

École Doctorale Mathématiques, Sciences de
l'Information et de l'Ingénieur

UdS — INSA — ENGEES

THÈSE

présentée pour obtenir le grade de

Docteur de l'Université de Strasbourg

Discipline: Sciences

Spécialité: Traitement du Signal et des Images

par

Swati Sharma

Estimation de l'atrophie cérébrale en imagerie par résonance magnétique : Application à la sclérose en plaques

English title: Brain atrophy estimation using magnetic resonance
imaging: Application to multiple sclerosis

Soutenue publiquement le 15 septembre 2011

Membres du jury

- Co-Directeur de thèse: **M. Jean-Paul Armspach**
LINC-IPB, Strasbourg.
- Co-Directeur de thèse: **M. Fabrice Heitz**
LSIIT, Illkirch.
- Rapporteur: **M. Grégoire Malandain**
INRIA, Sophia-Antipolis.
- Rapporteur: **M. Colin Studholme**
University of Washington, Seattle.
- Examineur: **M. Lucien Rumbach**
CHU, Besançon.
- Examineur: **M. François Rousseau**
LSIIT, Illkirch.

Acknowledgments

I would like to express my gratitude towards Dr. Grégoire Malandain and Prof. Colin Studholme for examining this thesis.

I am grateful for the financial support provided by Alsace region and ARSEP (December 2007-November 2010) and LFSEP (December 2010-November 2011).

Thanks are due to my advisers Dr. Jean-Paul Armspach and Prof. Fabrice Heitz for allowing me to pursue an interesting PhD problem, their guidance and constructive comments on my work throughout.

My heartfelt thanks to Dr. François Rousseau, who has also supervised this thesis. I am thankful for his invaluable scientific inputs, his availability, his enthusiasm and for the patience with which he went through (and is still going through!) the numerous manuscripts that I have written.

I have also greatly benefited from scientific discussions that I have had with Vincent Noblet and Benjamin Perret.

I deeply appreciate the efforts of Abdelkrim Belhaoua, Benjamin Perret and my mother, Nina, in proof reading parts of this thesis.

Many thanks to all members of MIV-LSIIT and IPB-LINC teams, particularly my fellow PhD students, Abdelkrim, Akram, Alban, Alice, Antoine, Benjamin, Benoit, Fatma, Félix, Hervé, Julien, Giorgos and Slim, with whom I have had many scientific as well as non-scientific discussions, who have helped me learn French in more ways than one, and who have contributed to an ambiance conducive to performing fruitful research.

The list of acknowledgements would not be complete without appreciating the encouragement of my entire family, especially, that of my husband, Fazal, without whose support and cooperation this thesis would not have been possible.

Contents

List of Abbreviations	v
Résumé de thèse	1
1 General Introduction	13
2 Background	17
2.1 The brain	17
2.1.1 White matter	18
2.1.2 Grey matter	19
2.1.3 Cerebrospinal fluid	19
2.2 Multiple sclerosis	20
2.2.1 Causes	20
2.2.2 Symptoms	21
2.2.3 Multiple sclerosis subtypes	21
2.2.4 Diagnosis and treatment	22
2.3 Brain atrophy	23
2.3.1 Brain atrophy in MS	25
2.3.2 Types of brain atrophy	25
2.3.3 Treatment of brain atrophy	25
2.3.4 Brain atrophy as a biomarker	26
2.4 Magnetic resonance imaging	26
2.4.1 MRI artefacts	28
2.4.2 Correction of artefacts	30
2.5 Brain atrophy estimation with MRI	38
2.5.1 Classification of existing atrophy estimation approaches	39
2.5.2 Atrophy estimation approaches	42
2.6 Brain extraction	53
2.7 Key issues in brain atrophy estimation	56
3 Evaluation of Existing Brain Atrophy Estimation Algorithms	58
3.1 Studies on evaluation of brain atrophy estimation algorithms	58
3.1.1 Evaluation criteria	58
3.1.2 Existing studies on validation of atrophy estimation methods	60
3.2 Evaluation framework	66
3.2.1 Simulation of atrophy: ground truth generation	67
3.2.2 Simulation of artefacts	70
3.3 Results	72
3.3.1 Results on simulation of atrophy and discussion	73

