HAL
open science

Evaluation of Brain Atrophy Estimation Algorithms using Simulated Ground-Truth Data

Swati Sharma, Vincent Noblet, François Rousseau, Fabrice Heitz, Lucien Rumbach, Jean-Paul Armspach

To cite this version:

Swati Sharma, Vincent Noblet, François Rousseau, Fabrice Heitz, Lucien Rumbach, et al.. Evaluation of Brain Atrophy Estimation Algorithms using Simulated Ground-Truth Data. Medical Image Analysis, 2010, 14 (3), http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6W6Y-4YDKK26-1&_user=4853051&_coverDate=06. 10.1016/j.media.2010.02.002 . hal-00477102

HAL Id: hal-00477102

https://hal.science/hal-00477102v1

Submitted on 28 Apr 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Evaluation of Brain Atrophy Estimation Algorithms using Simulated Ground-Truth Data

S. Sharma[*,a,b], V. Noblet[b], F. Rousseau[b], F. Heitz[b], L. Rumbach[a,c], J.-P. Armspach[a]

[a] *Laboratoire d'Imagerie et de Neurosciences Cognitives (LINC-UDS) FRE 3289, CNRS, Strasbourg, France.*
[b] *Laboratoire des Sciences de l'Image, de l'Informatique et de la Télédétection (LSIIT-UDS) UMR 7005, CNRS, Illkirch, France.*
[c] *Centre Hospitalier Universitaire, Service de Neurologie, Besançon, France.*

Abstract

A number of analysis tools have been developed for the estimation of brain atrophy using MRI. Since brain atrophy is being increasingly used as a marker of disease progression in many neuro-degenerative diseases such as Multiple Sclerosis and Alzheimer's Disease, the validation of these tools is an important task. However, this is complex, in the real scenario, due to the absence of gold standards for comparison. In order to create gold standards, we first propose an approach for the realistic simulation of brain tissue loss that relies on the estimation of a topology preserving B-spline based deformation fields. Using these gold standards, an evaluation of the performance of three standard brain atrophy estimation methods (SIENA, SIENAX and BSI-UCD), on the basis of their robustness to various sources of error (bias-field inhomogeneity, noise, geometrical distortions, interpolation artefacts and presence of lesions), is presented. Our evaluation shows that, in general, bias-field inhomogeneity and noise lead to larger errors in the estimated atrophy than geometrical distortions and interpolation artefacts. Experiments on 18 different anatomical models of the brain after simulating whole brain atrophies in the range of $0.2 - 1.5\%$ indicate that, in the presence of bias-field inhomogeneity and noise, a mean error of $0.64\% \pm 0.53$, $4.00\% \pm 2.41$ and

[*]Corresponding Author. Address: Institute Physique Biologique, 4, Rue Kirschleger, 67085 Strasbourg, France. Tel.: +33(0)390244039; Fax: +33(0)390244084.
E-mail address: swati.sharma@linc.u-strasbg.fr

1.79% $\pm$ 0.97 may be expected in the atrophy estimated by SIENA, SIENAX and BSI-UCD, respectively.

Key words: Brain atrophy simulation, Atrophy estimation, Evaluation.

---

1. Introduction

Brain atrophy is a common feature of many neuro-degenerative diseases such as Multiple Sclerosis (MS)[1] and Alzheimer's disease (AD) [2, 3]. The emergence of brain atrophy is related to a combination of multiple factors. For instance in MS, factors such as axonal injury, loss of myelin and others related to astrogliosis, remyelination and water-compartment change, influence the occurrence of atrophy in the brain in a combined fashion. Although it is difficult to determine the individual influence of these factors on atrophy, it has been shown to be a sensitive marker of disease evolution in several chronic diseases. Studies have reported annual global brain atrophy rates in MS (in both early and late stages) between 0.5% and 0.8% [4] and between 1% and 2.8% for AD patients [3]. It should be pointed out here that sensitive methods are required to accurately estimate the small atrophy that occurs in the brain.

Magnetic Resonance Imaging (MRI) acquisitions and image processing techniques have helped in this effort of quantification and provide ways of measuring the brain volume loss. In the literature, the currently available image processing methods for brain atrophy estimation have been differentiated based on several criteria. Pelletier *et al.* [5] suggest the classification of methods for brain atrophy analysis into two groups depending on their reliance on registration or segmentation. Registration-based methods include brain edge motion analysis ("Boundary Shift Integral" (BSI) [6], "Structural Image Evaluation, using Normalization, of Atrophy" (SIENA) [7, 8]), Voxel-Based Morphometry ("Statistical Parametric Mapping" (SPM) [9]), Template-Driven Segmentation ([10]) and Local Jacobian Analysis ([11, 12, 13]). Segmentation-based methods rely on different brain atrophy measurements such as the "Brain Parenchymal Fraction" (BPF) ([14, 15]) and "Brain to IntraCranial Capacity Ratio" (BICCR) ([16]). When using brain atrophy as a marker of disease evolution, two types of analyzes are conducted: longitudinal and cross-sectional. In longitudinal studies, repeated brain scans of the same individual are considered over time. These studies track disease progression by giving an estimate of the atrophy rate. Examples of widely

used methods for longitudinal studies are BSI [6] and SIENA [7]. Cross-sectional studies are used to know the state of the disease at a given time by providing an estimate of brain volume at one point in time. SIENAX [8] is a popular method utilized by cross-sectional studies. Another pertinent grouping classifies the methods based according to the estimated atrophy: global or local. A global atrophy estimation method, such as BSI [6] or SIENA [7], gives an estimate of overall decrease or rate of decrease in the brain volume. Global measures are used in population studies to separate the patient group from the normals or in cases when one just wants to know whether the brain has undergone an atrophy globally. Local measurements (Jacobian Integration [11], SPM [9]) make it possible to have a regional (or voxel by voxel) estimate of the atrophy in the brain. Local measures are interesting since they allow us to decipher which parts of the brain are more affected by atrophy in the course of a pathology.

The accuracy and reliability of brain atrophy estimation methods are affected by a number of factors. Non-destructive biologic factors such as inflammation, edema, steroid therapy, dehydration, alcohol consumption, cerebral vascular disease and normal aging [5] may contribute to a change in brain volume. Factors such as lesion growth or change in tissue intensity may lead to a change in the brain tissue appearance over time. MRI artefacts during acquisition also have a strong impact on brain atrophy measurements: motion artefacts, sequence variations, partial volume effect, noise in images and bias field inhomogeneity. Another source of error are geometrical distortions which result from errors in the gradient field strength and non-linearity of gradient fields in the MR scanner[17, 11]. Finally, other sources of errors are more method specific depending on the pre-processing algorithms that are used and include, for instance, inaccuracies in registration and segmentation of images.

Since brain atrophy is now being increasingly used as a marker of disease progression in many pathologies, the validation of these tools is a key problem. However, in the real scenario, the non-availability of the ground truth complicates the evaluation and comparison of these techniques [5]. In the literature, a number of attempts have been made to assess and compare these methods using real and simulated data. In Table 1, we list several studies on neuro-degenerative diseases, their evaluation frameworks and findings.

The most commonly used evaluation criteria are discussed in the following paragraph. Refer to Table 1 for a list of references where these criteria have been used in the literature.

1. *Same day scan-rescan*: Tests on two scans of the same subject obtained on the same day should yield zero atrophy. Ideally, the spread of the true brain volume changes in such cases should be negligible and the deviations from zero represent measurement uncertainties of the methods.
2. *Consistency*: This is a way of measuring the variability in atrophy estimation. Consistency can be measured using several approaches. A way of checking the consistency of a method is to measure volume loss for two (or more) baseline scans with respect to the same 1 year repeat scan. All such comparisons should lead to the same volume change estimation. Another approach is incremental atrophy summation comparison which is often used to validate a method in the absence of a ground truth. Consistency has to be interpreted with caution as a consistent method may not necessarily be accurate.
3. *Patient-control separation*: The basic idea is to carry out a population study between two groups that are known to be characterized by different atrophy rates. A reliable atrophy estimation method is expected to be able to reliably discriminate between the two groups in a statistical analysis. However, the information that can be gathered from such a metric is limited, for instance, we cannot infer anything about the error (accuracy) or bias in the measurement of atrophy.
4. *Evaluation using simulated ground truth*: This involves simulation of a known atrophy value in the brain. Recently, some methods have been proposed for the creation of ground truth data with simulated brain atrophy relying on bio-mechanical models [12, 13] or Jacobian maps [18, 19]. Such an evaluation framework is of great interest since it allows the assessment of method performance with respect to several sources of errors such as motion artefacts, noise or bias field inhomogeneity. This type of evaluation also makes it possible to provide a numerical estimate of the accuracy of a method in terms of the measured error. We will focus on this criterion in our study.

In this paper, we propose a topology preserving non-rigid registration based framework that enables us to simulate brain images with realistic atrophy (Section 2). Section 3 describes the framework we use to assess three popular freely available brain atrophy estimation methods (BSI-UCD, SIENA and SIENAX). The experimental setup is described in Section 4. Results of atrophy simulation on BrainWeb images [20] using the proposed

framework are presented in Section 5. In Section 6, we present an analysis of the performance of these methods on the basis of their robustness to bias field inhomogeneity, noise, geometrical distortions and interpolation artefacts using the simulated brain images. Influence of lesion load is also investigated. We also study the effect of anatomical variability on atrophy estimation by simulating the same atrophy on a cohort of 18 BrainWeb images.

2. Proposed approach for atrophy simulation

Designing methods that simulate realistic atrophy is of great importance for evaluating atrophy measurement techniques, since it is a way of generating ground truth data. In the literature, the currently available approaches for the simulation of brain atrophy can be classified into two groups: Jacobian-based methods and Bio-mechanical-based approaches. In [18], Karacali *et al.* have proposed a Jacobian-based method in which deformation fields are estimated in order to induce the desired volume variations in the regions of interest. This is done by minimizing the sum of squared differences between the Jacobian of the transformation and the desired level of atrophy at each voxel. An additional penalization term is also considered in order to prevent *corner Jacobians* from being negative to ensure that the estimated deformation field preserves topology. However, the penalization term cannot guarantee this property since this is only a necessary condition but not sufficient for ensuring topology preservation. Besides, the framework that is proposed initially is not capable of estimating large atrophy, thus requiring an iterative strategy that estimates the deformation in an incremental way, as a composition of several estimated transformations.

Pieperhoff *et al.* have recently presented a similar approach relying on "Local Volume Ratio (LVR)" [19]. LVR refers to the ratio of the distorted voxel volume in the source brain to the voxel volume in the target brain. Pieperhoff *et al.* use a similar cost function as proposed by Karacali *et al.* but LVR is used instead of the Jacobian. They also consider an additional regularization term that ensures that the estimated transformation is smooth. Unfortunately, none of these methods address the problem of enforcing skull invariance, which is a desirable property for the simulation of realistic brain atrophy.

The bio-mechanical-based approach proposed by Camara *et al.* [12] relies on a bio-mechanical model to deform brain tissues using a finite-element approach. A 3D mesh of labeled brain structures from an atlas is warped

Approach	Pathology Studied	Method(s) Evaluated	Criteria Used	Conclusions
Gunter *et al.*[21]	AD	GMM, BSI	Patient/Control separation	*Bias field correction had no effect on group separation, *GMM better than BSI for whole brain and ventricular measurements
Paling *et al.*[22]	AD (late-onset dementia)	BSI	Same-day scan-rescan, Patient/Control separation (Zero-atrophy metric)	*BSI accuracy lies between $2-4$ml (1σ), *Group separation ($p < 0.0001$)
Boyes *et al.*[11]	AD	BSI, JI	Same-day scan-rescan, Consistency, Patient/Control separation, Scaled Atrophy	*Both measures are well correlated, *JI better estimates simulated tissue loss
Smith *et al.*[23]	AD	SIENA, SIENAX, BSI	Using PBVC as metric, Patient/Control separation, Same-day scan-rescan, Incremental atrophy summation with first-last differencing	*BSI-SIENA (0.2% error rate), SIENAX-SIENA correlate well, *SIENA over-estimates atrophy as compared to BSI, *BSI-SIENA: similar sensitivity, *SIENAX: less sensitive in group separation
Camara *et al.*[13]	AD	BSI, SIENA, JI	Mean Absolute Difference of PBVC values with simulated atrophies	*SIENA-BSI correlate well, *FFD performed better than fluid registration
Pieperhoff *et al.*[19]	CBD	DFM	Specificity, Sensitivity (using simulated atrophy)	*Noise/Bias field inhomogeneity lead to different errors in volume change measurements, *Error of DFM comparable to SIENA, BSI

Table 1: Summary of the previous works on the evaluation of atrophy estimation approaches.
AD: Alzheimer's disease, **CBD**: Cortico-basal degeneration, **GMM**: Gradient Matching Method, **JI**: Jacobian Integration, **FFD**: Free-Form Deformations, **DFM**: Deformable Field Morphometry

onto the considered image using a fluid registration method. Atrophy is then simulated by a Finite Element Method (FEM) solver by considering a thermoelastic model of soft tissue deformation. This framework also incorporates the skull invariance constraint. Let us also notice that the interpolation strategy used to derive a dense deformation field from the set of displacement vectors in this work, does not ensure that the final transformation is a one-to-one mapping.

