement image. The process is described in detail in [38]; it is summarised here. The boundary contrast at the inner cortical surface in the MR image volumes used here approached the noise level (see Fig. 3 and Fig. 4a), and so this boundary was not enhanced in a gradient-based edge-enhancement image. An alternative approach was adopted based on the observation that, in the absence of significant inhomogeneity, the intensity of a voxel containing 50% WM and 50% GM should be consistent across each image. The segmentation algorithm applied in the earlier processing stages produced estimates of the mean intensities for GM and WM, allowing this intensity to be calculated. Taking this intensity as the definition of the inner cortical boundary, an edge-enhancement image was then produced containing the squared difference between the intensity of each voxel and this expected intensity, divided by the square of the image noise i.e. the log-likelihood that the voxel intensity was consistent with the expected edge intensity (see Fig. 4b). The image noise was estimated using the techniques
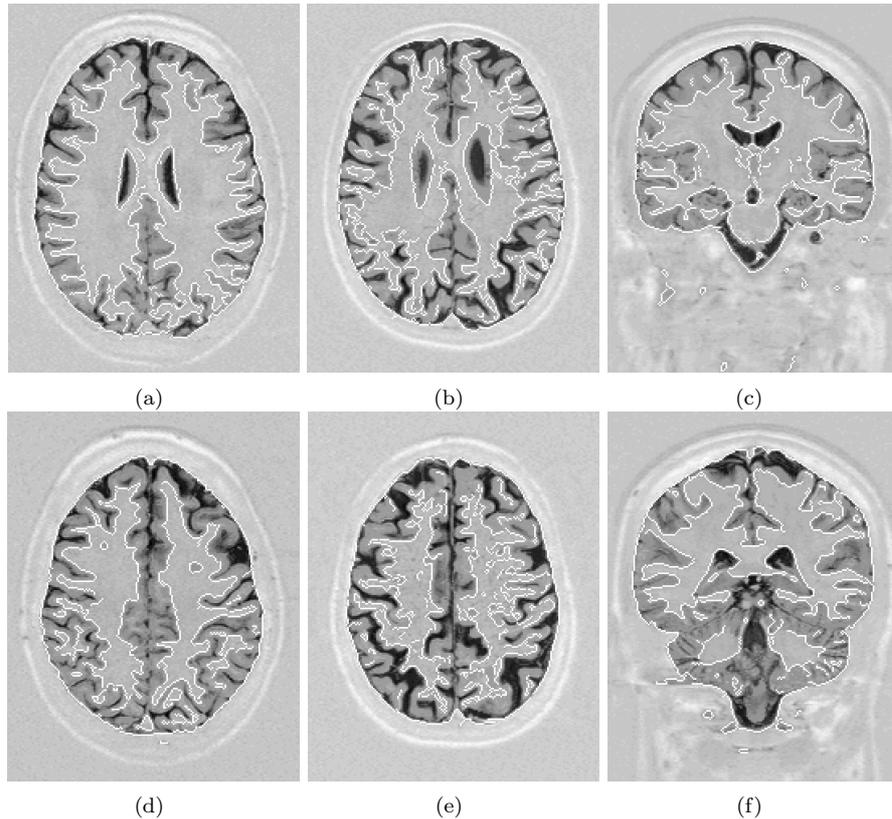
Figure 5: Example result images from the edge-detection algorithm, where the detected edges are represented as a white overlay on the images. Shown are 2 slices (chosen to illustrate the behaviour of the algorithm in the precentral gyrus) for each of the three subject groups in this study. (a) and (d) are from a 40 year old male in group 1, (b) and (e) are from an 81 year old female in group 3 and (c) and (f) are from a 45 year old male in group 2. Note that the skull/CSF boundary is erroneously detected as a grey/white matter boundary due to the similarity of grey-level intensity value of the partial volumed skull/CSF with the actual grey/white matter boundary value.

described by [27]. The Canny algorithm [5] was then applied to this edge-enhancement image in order to locate the boundary[1]. It has been shown [32] that edge detection algorithms of this form have typical positional accuracies of 0.1 pixels or less. Fig. 5 illustrates this process in example slices from three subjects, one from each group. Note that, due to the relative intensities of the various tissues in the images used, partial volume voxels on the CSF/skull boundary may have the same intensity as those on the GM/WM boundary, and so are also detected by this process.

The 3D normal to the surface at each voxel was then calculated. In order to reduce the noise on the calculation, 3D Gaussian smoothing was applied to the detected edges, using a kernel with a sigma of 1/2 voxel, truncated at 5 voxels. The kernel size was chosen in order to avoid introducing significant correlation between neighbouring voxels. The surface normal was then found by taking the local grey-level gradient in 3D, using spatial differentiation of the 6 adjacent voxels to the voxel containing the edge of interest. This procedure used the GM probability map from the segmentation to orient the search vector into the GM, also eliminating the false edge detections that did not lie on a GM boundary.

Measurements of the cortical thickness were made at each voxel on the inner cortical surface, using both the GM tissue volume maps and the GM/WM boundary locations and surface normals. A search was conducted along the normal direction in 3D, in 1mm steps (thus sampling at a resolution of approximately half the voxel size) until a second edge was detected on the GM volume map. An edge was defined as a point at which the GM contribution to the voxels dropped below 50% by volume, calculated using linear interpolation. The nature of the edge was then determined; in general, four cases could occur. First, at positions where the outer cortical surface was well delineated by an interface with CSF, a GM/CSF edge was detected; only in this case could this distance along the

---

[1]The Canny algorithm was applied with an upper threshold of 0.9 and a lower threshold of 0. One of the benefits of using a likelihood image as an edge enhancement image is that these values relate directly to the probability of a voxel being an edge. Note, however, that whilst the Canny parameters will affect the number of false positive and false negative edge detections, they cannot result in a systematic shift of the located boundary position, since this is dictated by the estimated edge intensity.
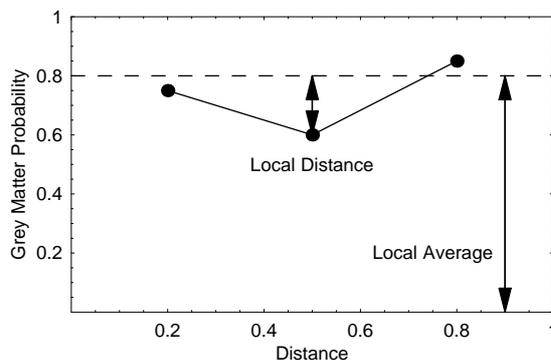
Figure 6: The calculation of the dip size. The dip size, or "local distance", is the difference between the grey matter volume probability at the position of interest and the average value of the grey matter volume probabilities of the positions 1mm either side ("local average"), along the direction of the surface normal to the grey/white matter edge.
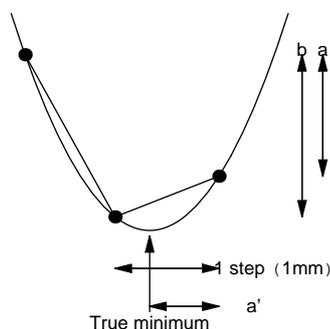

Figure 7: The calculation of the position of the trough of a dip, assuming that the dip can be modelled as a linear approximation to a quadratic curve. The fraction (a') of the step length (1mm) at which the dip occurs is equivalent to the fraction a/b.

search direction be used directly as the cortical thickness estimate. Second, at points such as those within deep sulci the CSF channel might be so narrow that, at the resolution of the MR images used here, it was represented only by partial volume voxel containing both CSF and GM, and thus visible as a dip in the GM volume rather than a definite edge. Therefore, the position of the greatest dip along the search orientation in the grey matter probability image was also recorded. Third, if the CSF channel were narrower still, or the dip obscured by noise, the outer cortical surface could be missed entirely. The search would then traverse both opposing sulcal banks and find the inner cortical surface again i.e. a GM/WM interface, or a dip in the GM volume introduced by WM partial voluming where there was a thin gyrus. Finally, a small number of searches, representing noise outliers on the calculation of the normal direction, reached unfeasibly long extents. These searches were terminated when they reached 20mm, a limit determined empirically in a calibration experiment described in Appendix 1.

A dip was defined as a position at which the GM volume was lower than the average of the GM volumes at the adjacent positions along the search direction. The value of the dip was calculated as the difference between the average and the value at the dip (local distance in Fig. 6), in order to take account of local intensities. The effect of placing a threshold on the minimum intensity change required to record a dip, in order to record only those that were statistically significant compared to the image noise, was investigated as described in Appendix 2. However, the resulting change in regional median cortical thickness values was found to be negligible, and so no such threshold was used in the final version of the algorithm. The exact position of the dip was defined as the minimum of the quadratic curve defined by the search point at which the dip was first located and those either side, as shown in Fig. 7. The position of the dip gave the centre of the CSF channel rather than the GM edge. However, the local distance was itself a tissue volume proportion, the volume of the voxel that was not composed of GM. Therefore, assuming that there were only two tissues in the voxel, that the CSF formed a contiguous block, and that this block was centred on the position of the dip, the actual GM edge was found by subtracting half of the local distance from the dip position, as shown in Fig. 8.
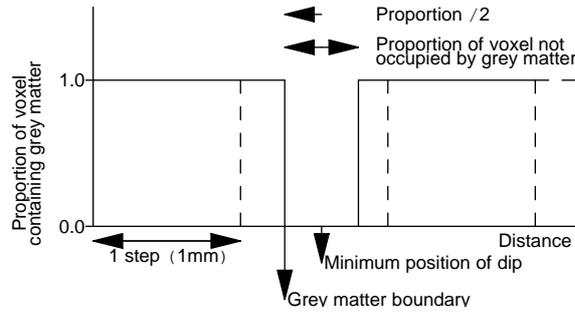
9

Figure 8: Diagram showing that the GM boundary at a partial volume dip is the distance in front of the calculated dip position equivalent to half of the proportion of the voxel not occupied by grey matter: ie, the distance is half of the value of the "local distance" value in Fig. 6.