3.3.2	Evaluation results of SIENA, SIENAX, BSI-UCD, JI and discussion	77
3.4	Comparison with other studies	91
3.5	Findings	93
4	Uncertainty Estimation for Longitudinal Brain Atrophy Measurement Algorithms with a Ground Truth	96
4.1	Uncertainty	97
4.2	Uncertainty estimation in image registration	97
4.2.1	Using a ground truth	97
4.2.2	Manipulating the similarity criterion or deformation field	99
4.2.3	Bayesian registration	102
4.2.4	Key issues in uncertainty estimation	107
4.3	Bias in estimated atrophy	108
4.4	Proposed framework for uncertainty estimation	112
4.4.1	The learning database	112
4.4.2	Confidence interval calculation	114
4.5	Results	116
4.5.1	Learning database construction	117
4.5.2	Gaussian assumption	117
4.5.3	Bias in atrophy estimations	119
4.5.4	Quantitative evaluation	119
4.6	Discussion and findings	124
5	Atrophy Estimation using Bayesian Non-Linear Registration	127
5.1	Modelling	127
5.1.1	Structural modelling	128
5.1.2	Direct relations between variables in DAG	132
5.1.3	Probability modelling	133
5.2	Maximum a posteriori estimation	135
5.2.1	Sampling schedule	135
5.3	Results	137
5.3.1	Bias field correction	138
5.3.2	Atrophy estimation	143
5.3.3	Simultaneous estimation of bias field inhomogeneity and noise	148
5.3.4	Consistency test	153
5.4	Discussion and findings	153
6	General conclusion and future perspectives	159
A	Appendix I	164
A.1	Legendre polynomials	164
A.1.1	Orthogonality property	165

B Appendix II	166
B.1 Experimental set-up	166
B.1.1 Image database	166
B.1.2 Bias field inhomogeneity correction	167
B.1.3 Simulation of atrophy	169
B.1.4 Brain atrophy estimation methods	169
B.1.5 Brain Segmentation	171
C Appendix III	174
C.1 MCMC algorithms	174
C.1.1 Metropolis-Hastings algorithm	174
C.1.2 Gibbs sampler	175
Publications	176
Bibliography	177

List of Abbreviations

AD	Alzheimer's Disease
ADNI	Alzheimer's Disease Neuroimaging Initiative
ANTS	Advanced Normalisation Tools
BET	Brain Extraction Tool
BICCR	Brain to IntraCranial Capacity Ratio
BPF	Brain Parenchymal Fraction
BSI	Boundary Shift Integral
CLADA	Cortical Longitudinal Atrophy Detection Algorithm
DBM	Deformation Based Morphometry
DFM	Deformation Field Morphometry
EM	Expectation Maximisation
FEM	Finite Element Model
FLAIR	Fluid Attenuated Inversion Recovery
FRAE	Fast Registration Accuracy Estimation
FSL	FMRIB Software Library
IBSR	Internet Brain Segmentation Repository
INU	Intensity Non-uniformity
JI	Jacobian Integration
LMMSE	Linear Minimum Mean Squared Error
MCMC	Markov Chain Monte Carlo
ML	Maximum Likelihood
MRI	Magnetic Resonance Imaging
MS	Multiple Sclerosis
NLM	Non Local Means
NRR	Non-rigid Registration

PDF Probability Distribution Function

PPMS Primary Progressive Multiple Sclerosis

PRMS Progressive Relapsing Multiple Sclerosis

RF Radio frequency

ROI Region of Interest

RRMS Relapsing/Remitting Multiple Sclerosis

SIENA Structural Image Evaluation, using Normalisation, of Atrophy

SPM Statistical Parametric Mapping

SPMS Secondary Progressive Multiple Sclerosis

TBM Tensor Based Morphometry

VBM Voxel Based Morphometry

Résumé de thèse

1 Introduction

Plusieurs maladies affectant le système nerveux central, dont la sclérose en plaques (SEP) [136], maladie d'Alzheimer [25, 138] et la schizophrénie [17], sont caractérisées par une diminution du volume cérébral. Ce processus est connu sous le nom d'atrophie cérébrale. C'est un terme général qui représente la diminution de la taille du cerveau due à une perte de neurones et de connexions entre eux. Des études cliniques suggèrent que la perte de matière cérébrale commence tôt au cours de maladies telles que la SEP. La mesure de l'atrophie du cerveau a récemment émergé comme un biomarqueur visible de progression de la maladie. Par conséquent, la communauté s'est fortement intéressée au problème de l'estimation du changement volumétrique du cerveau. Les outils automatisés, capables de détecter des changements dans le cerveau avec précision, sont devenus une nécessité car ils peuvent faciliter le diagnostic et le pronostic de ces maladies.