It is important to point out that the way tissue loss occurs in the brain is still not very well understood. It is more likely to be driven by physiological aspects rather than bio-mechanical ones. However, the use of bio-mechanical modeling may be interesting to account for the consequences of tissue loss on the brain shape, for instance, on the orientation of the gyri. Although approaches based on bio-mechanical models of brain tissues may appear attractive, one can wonder whether such models are actually more relevant than the one proposed by Karacali *et al.* [18] or Pieperhoff *et al.* [19].

Here, we present an alternative to the methods discussed above. The proposed approach estimates a deformation field that preserves topology so that the Jacobian is at each voxel, as close as possible to the desired local level of atrophy. Topology preservation ensures that connected structures remain connected and that the neighborhood relationship between structures is maintained, which are desired properties when simulating atrophy. By enforcing this constraint, the space of possible solutions is restricted to physically acceptable deformations. It also prevents the appearance or disappearance of existing or new structures. For instance, topology preservation also has its disadvantages as it restricts closed structures such as sulci from opening up when atrophy is simulated. Contrary to Karacali *et al.* [18] who consider the sum of squared differences between the Jacobian of the transformation and the desired level of atrophy, we consider the logarithm of the Jacobian so that dilations ($1 < J < +\infty$) and contractions ($0 < J < 1$) have a similar influence on the objective function. The interest of considering the logarithm of the Jacobian has already been highlighted in [24]. Besides, additional constraints are introduced in order to make sure that the skull remains invariant by the estimated transformation. An overview of the existing atrophy simulation approaches and of the proposed approach is presented in Table 2.

First, we describe the B-spline based deformation model that is considered and then the associated optimization problem.

2.1. The multi-resolution deformation model

Let $\boldsymbol{s} \triangleq [x, y, z]^t \in \Omega \subset \mathbb{R}^3$. The mapping $\boldsymbol{h} : \Omega \mapsto \Omega$ writes $\boldsymbol{h}(\boldsymbol{s}) = \boldsymbol{s} + \boldsymbol{u}(\boldsymbol{s})$, where $\boldsymbol{u}$ is the displacement vector field. We consider a decomposition of the displacement vector field $\boldsymbol{u}$ over a sequence of nested subspaces $V_1 \subset \ldots \subset V_l \subset V_{l+1} \subset \ldots$, defining a multi-resolution approximation of $\boldsymbol{u}$. Space V_1 defines the coarsest scale representation. Any deformation field at scale l may also be expressed as a deformation field at a finer scale $l+1$. A basis of V_l may be generated from a scaling function Φ. To handle a 3-D deformation field, three multi-resolution decompositions are considered, one for each component of the displacement. Every element of the basis of V_l writes

$$\begin{aligned} \Phi^l_{i,j,k}(x, y, z) &= \Phi^l_i(x)\ \Phi^l_j(y)\ \Phi^l_k(z) \\ &= 2^{3l/2}\ \Phi\left(2^l x - i\right)\ \Phi\left(2^l y - j\right)\ \Phi\left(2^l z - k\right). \end{aligned}$$

At scale l, *i.e.,* in space V_l, the displacement field $\boldsymbol{u}^l$ is parametrized by the vector $\boldsymbol{a}^l \triangleq \left\{a^l_{x;i,j,k}, a^l_{y;i,j,k}, a^l_{z;i,j,k}\right\}$ as (see [25] for more details):

$$\boldsymbol{u}^l(x, y, z) = \begin{bmatrix} u^l_x(x, y, z) \\ u^l_y(x, y, z) \\ u^l_z(x, y, z) \end{bmatrix} = \begin{bmatrix} \sum_{i,j,k} a^l_{x;i,j,k} \Phi^l_{i,j,k}(x, y, z) \\ \sum_{i,j,k} a^l_{y;i,j,k} \Phi^l_{i,j,k}(x, y, z) \\ \sum_{i,j,k} a^l_{z;i,j,k} \Phi^l_{i,j,k}(x, y, z) \end{bmatrix}. \quad (1)$$

First degree polynomial B-spline scaling functions Φ have been considered in the current implementation [25]. We refer the reader to [25, 26] for additional details on the deformation model.

2.2. Optimization problem

Let $\Omega_J \subset \Omega$ be the area where the desired simulated atrophy level $J(\boldsymbol{s})$ (the value of the Jacobian at each voxel $\boldsymbol{s} \in \Omega_J$) is user-specified. To estimate the corresponding deformation field $\boldsymbol{u}$, we consider the following objective function:

$$E_{\boldsymbol{u},J,\lambda} = \int_{\Omega_J} \left|\log\left(J_{\boldsymbol{u}}(\boldsymbol{s})\right) - \log\left(J(\boldsymbol{s})\right)\right|^2 d\boldsymbol{s} + \lambda\, C \int_{\Omega} E_{Reg}(\boldsymbol{u}(\boldsymbol{s}))\, d\boldsymbol{s}, \quad (2)$$

where $J_{\boldsymbol{u}}$ stands for the Jacobian of $\boldsymbol{u}$, E_{reg} is a regularization term that ensures that the estimated transformation is smooth, λ is the weight of the

regularization term and C is a scaling factor computed at the beginning of each scale so that the data term and regularization term are comparable [26]. Among the many regularization terms proposed in the literature, we choose membrane energy.

$$\begin{aligned} E_{Reg}\left(\mathbf{u}(\mathbf{s})\right) = & \left(\frac{\partial u_x(\mathbf{s})}{\partial x}\right)^2 + \left(\frac{\partial u_x(\mathbf{s})}{\partial y}\right)^2 + \left(\frac{\partial u_x(\mathbf{s})}{\partial z}\right)^2 + \left(\frac{\partial u_y(\mathbf{s})}{\partial x}\right)^2 + \left(\frac{\partial u_y(\mathbf{s})}{\partial y}\right)^2 \\ & + \left(\frac{\partial u_y(\mathbf{s})}{\partial z}\right)^2 + \left(\frac{\partial u_z(\mathbf{s})}{\partial x}\right)^2 + \left(\frac{\partial u_z(\mathbf{s})}{\partial y}\right)^2 + \left(\frac{\partial u_z(\mathbf{s})}{\partial z}\right)^2. \end{aligned} \tag{3}$$

Contrary to Karacali *et al.* [18] who consider an additional term that prevents the violation of the topology preservation condition, we directly solve the following constrained optimization problem, which ensures exact topology preservation in the underlying continuous domain:

$$\hat{\mathbf{u}} = \arg \min_{\forall s \in \Omega,\ 0 < J_{\mathbf{u}}(s) < +\infty} E_{\boldsymbol{u},J,\lambda}. \tag{4}$$

The procedure for solving this optimization problem is quite involved and is detailed in [25]. The estimation is done in a hierarchical way until the desired final scale l_f, the optimization procedure at scale l being initialized with the solution at scale $l-1$. Instead of considering a gradient descent algorithm as in [25], we use the Levenberg-Marquardt optimization procedure [27] in order to improve the convergence rate.

To simulate realistic atrophy, it is desirable to enforce the skull to remain invariant under the simulated transformation. This constraint has been considered in [12] but not in [18]. In our framework, a simple technique is used to ensure that the deformation field vanishes on the skull. This is accomplished by carrying out the optimization procedure only for those B-spline parameters that do not affect the skull, while setting the other parameters to zero. This limits the optimization as depending on the resolution voxels far away from the skull are fixed. Nonetheless, we adopt this approach because it is capable of completely guaranteeing that the skull will remain in place after the application of the estimated transformation.

Finally, to obtain the warped image, it is more convenient to consider the backward transformation so that standard interpolation techniques can be used for the regularly sampled data. Inversion is performed by using the algorithm described in [28].

Study	Approach for Atrophy Simulation	Topology Preservation	Skull Invariance	Evaluated Methods	Simulated artefacts
Camara *et al.* [12]	Bio-mechanical framework	Not guaranteed	Yes	*SIENAX	*Rician Noise *Rotational Motion *Pulsatile Motion
Camara *et al.* [13]	Bio-mechanical framework	Not guaranteed	Yes	*SIENA *BSI *B-spline Free Form Deformations *Non-rigid Fluid Registration	*Rician Noise
Karacali *et al.* [18]	Jacobian-based framework	Consider a necessary but not sufficient condition	No	-	-
Pieperhoff *el al.* [19]	Jacobian-based framework	Consider a necessary but not sufficient condition	No	*SIENA *BSI *Elastic registration	*Noise *Bias field inhomogeneity
Present Study	Jacobian-based framework	Yes	Yes	*SIENA *BSI *SIENAX	*Gaussian Noise *Bias field inhomogeneity *Geometrical distortions *interpolation artefacts *Presence of lesions

Table 2: A survey of works on evaluation of brain atrophy methods using a simulated ground truth.

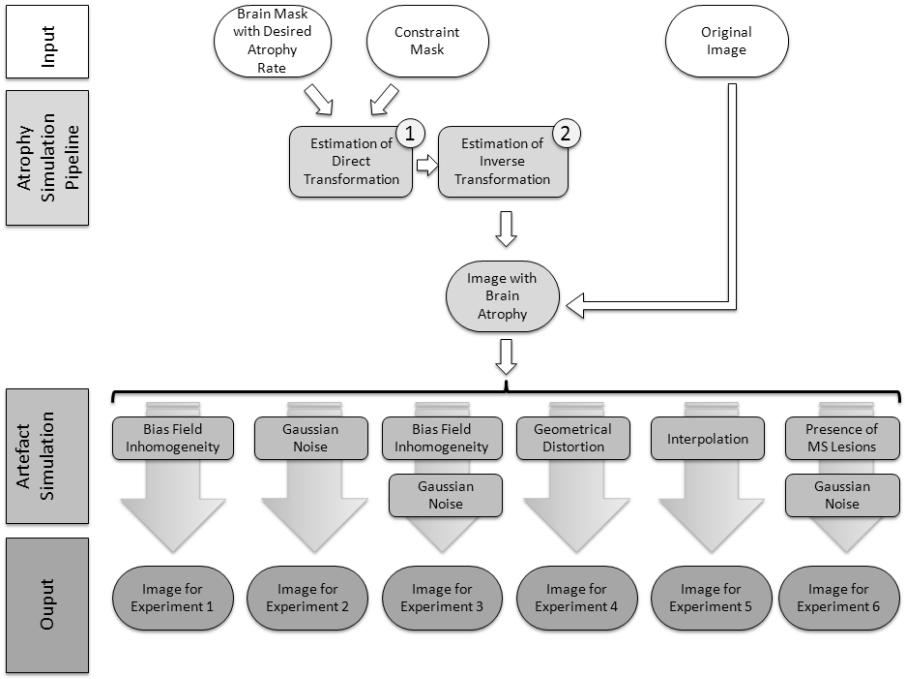


Figure 1: Proposed atrophy simulation framework and the evaluation framework showing the addition of various artefacts to the simulated images in order to carry out the validation procedure. This figure shows the various steps involved in the simulation of atrophy (Section 2) followed by the addition of various artefacts (Section 3). Here "Output" refers to the images that we simulate for various evaluation experiments. Experiment 1, 2, 3, 4, 5 and 6 refers to the experiments for evaluation of atrophy estimation algorithms on the basis of bias field inhomogeneity, noise, bias field inhomogeneity and noise, geometrical distortions, interpolation artefacts and effect of lesions, respectively.

3. Evaluation Framework

3.1. Simulation of artefacts

In order to conduct a realistic evaluation of the methods, we create a set of images (with different artefacts) from the baseline BrainWeb image[1] and images on which atrophy is simulated. We use BrainWeb images in our experiments since the ground truth segmentation of the brain tissues is available. Our evaluation framework is depicted in Fig 1.

[1] `http://www.bic.mni.mcgill.ca/brainweb/`

The simulation procedures of the artefacts that are considered in this paper are desribed below:

- *Bias Field Inhomogeneity*: It arises due to inhomogeneity in RF and results in a smooth multiplicative variation of the RF levels throughout the image. Here, the baseline image as well as the atrophied images are degraded using two different intensity non-uniformity (INU) fields (20% INU) available with BrainWeb. Let B be the bias field. The image (Y) degraded by the bias field (B) is simulated as $Y(s) = X(s).B(s)$ for each voxel s and the artefact-free image X.