Once the position of the edge/dip boundary had been determined, the value of the voxel in the original grey level image was assessed to determine whether the edge was a GM/CSF or GM/WM boundary. If the value was less than the mean grey level value it was a CSF boundary, and the edge/dip position remained intact. Otherwise, it was a WM boundary. It was therefore assumed that the opposing banks of a sulcus had been traversed. In this case, it was assumed that the actual thickness was half of the measured thickness i.e. that the opposing sulcal banks were of the same thickness. Whilst this will clearly be true on average over the whole cortex, it may not true on a regional basis. For example, it is well established that the anterior bank of the central sulcus is thinner than the posterior bank (e.g. [16]). However, dividing the thickness by two at such points was the only possible data-driven approach, since there was no information available from the data to support an alternative interpretation. Note that the same problem is present for model-driven approaches; the measurements produced at such points are entirely driven by the constraints applied during the model-fitting process. The benefit of the data-driven approach is that the division by two is only applied at these information-free points; the constraints applied in model-based approaches affect the measurement at all points within the image volume (see [24] for a discussion of this issue).

All of the processing described so far was based as closely as possible on individual voxel measurements. However, the fact that the surface of the cortex and the boundary between tissues can both be expected to be smooth surfaces was exploited in order to increase the stability of individual cortical thickness measurements. This was implemented by computing the median of the thickness measurements in five-point windows along the edge strings produced by the edge-detection algorithm, and comparing this to the actual thickness measurement at the centre of the window. If the difference between the two exceeded a set threshold, defined as 4mm, since this is expected to be the mean thickness in the thickest parts of the cortex [16], then the measured thickness was replaced by the median. In the majority of cases, this resulted in the identification and removal of spurious edges. In the 18 subject subset of data, the frequency with which edge points that had no measured thickness were assigned a non-zero value as a result of this process was 0.035%, the frequency with which edge points that had a measured thickness were replaced with zero was 0.085%, and 4.66% of existing values were changed. These frequencies were sufficiently low to prevent biasing the median thickness estimates. Note that this is not equivalent to imposing a thickness constraint, which may result in bias, since spuriously large thickness values were replaced with the local median, rather than a pre-determined thickness.

Numerical comparisons of the results against either gold standard, manual measurements, or the results from other cortical thickness algorithms, cannot proceed without defining a parcellation of the cortex or a set of points at which to perform the comparison. Therefore, validation was based on regional cortical thickness measurements, necessitating the registration of the data into a common stereotaxic space. The up-interpolated image volumes were automatically registered to an average brain in the Talairach space [45] generated from the Talairach Daemon [19], additionally providing structural labels for subsequent regional parcellation of the thickness measurements. A 9-parameter similarity transform was used, consisting of translation, rotation and scaling; the objective criterion used was a χ^2 similarity metric based on the alignment between edges in the images. Unlike the original definition for the location of this space, the brains were not registered specifically using the line connecting the anterior and posterior commisures. Instead, features from the entire brain were used in order to obtain a registration where the alignment of the brain boundary (hence alignment of the cortical structures to the correct labels in the Talairach atlas) was as accurate as possible. As a precaution only sulcal/gyral structures and larger were singled out for analysis, as smaller structures ($< 1cm$) may not be accurately located in all brains. The Talairach space was used in order to facilitate comparisons of thickness measurements in normal subjects to results from

the literature, particularly the extensive numerical data published by [16]. However, several standard spaces have been described in the literature, and the exact one used is irrelevant from an algorithmic viewpoint. [23] have demonstrated that differences in cortical thickness measurements between subject groups may be exaggerated if the measurements are performed on scaled data. Therefore, the registration result was used only to transform the final thickness measurements into the Talairach space, in order to allow comparisons between subjects: all processing stages leading to the thickness measurement were performed in the native space, in order to avoid introducing such errors. The thickness values of the edges found in these regions were entered into histograms (from 0-10mm, 50 bins) of the regions and the median and interquartile ranges (25% and 75%) of the histograms determined. The regions for which the cortical thickness was estimated are given in Table 2. Note that they are listed as gyral regions, but do cover the corresponding sulci.

The proposed algorithm is completely automated and, partly as a result of this, is relatively efficient in terms of processor time requirements compared to model-based alternatives. The images volumes described in 2.1 were processed on a 1GHz Xeon processor with 4GB of RAM. Execution times ranged from 6-10 minutes with an average of 8 minutes. These figures compare favourably to the highly efficient approach proposed by [50], who quoted execution times of 3 minutes for 3D image volumes on an 800MHz Pentium 3 processor, and are roughly an order of magnitude smaller than the execution times of model-based algorithms.

2.4 Validation

Validation of the algorithm was performed primarily through comparison of cortical thickness measurements obtained from the subject group described in Section 2.1 with the algorithm described here to previous measurements from the literature obtained using alternative algorithms. There were two main reasons for this. First, the definition of a gold-standard for cortical thickness is problematic, due to the inherent ambiguity in defining the distance between two curved surfaces, and the prohibitive time requirements of manually measuring thickness at each voxel across the cortical surface. This suggests validation through comparison of results from different algorithms, seeking some consistency in order to define a "bronze standard". However, the issue of differences in the definition of thickness remains. Systematic differences between the results from any two algorithms could be attributed to differences in the definition of the quantity to be measured, rather than systematic errors in either algorithm. Therefore, definitive conclusions can only be drawn when the results from multiple algorithms agree, using the preponderance of the evidence to define a "bronze standard", and then comparing results from individual algorithms to this. Performing a typical algorithmic "shoot-out" i.e. implementing/re-implementing multiple algorithms and running them on the same data sets, may therefore be redundant if a sufficient body of results is available in the literature to establish a bronze standard. Consistency between the results would indicate that demographic differences between the subject groups used in each study are not significant compared to random errors. Validation through comparison to previously published measurements also has three main advantages. First, it considerably reduces the time required for validation. Second, it eliminates the possibility of sub-optimal use of unfamiliar, off-site software. Third, it avoids the issue of whether to re-estimate any free parameters, such as objective function weights, and the effects of this on the accuracy of the results. In order to perform such an evaluation, an independent measure of the errors on the regional median thickness values produced by the algorithm was required. This was obtained through a scan/rescan study in four individuals and a comparison between measurements in the left and right hemispheres of a single individual, as described in Appendices 1 and 2.

In order to investigate the possibility of systematic errors due to varying slice thickness or orientation, the thickness measurements from subject groups 1 and 2 were compared. The MR image volumes from group 1 consisted of 3mm thick axial slices, whereas those from group 2 consisted of 4mm thick coronal slices. The groups also differed slightly in mean age and, since cortical thickness generally decreases with age, this factor had to be taken into account in the analysis. Therefore, a linear fit to the group 1 vs. group 2 regional thickness measurements from all regions was produced. The deviation of this regression line from the line $x = y$ accounted for age-related variation in the mean cortical thickness across the whole brain. Measurements from each lobe were then compared to this fit. Gyri tend to be deeper than they are thick and so each lobe has a preferential thickness measurement direction. For example, the majority of the vectors along which thickness is measured in the parietal lobe will be aligned more with the anterior-posterior direction than with the inferior-superior direction. Therefore, if slice direction or thickness introduce a bias into the cortical thickness measurement, this should manifest itself as a consistent offset between the regional measurements in each lobe and the whole-brain regression line, varying according to the location of the lobe around the brain and thus its preferential thickness measurement direction. The drawback of this technique is that any difference in the age-related rate of cortical thickness change between lobes would also manifest as a consistent offset, on a lobe-by-lobe basis, between the measurements and the whole-brain regression line. Therefore, whilst the absence of consistent offsets would indicate that any effects of scan geometry were insignificant compared to random errors, their presence could indicate either variation in age-related

behaviour between lobes or systematic errors introduced by scan geometry, requiring further investigation in order to determine which effect was present.

The second stage of validation investigated the possibility of systematic errors in the thickness measurements through comparison with an alternative, model-based algorithm (ASP). Kabani et al. [16] have published an extensive set of regional cortical thickness measurements in numerical form, together with gold-standard data produced by manual markup, in a set of 40 brain MRI volumes from a group aged 18-40. Half were used for assessment of the left-hand regions, and the other half for the right-hand regions. They applied a dual-surface deformation algorithm to determine the WM/GM and GM/CSF surfaces of the cortical ribbon, and used the Talairach parcellation to define regions in which thickness measurements were made, as in the present study. Cortical thickness measurements were then produced both automatically and manually by a trained neuroanatomist, who marked out the correspondences between the two surfaces. Group averages from both the manual and automatic methods were presented. Validation of the algorithm presented here was performed through comparison with these results, using the manual measurements as a gold standard for comparison with both the automatic results presented by Kabani et al. and those produced by the algorithm presented here. In order to provide an approximately age-matched sample, the comparison was performed using the 13 young normals from this study (aged 19-53). Measurements were made in both hemispheres, so that the effective group sizes were roughly equal. Since the Kabani et al. manual measurements were used as the gold standard both for the Kabani et al. automatic results and the automatic results from the data used in this study, designing the comparative analysis in this way incorporated all possible forms of systematic difference between the results, including systematic errors in one or other of the algorithms, demographic differences between the subject groups, and differences in the thickness measure. If any of these were present, they would manifest as a systematic difference from the gold standard in the results from this study. Therefore, the weakness of this design is that it cannot ascribe such differences to a particular source: its strength is that, if no such systematic difference is observed, all of these possibilities are eliminated simultaneously (assuming that the probability of multiple systematic errors that cancel is low). In addition, it should be noted that the measurements in the Kabani et al. study were derived from single points in each region, whereas median thicknesses for whole regions were used in the present study. The random errors on the results cannot therefore be compared directly.