En fournissant un bon contraste entre les différents tissus, l'imagerie par résonance magnétique (IRM) est une technique d'imagerie non-invasive prédominante pour étudier les changements dans le cerveau. Par conséquent, le développement de programmes de traitement d'images permettant la manipulation des images d'IRM, offrent des moyens de mesurer la perte de volume du cerveau au cours d'une maladie.

Le but de cette thèse est d'étudier les méthodes de traitement d'image permettant l'estimation de l'atrophie cérébrale chez les patients atteints de la SEP. L'analyse d'images IRM acquises à des temps différents permet d'obtenir une estimation longitudinale de l'atrophie cérébrale.

Trois grandes parties définissent les travaux réalisés dans cette thèse. Tout d'abord, sur la base de leur robustesse à diverses sources d'erreur (inhomogénéité en intensité, bruit, distorsions géométriques, artefacts d'interpolation et présence de lésions), une évaluation des principales approches d'estimation de l'atrophie cérébrale est réalisée à l'aide de simulations d'une vérité terrain. Ensuite, une analyse statistique est effectuée afin d'estimer les incertitudes associées à l'atrophie mesurée. Fondé sur l'utilisation d'une base d'apprentissage, un cadre générique est proposé pour construire des intervalles de confiance de l'atrophie estimée dans une région d'intérêt du cerveau. Enfin, un cadre bayésien est proposé pour l'estimation conjointe d'une transformation non-rigide (à partir de laquelle l'atrophie est calculée) et du champ d'inhomogénéité en intensité présent dans les images. Cette approche bayésienne couplée aux techniques MCMC (pour Markov Chain Monte Carlo) four-

nit un cadre rigoureux pour la construction d'intervalles de confiance de l'atrophie cérébrale.

Ce résumé est organisé en cinq sections. Des méthodes existantes largement utilisées pour l'estimation de l'atrophie cérébrale sont décrites dans la section qui est suivie par une brève description du cadre établi pour l'évaluation des méthodes de l'estimation de l'atrophie du cerveau et de l'estimation des mesures d'incertitude dans les sections et . Enfin, l'approche bayésienne consacrée à l'estimation simultanée de l'atrophie cérébrale et du champ d'inhomogénéité en intensité est décrite dans la section . Les conclusions de cette thèse sont présentées dans la section .

2 État de l'art

Les méthodes actuellement disponibles pour l'estimation de l'atrophie du cerveau se différencient en fonction de critères tels que leur dépendance au recalage ou à la segmentation [106], le type de mesure (globale ou locale) de l'atrophie.

Les principales méthodes fondées sur une utilisation du recalage sont: BSI (pour "Boundary Shift Integral") [44], SIENA (pour "Structural Image Evaluation using Normalisation, of Atrophy") [140, 143], VBM (pour "Voxel based morphometry") [5]), et enfin, celles qui utilisent une analyse locale du jacobien du champ de déformation [18, 21, 20, 171, 170, 59]. Les approches reposant sur une étape de segmentation s'appuient sur différentes mesures d'atrophie telles que la fraction du parenchyme du cerveau (BPF pour Brain Parenchym fraction) [57, 131] et BICCR (pour "Brain to intracranial capacity ratio") [31].

Lorsque on utilise la mesure de l'atrophie du cerveau comme un marqueur de l'évolution de maladies, deux types d'analyses peuvent être effectuées: longitudinales ou transversales. Dans les études longitudinales, plusieurs images IRM du cerveau d'un même individu, acquises à des temps différents, sont considérées. Ces études suivent la progression de la maladie en donnant une estimation de l'atrophie cérébrale. On peut citer comme exemple de méthodes longitudinales BSI [44] et SIENA [140]. Les études transversales sont utilisées pour connaître l'état de la maladie en fournissant une estimation du volume du cerveau à un temps donné. SIENAX [143] est une méthode couramment utilisée pour ce type d'études.