- *Noise*: RF emission from the patient's body and the measurement chain of the MR scanner are responsible for noise in the images. It results in an irregular granular pattern on the image[2]. In this work, a Gaussian noise is added to all brain scans, such that a signal to noise ratio (SNR) of 15dB is achieved. Gaussian noise provides a good approximation to the noise present in MRI when signal-to-noise ratio is larger than 2 dB[29]. Note that this approximation holds in all our simulations (even for the CSF region which may have a low intensity in T1 images as compared to grey or white matter) because a signal-to-noise ratio greater than 2 dB is maintained.

- *Geometrical Distortions*: They result from errors in gradient field strength and non-linearity of gradient fields in the MR scanner. It introduces image deformations other than the actual anatomical changes in the brain and this affects the accuracy of the atrophy estimation procedures. Assuming that the undistorted brain image is available, measuring geometrical distortions means performing a point to point registration between the distorted image 3D coordinate x' and the undistorted image 3D coordinate x. Let T be the transformation that represents the geometrical distortions then:

$$x' = Tx \tag{5}$$

  Holden *et al.* [17] suggest that geometrical distortions can be modeled as a fourth degree polynomial as follows:

$$x' = P_0 + P_1 x + P_2 x^2 + P_3 x^3 + P_4 x^4 \tag{6}$$

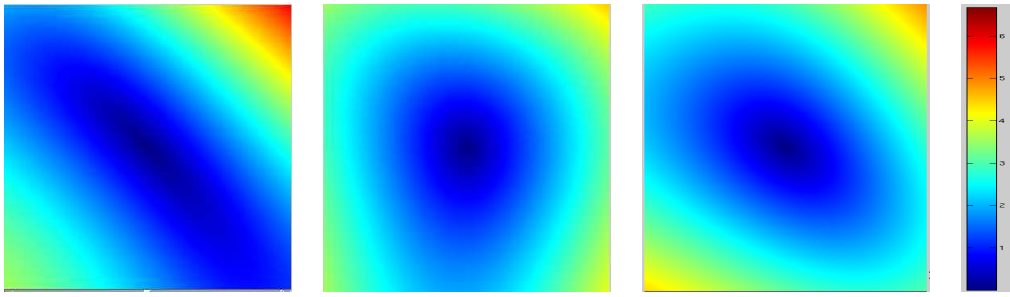
[2] `http://www.e-mri.org/quality-artefacts/signal-to-noise-ratio.html`

where P_d refers to the coefficient of degree d of the polynomial. To simulate these distortions, we need a rough numerical idea of the geometrical distortions that occur in the real scenario. Due to the unavailability of any undistorted image, we resort to registering two brain scans of a healthy subject that were taken on the same day. The registration parameters represent the geometrical distortions as well as any rotation and/or translation due to head movement between the two scans. We use "Automated Image Registration (AIR) [30]" algorithm (version 5.2.5) to perform the registration. Since AIR is a polynomial based registration, we directly obtain the registration parameters in the form of polynomial coefficients, for the simulation of geometrical distortions. Note that only relative geometrical distortions between the two scans are estimated. This artefact is simulated only on the baseline image that is compared with other atrophied images that are free of any geometrical distortions.

The simulated field is illustrated in Fig 2. This field induces a mean displacement of 1.57 ± 0.66 mm in the whole head area and of 1.22 ± 0.54 mm in the brain area. In order to make sure that the simulated field is realistic, we compared our geometrical distortion field to the one in [31] who demonstrate a phantom based geometrical distortion field which is determined based on $T1$ weighted acquisitions. The range of displacements in [31] is $0 - 5$ mm gradually increasing from the center to the periphery. Our displacement field varies smoothly in the range $0 - 6.5$ mm, also showing an increase from the center to the periphery. As a result of the simulated transformation, the brain volume of the baseline scan undergoes a decrease of 2% globally.

- *Interpolation Noise*: The idea is to investigate and quantify the change in the estimation of atrophy due to the introduction of an extra interpolation step. This is important because interpolation is used during registration which is an inevitable step for most of the longitudinal atrophy estimation techniques. To simulate this artefact, we rotate all the images by 1° using fifth order B-spline interpolation.

- *Presence of Lesions*: Lesions in the brain may arise due to a number of factors such as stroke, tumors, arteriovenous malformations, MS, injury or congenital brain abnormalities. The presence of lesions in the brain may have an adverse effect on the analysis of atrophy. For example,



(a) (b) (c)

Figure 2: The modulus (in mm) of the simulated geometrical distortion field (a) Coronal (b) Sagittal (c) Axial views. The simulated deformations vary in the range of $0 - 6.5$ mm.

the mis-classification of white matter lesions as CSF/gray matter or partial volume effects may lead to inaccuracies in volume estimation. To evaluate the effect of lesions, we use the MS database of BrainWeb containing a normal brain and images with mild, moderate and severe degrees of lesions. Due to the unavailability of noise-free images with lesions, the experiments are performed on normal as well as images with lesions that are degraded with 3% of noise [20]. Since the images under test contain the same amount of noise, it can be assumed that any variability between the results with different lesion loads is due to the presence of lesions only.

3.2. Evaluated Brain Atrophy Estimation Methods

Several approaches have been proposed in the literature to estimate brain atrophy. Popular methods include "Boundary Shift Integral" (BSI) [6], "Structural Image Evaluation, using Normalization, of Atrophy", SIENA [7, 8], for longitudinal evaluation and SIENAX [8] for cross-sectional measurements. We choose to study these methods since they are freely available and have been used in a number of studies earlier [21, 22, 11, 23, 13]. SIENA, SIENAX and BSI are described briefly in the following section.

3.2.1. SIENA

SIENA is a widely used automated tool that estimates longitudinal brain atrophy [7, 8]. This algorithm begins by extracting the brain and skull from the baseline and repeat scans using the Brain Extraction Tool (BET) [32]. The brains are then affine registered using the skull images to constrain the

scaling. The affine transformation, thus obtained, is first divided into two halfway transformations that are then applied to the brain scans to re-sample them into a common space. This avoids asymmetric blurring of the scans due to interpolation. Next, tissue type segmentation is performed in order to find brain/non-brain edge points [33]. At this step, bias field correction is also applied. The percentage brain volume change (PBVC) is then computed by estimating the mean perpendicular edge displacement over the entire brain, from one time point to another. The key step of this algorithm is to estimate the displacement of the edges between the scans. Errors in segmentation and registration of the edge points thus affect the quality of the results obtained using SIENA.

3.2.2. SIENAX

SIENAX [8] attempts to estimate cross-sectional atrophy using a single time point scan. Unlike SIENA that indicates the rate of disease progression, this tool gives an estimate of the current extent of disease progression [23]. In this study, SIENAX is used for longitudinal measurements. Brain volume is estimated for each of the two given scans and then the percentage decrease in brain volume from time point one to two is calculated.

To begin with, the brain and skull images are extracted from the given scan. The MNI152 standard brain is affine registered to the given brain, using the extracted skull and the standard brain skull to constrain the scaling. At this step, a volumetric scaling factor is calculated which is required for brain normalization at the final stage of the algorithm. Next, tissue type segmentation [33] is performed on the original (unregistered) extracted brain, up to a sub-voxel accuracy. The total brain volume is derived from this segmentation result. To normalize for head size, the brain volume is multiplied with the volumetric scaling factor derived earlier. In all our analyzes with SIENAX unnormalized brain volume is used as the real brain volume is required for comparison with the ground truths. It may be noted that for SIENAX the brain volume change is directly calculated from the segmentation; so inaccuracies in segmentation, (for instance, mis-classification of lesions) affect the performance of this method.

3.2.3. BSI-UCD

Another popular tool for estimating cerebral atrophy from two time point brain scans is BSI [6]. In this approach, the repeat scan and the baseline scan are registered using an affine transformation. In the original BSI implemen-

tation [6], an affine registration procedure that determines the spatial scaling factors based on the cranial surface between the two scans is used. The registered scans are then intensity normalized in order to compare the intensity values. At this stage, the brain is segmented using an iterative morphological technique, where the sizes of the morphological operators is selected manually. The idea is to remove all non-brain tissue including CSF, in order to define a boundary region on which the boundary shift integral is computed. The brain atrophy is estimated by calculating the amount by which the boundary of the brain tissue has moved over a period of time.

Due to the unavailability of the original BSI implementation, we use the BSI implementation developed by Imaging of Dementia and Aging lab, University of California, Davis (BSI-UCD) [3]. The calculation of the boundary shift integral is done in exactly the same way as described in [6] but different pre-processing algorithms are used. The bias correction of BSI-UCD [4] is a template based bias correction procedure in which a template, which is assumed to be bias-free, is compared to the subject image. The corresponding voxel intensities are compared in the template and the subject image in order to identify the non-uniformities. An affine registration method is also a part of the package[5]. The affine registration procedure of BSI-UCD does not estimate spatial scaling factors using the cranial surface in order to compensate for spatial scaling. In order to avoid any confusion, we will refer to the BSI implementation that we use as BSI-UCD in the paper.

4. Experimental Setup

In this paper, we use the SIENA and SIENAX implementations available as a part of the FMRIB Software Library (FSL)[6] version 4.1 [34]. As mentioned earlier, we use the BSI implementation developed by Imaging of Dementia and Aging lab, University of California, Davis (BSI-UCD). The bias correction and registration (with Cross-Correlation matching criterion) provided by the BSI-UCD package are used in all the experiments with BSI unless mentioned otherwise. While the implementations of SIENA and SIENAX are completely automated, the implementation of BSI-UCD

[3] `http://neuroscience.ucdavis.edu/idealab/software/index.php`
[4] `http://neuroscience.ucdavis.edu/idealab/software/bias_correction.php`
[5] `http://neuroscience.ucdavis.edu/idealab/software/linear_coreg.php`
[6] `http://www.fmrib.ox.ac.uk/fsl/fsl/list.html`

requires manual intervention, in order to obtain a gray-white matter mask, for delineating the brain boundaries on which the boundary shift integral is calculated. This problem is automatically alleviated in our case since the gray-white matter mask of the baseline image is available (with BrainWeb). This mask can then be warped, using the transformation estimated for atrophy simulation, for obtaining masks of the atrophied images. Moreover, the Brain Extraction Tool (BET) [32] of the FSL library is used to perform brain extraction for BSI-UCD, SIENA and SIENAX in all our experiments. Note that two types of masks are required by BSI-UCD. One is the "brain mask" (Fig. 3(b)) such as the one obtained as an output of BET after the removal of the non-brain tissue. This mask is used by the bias correction and the registration steps of the BSI-UCD algorithm. Another mask defining the gray-white matter (the atrophy affected areas of the brain) boundaries (Fig. 3(c)). This is used by the final step of the BSI-UCD algorithm (atrophy calculation using movement of boundaries with time).

The experiments are conducted using the default parameters of SIENA, SIENAX and BSI-UCD so that the comparison is not biased as a result of manual intervention. For SIENAX, the "unnormalized" brain volume estimates are used for all subsequent calculations and comparisons.

The time required for one experiment of simulation of atrophy using the proposed approach is 2 hours. The computational times (including the time required for pre-processing) when performing one experiment with SIENA, SIENAX and BSI-UCD are approximately 1 hour, 20 minutes and 1 hour, respectively. All the experiments are conducted on an Intel Dual Core 2.40 GHz processor with images of size $256 \times 256 \times 256$.

A summary of the brain extraction, registration and bias field correction algorithms used for SIENA, SIENAX and BSI in this work is illustrated in the Table 3.

Table 3: Summary of the brain extraction, registration and bias field correction algorithms used for SIENA, SIENAX and BSI in this work.

	SIENA	SIENAX	BSI
Brain extraction	BET[32]	BET[32]	BET[32]
Registration Type	Affine(FLIRT [35])	Affine(FLIRT [35])	Affine[5]
Similarity Criterion for Registration	Correlation Ratio	Correlation Ratio	Cross-correlation
Final Interpolation	Tri-linear	Tri-linear	Tri-linear
Bias-Field Correction	With Segmentation (EM based)[33]	With Segmentation (EM based)[33]	Template based[4]

(a) (b) (c)

Figure 3: (a) A T1-weighted image (b) Brain mask (extracted by BET) and (c) Gray-White matter mask of the brain computed from the T1-weighted image shown in (a).

5. Results on Simulation of Atrophy and Discussion

In this section, we study the performance of the proposed atrophy simulation algorithm. First, we investigate the influence of considering the logarithm of the Jacobian in the objective function (*Log*-norm) instead of the standard sum of squared differences (L_2-norm). To this end, we simulate several rates of uniform atrophy (20% and 50%) and hypertrophy (25% and 100%) without considering any invariance constraint. Histograms of the Jacobian values of the estimated deformation fields are represented in Fig. 4. We can see that the use of the logarithm for simulating a large atrophy (Fig. 4(b)) yields a smaller dispersion of Jacobian values as compared to the L_2 norm. This dispersion increases when simulating large hypertrophies (Fig. 4(d)). When simulating small atrophies or hypertrophies, the use of the logarithm or the sum of squared differences tends to yield similar results (Fig. 4(a) and (c)). The distributions of the simulated Jacobian values also highlight the fact that using the *Log*-norm yields to a constant relative dispersion of the Jacobian values (the dispersion varies linearly with the desired Jacobian value) which is not the case with the L_2 norm. Hence, the *Log*-norm is more consistent as opposed to the L_2 norm.