Care must be taken in the comparison of random errors between different algorithms due to the number of factors that contribute. The algorithm described here was primarily intended to produce regional median thicknesses. The error on the median is therefore related to the error on the thickness measurements at individual points, weighted by the variation in thickness across the region and by the inverse of the square-root of the number of measurements made in the region. When measurements are averaged across groups of subjects, the error is further weighted by the inter-subject thickness variation and the inverse of the square-root of the number of subjects in the group. Registration error, i.e. errors on the placement of the region boundaries, must also be taken into account. Therefore, a third stage of validation was performed, focused on random errors, and limited to studies that quoted whole-brain average thickness measurements from individual subjects (or the data necessary to reconstruct them, such as the range of thickness measurements across the brain), eliminating concerns related to registration and to averaging across subjects. Random errors from the present study were compared to those presented by [13], who used the Freesurfer algorithm, and [29], who used an intensity-based technique similar to that presented here, although the shortest inter-surface distance was used as the thickness measure.

The fourth stage of evaluation focused on the comparison of age-related changes in cortical thickness, again through comparison with previously published results. Two further comparative studies were performed, limited to average thickness measurements in the precentral gyrus, and to average thickness measurements for the whole brain, since the greatest amount of published data was available for these cases.

In the final stage of validation the correlation of cortical thickness with sex, intracranial volume and age were calculated. This allowed a qualitative comparison against the conclusions drawn in studies using alternative cortical thickness measurement algorithms.

3 Results

3.1 Effects of Varying Slice Thickness and Orientation

Fig. 9 shows the results of comparing the thickness measurements from subject groups 1 and 2, on a lobe-by-lobe basis, to the regression line for all measurement regions. The regression line takes account of the differences in mean age between the two groups. Any systematic errors introduced by slice direction or thickness should manifest as consistent offsets between the group 1 vs. group 2 measurements and the regression line, varying according to
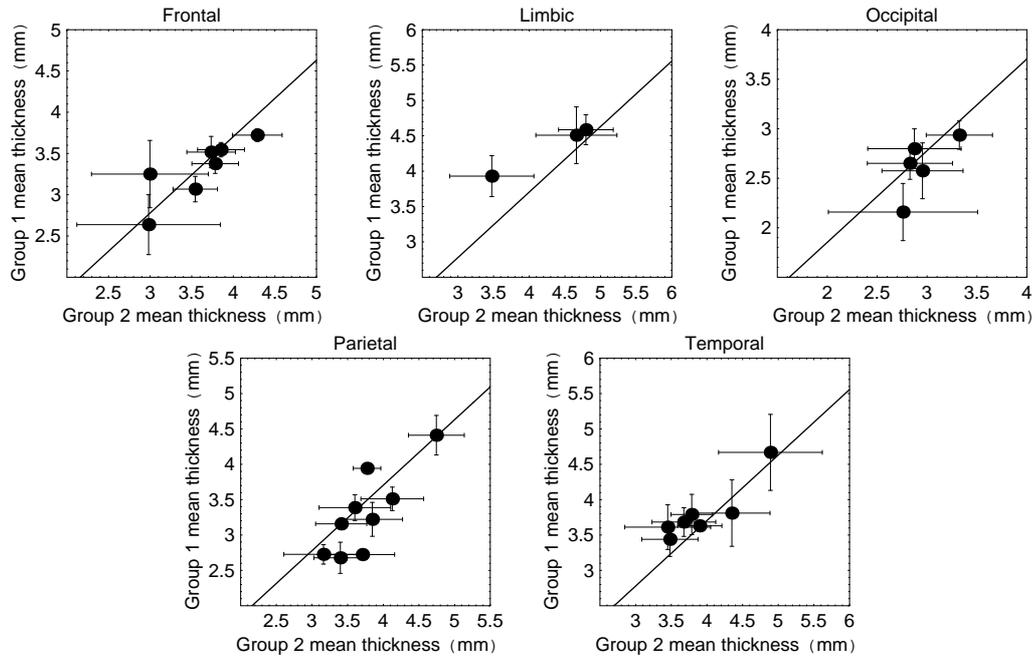
Figure 9: Group 2 vs. Group 1 thickness measurements in each lobe. The solid line shows the whole-brain regression line.
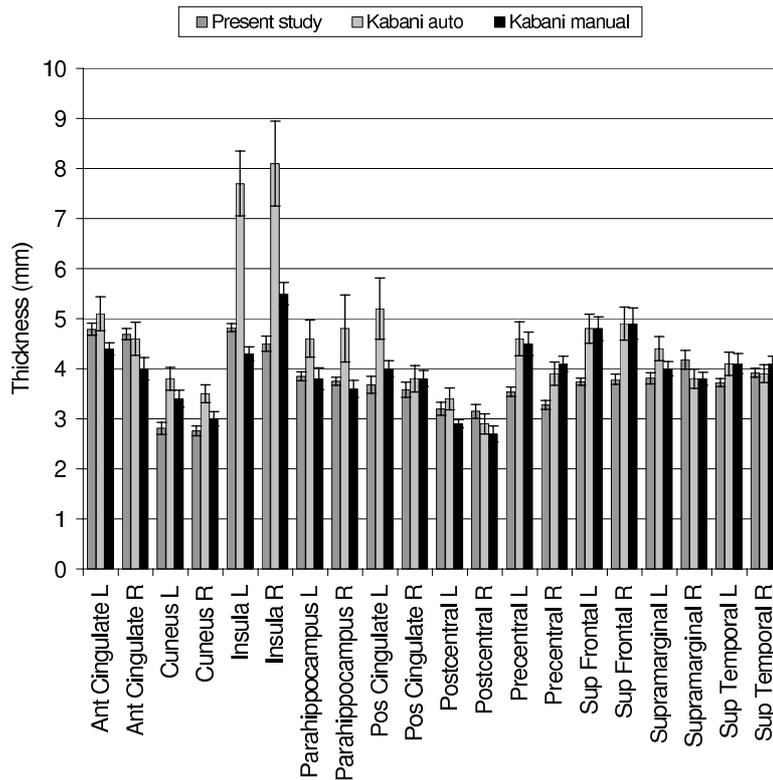

Figure 10: Comparison of the cortical thickness measurements in 20 regions presented by [16], measured from a group of 40 normal subjects using both manual and automatic techniques (see main text for details), to the measurements in the same regions from a group of 13 subjects using the algorithm presented here.

the location of the lobe around the brain. No statistically significant, consistent offsets between the group 1 vs group 2 thickness measurement and the whole-brain regression line are observed, indicating that any systematic errors introduced by variations in slice thickness or direction into the thickness measurements produced by this
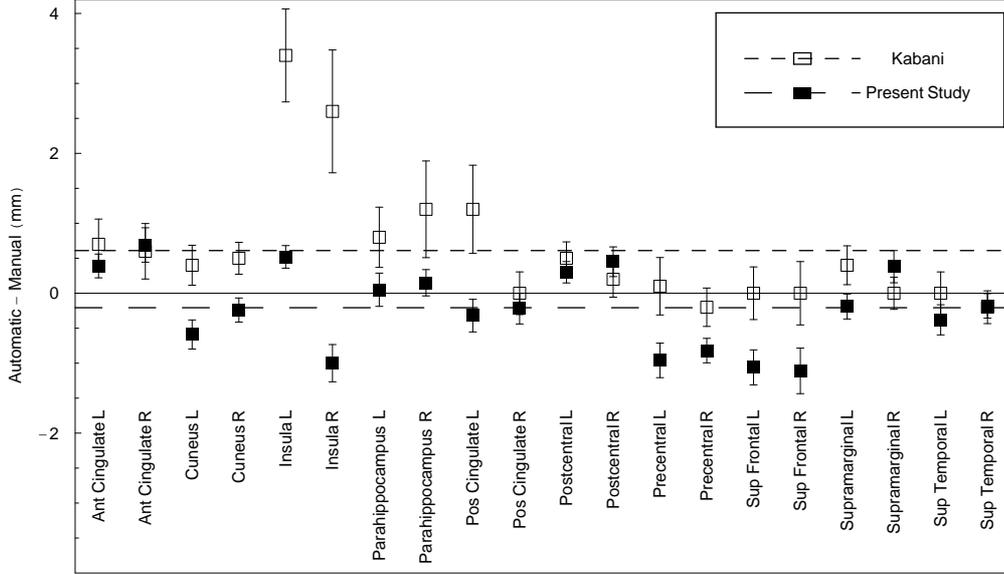
Figure 11: The differences in regional cortical thickness between the manual measurements presented by [16], and the automatic measurements presented by the same authors and the present study. The dashed lines show the mean differences across all cortical regions.