Une méthode globale d'estimation atrophie, telle que BSI [44] ou SIENA [140], donne une estimation globale du taux de diminution du volume cérébral. Dans la SEP, les mesures de l'atrophie du cerveau entier sont parfois utilisées comme biomarqueurs de la progression de la maladie [132]. Les mesures locales (TBM (pour Tensor based morphometry) [18, 21, 20, 171, 59], VBM [5]) fournissent une estimation locale (voxel par voxel) de l'atrophie du cerveau. Ces mesures locales sont intéressantes car elles nous permettent de déchiffrer quelles parties du cerveau

sont les plus touchées par l'atrophie dans le cadre d'une pathologie.

Penchons-nous sur quelques méthodes populaires d'estimation de l'atrophie cérébrale.

SIENA Cet algorithme commence par l'extraction du cerveau et du crâne des images en utilisant l'outil BET (pour "Brain extraction tool" du logiciel FSL) [139]. Les 2 images sont ensuite recalées en utilisant une transformation affine corrigée par un facteur d'échelle prenant en compte l'invariance du crâne. La transformation affine ainsi obtenue, est d'abord décomposée en deux transformations "à mi-chemin" qui sont ensuite appliquées aux deux images ré-échantillonnées dans un espace commun. Cela évite des flous asymétriques dans les images dus à l'interpolation. Ensuite, une segmentation est effectuée afin de trouver les bords du cerveau [175]. À cette étape, une correction d'inhomogénéité en intensité est également appliquée. Enfin, un tableau 1D est rempli avec des valeurs d'intensité du profil perpendiculaire au bord du cerveau de l'image de référence. En utilisant une interpolation tri-linéaire, ces valeurs sont échantillonnées avec une précision sous "voxelique". Le nombre d'éléments de ce tableau est prédéterminé (± 3 mm du bord), mais il peut aussi changer de région en région selon la présence d'un autre bord. Un autre tableau 1D est également construit avec des valeurs d'intensité de la deuxième image IRM. Le mouvement du bord du cerveau est désormais estimé en calculant le décalage relatif, entre les tableaux, qui produit la corrélation maximale.

SIENA est une méthode entièrement automatisée qui rend possible l'analyse longitudinale. Smith *et coll.* [141] rapportent une précision de 0,15% à l'aide d'une évaluation de type "scan-rescan".

BSI Un autre outil populaire pour estimer l'atrophie cérébrale à partir de deux images longitudinales est "Boundary Shift Integral" (BSI) [44]. Dans cette approche, les deux images sont recalées en utilisant une transformation affine. Dans la mise en œuvre originale de BSI, une procédure de recalage affine permet de déterminer les facteurs d'échelle spatiale en utilisant la surface du crâne comme invariant. Les images sont ensuite normalisées en intensité et segmentées en utilisant une approche morphologique itérative, où la taille des opérateurs est sélectionnée manuellement. L'idée est de définir une région d'intérêt sur laquelle est calculée l'indice BSI.

BSI est un outil semi-automatique pour estimer l'atrophie longitudinale. Freeborough *et coll.* [44] constatent que les indices BSI pour le cerveau et le LCR sont fortement corrélés ($r = 1,000$ et $r = 0,999$) avec les changements volumétriques simulés.

SIENAX [143] est un algorithme de segmentation fournissant une estimation transversale de l'atrophie à partir d'une seule image IRM. Contrairement à SIENA

qui indique le taux progression de la maladie, cet outil donne une estimation "courante" du volume cérébral. Le cerveau et le crâne sont tout d'abord extraits de l'image. Ensuite, l'image MNI 152 est recalée de façon affine sur l'image du patient en utilisant le crâne comme invariant pour contraindre le facteur d'échelle. Enfin une segmentation [175] est effectuée sur l'image originale. Celle-ci prend en compte l'estimation du volume partiel, donnant une précision supérieure à une segmentation volumétrique dure. Le volume total du cerveau est obtenu à partir de ce résultat de segmentation. Pour normaliser la taille de la tête, le volume du cerveau est multiplié par le facteur d'échelle calculé avec l'image MNI 152. Smith *et coll.* [141] affirment obtenir une précision de 0,5 – 1% du volume du cerveau avec SIENAX.