A quantitative analysis of the ability of the proposed algorithm to simulate the desired atrophy has also been done (see Table 4). Simulations have been done by considering a uniform atrophy over gray and white matter using the Brainweb image. In all our experiments, we use 7 levels in the multi-resolution deformation model. The final scale amounts to considering a regular grid of control points with a spacing of $2 \times 2 \times 2$ voxels (each con-

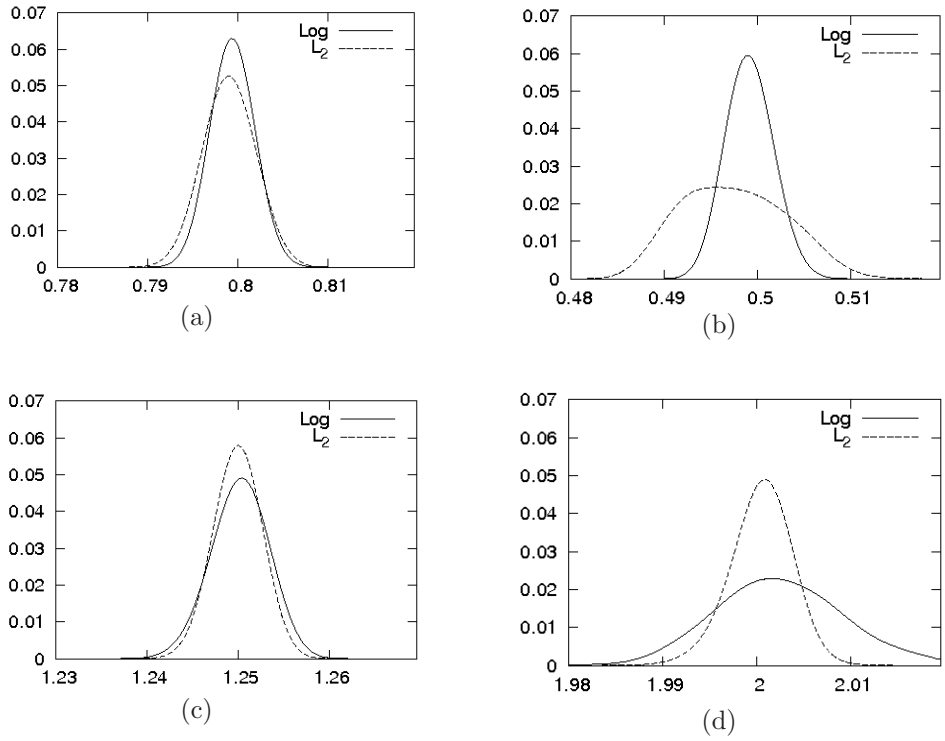


Figure 4: Distributions of Jacobian values obtained when simulating uniform atrophies [(a) 20% ($J = 0.8$) and (b) 50% ($J = 0.5$)] and hypertrophies [(c) 25% ($J = 1.25$) and (d) 100% ($J = 2$)].

trol point corresponding to the center of a B-spline function). We can notice that, on the average, the desired atrophy is well achieved without and with the skull constraint. It can also be seen that, without the skull constraint, the standard deviation of the Jacobian values decreases when the desired atrophy rate increases. An inverse trend is observed when considering the skull constraint. This is due to the fact that the Jacobian values tend to be equal to one in the neighborhood of the skull and do not vary abruptly in order to warrant topology preservation. Thus, the Jacobian values of the voxels located on the brain boundary are far from the desired atrophy. Fig. 5 elucidates this point using an error map (the absolute difference between the desired Jacobian map and the one obtained by our algorithm).

Notice that the proposed simulation algorithm can easily achieve very

Table 4: Influence of considering the skull constraint on the mean and standard deviation of Jacobian values of the simulated deformation fields.

Desired Atrophy	Without skull constraint	With skull constraint
10%(J=0.9)	0.9017 ± 0.0021	0.9032 ± 0.0115
20%(J=0.8)	0.8015 ± 0.0021	0.8019 ± 0.0202
30%(J=0.7)	0.7011 ± 0.0020	0.7032 ± 0.0349
40%(J=0.6)	0.6008 ± 0.0018	0.6025 ± 0.0797
50%(J=0.5)	0.5005 ± 0.0017	0.5140 ± 0.2088

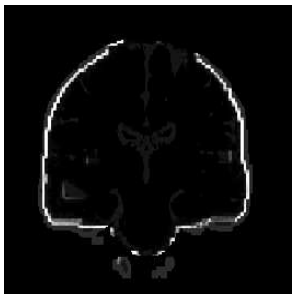


Figure 5: Error map illustrating the absolute difference between the desired and the obtained Jacobian maps for 50% ($J = 0.5$) of atrophy. The gray levels (from black to white) depict the errors in the range of $0 - 5.5$ (from low to high). Note the high error near the skull surface. Also, one must remember that an atrophy of 50% in the brain is an extreme case of brain tissue loss.

high atrophy, contrary to the method proposed in [18], which requires the estimation of a large atrophy in an incremental way. For example, it has been possible to simulate a uniform atrophy of 99.9%($J = 0.001$), without the skull constraint, with an obtained average Jacobian value of 0.00106 ± 0.000716. Such an atrophy rate is obviously unrealistic, but it highlights the ability of the proposed method to converge to the desired solution, even for very low Jacobian values, while still preserving the topology.

An example of a simulated atrophied image is shown in Fig. 6.

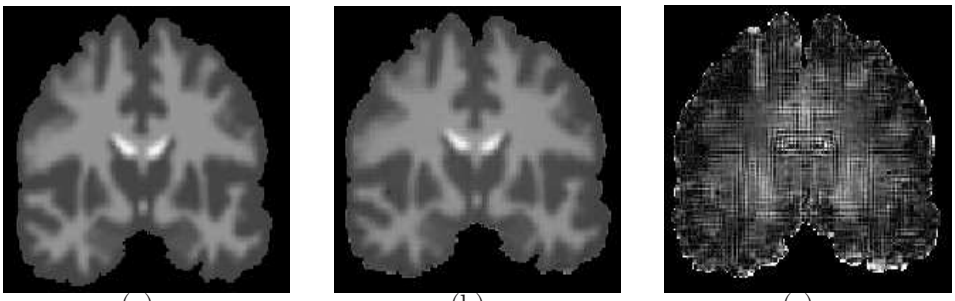
Table 5: Table showing the desired and the obtained mean and standard deviation values of the Jacobian for various parts of a brain in which non-uniform atrophies are simulated using the proposed atrophy simulation approach. The desired and the achieved atrophy values are also shown.

Area of interest	Desired Jacobian	Obtained Jacobian	Desired Atrophy	Achieved Atrophy
Ventricles	1.0408 ± 0.0195	1.0444 ± 0.0110	+4.08%	+4.44%
Cortex and cerebellum	0.9520 ± 0.0086	0.9562 ± 0.0092	−4.80%	−4.38%
Hippocampus	0.9648 ± 0.0096	0.9682 ± 0.0101	−3.52%	−3.18%
White Matter	0.9911 ± 0.0141	0.9936 ± 0.0280	−0.89%	−0.64%
Brain	0.9675 ± 0.0220	0.9673 ± 0.0658	−3.25%	−3.27%

Experiments have also been conducted for the simulation of non-uniform

(a) (b) (c)

Figure 6: (a) Original BrainWeb image (b) Simulated image with 10% of globally uniform atrophy (c) Difference between images (a) and (b). Note that there is no deformation on the skull between (a) and (b).



(a) (b) (c)

Figure 7: Simulation of non-uniform atrophies (a) Desired Jacobian map (b) Jacobian map obtained by using our algorithm (c) Error map. The gray levels (from black to white) depict the values of the Jacobian (from low to high) in the range of $0.94 - 1.08$ for the maps shown in (a) and (b) and for the error map (c) in the range of $0 - 0.12$.

atrophies. To this end, we utilize the "Internet Brain Segmentation Repository" [7]. This database provides manual segmentation of the brain in several brain structures. We simulate different atrophies in some structures of the same brain including the hippocampus, the cortex, the cerebellum, the ventricles and in the rest of the brain without considering the skull constraint. In order to ensure that the simulated atrophy rates corroborate with the atrophies in real cases, we refer to [36] and [13], which mention atrophy rates observed in various parts of the brains of Alzheimer's disease patients. We

[7] `http://www.cma.mgh.harvard.edu/ibsr/`

(a) (b) (c)

Figure 8: Simulation of atrophy in a brain with an increase in the MS lesion size: (a) Original BrainWeb image with moderate lesions (b) Image with 10% of global atrophy and 100% of MS lesion increase (c) Difference between images (a) and (b).

illustrate the desired and the obtained Jacobians with the corresponding error map in Fig. 7. Table 5 shows the desired and obtained Jacobian values for various areas of interest with the corresponding atrophy rates. It can be seen that the algorithm is able to reach the desired atrophy accurately.

The proposed simulation framework is also versatile and can be used to simulate a more complicated pattern of atrophy. For instance, it can be used to simulate simultaneously a global brain atrophy and a change in a given pathological area such as multiple sclerosis lesion evolution or tumor growth. The BrainWeb MS moderate lesion database is used for this experiment. The simulation is done by first simulating the hypertrophy in MS lesion area while constraining the gray matter and the cerebrospinal fluid to be invariant in order to ensure that lesions evolve inside the white matter. Then, global atrophy is simulated, while only constraining the skull to be invariant, and the estimated deformation field is combined with the previous one. Notice that the global atrophy will modify the hypertrophy rate of the lesions. This has to be taken into account when simulating MS lesion increase in order to get the desired volume change. In Fig. 8, we present a simulation of 10% of global brain atrophy and 100% of MS lesion volume increase.

6. Evaluation Results of SIENA, SIENAX and BSI-UCD and Discussion

Our evaluation is divided into two parts. The first part consists of the simulation of a number of atrophies on a single normal brain image from BrainWeb to study the robustness of SIENA, SIENAX and BSI-UCD to

various artefacts. In the second part, we present results with these algorithms for 18 different normal brains (from BrainWeb [37]) to examine the effect of anatomical variability on these methods. Atrophy is simulated using the *Log*-norm in both the cases. The experiments are detailed in the following sections. In the forthcoming discussion, we would use the term "error" to represent the absolute difference between the ground truth and the estimated Percentage Brain Volume Change (PBVC), unless stated otherwise.

6.1. Experiments on One Normal BrainWeb Image

For this part of the experiments, the results are obtained by simulating brain atrophy ranging between 0-1% (step size 0.1%) and 1-10% (step size 1%). Although, the brain atrophy range of 0-1% is more relevant to neuro-degenerative pathologies, we also present results for larger brain volume changes, in order to better assess the accuracy of these methods. Fig. 9(a) shows the percentage brain volume change (PBVC) between any two brain scan pairs, $B1$ and $B2$, such that the simulated atrophy on brain $B1$ is less than that of $B2$ (for the simulated atrophy ranges of 0-1% and 1-10%). For example, if the simulated atrophy on $B2$ is 10%, then it is compared with all simulated atrophies of $1 - 9\%$ (on $B1$). Fig. 9(a) compares the methods under consideration with respect to the ground truth, for the noise-free case.

6.1.1. Effect of Bias Field Inhomogeneity and Noise

In order to comprehend the effect of these artefacts, we create three sets of images using the baseline BrainWeb image and the simulated images with atrophy.

1. The baseline image as well as the atrophied images are degraded using different intensity non-uniformity (INU) fields (20% INU) available with BrainWeb.
2. A second set of images is obtained by adding Gaussian noise to all the brain scans, such that a signal to noise ratio (SNR) of 15dB is achieved.
3. We create a third set of images that are degraded by bias field inhomogeneity followed by noise using the same parameters as in 1 and 2.

Fig. 9 (b-d) show the PBVC estimation for SIENA, SIENAX and BSI-UCD for observations degraded with bias field, noise, both bias field and noise, respectively. The absolute errors in the estimated PBVC for the three methods are depicted using boxplots in Fig. 10 for the noise-free case and

for images degraded with bias-field inhomogeneity and noise. Note that the boxplots are shown separately for atrophy ranges of $0 - 1\%$ and $1 - 10\%$. This is done to capture the trends in the observed errors properly as Fig. 9 clearly shows that the errors in the estimated PBVC are dependent on the simulated atrophy value.