Lobe	Region	Group 1 (n=4)	Group 2 (n=9)	Group 3 (n=106)
Frontal	Rectal Gyrus	3.25 +/- 0.41	3.00 +/- 0.70	3.03 +/- 0.61
	Orbital Gyrus	2.64 +/- 0.36	2.98 +/- 0.86	2.28 +/- 0.58
	Precentral Gyrus	3.07 +/- 0.15	3.54 +/- 0.27	2.50 +/- 0.27
	Inferior Gyrus	3.52 +/- 0.19	3.73 +/- 0.29	3.03 +/- 0.31
	Middle Gyrus	3.38 +/- 0.12	3.78 +/- 0.28	2.91 +/- 0.30
	Superior Gyrus	3.54 +/- 0.09	3.85 +/- 0.28	3.11 +/- 0.33
	Medial Gyrus	3.72 +/- 0.08	4.29 +/- 0.30	3.27 +/- 0.31
Limbic	Posterior Cingulate	3.93 +/- 0.29	3.47 +/- 0.59	3.01 +/- 0.38
	Anterior Cingulate	4.59 +/- 0.21	4.79 +/- 0.38	4.32 +/- 0.50
	Subcallosal Gyrus	4.51 +/- 0.40	4.66 +/- 0.57	3.43 +/- 0.65
Occipital	Inferior Gyrus	2.16 +/- 0.29	2.76 +/- 0.74	2.10 +/- 0.37
	Lingual Gyrus	2.57 +/- 0.28	2.95 +/- 0.41	2.31 +/- 0.26
	Middle Gyrus	2.80 +/- 0.20	2.87 +/- 0.47	2.52 +/- 0.29
	Superior Gyrus	2.94 +/- 0.14	3.32 +/- 0.33	2.51 +/- 0.44
	Cuneus	2.65 +/- 0.16	2.83 +/- 0.43	2.25 +/- 0.25
Parietal	Insula	4.41 +/- 0.28	4.74 +/- 0.39	3.31 +/- 0.41
	Angular Gyrus	3.39 +/- 0.18	3.60 +/- 0.50	2.83 +/- 0.36
	Supramarginal Gyrus	3.51 +/- 0.16	4.13 +/- 0.44	2.85 +/- 0.32
	Cingulate Gyrus	3.94 +/- 0.06	3.77 +/- 0.19	3.57 +/- 0.32
	Inferior Lobule	3.22 +/- 0.24	3.85 +/- 0.42	2.70 +/- 0.29
	Superior Lobule	2.73 +/- 0.14	3.16 +/- 0.55	2.50 +/- 0.29
	Paracentral Lobule	2.72 +/- 0.04	3.71 +/- 0.45	2.38 +/- 0.34
	Postcentral Gyrus	2.68 +/- 0.22	3.40 +/- 0.37	2.32 +/- 0.23
	Precuneus	3.16 +/- 0.08	3.41 +/- 0.36	2.76 +/- 0.30
Temporal	Transverse Gyrus	3.44 +/- 0.24	3.48 +/- 0.39	2.88 +/- 0.46
	Uncus	4.67 +/- 0.53	4.89 +/- 0.73	4.02 +/- 0.55
	Fusiform Gyrus	3.81 +/- 0.47	4.35 +/- 0.53	3.34 +/- 0.42
	Inferior Gyrus	3.61 +/- 0.32	3.45 +/- 0.60	3.39 +/- 0.39
	Parahippocampal Gyrus	3.79 +/- 0.28	3.79 +/- 0.29	3.50 +/- 0.33
	Middle Gyrus	3.28 +/- 0.20	3.68 +/- 0.44	3.38 +/- 0.35
	Superior Gyrus	3.63 +/- 0.08	3.90 +/- 0.31	3.18 +/- 0.32

Table 2: The mean (+/- standard deviation) of the median thickness values for each of the 31 cortical regions examined, organised by lobe, for each of the three subject groups (n = number of subjects in group).

technique are not significant compared to the random errors.

3.2 Comparison to the Results of Kabani et al.

Fig. 10 shows a comparison of the manual and automatic results presented by [16] to thickness measurements in the same regions produced using the algorithm presented here. Fig. 11 provides a more quantitative comparison, showing the difference between the manual measurements and the results from each algorithm. The mean differences across all regions are $0.61 \pm 0.43mm$ for the ASP algorithm and $-0.21 \pm 0.22mm$ for the algorithm presented here. Neither difference is statistically significant, and so neither algorithm shows evidence of systematic errors based on these results. As stated above, the random errors from the two studies are not comparable due to differences in
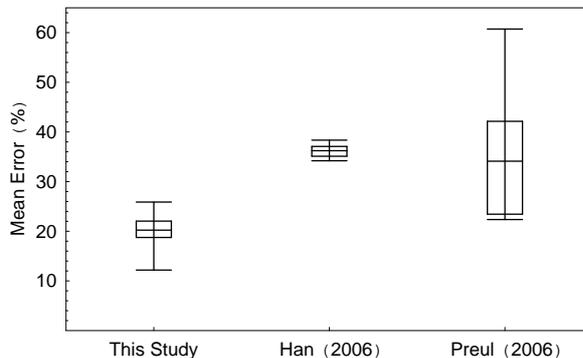
Figure 12: Percentage errors on whole brain mean cortical thickness measurements from individual subjects, presented as box-and-whisker plots, for the present study and the results reported by [29] and [13].

Parameter	This study	Han et al., 2006 [13]	Preul et al., 2006 [29]
Number (males)	119 (52)	15 (8)	525 (263)
Mean Age (range)	70.3 (19-86)	69.5 (66-81)	27.9 (17-68)
Software	Intensity-based	Freesurfer	Intensity-based
Thickness measure	Normal distance	Vertex-vertex	Shortest distance

Table 3: Details of the individual studies incorporated into the comparison of random errors.

the study designs.

3.3 Comparison of Random Errors

Fig. 12 shows the results of the comparison of random errors between the present study and those presented by [13] and [29]. Details of the subjects included in each study are given in Table 3. Since the ages of the subjects in each study vary considerably, the results are presented in the form of percentage errors, in order to account for the effect of age-related thinning of the cortex. In the case of the present study, and that presented by Han et al., the error referred to is the standard deviation of cortical thickness measures over the whole cortex calculated from individual subjects. The results are presented in the form of a box-and-whisker plot i.e. showing the mean, upper and lower quartiles, maximum and minimum of the errors from the individual subjects. In the case of the Preul et al. study, results were presented by decade. Therefore, these were corrected by multiplying by the square-root of the number of subjects in each group, to account for the reduction in error due to averaging across subjects, and by dividing by the range of cortical thickness measurements in each group, to account for the increase in error due to inter-subject variation. Based on these data, the algorithm presented here provides significantly ($p < 0.01$) lower random errors than the ASP algorithm, and lower errors than the alternative intensity-based algorithm used by Preul et al., although the difference is not significant ($p > 0.05$). For comparison, [17] published data showing the reduction in RMS errors between the fitted cortical surfaces and the mean surfaces over 16 scans of one pediatric brain between ASP and CLASP. They found average reductions of 9.3% and 44.8% in the errors on the WM and GM surfaces respectively. These values correspond closely to the difference in random errors between the results from Han et al. and the present study shown in Fig. 12.

3.4 Comparison of Thickness Measurements

Fig. 13 shows the average cortical thickness measurements for the precentral gyrus in 119 subjects produced using the algorithm presented here, plotted against age. A quadratic fit to the data is shown: the dashed curves either side of the fit show the upper and lower standard error bounds. A significant ($P < 0.0001$) reduction of cortical thickness with age is observed. In previous work [37] we found similar dependencies in other cortical regions. Also shown are a number of measurements of precentral gyrus thickness from the literature. The details of the studies involved are given in Table 4. The results presented by [9] were measured manually post-mortem: brain volume decreases by approximately 10% during post-mortem fixation [31]. However, the thickness measurement was performed only on the gyral crown, which is known to be thicker than the sulcal fundi [24]. Similarly the presentation of the results in [43] and [47] as projections onto the outer cortical surface prevented identification of

Reference	No. subjects	Age range (years)	Algorithm type
Kabani et al., 2001 [16]	40	18-40	Model based
Von Economo, 1929 [9]	-	30-40	Manual measurement
Sowell et al., 2004 [43]	45	5-11	Intensity based
Tosun et al., 2004 [49]	105	59-84	Model based
Fischl and Dale, 2000 [10]	30	20-37	Model based
Thompson et al., 2005 [47]	40	18-48	Intensity based
MacDonald et al., 2000 [24]	150	18-40	Model based
Salat et al., 2004 [35]	106	18-93	Model based
Present study	119	19-86	Intensity based

Table 4: Details of the individual studies used in the comparison of precentral gyrus thickness measurements.
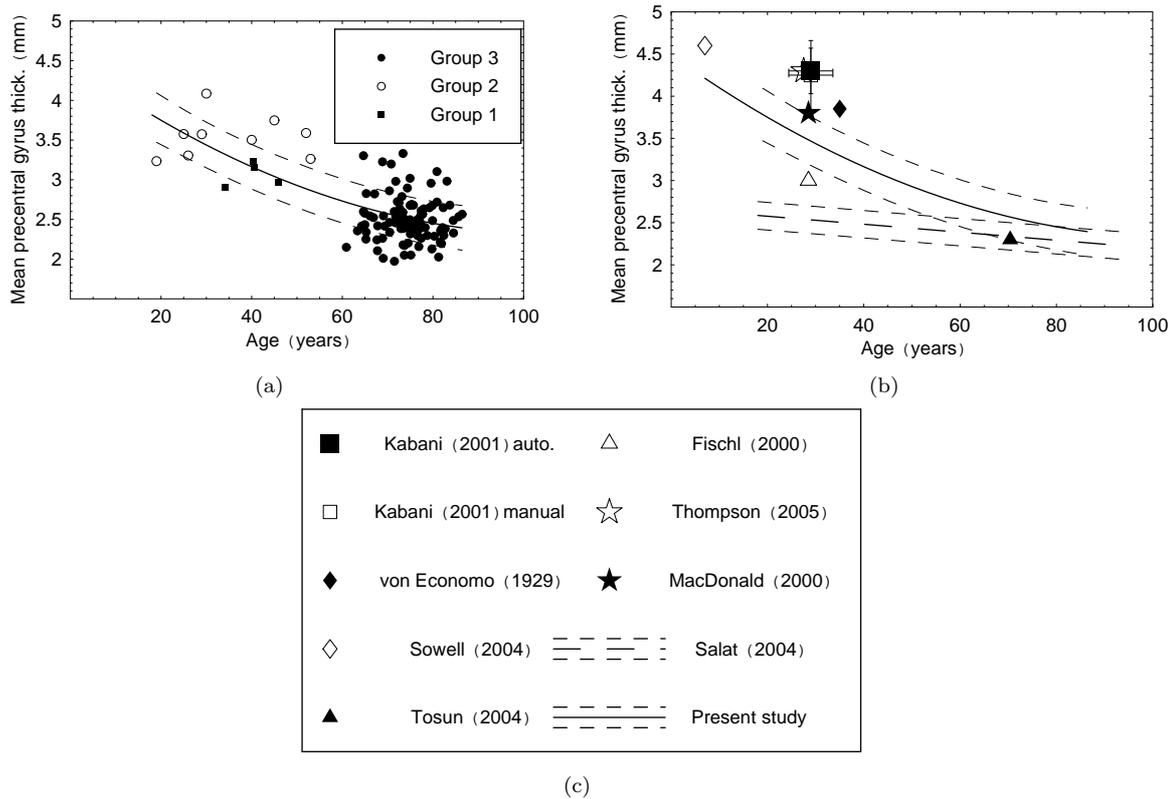