Voxel (VBM) or tensor (TBM) based morphometry VBM [5] est dédié à l'analyse des différences locales dans le cerveau en utilisant une approche statistique. La procédure consiste à normaliser les images de tous les sujets de l'étude dans le même espace, suivie par une segmentation de la matière grise à partir des images spatialement normalisées suivi par une étape de lissage de la matière grise. Le modèle linéaire général [89] fournit un cadre pour l'application de tests statistiques afin d'identifier les régions de matière grise qui sont significativement liées à des paramètres de l'étude tels que l'âge ou des critères pathologiques. Les images sont analysées voxel par voxel avec des tests statistiques (des tests non paramétriques peuvent également être appliqués comme dans [122]).

Enfin, les méthodes fondées sur une utilisation d'algorithmes de recalage non-rigide constituent une autre classe conséquente pour l'estimation de l'atrophie cérébrale. Des cartes du changement volumétrique peuvent être obtenues sous forme de cartes scalaires (jacobien du champ de déformation), celles-ci représentant alors les changements entre 2 images consécutives. L'analyse de ces cartes voxel par voxel ou dans une région d'intérêt est appelé TBM (pour Tensor-based morphometry). Si les statistiques sont effectuées sur le champ de déformation directement, alors cette approche est plus communément appelée "deformation based morphometry" (DBM) ou "deformation field morphometry" (DFM) [18, 45, 109].

3 Évaluation des méthodes d'estimation de l'atrophie du cerveau

La validation des algorithmes d'estimation de l'atrophie cérébrale est une tâche cruciale afin que ces mesures soit de plus en plus employées comme biomarqueurs dans la progression de maladies neuro-dégénératives. Néanmoins, il est difficile d'évaluer ces méthodes lorsque aucune vérité terrain n'est disponible. Des travaux sur l'évaluation de ces algorithmes comprennent [105, 18, 50, 142, 21, 20, 109, 34].

Dans cette section, un cadre pour évaluer trois méthodes de l'estimation de l'atrophie (BSI-UCD [44], SIENA [140, 143] SIENAX [143]) est décrit. La vérité

terrain est créée à l'aide d'un schéma de recalage non-rigide qui préserve la topologie et utilisant un modèle de déformation B-spline. Des contraintes supplémentaires ont été introduites pour assurer l'invariance du crâne dans l'image simulée. Cette approche permet de générer efficacement un champ de déformation correspondant à une carte scalaire de jacobien donnée, tout en tenant compte des contraintes d'invariance.

Ensuite, les influences de l'inhomogénéité en intensité, du bruit, des distorsions géométriques, des méthodes d'interpolation et de la charge lésionnelle sur les mesures d'atrophie du cerveau sont analysées. Dans une autre série d'expériences, la conséquence de la variabilité anatomique sur l'estimation de l'atrophie en simulant des atrophies sur une cohorte de 18 images Brainweb est examinée. Enfin, des expérimentations ont également été menées afin de comparer un algorithme de recalage déformable (Noblet *et coll.* [98]) avec SIENA, SIENAX et BSI-UCD.

Résultats et discussion

Différents degrés d'atrophie (0 – 1%) ont été simulés sur une image de la base Brainweb. Dans ce cas, l'erreur moyenne du changement de volume du cerveau en pourcentage (PBVC pour Percentage of Brain Volume Change) estimé pour SIENA était de $0,06 \pm 0,04\%$ pour des images non bruitées et $0,35 \pm 0,3\%$ pour des images dégradées par un bruit gaussien (avec un rapport signal sur bruit de 15dB rapport) et un champ d'inhomogénéité en intensité (20% INU). Les erreurs sont beaucoup plus élevées pour SIENAX et BSI-UCD.

Des expériences complémentaires sur les 18 Brainweb images ont indiqué qu'en présence du champ d'inhomogénéité en intensité et de bruit, une erreur moyenne de $0,64 \pm 0,53$ peut être attendue dans l'atrophie estimée par SIENA. Ceci contraste avec les résultats obtenus pour le cas sans bruit pour SIENA ($0,09 \pm 0,07\%$). Les erreurs obtenues avec SIENAX et BSI-UCD étaient considérablement plus élevées par rapport à SIENA. En revanche, il est apparu qu'une approche reposant sur l'utilisation d'algorithme de recalage non-rigide permet d'obtenir des erreurs plus faibles que SIENA.