For low atrophy values (less than 1%), SIENA, SIENAX and BSI-UCD show a similar performance in terms of error in PBVC measured with respect to the ground truth (See Fig. 9(a)). Fig. 10(a) illustrates that, for the noise-free case, the maximum errors in the estimated PBVC values for SIENA, SIENAX and BSI-UCD are 0.15%, 0.24% and 0.34%, respectively for the low atrophy values. Also, Fig. 9(a) shows that SIENA overestimates the atrophy, while BSI-UCD underestimates it for large atrophy values when no artefact is added. For this range of atrophy, Fig. 10(b) shows that SIENAX (maximum error 1%) outperforms SIENA and BSI-UCD (maximum error 3.7%).

As can be seen from Fig. 9, the introduction of bias field inhomogeneity (b) and both noise and bias field inhomogeneity (d) leads to a visible increase in errors for all the three methods. The maximum errors obtained with SIENA, SIENAX and BSI-UCD for the atrophy range $0-1\%$ and $1-10\%$ are 1.55%, 5.60%, 3.35% and 5.70%, 5.83%, 9.22%, respectively when using images corrupted with bias-field inhomogeneity and noise (Fig. 10(c-d)). The errors that are discerned here are very large as compared to the simulated values of atrophy. A prominent reason for this is incorrect brain extraction while using BET due to the addition of bias field inhomogeneity, mostly at the boundary of the brain. In our observation, the addition of noise over bias field inhomogeneity leads to a significant change in the brain extraction as compared to the brain extraction performed when only bias field inhomogeneity is added. However, for the observations that are degraded with Gaussian noise only, we do not observe any gross errors in the extraction of brain. Brain extraction is crucial for SIENAX because segmentation is performed on the brain image, that is directly related to the calculation of the brain volume. Although, SIENA uses a combined brain mask from the two examinations to evaluate the change in brain volume, it is seen that non-brain areas are included in the calculations if they were included in one of the examinations. This is a reason of the degradation of the performance of SIENA since it leads to errors in detection of the brain/non-brain boundaries. It is evident that registration of the two examinations is an important step for SIENA and BSI-UCD because it determines the relative edge position

(for SIENA) and corresponding intensity values (for BSI-UCD). We would like to point out that bias field correction integrated with the segmentation algorithm FAST [33] is performed for SIENA and SIENAX. In our experiments on BSI-UCD, a bias field correction provided with the implementation of BSI-UCD is applied after the registration has been performed. Since the bias field correction step comes after the brain extraction and registration steps are performed, the error in these steps possibly propagate to the final result. We note that, for BSI-UCD too, brain extraction is an important step since the final result depends on getting a good mask of the brain. Our experiments also suggest that an improper extraction of the boundary of the brain leads to a mis-calculation of the boundary shift integral ($\sim 2-3\%$ in some cases). Note that for a pathology like MS, where the annual atrophy is small, an error of 0.5% in PBVC, for instance, is quite significant, specially when it is not consistent over repeated measurements.

A striking observation is that addition of Gaussian noise leads to a decrease in the mean absolute errors in the measurements of SIENA and SIENAX (Table 7). As described previously, SIENA and SIENAX algorithms are sensitive to the extraction of brain boundaries. When we compare the brains extracted by BET, for the noise-free and noise-only case, the extracted brains of some of the noise-free images contain some parts of the brain in addition (such as some extra length of the brain stem, CSF), which are not present in the noise-only extracted brains in some images. The extracted brains for the noise-only case are closer to the BrainWeb ground-truth as compared to the ones in the noise-free case. Noise may affect the subsequent steps also. For BSI-UCD, the errors in the PBVC values show an increase of around 0.5% and 1% for simulated atrophy of less than 1% and between 1% and 10%, respectively, with respect to the noise-free case. Although, BET is used with BSI-UCD also, an increase in the error is seen for BSI-UCD for the noise-only case which may be due to the pre-processing steps or due to the use of image intensities directly for the calculation of the boundary shift integral. Note that such phenomena are random and may arise due to the noise that is added. It is also informative to determine whether these differences are statistically significant. A two-sample paired t-test indicates that for SIENA ($p < 0.001$) and BSI ($p < 0.001$) there are significant differences in the absolute errors with respect to the ground truth for the noise-free and noise-only case at the 95% level. For SIENAX these differences are not significant ($p = 0.1258$) at the 95% level.

6.1.2. Effect of Geometrical Distortions

Geometrical distortions are simulated as described in section 3.1. Fig. 11 shows the PBVC comparison for SIENA, SIENAX and BSI-UCD, respectively with the ground truth when the geometrically distorted BrainWeb image is compared with simulated images for different atrophy rates. For our experiments, we simulate 4th degree polynomial distortions. SIENA constrains the affine registration, which corresponds to a 1st degree polynomial, using the skull, attempting to correct for the geometrical distortions. SIENAX uses the same registration as SIENA. However, the purpose of registration for SIENAX is to calculate the scaling factor to normalize the brain volume with respect to the MNI brain atlas. As mentioned earlier, in our study only unnormalized results are considered. With the affine registration of BSI-UCD, one can expect that the scaling distortions are somewhat corrected. On comparing the three methods, we find that SIENA is least affected by the simulated distortions: when compared to the ground-truths, a mean absolute error of 0.07% is observed in the PBVC values. SIENAX and BSI-UCD digress more from the ground truth where mean absolute errors of 1.68% (SIENAX) and 0.82% (BSI-UCD) in the estimated PBVC are observed. The mean absolute errors are calculated for the simulated atrophy range of $0-1\%$. The fact that an additional volume decrease of 2% is induced due to the simulation of these distortions indicates that the distortion correction of SIENA is effective. An increase in errors for SIENAX with respect to the noise-free case is expected because the distortions are not rectified. As for BSI-UCD, the error indicates that the geometrical distortions can not be corrected for only using an affine transformation. The annual atrophy rate in MS and other neuro-degenerative diseases is comparable to the estimation error caused by geometrical distortions. Hence, correction of such distortions is crucial in order to guarantee the accuracy of these measurement techniques.

6.1.3. Effect of Interpolation

The effect of introducing an additional interpolation step to the images on the PBVC obtained with the three methods is shown in Fig. 12. We see that SIENA and SIENAX remain generally unaffected, when compared to the noise-free case shown in Fig. 9(a). For both these methods, the error in estimation is less than 0.23% for the atrophy range of 0-1%. For the higher

range (1-10%), the error is less than 3.4% and 1% for SIENA and SIENAX, respectively. While the error in PBVC increases for BSI-UCD for atrophy values lower than 1% (maximum error 1.1), it decreases for atrophy values between 1-10% (maximum error 1.7%), as compared to the noise-free case shown in Fig. 9(a).

6.1.4. Effect of Lesions

To determine the effect of presence of lesions on atrophy, SIENA, SIENAX and BSI-UCD algorithms are run for 6 cases (ranging from "normal" brain to "severe" lesions). These are illustrated in Table 6. In all these cases, no additional atrophy is simulated. Hence, it is expected that ideally the atrophy estimated between the various cases listed in Table 6 is close to zero. The deviations from zero thus represent the change in estimated atrophy due to the presence of lesions. It can be noticed from the Table 6 that lesions can lead to a significant non-desired change of up to 0.2% when comparing a normal brain with the same brain with lesions using SIENA. SIENAX is more affected by the presence of lesions when a normal brain is compared with the same brain with lesions (maximum PBVC is $\sim 0.4\%$). However, when two brains with lesions are compared the errors are lesser for both SIENA and SIENAX as depicted in Table 6. A maximum error of 0.78% is observed when testing the influence of lesions on the atrophy estimated by BSI-UCD. As we pointed out earlier, a gray-white matter mask must be provided to BSI-UCD in order to define the brain boundaries on which the boundary shift integral is calculated, which in our case is the ground truth mask. The images that are used for this experiment contain some lesions close to brain boundaries which can affect the atrophy calculation. However, only the presence of lesions is not expected to have a large effect on the estimated atrophy by BSI-UCD in the experiments here. Note that the experiments are conducted with images degraded by noise, which is mainly responsible for the poor performance of BSI-UCD (see Table 6). Our experiments on images degraded with noise have already illustrated the sensitivity of BSI-UCD towards presence of noise (see Table 7).

Table 7 summarizes the results corresponding to various sources of error discussed in this section.

Table 6: PBVC obtained using SIENA, SIENAX and BSI-UCD for establishing the effect of lesions on atrophy measurements. In the first column "Normal-Mild" means that atrophy is estimated between BrainWeb normal image and the mild lesion case

Case	SIENA	SIENAX	BSI-UCD
Normal-Mild	0.1177	0.4574	0.3632
Normal-Moderate	0.2070	0.4309	0.4245
Normal-Severe	0.1487	0.3868	0.4573
Mild-Moderate	-0.0078	-0.0263	0.7757
Mild-Severe	-0.0246	-0.0702	0.7121
Moderate-Severe	-0.0589	0.0439	0.7696

Table 7: Summary of the results discussed in this section for the simulated atrophy range $0 - 1\%$ in one BrainWeb image. This table illustrates the mean error in the estimated PBVC for various artefacts. Note that, for presence of lesions, the error represents the non-desired change that is observed when comparisons are done using different versions of the same brain with varying lesion load. Refer to section 6.1.4 for details.

Artefact	SIENA	SIENAX	BSI-UCD
Noise-free	0.0615 ± 0.0407	0.0815 ± 0.0710	0.1072 ± 0.0899
Bias Field Inhomogeneity	0.2940 ± 0.4343	0.8420 ± 0.7502	1.0412 ± 0.3827
Noise	0.0292 ± 0.0226	0.0673 ± 0.0433	0.4400 ± 0.1082
Bias Field Inhomogeneity and Noise	0.3492 ± 0.3812	2.0277 ± 1.4622	0.9131 ± 0.7993
Interpolation artefacts	0.0628 ± 0.0401	0.0814 ± 0.0539	0.2209 ± 0.1619
Geometrical Distortions	0.0745 ± 0.0562	1.6840 ± 0.1233	0.8273 ± 0.3495
Presence of Lesions	0.0941 ± 0.0772	0.2359 ± 0.2089	0.5837 ± 0.1886

6.2. Experiments on Multiple Normal Brains

In this section, we present the results of our experimentation on several anatomical models of normal brains provided by BrainWeb [37]. The BrainWeb images have been obtained from the brain images of normal adults. Each image is created by registering and averaging T1, T2 and PD weighted MRI images for each subject. A fuzzy minimum distance classification is used to separate the white matter, gray matter, CSF and fat from the average volumes for each case.

These experiments are rendered on 18 normal brains by simulating two patterns of atrophy: uniform and non-uniform over the brain.

- Uniform atrophy case: For every subject, 4 atrophy levels, 0.2%, 0.5%, 1% and 1.5%, are simulated in a globally uniform fashion in the brain.
- Non-uniform atrophy case: In every brain, a slowly varying atrophy in the gray matter and a uniform atrophy in the white matter is

simulated. Three such patterns are simulated using three different slowly varying fields. The simulated gray matter and white matter atrophies are (range of gray matter atrophy/ white matter atrophy/ global atrophy): $0.4\% - 0.8\%/0.2\%/0.25\%$, $0.4\% - 0.6\%/1.5\%/0.3\%$ and $0.3\% - 0.7\%/0.5\%/0.57\%$ which we will refer to as A, B and C, respectively in the following discussion.

In both cases, we choose these values of atrophy so that the whole brain simulated atrophies are consistent with the atrophies observed in many typical neuro-degenerative diseases ($0.5\% - 1\%$). The atrophied images are compared with the baseline images for all the 18 anatomies. Two comparisons are performed, one involving noise-free images and the other images degraded with bias field inhomogeneity and noise. The parameters of the applied bias field inhomogeneity and noise are the same as in Section 6.1.1.

Fig. 13(a-d) illustrate the outcome of this experiment, for noise-free (a), (c) and "noisy" images (b), (d) for the uniform and non-uniform atrophy cases. The figure depicts the mean and standard deviation of absolute error in the estimated PBVC with respect to the ground truth for the 18 brains. The mean error and the standard deviation for a particular value of simulated atrophy are calculated as follows:

Let $E_{ki} =\mid PBVC_{GTki} - PBVC_{Estimatedki} \mid$ where $k = 1 \ldots 18$ and $i = 0.2\%, 0.5\%, 1\%, 1.5\%; A, B, C$.

$$\mu_i = \overline{E_{ki}} \tag{7}$$

$$\sigma_i = \sqrt{\frac{1}{n}\sum_{k=0}^{n}(E_{ki} - \mu_i)^2} \tag{8}$$

where $PBVC_{GTki}$ and $PBVC_{Estimatedki}$ are the ground truth and estimated PBVC, respectively. n represents the number of experiments (in this case $n = 18$). Ideally, one expects the mean error and standard deviation to be close to zero. Notice that the error in the estimation of PBVC is dependent on the amount of simulated atrophy for the noise-free case, Fig. 13(a) the error being higher for higher values of atrophy. The error in PBVC increases in the following order : SIENA, SIENAX and BSI-UCD. However, the standard deviation is the lowest for SIENA and of a similar order for SIENAX and BSI-UCD. A different trend is seen for the noisy case shown in Fig. 13(b). In this case, the error in the PBVC increases in the following order : SIENA,

BSI-UCD and SIENAX. The standard deviation is much higher for SIENAX as compared to SIENA and BSI-UCD. Similar results can be identified in the non-uniform atrophy case Fig. 13(c) and (d) with the exception that the performance of SIENA is comparable to that of SIENAX when non-uniform atrophies are simulated and noise-free images are used for conducting the experiments (Fig. 13(c)).