Figure 13: Comparison of precentral gyrus thickness measurements. a) Measurements from the present study, quadratic fit (solid line), and upper and lower standard errors (dashed lines). b) Fit to the data from the present study compared to results from the literature. c) Key to the previous studies: details of each are given in Table 4.

the thickness in the sulcal fundi. These three data therefore represent upper limits on the average thickness in the region. Overall, the studies represented in Fig. 13 represent the widest possible range of methods for defining the inner and outer cortical surfaces, the thickness measure, and the presentation of the results.

All results except those from [16] were read from graphical representations, and of these all except [35] and [9] presented results as views of the outer cortical surface, in some cases partially inflated to reveal the sulcal fundi, with colour coding to represent the thickness at each point. This method of data display is popular in the literature as it avoids the need for parcellation of the data into particular regions. However, the calculation of regional average thicknesses from such representations is difficult and the calculation of errors on the averages impossible. Hence, these points are shown without error bars as their precision is unknown. However, given the results of the comparison of random errors presented in Section 3.3, if we assume that the random errors from the other studies represented in the comparison are equal to or greater than those from the present study, then there is no statistically significant difference between any of these results and our own. The remaining paper included in this study, [35], was the only one to study variation in cortical thickness with age, and the only model-based study to

Reference	No. subjects	Age range (years)	Algorithm type
Present study	119	19-86	Intensity based
Magnotta et al., 1999 [25]	148	18-82	Model based
Preul et al., 2006 [29]	97	17-68	Intensity based
Salat et al., 2004 [35]	106	18-93	Model based
Shaw et al., 2007 [40]	166	8-17	Model based

Table 5: Details of the individual studies used in the comparison of whole brain rate of thickness change.
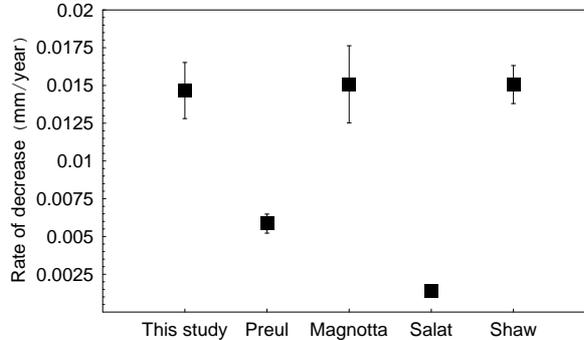

Figure 14: Comparison of the average rate of whole-brain cortical thickness decrease with age. Details of the individual studies included are given in Table 5 and in the main text.

cover the whole age range between adolescence and senescence. A significant disagreement between this study and the others included in the comparison can be seen, with [35] suggesting a much lower rate of cortical thickness change with age.

Fig. 14 shows the comparison of average rates of whole-brain thickness decrease with age. Data from the present study were fitted with a linear function in order to allow an accurate comparison to the previous studies, which also used linear functions. The previous study gave a rate of $-0.0147 \pm 1.86 \times 10^{-3} mm\ year^{-1}$, consistent with the results of the studies by [25] and [40], the latter of which focused on a much younger age range. Significantly ($P < 0.01$) lower rates of change were found by [35] and [29], the latter of which used an intensity-based technique similar to the one used in the present study, although the shortest distance between the inner and outer cortical surfaces was used as the thickness measure. The results from the [35] and [29] are also significantly ($P < 0.01$) different from each other. For comparison, age-related changes in GM density, a quantity closely related to cortical thickness, were measured by [42] in a group of 176 subjects aged 7 to 87 years: the average, proportional rate of change of GM density in the precentral gyrus was $0.593\%\ year^{-1}$, consistent with the $0.567 \pm 0.270\%\ year^{-1}$ rate of change in cortical thickness found in the present study.

3.5 Cortical Thickness, Intracranial Volume, Sex and Age

An analysis of variance (ANOVA) was performed for each region to determine whether sex, intracranial volume and age were significant in predicting group variance in median thickness in the 119 subjects included in this study. For simplicity, intracranial volume was represented by a related quantity, the volume of the bounding box on the interior of the skull. This has been shown to be directly proportional to the true intracranial volume [4], but is significantly easier to calculate. For this subject group there was no significant difference in age with sex at the 95% significance level (p=0.1287, Adj R-sq = 0.01124). As in previous studies (e.g. [6, 4]), intracranial volume was not significantly correlated with age (p=0.143, Adj R-sq = 0.01) and the correlation was even lower when sex was taken into account. Sex itself was highly significant in predicting intracranial volume (p=1.41x10-10, Adj R-sq for the whole model = 0.305), with females having smaller heads than males; again, this has been found in previous studies (e.g. [6, 4]).

The model used in the ANOVA took the form median thickness = sex + intracranial volume + age, where sex was a factor and intracranial volume and age were covariates; the results are shown in Table 6. Type 1 sum of squares were used, meaning that the order in which the variables are inserted in the model was important; the variance accounted for by intracranial volume is that above and beyond the variance accounted for by sex, and the variance

Lobe	Region	Sex Constant	p-value	TIV Slope (x10-7)	p-value	Age Slope	p-value
Frontal	Rectal Gyrus	0.213	**0.020**	-2.836	0.362	-0.008	0.07
	Orbital Gyrus	0.187	0.200	-0.759	0.973	-0.018	**1.42e-05**
	Precentral Gyrus	0.061	0.792	-1.043	0.921	-0.020	**5.55e-16**
	Inferior Gyrus	0.036	0.653	-1.668	0.489	-0.014	**1.74e-09**
	Middle Gyrus	0.080	0.244	-1.88	0.433	-0.017	**1.52e-12**
	Superior Gyrus	0.099	**0.046**	-2.933	0.128	-0.016	**5.10e-11**
	Medial Gyrus	0.076	0.171	-3.012	0.149	-0.020	**7.77e-16**
Limbic	Posterior Cingulate	0.013	0.592	0.280	0.651	-0.013	**3.12e-05**
	Anterior Cingulate	0.337	**1.4e-5**	-3.041	0.227	-0.013	**4.62e-05**
	Subcallosal Gyrus	0.017	0.311	-0.628	0.466	-0.031	**6.36e-11**
Occipital	Inferior Gyrus	0.074	0.219	-2.091	0.387	-0.009	**2.89e-03**
	Lingual Gyrus	0.090	0.138	-1.169	0.623	-0.012	**5.72e-08**
	Middle Gyrus	0.097	0.107	-0.684	0.778	-0.006	**4.49e-03**
	Superior Gyrus	0.043	0.772	-1.045	0.820	-0.012	**1.66e-04**
	Cuneus	0.083	0.177	-1.044	0.676	-0.011	**5.03e-08**
Parietal	Insula	0.172	0.187	-1.766	0.900	-0.034	**2.00e-16**
	Angular Gyrus	0.145	**0.039**	-2.234	0.387	-0.018	**1.17e-10**
	Supramarginal Gyrus	0.119	0.090	-3.164	0.208	-0.025	**2.00e-16**
	Cingulate Gyrus	0.094	**0.009**	-2.843	0.061	-0.007	**3.65e-04**
	Inferior Lobule	0.118	0.221	-1.449	0.773	-0.022	**2.00e-16**
	Superior Lobule	0.136	**0.036**	-1.332	0.619	-0.014	**3.15e-09**
	Paracentral Lobule	0.044	0.705	-2.681	0.381	-0.023	**4.47e-14**
	Postcentral Gyrus	0.046	0.633	-1.932	0.423	-0.019	**2.22e-16**
	Precuneus	0.060	0.182	-2.448	0.186	-0.013	**4.07e-09**
Temporal	Transverse Gyrus	0.158	0.179	-0.422	0.868	-0.016	**3.78e-07**
	Uncus	-0.073	0.809	-6.654	**0.026**	-0.021	**3.45e-07**
	Fusiform Gyrus	0.151	**0.006**	-5.239	**0.027**	-0.020	**1.28e-09**
	Inferior Gyrus	0.008	0.432	-2.287	0.244	-0.003	0.37
	Parahippocampal Gyrus	0.181	**0.001**	-1.304	0.500	-0.007	**1.11e-03**
	Middle Gyrus	0.126	**0.040**	-1.282	0.547	-0.007	**4.96e-03**
	Superior Gyrus	0.156	**0.028**	-1.110	0.804	-0.017	**1.82e-12**

Table 6: The probabilities that sex, intracranial volume (TIV) and age are significant in predicting variance in median regional thickness. Values in bold are significant at the p=0.05 level. Also shown are the slopes for the regression of median thickness against intracranial volume and age in the model, and the constant offset in thickness according to sex. Offsets are for the females, relative to the zero constants for males.

accounted for by age is additional to that already accounted for by sex and intracranial volume. A significance value of p=0.05 was used. Note that 31 regions have been analysed, and that it is not expected that the cortical thickness in each region will be independent of all others. Sex was significant in accounting for the variance seen in cortical thickness in ten regions (rectal and superior frontal gyri; anterior cingulate gyrus; angular and cingulate parietal gyri and superior parietal lobule; and the fusiform, parahippocampal, middle and superior gyri of the temporal lobe). In these ten cases, females have greater cortical thickness than males. Indeed, as can be seen from column two of Table 6, only the uncus shows a thinner median thickness in females compared to males and there is an average global difference of 0.02mm. Intracranial volume is significant in accounting for the variance in two regions (the uncus and fusiform gyrus of the temporal lobe): in both cases, the larger the intracranial volume, the smaller the thickness (although the correlation is very shallow). There is a highly significant correlation in most regions of median thickness against age. Exceptions to this are the rectal gyrus of the frontal lobe and the inferior temporal gyrus.