As we said before, the addition of bias field inhomogeneity and noise is responsible for errors in brain extraction. Particularly, when the atrophy under consideration is small (like in our experiment here), incorrect brain extraction can lead to misleading results. A disproportionate increase in error for SIENAX can be explained as following. Different bias field inhomogeneities are added to all the 18 baseline images and the simulated images. It is observed that whenever a strong bias field is present in a part of the image, some parts of the brain (corresponding to a volume of around 10000 mm^3 in some brains) are not detected by BET. The bias fields added to some of the simulated images also lead to the inclusion of non-brain parts. Hence, the PBVC estimated by SIENAX is erroneous, which is reflected by the mean of the error. Two factors contribute to the change in error in BSI-UCD: incorrect extraction of brain and perturbation of intensity values due to the addition of noise and bias field inhomogeneity. As mentioned in Section 6.1.1, the increase in error of SIENA can be attributed to the inclusion of non-brain parts, while finding the brain/non-brain boundary. Although similar in nature, SIENA does not lose performance in the noisy case as much as BSI-UCD due to the use of edges instead of intensity values to find the PBVC and also the fact that the final PBVC given by SIENA is the average of PBVC in both directions (first time point to second time point and the reverse), which can compensate to an extent for small errors in one of the directions.

It is also interesting to calculate the overall mean and the standard deviation of error in the estimated atrophy using these 18 cases for the simulated atrophy range $0.2-1.5\%$ (uniform atrophy case) and for the range $0.2-0.57\%$ (non-uniform atrophy case). This provides an idea of the accuracy and precision of the methods under consideration. The overall mean and standard deviation of error of the methods for the different cases are tabulated in Table 8. It is observed that SIENA performs better than SIENAX and BSI-UCD when noise-free and noisy images are used since it shows lowest mean and standard deviation of error. SIENAX performs better than BSI-UCD in the noise-free case. However, BSI-UCD outperforms SIENAX in the noisy case.

The same trend can be noticed for the noise-free uniform and the non-uniform atrophy cases, except that the performances of SIENA as well as SIENAX are comparable in the non-uniform atrophy case. Also, note that the absolute error is less, when non-uniform atrophy is simulated, for SIENAX and BSI-UCD, with respect to the corresponding error values in the uniform atrophy case. A two sample t-test between the absolute errors observed when non-uniform and uniform atrophies are simulated for the same simulated atrophy range ($0.2\% - 0.5\%$) shows that for all the three methods these differences are significant (at the 95% significance level, $p << 0.001$).

Table 8: Overall mean and standard deviation of absolute error in PBVC of SIENA, SIENAX and BSI-UCD for uniformly and non-uniformly simulated atrophy. Noisy observations are images degraded with noise and bias field inhomogeneity.

	Uniform Atrophy		Non-uniform Atrophy	
Method	Noise-free Observations	Noisy Observations	Noise-free Observations	Noisy Observations
SIENA	0.0944 ± 0.0656	0.6382 ± 0.5306	0.1545 ± 0.0948	0.4472 ± 0.3074
SIENAX	0.4492 ± 0.2210	4.0073 ± 2.4087	0.1510 ± 0.0629	3.4483 ± 1.7303
BSI-UCD	1.1881 ± 0.2418	1.7948 ± 0.9743	0.6715 ± 0.1891	1.5255 ± 0.6828

As mentioned earlier in this section, the errors in the estimated atrophy by SIENAX in the "noisy" case show an abrupt increase (as compared to the noise-free case) due to incorrect segmentation of brain by BET. In order to gain a better insight into this problem, voxel by voxel error in brain segmentation for the 18 baseline images is investigated. Fig. 14 shows the voxel by voxel mean error for the noise-free and noisy baseline images. It can be seen that in the noisy case the mean error is high near the cerebellum since this part is not taken into account by BET. This is due to a strong bias-field inhomogeneity in that area in some of the images. We also observed that the skull is not properly stripped for one of the cases and can be distinctly seen in Fig. 14(b). In BET, the detection of the brain boundary is dependent on local calculation of an intensity threshold, which is disturbed by the perturbation of intensity values, due to the existence of inhomogeneity in intensities. The results show that bias-field inhomogeneity is a major factor that contributes to the errors in brain segmentation.

We also conduct some additional experiments on BSI-UCD, in order to explain the high error that is observed. Studies on BSI (such as [13]) have

stated the mean absolute error to be around 0.2% which is lower than what can be seen in our analysis (See Table 8). Three experiments are performed with BSI-UCD using noise-free images on which non-uniform atrophy is simulated. To ascertain the effect of pre-processing (registration and bias-field correction) on the final BSI-UCD calculation, we skip these steps from the BSI-UCD analysis in the first experiment. These steps can be passed over as the images are bias free and the BrainWeb images are already registered to the same space. For the second experiment, the BSI-UCD algorithm is run normally as has been done for all our experiments with BSI-UCD. Finally, an experiment is conducted by replacing the registration of the BSI-UCD with an in-house Mutual Information based affine registration algorithm. These results are demonstrated in Fig. 15. It can be seen that when no pre-processing is used the BSI-UCD algorithm works better than the other two cases where images are pre-processed. This is expected since the BrainWeb images are already registered. The mean absolute error in the measured atrophy for the first experiment (no pre-processing) is 0.15%, which is in accordance with what has been reported in [13]. There is a reduction in the error with our in-house affine registration as compared to the registration of BSI-UCD, indicating that registration is a critical step when employing BSI-UCD and registration parameters should be carefully selected. We have also performed experiments that test the influence of the bias-field correction algorithm on the estimated error. Results, obtained on the noise-free case, show that there is no significant difference between the results when the bias-field correction is not performed.

7. Comparison with other studies

In this section, we compare the results of evaluation with other works. Camara *et al.* [13], who utilize the same criterion as in our study, report the mean absolute errors in the estimated atrophy as 0.27% and 0.22% for SIENA and BSI, respectively. These values do not agree with the mean absolute errors that we obtain in our work (SIENA: 0.64%, BSI-UCD: 1.79%) in a real scenario (where experiments are conducted after a deliberate addition of bias-field inhomogeneity and noise after the simulation of atrophy). The experiments are performed on real images. We believe that the difference in the quality of the images could be responsible for the differences in the results. As we have said before, bias-field inhomogeneity along with noise can degrade the performance of SIENA, SIENAX and BSI-UCD in a large way.

Also, the results may vary depending on the implementation of BSI used in the simulations. It is shown in the previous section that the registration algorithm should be carefully chosen so as to guarantee a good performance by BSI.

Let us compare our results with noise-free observations to the results of Camara *et al.* [13]. In our analysis, a mean absolute error of 0.09% is observed when using noise-free observations, which is less than what was obtained by Camara *et al.* (0.27%) for SIENA. We obtain higher mean absolute errors for BSI (Current study: 0.67%, Camara: 0.23%). Note that the mean absolute errors can only be compared to those studies who use simulated ground truths as us. The employment of other criteria of evaluation by Smith *et al.* [23] and Boyes *et al.* [11] makes it incorrect to compare their mean absolute errors with our estimates.

Our results also illustrate that SIENA has a tendency of overestimating atrophy while BSI-UCD underestimates it and that the error in the measured atrophy is larger for higher values of atrophy (See Fig. 9). The least squares fitted scaling factors are GT(Ground Truth)=SIENA$*$0.89; GT=BSI-UCD$*$1.88; SIENA=BSI-UCD$*$2.15 (for noise-free observations). A similar trend has been observed in Smith *et al.* [23] (SIENA=BSI$*$1.20) and Camara *et al.* [13] (GT=SIENA$*$0.90; GT=BSI$*$1.18; SIENA=BSI$*$1.29). The scaling factors in our study with BSI-UCD have a higher value than the other studies. This can be explained by the fact that we use a different implementation of BSI and the quality of images too is not the same. We obtain a weak correlation of $r = 0.42, p = 0.0015$ between SIENA and BSI-UCD while other studies show a much better correlation (Smith: $r = 0.87$, $p < 0.0001$, Camara: $r = 0.97$). Our results indicate a better correlation between the atrophy measured by SIENA and SIENAX ($r = 0.89$; $p < 0.0001$) as compared to Smith et al. ($r = 0.71$; $p < 0.0001$).

8. Conclusions

In this paper, we evaluated the performance of three popular methods for the estimation of cerebral atrophy using gold standards. To create the gold standards, we have proposed a topology preserving scheme to simulate atrophy using a B-spline based deformation model. Additional constraints were introduced to ensure that the skull remains invariant in the simulated image. The proposed framework for atrophy simulation can efficiently generate a deformation field that fits a given Jacobian map, while taking into account the

invariance constraints. The ability of the method to simulate uniform and non-uniform atrophies accurately was demonstrated using various examples.

We also assessed the performance of three freely available algorithms (SIENA, SIENAX and BSI-UCD) by simulating atrophy on BrainWeb images. Our analysis consisted of two steps: the simulation of atrophies on a single BrainWeb image to examine the robustness of the three methods to bias field inhomogeneity, noise, geometrical distortions and interpolation artefacts; statistical analysis of the results obtained on 18 different anatomical models of the brain. From the various tests that were performed, we draw the following conclusions:

- Experiments in the presence of bias-field inhomogeneity and noise
  - Our experiments related to a single BrainWeb image on which a number of atrophies in the range $0-1\%$ were simulated, showed that the mean error in the estimated PBVC for SIENA was $0.06\%\pm 0.04$ and $0.35\%\pm 0.38$ for noise-free and images degraded with bias-field inhomogeneity and noise, respectively. The errors were much higher for SIENAX and BSI-UCD.
  - Complementary experiments on 18 different BrainWeb images, where uniform atrophy was simulated, indicated that, in the presence of bias-field inhomogeneity and noise, a mean error of $0.64\%\pm 0.53$ may be expected in the atrophy estimated by SIENA. This is contrastingly high as compared to the results for the noise-free case for SIENA ($0.09\%\pm 0.07$). The errors obtained with SIENAX and BSI-UCD were considerably high as compared to SIENA. These errors were also larger than the overall errors (mean absolute differences) obtained by Camara *et al.* [13] for SIENA and BSI ($\sim 0.2\%$).

  Both sets of experiments showed that, SIENA is the best performer with respect to the error in the estimated PBVC in the noise-free case as well as when the images are degraded with bias field inhomogeneity and noise. The errors that we observed here are comparable to the whole brain annual atrophy rates ($0.5-2.8\%$) that have been reported for various pathologies. This highlights the need for more sensitive methods.

  We also concluded that bias field inhomogeneity and noise were responsible for incorrect brain extraction that considerably affected the

accuracy of all the methods. A good bias field correction algorithm is essential in order to improve the performance of the methods. The two sets of tests that we performed also indicated that SIENA and BSI-UCD are capable of estimating the longitudinal atrophy more accurately than SIENAX in a real scenario, where the images are corrupted with bias field inhomogeneity and noise. We would like to remind the reader that SIENAX has been developed for performing cross-sectional studies. Hence, one has to be careful when interpreting the longitudinal atrophy estimated through SIENAX.

- Geometrical distortions lead to mean absolute errors of around 0.07% in SIENA, 0.82% in BSI-UCD and 1.68% in SIENAX.

- Interpolation artefacts did not have a noticeable impact on the results of SIENA and SIENAX as compared to the noise-free case.

- The presence of lesions on atrophy estimation by SIENA, SIENAX and BSI-UCD was also evaluated. A maximum error of 0.2%, 0.45% and 0.46% was observed with SIENA, SIENAX and BSI-UCD, respectively, when comparing image of a normal brain with the same brain with lesions.