4   Discussion

The validation of the algorithm was performed primarily through comparison to previous measurements from the literature, due to the reasons outlined in Section 2.4. The validation procedure had three main aims. The first was to identify any systematic error in the algorithm. The second was to compare the random errors to those produced by alternative algorithms. This is particularly relevant to model based algorithms since, whilst the data-driven approach adopted here has advantages such as reduced processor time requirements, the absence of the regularising effect of the cortical surface models might be expected to result in higher random errors. The third was to confirm that the algorithm is capable of detecting cortical thickness differences across age, sex, head size etc.

Given that anisotropic data was used in this study, it was necessary first to investigate whether the slice orientation and thickness had any effect on the cortical thickness measurement. The comparison of regional median thickness measurements from subject groups 1 and 2 shown in Fig. 9 to the whole-brain regression line indicates no systematic variation in the measurements according to the location of the measurement region around the brain, and thus the dominant orientation of the measurement vectors. This indicates that any systematic errors introduced by scan geometry are not significant compared to the random errors. The comparison to the manual and automatic measurements presented by [16] also fails to detect any systematic error in the results. The [16] manual measurements are factored into the automatic results from both studies in Fig. 11, and so any source of systematic difference (demographic differences, differences due to the choice of points at which to make the measurements) would show up as a consistent offset from zero for the differences between the manual measurements and those from this study. No such offset is seen: the results from the present study are on average closer to zero difference

than the results from Kabani et al. However, the random errors on the two studies are not comparable due to differences in study design.

The comparison of random errors between the present study and those presented by [13] and [29] indicated that the algorithm presented here produces significantly lower random errors than ASP, as used in the former, and lower random errors than the data-driven approach used in the latter, although this difference was not significant. This is unexpected due to the lack of the regularising effect of cortical surface models data-driven approaches. However, the results presented by [17] indicate that the adaptations incorporated into the CLASP algorithm could provide equal accuracy to the algorithm presented here, albeit at the expense of considerably greater processor time requirements. Therefore, we conclude only that the technique presented suffers no disadvantage in accuracy compared to model-based alternatives.

The study of precentral gyrus thickness measurements allows the comparison of the absolute values of cortical thickness measurements from a variety of algorithms, albeit limited to a single cortical region. No significant difference was observed between the majority of the results included, indicating that demographic differences between the subject groups used and differences in the definition of thickness are not significant compared to the random errors. This strongly suggests that the accuracy these algorithms is dominated by the accuracy of the initial segmentation, rather than the details of the cortical thickness measurement. The results from [35] are the one exception, showing a much lower rate of age-related change than is supported by the other studies. The same effect is seen in the comparison of whole-brain average rates of thickness change. The results presented by [25] and [40] are consistent with those from the present study, but the results presented by Salat et al. show a considerably lower rate of thickness change with age. We observe that the possibility of bias identified by [24] in model-based algorithms, such as that used in the [35] study, could explain this effect: the constraints applied to prevent self-intersection of the models and to aid in the modelling of the surfaces in tightly folded gyri would tend to bias the model towards a fixed thickness, reducing the statistical power to age-related changes. Furthermore, the bias would only manifest itself when a considerable mismatch was present between the effect of the constraints and the true thickness, and could therefore vary according to the age range of the subjects or demographic differences, explaining why they are not observed in all model-based studies. However, the present study is unable to ascribe the low rate of age-related thickness change seen in the [35] study to any particular cause: we can merely observe that it is inconsistent with the consensus of measurements from the literature, and with our own results. The rate of thickness change in the data presented by [29], derived using an intensity-based technique, is also inconsistent with the other studies. That study used the minimum inter-surface distance as the thickness measure. Again, we observe that the use of the minimum distance as the thickness measure allows the directions of the vectors along which thickness is measured to change between subjects in such a way that minimises the detected rate of thickness change. However, the present study is unable to confirm whether this effect is in fact present. The inconsistencies in measured rates of thickness change indicate that further study is required in order to elucidate the differences between the various cortical thickness measurement algorithms in common usage.

In the results from the present study, median thickness almost universally showed a negative correlation with age, whereas intracranial volume had little effect. The lack of correlation between median thickness and intracranial volume supports the view that subjects with larger heads, hence larger brains, have an increased white matter proportion, assuming that the surface area of the GM/WM boundary is directly proportional to the surface area of the interior of the skull. The sex of the individual was significant in 10 of the 31 regions, although there was no clear trend with cerebral lobe. The finding that females have a greater cortical thickness than males, by on average 0.02mm over the whole brain is consistent with a study by [29] who found a 0.07mm difference. [14] found, using CLASP, a significantly greater cortical thickness in females over the left frontal parietal and occipital lobes, as well as in the right parietal lobe in corrected space, although this was reduced to only the left parietal lobe when native space was used (as in the technique presented here). [23] also demonstrate a near-global increase in thickness in females (although no difference values are quoted), both in stereotaxic and native space, albeit with less significance in the native space. The reduction of GM thickness with age assumes that all subjects had approximately equal GM thicknesses as young adults, and shows that volume reduction in the grey matter due to atrophy can be observed as a reduction in thickness. Whether there is a correlation of the regional GM/WM boundary surface area with age remains to be investigated, although regional surface area measurements would be highly dependent upon the accuracy of the stereotaxic atlas and registration of the data to this atlas.

4.1   Limitations of the Study

The algorithm presented here was designed primarily to produce regional median thickness estimates. This approach was deliberately chosen in order to maximise the statistical power of the technique whilst minimising the amount of smoothing or regularisation applied prior to obtaining the point-wise thickness measurements. In addition, the variability in brain structure brought about by cortical folding is less significant to the measurements.

However, this approach does impose the requirement for a volumetric registration, and so the technique is not inherently surface-based. It would therefore not be applicable to a dense, point-wise comparison of thickness measurements from different subjects in its current form i.e. without a method of extracting a surface mesh. In addition, a specific parcellation of the data is required in order to define regions in which median measurements are made, and this may have drawbacks in some cases. For instance, in normal development the anterior superior temporal gyrus increases in thickness whilst the posterior superior temporal gyrus decreases. Therefore, using one superior temporal gyrus ROI would miss this trend. However, other authors have demonstrated the use of deformable models in order to define homologous points on the cortical surfaces across different subjects and using these in registration, for defining the points at which thickness measurements are made, and to define a surface on which the results may be displayed, whilst still using only image intensities in the thickness measurement itself [43, 22], applying some regional smoothing to the thickness measurements in order to remove outliers. The same approach could be applied here, although this was not investigated in the current study and would considerably increase processor time requirements.

The present study was also limited to normal subjects, and so more work is now needed to extend these measurements to disease groups. However, the ability of the algorithm to detect significant changes in cortical thickness with age implies the ability to detect pathological changes. [12] present results from several studies on rates of change of the volume of various structures within the brain in normals and Alzheimer's disease, demonstrating that pathological changes may occur up to an order of magnitude faster than age-related changes. Changes in GM volume would be expected to produce a change in thickness, although the relationship might not be simple if changes occur simultaneously in WM. Therefore, the ability to detect significant age related change confirms the ability to detect pathological changes. In addition, several studies (listed in the Introduction) have demonstrated correlations between disease states and cortical thickness. On the basis of the comparison of random errors to the studies presented by [13] and [29], the algorithm presented here produces lower random errors, and so should also be capable of quantifying such correlations.

5   Conclusion

Cortical thickness measurement algorithms can be coarsely divided into two groups: those that represent the inner and outer cortical surfaces using deformable models, and those that are based solely on image intensities. This paper has presented an algorithm that falls into the second group. The accuracy of the algorithm has been confirmed by comparing measurements of cortical thickness in 119 normal subjects to previously published results. These comparisons, within the limitations of the number of studies included, indicate that the algorithm is at least as accurate than those presented in the literature with no evidence of systematic error. We therefore conclude that the data-driven approach presented here has no disadvantage in terms of accuracy, but may have advantages in terms of freedom from model bias, elimination of free parameters in the objective function that must be determined empirically in model-based algorithms, and a reduction in the processor time required. The algorithm may therefore be more suited to wide-scale clinical utilisation than the alternatives based on cortical surface models.

Appendix 1: Calibration of the Search Termination Criterion

As described in Section 2.3, a small proportion of searches along the normal vector to the inner cortical surface failed to locate the outer cortical surface within a reasonable distance. These searches were terminated when they exceeded a threshold distance, determined via a calibration experiment. Cortical thickness measurements were made in one subject (male, age 40 years) with the search termination threshold varying from 5 to 40 mm. The number of unterminated searches performed and the subsequent regional median thickness measurements in 31 cortical regions are shown as a function of termination threshold in Fig. 15. No change is apparent in the number of unterminated searches above a threshold of approximately 15mm, indicating that at this point all searches have either terminated on an edge or intensity dip, or will never terminate. This indicates that the searches still proceeding at this point are spurious. The search termination criterion used in the final version of the algorithm was set to 20mm, the the point at which the number of proceeding searches fell below 0.1% of the total number of searches performed across all cortical regions. In practice, 0.06% of searches were still proceeding at this point; this was a low enough proportion to ensure that the termination process did not bias subsequent regional median thickness calculations.
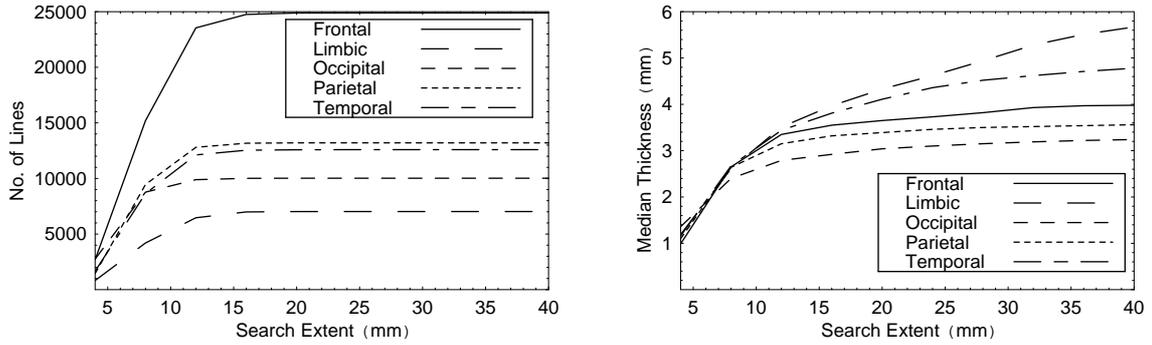
Figure 15: The dependence of the number of search vectors (a) and the resultant regional median thicknesses (b) on the limit applied to the search along the normal vector to the inner cortical surface, for a single individual (male, age=40).
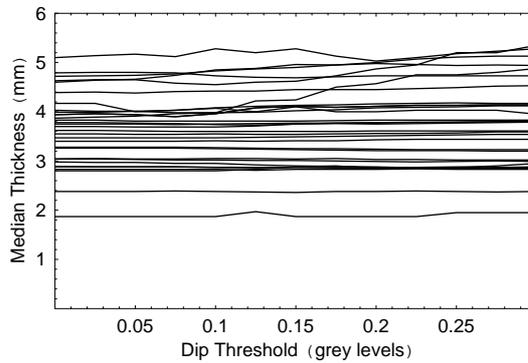

Figure 16: The dependence of regional median thickness calculations on applying a threshold to the definition of a dip in intensity during the search along the normal vector to the inner cortical surface, for a single individual (male, age=40).

Appendix 2: Evaluating the Effects of Placing a Threshold on the Dip Detection

As described in Section 2.3, the search along the normal vector to the inner cortical surface considers dips in intensity in order to increase the sensitivity to the outer cortical surface at points where it is obscured by partial voluming. In this context a dip was defined as a point along the search at which the intensity was lower than at the points on either side. This procedure raises the possibility of placing a threshold on the dip detection process, in order to avoid false detections due to image noise.

In order to test this process, regional median cortical thicknesses were measured in a single subject (male, age 40 years) in 31 regions, with a threshold placed on the minimum absolute value of the intensity dip, in terms of the proportion of the difference between the mean intensities of GM and WM as determined by the segmentation stage of the algorithm. An intensity change greater than 50% of this value was defined as an edge, and so the threshold was varied between 0 and 30%. The results are shown in Fig. 16 for all regions. As can be seen, the median thickness measurement in insensitive to the imposition of a threshold on the dip calculation; therefore, no threshold was used in the proposed algorithm.

Appendix 3: Reproducibility of the Measurement

Four young normals from Group 2 (see Table 1) were scanned twice, with the second scan occurring within 4-21 days of the first and at approximately the same time of day (mid-morning) in all cases. All eight datasets were processed independently, and the results compared across 31 cortical regions, averaging the measurements from the left and right hemispheres. The results are shown in Fig. 17.
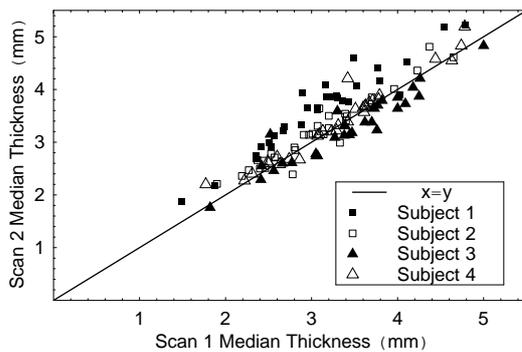
Figure 17: Median regional thicknesses for 31 cortical regions (averaged across the left and right hemispheres) in repeated scans of four young normals, compared to the line $x = y$.
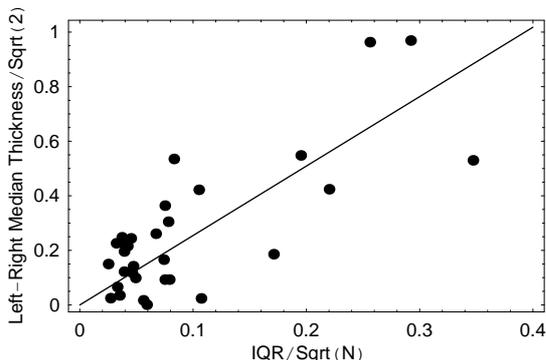

Figure 18: Absolute values of the difference between median regional thicknesses in the left and right hemispheres of a single individual (male, age=40 years), plotted against the interquartile range of the regional histogram. The regression line was used to calibrate the calculation of random errors.

The results were used to evaluate the systematic and random errors on the median thickness calculations. The systematic error on the measurements was calculated as the extent of the deviation of the regression coefficients for each subject from the line of equality, calculated as the RMS of the differences divided by $\sqrt{2}$ to account for the fact that two measurements have been taken. The mean systematic error on the measurement in a given individual was 6.2%. The mean random error on any given region in any given individual, calculated from the scatter about the regression line for that individual, was 0.06mm, and so the systematic errors were small compared to the random errors.

Appendix 4: Calibration of the Error Calculation

In the absence of a gold-standard cortical thickness measurement to compare with the results of the proposed algorithm, which is problematic to obtain for the reasons listed in 2.4, it is difficult to quantify the errors on the regional median cortical thickness measurements. The reproducibility study described in Appendix 3 involved comparisons between multiple images of the same subject, and so does not fully consider the effects of registration error in comparisons between multiple subjects, who will have different brain shapes and therefore present a more difficult registration problem. It therefore provides a lower bound on the error. A second study was therefore performed to provide an upper bound on the error. The error on the calculation of a median thickness from a regional histogram will be dictated by the width of that histogram, divided by the square-root of the number of samples entered into it. However, the histogram is expected to contain outliers, and so a robust measure of the width is required: the inter-quartile range was used for this purpose. Median regional cortical thickness measurements were calculated in 31 regions for one subject (male, age 40 years) and the difference between the thicknesses in equivalent regions in the left and right hemispheres was found. Since the left and right hemispheres are not expected to be identical, this quantity can provide an upper bound on the measurement error. Figure 18 shows the left-right thickness difference, divided by the square-root of 2 to account for the fact that two measurements have been

taken, plotted against the inter-quartile range of the regional histogram divided by the number of samples entered into it. The gradient of the regression line was 2.54; this value relates the upper bound on the random error in a region to the statistical properties of the histogram from which it was derived. The coefficient of determination was 0.80, indicating that the errors are, to a good approximation, dominated by these statistical properties. The mean error on regional median thickness measurements, calculated using this method and averaged across the left and right hemispheres, was 0.12mm. This differs from the lower bound obtained from the reproducibility study by a factor of 2: considering that these are error calculation, this is a relatively close agreement. Therefore, the upper error bound was used to determine the random errors on all regional median thickness measurements described in this study from the inter-quartile range and number of samples in the regional histograms from which they were calculated.

Acknowledgements

References

[1] J Annese, A Pitiot, I D Dinov, and A W Toga. A myelo-architechtonic method for the structural classification of cortical areas. *NeuroImage*, 21(1):15–26, 2004.

[2] M Ballmaier, E R Sowell, P M Thompson, A Kumar, H Lavretsky, S E Welcome, H DeLuca, and A W Toga. Mapping brain size and cortical grey matter changes in elderly depression. *Biol Psychiatry*, 55(4):382–389, 2004.

[3] K Brodmann. *Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prenzipien dargestellt auf Grund des Zellaufbaus.* Springer, Barth, Leipzig, Germany., 1909. Translated by Laurence Garey as Localisation in the Cerebral Cortex (1994), London: Smith-Gordon, new edition 1999, London: Imperial College Press.

[4] P A Bromiley, N A Thacker, and A Jackson. Trends in brain volume change with normal ageing. In *Proc. MIUA'05*, pages 247–250, 2005.

[5] J F Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986.

[6] E Courchesne, H J Chisum, J Townsend, A Cowles, J Covington, B Egaas, M Harwood, S Hinds, and G A Press. Normal brain development and aging: Quantitative analysis at in vivo mr imaging in healthy volunteers. *Radiology*, 216:672–682, 2000.

[7] A M Dale, B Fischl, and M I Sereno. Cortical surface-based analysis i: segmentation and surface reconstruction. *NeuroImage*, 9(2):174–194, 1999.

[8] K L Double, G M Halliday, J J Kril, J A Harasty, K Cullen, W S Brooks, H Creasey, and G A Broe. Topography of brain atrophy during normal aging and alzheimer's disease. *Neurobiol Aging*, 17(4):513–521, 1996.