The bottlenecks for SIENA are registration of the two given examinations and segmentation of the boundary voxels (that is affected by the accuracy of the brain extraction). Since SIENA measures atrophy by measuring the displacement of the brain surface edge points after registering the two brains, it needs an accurate registration algorithm. Accurate segmentation would better localize the brain edges and thereby improve the accuracy of SIENA. The critical steps for BSI-UCD are the registration of the two brain examinations and the manual extraction of a gray-white matter mask to determine the boundaries of the brain on which the volume change is calculated. For SIENAX, the bottleneck is the brain extraction stage. Since, SIENAX estimates brain volume directly by counting the number of brain voxels, a better brain extraction method would improve the accuracy of SIENAX (see for instance, [38]). In our opinion, brain atrophy estimation is still an open issue and accurate algorithms are needed to measure the small atrophy that occurs in neuro-degenerative diseases.

9. Acknowledgments

We are thankful to Alsace Region and ARSEP for supporting this study. We also thank Dr. Evan Fletcher (University of California, Davis) for his support on the BSI-UCD software.

References

[1] J. H. Simon, Brain atrophy in multiple sclerosis: What we know and would like to know, Multiple sclerosis (Houndmills, Basingstoke, England) 12 (6) (2006) 679–687.

[2] D. Chan, N. C. Fox, R. Jenkins, R. I. Scahill, W. R. Crum, M. N. Rossor, Rates of global and regional cerebral atrophy in AD and frontotemporal dementia, Neurology 57 (2001) 1756–1763.

[3] J. D. Sluimer, H. Vrenken, M. A. Blankenstein, N. C. Fox, F. Scheltens, P. Barkhof, W. M. van der Flier, Whole-brain atrophy rate in Alzheimer's disease: identifying fast progressors, Neurology 70 (19) (2008) 1836–1841.

[4] V. M. Anderson, N. C. Fox, D. H. Miller, MRI measures of brain atrophy in multiple sclerosis, Journal of Magnetic Resonance Imaging 23 (2006) 605–618.

[5] D. Pelletier, K. Garrison, R. Henry, Measurement of whole-brain atrophy in multiple sclerosis, Journal of NeuroImaging 14 (3) (2004) 11S–19S.

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

[7] S. M. Smith, N. De Stefano, M. Jenkinson, P. M. Matthews, Normalized accurate measurement of longitudinal brain change, Journal of Computer Assisted Tomography 25 (3) (2001) 466–475.

[8] S. M. Smith, Y. Zhang, M. Jenkinson, J. Chen, P. M. Matthews, A. Federico, N. De Stefano, Accurate, robust and automated longitudinal and cross-sectional brain change analysis, NeuroImage 17 (1) (2002) 479–489.

[9] J. Ashburner, K. Friston, Voxel-based morphometry: the methods, NeuroImage 11 (2000) 805–821.

[10] C. R. Guttmann, S. K. Benson, R.and Warfield, X. Wei, M. Anderson, C. Hall, K. Abu-Hasaballah, J. Mugler, L. Wolfson, White matter abnormalities in mobility-impaired older persons, Neurology 54 (6) (2000) 1277–1283.

[11] R. G. Boyes, D. Rueckert, P. Aljabar, J. Whitwell, J. M. Schott, D. L. G. Hill, N. C. Fox, Cerebral atrophy measurements using jacobian integration: Comparison with the boundary shift integral, NeuroImage 32 (2006) 159–169.

[12] O. Camara, M. Schweiger, R. Scahill, W. Crum, B. Sneller, J. Schnabel, G. Ridgway, D. Cash, D. Hill, N. Fox, Phenomenological model of diffuse global and regional atrophy using finite-element methods, IEEE Transactions on Medical Imaging 25 (11) (2006) 1417–1430.

[13] O. Camara, , J. A. Schnabel, G. R. Ridgway, W. R. Crum, A. Douiri, R. I. Scahill, D. L. G. Hill, N. C. Fox, Accuracy assessment of global and local atrophy measurement techniques with realistic simulated longitudinal Alzheimer's disease images, NeuroImage 42 (2) (2008) 696–709.

[14] M. A. Horsfield, M. Rovaris, M. A. Rocca, P. Rossi, R. H. Benedict, M. Filippi, R. Bakshi, Whole-brain atrophy in multiple sclerosis measured by two segmentation processes from various MRI sequences, Journal of Neurological Sciences 216 (1) (2003) 169–177.

[15] J. Sharma, M. P. Sanfilipo, R. H. Benedict, B. Weinstock-Guttman, F. E. r. Munschauer, R. Bakshi, Whole-brain atrophy in multiple sclerosis measured by automated versus semiautomated MR imaging segmentation, American Journal of Neuroradiology 25 (6) (2004) 985–996.

[16] D. L. Collins, J. Montagnat, A. P. Zijdenbos, A. C. Evans, D. L. Arnold, Automated estimation of brain volume in multiple sclerosis with BICCR, in: Lecture Notes In Computer Science, Vol. 2082, 2001, pp. 141–147.

[17] M. Holden, Registration of 3D serial MR brain images, Ph.D. thesis, University of London (2001).

[18] B. Karacali, C. Davatzikos, Simulation of tissue atrophy using a topology preserving transformation model, IEEE Transactions on Medical Imaging 25 (5) (2006) 649–652.

[19] P. Pieperhoff, M. Sudmeyer, L. Homke, K. Zilles, A. Schnitzler, K. Amunts, Detection of structural changes of the human brain in longitudinally acquired MR images by deformation field morphometry: Methodological analysis, validation and application, NeuroImage 43 (2) (2008) 269–287.

[20] C. A. Cocosco, V. Kollokian, R. K.-S. Kwan, A. C. Evans, Brainweb: Online interface to a 3D MRI simulated brain database, NeuroImage 5 (4) (1997) S425.

[21] J. L. Gunter, M. M. Shiung, A. Manduca, C. R. Jack, Methodological considerations for measuring rates of brain atrophy, Journal of Magnetic Resonance Imaging 18 (2003) 16–24.

[22] S. M. Paling, E. D. Williams, R. Barber, E. J. Burton, W. C. Crum, N. Fox, J. T. O'Brien, The application of serial MRI analysis techniques to the study of cerebral atrophy in late-onset dementia, Medical Image Analysis 8 (2003) 69–79.

[23] S. M. Smith, A. Rao, N. De Stefano, M. Jenkinson, J. M. Schott, P. M. Matthews, N. C. Fox, Longitudinal and cross-sectional analysis of atrophy in Alzheimer's disease: Cross validation of BSI, SIENA and SIENAX, NeuroImage 36 (2007) 1200–1206.

[24] A. D. Leow, I. Yanovsky, A. D. Chiang M. C.and Lee, A. D. Klunder, A. Lu, J. T. Becker, S. W. Davis, A. W. Toga, P. M. Thompson, Statistical properties of jacobian maps and the realization of unbiased large-deformation nonlinear image registration, IEEE Transactions on Medical Imaging 26 (6) (2007) 822–832.

[25] V. Noblet, C. Heinrich, F. Heitz, J.-P. Armspach, 3-D deformable image registration: a topology preservation scheme based on hierarchical deformation models and interval analysis optimization, IEEE Transactions on Image Processing 14 (5) (2005) 553–566.

[26] V. Noblet, C. Heinrich, F. Heitz, J.-P. Armspach, Retrospective evaluation of a topology preserving non-rigid registration method, Medical Image Analysis 10 (3) (2006) 366–384.

[27] W. Press, S. Teukolsky, W. Vetterling, B. Flannery, Numerical recipes in C. The art of scientific computing, 2nd Edition, Cambridge University Press, 1992.

[28] V. Noblet, C. Heinrich, F. Heitz, J.-P. Armspach, Accurate inversion of 3-D transformation fields, IEEE Transactions on Image Processing 17 (10) (2008) 1963–1968.

[29] H. Gudbjartsson, S. Patz, The rician distribution of noisy MRI data, Magnetic Resonance in Medicine 34 (6) (1995) 910–914.

[30] R. P. Woods, S. T. Grafton, C. J. Holmes, S. R. Cherry, J. C. Mazziotta, Automated image registration: I. general methods and intrasubject, intramodality validation, Journal of Computer Assisted Tomography 22 (1998) 139–152.

[31] Z. Caramanos, V. S. Fonov, S. J. Francis, S. Narayanan, D. L. Collins, D. L. Arnold, Effects of z-shift-associated gradient-distortions on siena-generated measures of brain atrophy, in: MICCAI workshop on Medical Image Analysis on Multiple Sclerosis (MIAMS08), 2008, pp. 13–24.

[32] S. M. Smith, Fast robust automated brain extraction, Human Brain Mapping 17 (3) (2002) 143–155.

[33] Y. Zhang, M. Brady, S. M. Smith, Segmentation of brain MR images through a hidden markov random field model and the expectation maximization algorithm, IEEE Transactions on Medical Imaging 20 (1) (2001) 45–57.

[34] S. M. Smith, M. Jenkinson, C. F. Woolrich, M. W.and Beckman, T. E. J. Behrens, H. Johansen-Berg, P. R. Bannister, et. al, Advances in functional and structural MR image analysis and implementation as FSL, NeuroImage 23 (S1) (2004) 208–219.

[35] M. Jenkinson, S. S. M., A global optimisation method for robust affine registration of brain images., Medical Image Analysis 36 (2001) 143–156.

[36] P. M. Thompson, K. M. Hayashi, G. De Zubicaray, A. L. Janke, S. E. Rose, J. Semple, D. Herman, M. S. Hong, S. S. Dittmer, D. M. Doddrell, A. W. Toga, Dynamics of gray matter loss in Alzheimer's disease, The Journal of Neuroscience 23 (2003) 994–1005.

[37] B. Aubert-Broche, M. Griffin, G. B. Pike, A. C. Evans, D. L. Collins, 20 new digital brain phantoms for creation of validation image data bases, IEEE Transactions on Medical Imaging 25 (11) (2006) 1410–1416.

[38] M. Battaglini, S. M. Smith, S. Brogi, N. De Stephano, Enhanced brain extraction improves the accuracy of brain atrophy estimation, NeuroImage 40 (2008) 583–589.

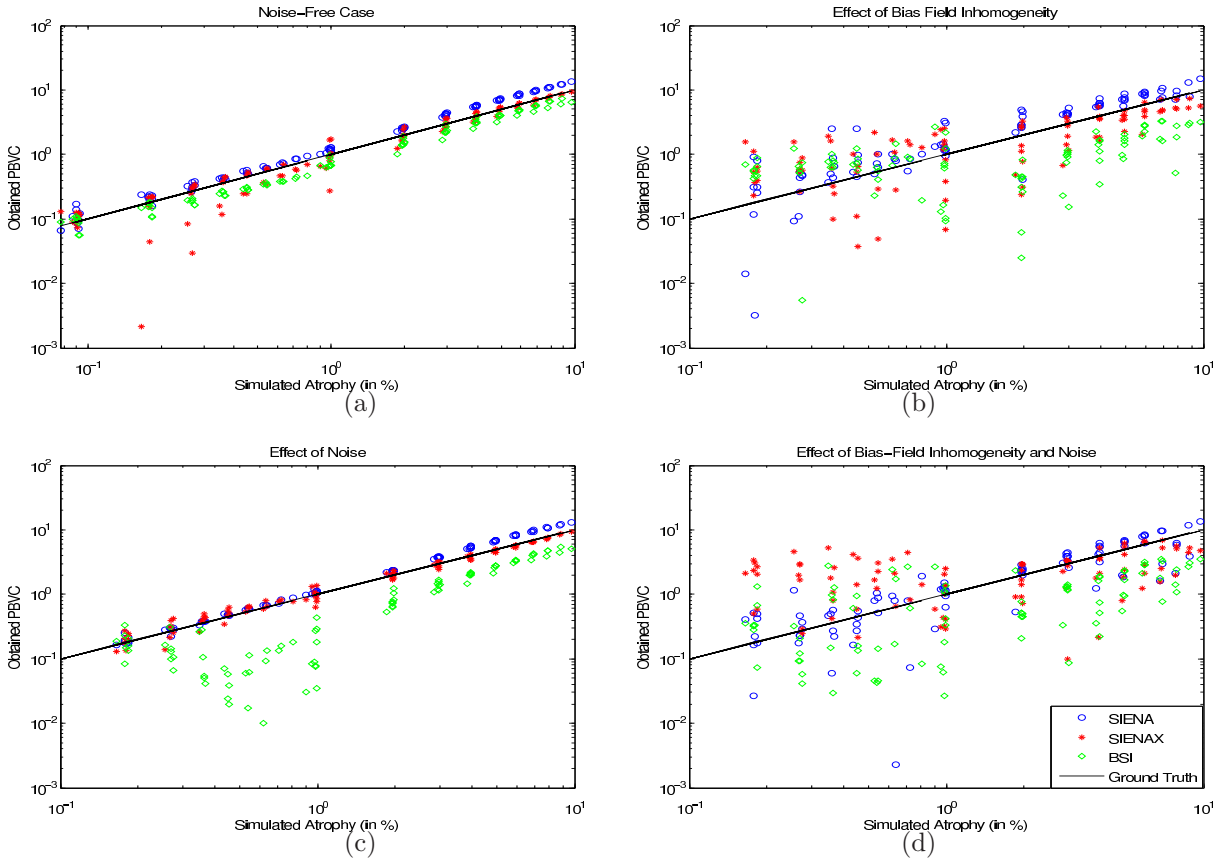


Figure 9: PBVC comparison of SIENA, SIENAX and BSI-UCD with the ground truth for (a) Noise-free images and those degraded by (b) Bias field inhomogeneity only (c) Noise only and (d) Bias field inhomogeneity and Noise. (Refer to section 6.1.1.)