[9] C Von Economo. *The Cytoarchitectonics of the Human Cerebral Cortex.* Oxford University Press, Oxford, 1929. Translated by S. Parker.

[10] B Fischl and A M Dale. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proc Natl Acad Sci U.S.A.*, 97:11050–11055, 2000.

[11] B Fischl, M I Sereno, and A M Dale. Cortical surface-based analysis ii: Inflation, flattening and a surface-based coordinate system. *NeuroImage*, 9(2):195–207, 1999.

[12] N C Fox, S Cousens, R Scahill, R J Harvey, and M N Rossor. Using serial registered brain magnetic resonance imaging to measure disease progression in alzheimer disease: Power calculations and estimates of sample size to detect treatment effects. *Arch Neurol*, 57:339–344, 2000.

[13] X Han, J Jovicich, D Salat, A van der Kouwe, B Quinn, S Czanner, E Busa, J Pacheco, M Albert, R Killiany, P Maguire, D Rosas, N Makris, A Dale, B Dickerson, and B Fischl. Reliability of mri-derived measurements of human cerebral cortical thickness: The effects of field strength, scanner upgrade and manufacturer. *NeuroImage*, 32(1):180–194, 2006.

[14] K Im, J-M Lee, J Lee, Y-W Shin, I Y Kim, J S Kwon, and S I Kim. Gender difference analysis of cortical thickness in healthy young adults with surface-based methods. *NeuroImage*, 31(1):31–38, 2006.

[15] S E Jones, B R Buchbinder, and I Aharon. Three-dimensional mapping of cortical thickness using laplace's equation. *Hum Brain Mapp*, 11(1):12–32, 2000.

[16] N Kabani, G Le Goualher, D MacDonald, and A C Evans. Measurement of cortical thickness using an automated 3d algorithm: A validation study. *NeuroImage*, 13(2):375–380, 2001.

[17] J S Kim, V Singh, J K Lee, J Lerch, Y Ad-Dab'bagh, D MacDonald, J M Lee, S I Kim, and A C Evans. Automated 3-d extraction and evaluation of the inner and outer cortical surfaces using a laplacian map and partial volume effect classification. *NeuroImage*, 27(1):210–221, 2005.

[18] G R Kuperberg, M R Broome, P K McGuire, A S David, M Eddy, F Ozawa, D Goff, W C West, S C R Williams, A J W van der Kouwe, D H Salat, A M Dale, , and B Fischl. Regionally localized thinning of the cerebral cortex in schizophrenia. *Arch Gen Psychiatry*, 60:878–888, 2003.

[19] J L Lancaster, J L Summerlin, L Rainey, C S Freitas, and P T Fox. The talairach daemon, a database server for talairach atlas labels. *NeuroImage*, 5(4):S633, 1997.

[20] J P Lerch and A C Evans. Cortical thickness analysis examined through power analysis and a population simulation. *NeuroImage*, 24(1):163–173, 2005.

[21] G Lohmann, C Preul, and M Hund-Georgiadis. Morphology-based cortical thickness estimation. In C J Taylor CJ and J A Noble, editors, *Proc. 18th International Conference on Information Processing in Medical Imaging, IPMI'03*, pages 89–100, Ambleside, UK, 2003. Volume 2732 of Lecture Notes in Computer Science, 89-100. Springer.

[22] E Luders, K L Narr, P M Thompson, D E Rex, L Janke, and A W Toga. Hemispheric asymmetries in cortical thickness. *Cereb Cortex*, 16(8):1232–1238, 2006.

[23] E Luders, K L Narr, P M Thompson, D E Rex, R P Woods, H DeLuca, L Jancke, and A W Toga. Gender effects on cortical thickness and the influence of scaling. *Hum Brain Mapp*, 27(4):314–324, 2006.

[24] D MacDonald, N Kabani, D Avis, and A C Evans. Automated 3-d extraction of inner and outer surfaces of cerebral cortex from mri. *NeuroImage*, 12(3):340–356, 2000.

[25] V A Magnotta, N C Andreasen, S K Schultz, G Harris, T Cizadlo, D Heckel, P Nopoulos, and M Flaum. Quantitative in vivo measurement of gyrification in the human brain: changes associated with aging. *Cereb Cortex*, 9(2):151–160, 1999.

[26] M I Miller, A B Massie, J T Ratnanather, K N Botteron, and J G Csernansky. Bayesian construction of geometrically based cortical thickness metrics. *NeuroImage*, 12(6):676–687, 2000.

[27] S I Olsen. Estimation of noise in images: An evaluation. *Graphical Models and Image Processing*, 55:319–323, 1993.

[28] M Pokric, N A Thacker, M L J Scott, and A Jackson. Multi dimensional medical image segmentation with partial voluming. In *Proc. MIUA'01*, pages 77–81, 2001.

[29] C Preul, M Hund-Georgiadis, B U Forstmann, and G Lohmann. Characterization of cortical thickness and ventricular width in normal ageing: a morphometric study at 3 Tesla. *J Magn Reson Imaging*, 24(3):513–519, 2006.

[30] C Preul, G Lohmann, M Hund-Georgiadis, T Guthke, and D Yves von Cramon. Morphometry demonstrated loss of cortical thickness in cerebral microangiopathy. *J Neurol*, 252:441–447, 2005.

[31] R Quester and R Schroder. The shrinkage of the human brain stem during formalin fixation and embedding in paraffin. *J Neurosci Methods*, 75(1):81–89, 1997.

[32] P Rockett. The accuracy of sub-pixel localisation in the canny edge detector. In *Proc. BMVC'99*, pages 392–401, 1999.

[33] H D Rosas, A K Liu, S Hersch, M Glessner, R J Ferrante, D H Salat, A van der Kouwe, B G Jenkins, A M Dale, and B Fischl. Regional and progressive thinning of the cortical ribbon in huntingdon's disease. *Neurology*, 58(5):695–701, 2002.

[34] M Sailer, Fischl, D Salatand C Tempelmann, M A Schonfeld, E Busa, N Bodammer, H-J Heinze, and A Dale. Focal thinning of the cerebral cortex in multiple sclerosis. *Brain*, 126(8):1734–1744, 2003.

[35] D H Salat, R L Buckner, A Z Snyder, D N Greve, R S R Desikan, E Busa, J C Morris, A M Dale, and B Fischl. Thinning of the cerebral cortex in aging. *Cereb Cortex*, 14(7):721–730, 2004.

[36] G Sapiro. *Geometric partial differential equations and image analysis*. Cambridge University Press, Cambridge, 2001.

[37] M L J Scott and N A Thacker. Regional cerebral cortical thickness measurement using MRI. In *Proc. MIUA'05*, pages 7–10, 2005.

[38] M L J Scott and N A Thacker. Robust tissue boundary detection for cerebral cortical thickness estimation. In *Proc. MICCAI'05, LNCS 3750*, pages 878–885. Springer Berlin / Heidelberg, 2005.

[39] L D Selemon, G Rajkowska, and P S Goldman-Rakic. Evidence for progression in frontal cortical pathology in late-stage Huntington's disease. *J Comp Neurol*, 468(2):190–204, 2004.

[40] P Shaw, J Lerch, D Greenstein, W Sharp, L Clasen, A Evans, J Geidd, F X Castellanos, and J Rapoport. Longitudinal mapping of cortical thickness and clinical outcome in children and adolescents with attention-deficit/hyperactivity disorder. *Arch Gen Psychiatry*, 63:540–549, 2007.

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

[42] E R Sowell, B S Peterson, P M Thompson, S E Welcome, A L Henkenius, and A W Toga. Mapping cortical change across the human life span. *Nat Neurosci*, 6(3):309–315, 2003.

[43] E R Sowell, P M Thompson, C M Leonard, S E Welcome, E Kan, and A W Toga. Longitudinal mapping of cortical thickness and brain growth in normal children. *J Neurosci*, 24(38):8223–8231, 2004.

[44] E R Sowell, P M Thompson, K D Tessner, and A W Toga. Mapping continued brain growth and gray matter density reduction in dorsal frontal cortex: Inverse relationships during postadolescent brain maturation. *J Neurosci*, 21(22):8819–8829, 2001.

[45] J Talairach and P Tournoux. *Co-planar stereotaxic atlas of the human brain. 3-dimensional proportional system: An approach to cerebral imaging.* Thieme Medical Publishers, New York, 1988.

[46] P M Thompson, K M Hayashi, E R Sowell, N Gogtay, J N Geidd, J L Rapoport, G I de Zubicaray, A L Janke, S E Rose, J Semple, D M Moddrell, Y Wang, T G M van Erp, T D Cannon, and A W Toga. Mapping cortical change in Alzheimer's disease, brain development, and schizophrenia. *NeuroImage*, 23:S2–S18, 2004.

[47] P M Thompson, A D Lee, R A Dutton, J A Geaga, K M Hayashi, M A Eckert, U Bellugi, A M Galaburda, J R Korenberg, D L Mills, A W Toga, and A L Reiss. Abnormal cortical complexity and thickness profiles mapped in Williams syndrome. *J Neurosci*, 25(16):4146–4158, 2005.

[48] A W Toga, P M Thompson, and E R Sowell. Mapping brain maturation. *Trends Neurosci*, 29(3):148–159, 2006.

[49] D Tosun, M E Rettmann, X Han, X Tao, C Xu, S M Resnick, D L Pham, and J L Prince. Cortical surface segmentation and mapping. *NeuroImage*, 23:S108–S118, 2004.

[50] A J Yezzi and J L Prince. An eulerian pde approach for computing tissue thickness. *IEEE Trans Med Imag*, 22(10):1332–1339, 2003.