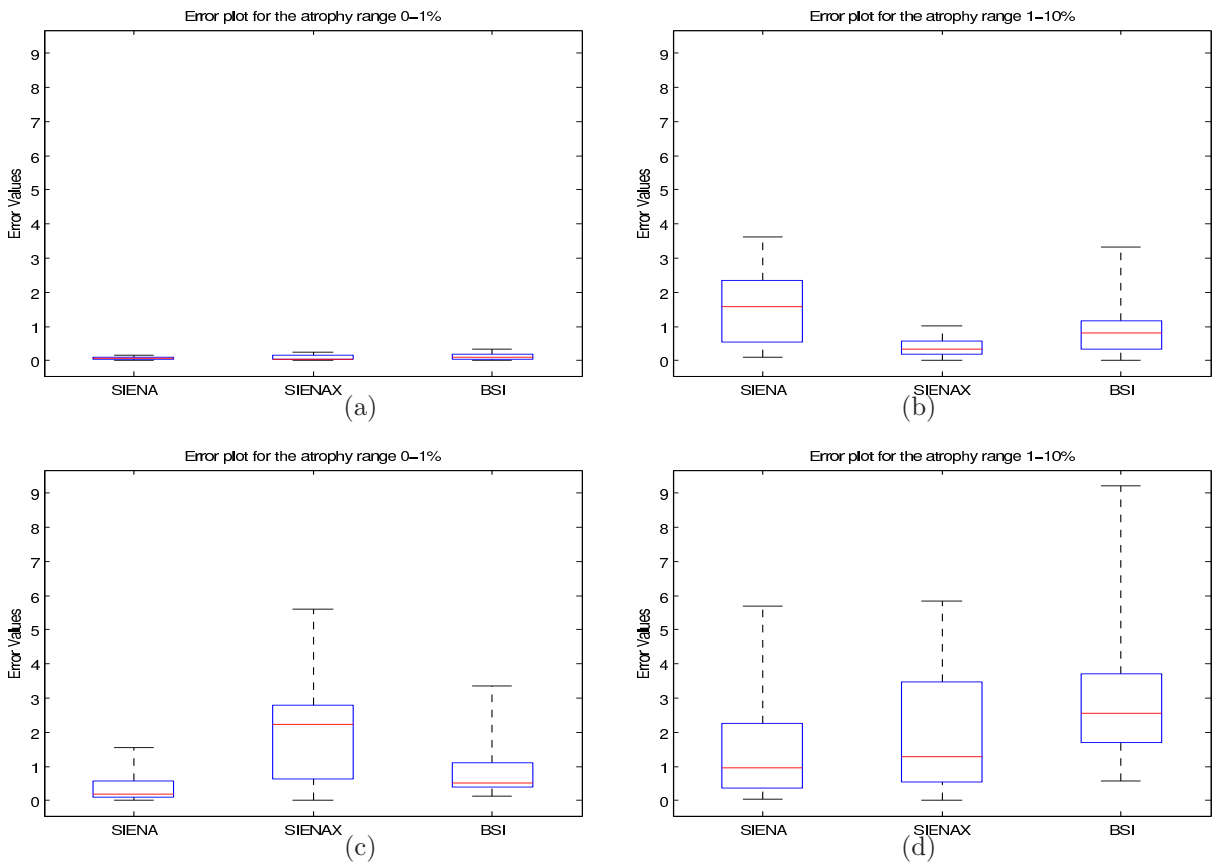


Figure 10: Boxplot showing absolute errors in the estimated PBVC by SIENA, SIENAX and BSI-UCD with respect to the ground truth for (a-b) Noise-free images and (c-d) Those degraded by Bias field inhomogeneity and Noise. The ends of the boxes show the lower quartile and upper quartile values. The red line denotes the median. The lines extending from each end of the boxes show the maximum and minimum values of the data.

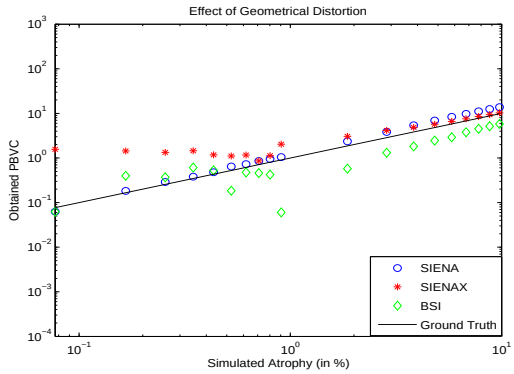


Figure 11: PBVC comparison for SIENA, SIENAX and BSI-UCD, respectively with the ground truth when the geometrically distorted BrainWeb image is compared with simulated images for different atrophy rates.

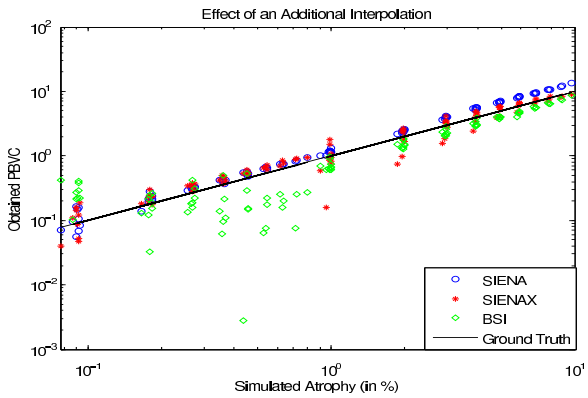


Figure 12: PBVC comparison of SIENA, SIENAX and BSI-UCD, respectively, with the ground truth when images are rotated by 1° to add an extra interpolation (B-Spline 5th order) step.

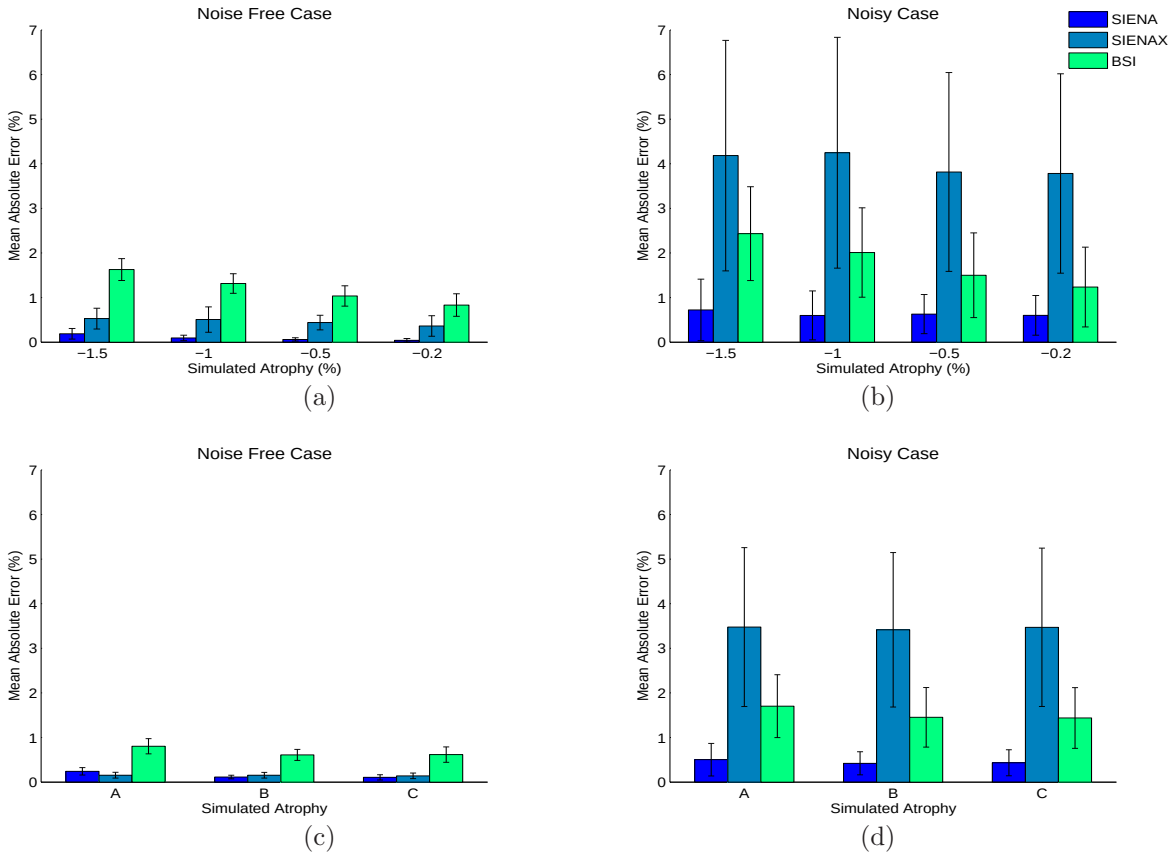
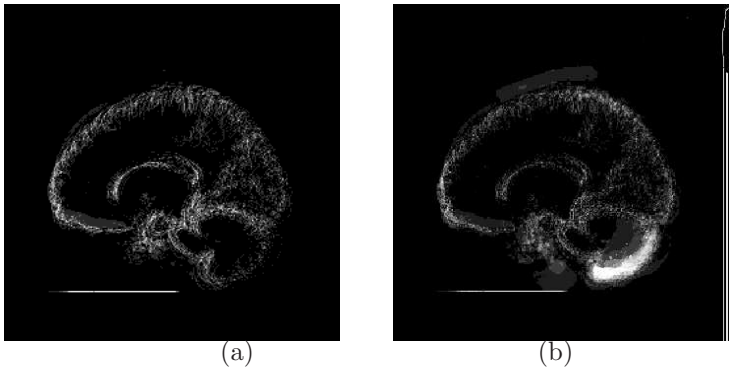


Figure 13: Bar plot showing the mean absolute error and its standard deviation in the estimated PBVC (in percentage) with respect to the ground truth for SIENA, SIENAX and BSI-UCD for the Uniform atrophy case using (a) Noise-free images (b) With images degraded with bias field inhomogeneity and noise and for the Non-uniform atrophy case using (c) Noise-free images (d) With images degraded with bias field inhomogeneity and noise. A (-0.25%), B (-0.3%) and C (-0.57%) are different patterns of simulated non-uniform atrophy.



(a) (b)

Figure 14: Mean error in the segmentation of brain for (a) the Noise-free case. (b) when using Images degraded by bias field inhomogeneity and noise for the 18 baseline images. The gray levels (from black to white) depict the mean error values in the range of $0.2282 - 0.6702$ and $1.5986 - 6.4160$ in images shown in (a) and (b) respectively.

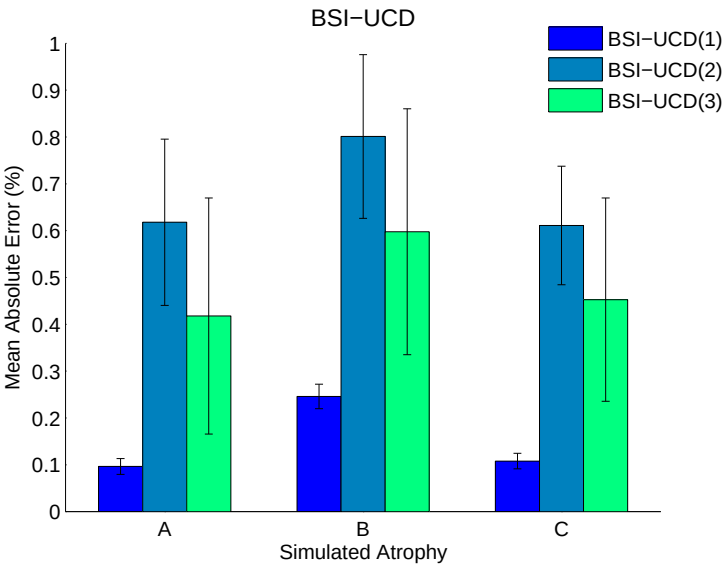


Figure 15: Bar plot showing the mean absolute error and the standard deviation in the estimated PBVC (in percentage) when the BSI-UCD algorithm is run with no pre-processing (BSI-UCD(1)), normal BSI-UCD algorithm (BSI-UCD(2)) and by replacing the registration algorithm of BSI-UCD with an in-house affine registration algorithm (BSI-UCD(3)).