Les différentes expériences (lorsque les images sont dégradées ou non par du bruit et un biais en intensité) ont montré que SIENA fournit les meilleurs résultats en terme de PBVC que SIENAX et BSI. Cependant, les expériences incluant JI ont montré que cette méthode semble encore plus performante que SIENA. Les erreurs d'estimation que nous avons pu observer sont comparables au taux annuel d'atrophie (0.5 – 2.8%) de différentes pathologies neuro-dégénératives. Ce résultat démontre qu'il est nécessaire de développer des méthodes d'estimation de l'atrophie cérébrale plus précises.

Nous avons également observé que les artefacts liés à l'inhomogénéité en in-

intensité et au bruit perturbent fortement l'étape d'extraction du cerveau, ce qui affecte directement les estimations de SIENA, SIENAX et BSI. L'algorithme de recalage non rigide utilisait comme critère de similarité la différence des intensités. Ainsi, cette approche (appelée JI pour Jacobian integration) s'est révélée sensible aux inhomogénéités d'intensité. Un algorithme de correction de biais est essentiel afin d'améliorer la performance de ces méthodes. L'ensemble des tests que nous avons effectués a aussi indiqué que SIENA, BSI-UCD et JI sont capables d'estimer l'atrophie longitudinale de façon plus précise que SIENAX dans un scénario réel, où les images sont corrompues par un champ d'inhomogénéité en intensité et du bruit. Nous tenons à rappeler que SIENAX a été développé originellement pour être appliqué à des études transversales. Par conséquent, il est nécessaire de rester prudent dans l'interprétation de l'atrophie longitudinale estimée par SIENAX.

Les distorsions géométriques conduisent à des erreurs absolues d'environ 0,07% pour SIENA, 0,82% pour BSI-UCD et 1,68% pour SIENAX. Les artefacts d'interpolations n'ont pas eu d'impact notable sur les résultats de SIENA et SIENAX par rapport au cas sans bruit.

4 Estimation de l'incertitude en utilisant une vérité terrain

L'analyse des propriétés statistiques des cartes de changements volumétriques du cerveau, telles que les cartes de jacobien, est une étape importante dans les études de type TBM ou DBM. Toutefois, la précision des cartes du changement volumétrique est discutable quand ils sont obtenus à partir d'images corrompues par du bruit ou des variations en intensité. En outre, des facteurs tels que les inexactitudes des procédures de recalage et de segmentation peuvent également entraîner des estimations biaisées, même en l'absence d'artefacts dans les images. Les différentes études de validation [142, 20, 109] dont la notre ont confirmé l'existence de ce type de biais dans les estimations de l'atrophie. Un autre point à prendre en compte concerne l'erreur de modélisation liée au modèle mathématique utilisé. Pour ces raisons, il devient difficile de séparer les changements anatomiques réels de ceux induits par ces artefacts. Il est donc important de rendre ces cartes de changement plus fiables en fournissant également une estimation des incertitudes.

Dans cette partie, l'objectif est de développer un cadre générique pour l'estimation des incertitudes des mesures longitudinales de l'atrophie du cerveau au moyen d'intervalles de confiance construits *pour toute méthode d'estimation d'atrophie*. Des exemples de simulations d'atrophie sur différents sujets nous permettent d'estimer ces incertitudes qui sont ensuite stockées pour des mesures futures dans une base d'apprentissage. Ainsi, pour un jeu d'images et une méthode d'estimation donnés, il est possible de déterminer l'atrophie sous-jacente en utilisant une mesure de distance (ici, un z-score) avec la base d'apprentissage. Cela nous permet d'utiliser les erreurs

correspondant à l'atrophie de la base sélectionnée pour construire les intervalles de confiance de notre estimation.

La nouveauté de cette approche réside dans sa capacité à intégrer à la fois des erreurs découlant des images et de la méthode d'estimation. Les atrophies simulées qui composent la base de données d'apprentissage peuvent être choisies sur la base de la maladie et de la région d'intérêt à l'étude. Au vue des performances de la méthode de recalage non-rigide (RNR) de Noblet *et coll.* [98] par rapport à SIENA, SIENAX et BSI (section), et de part sa flexibilité de manipulation (mesure d'atrophie dans n'importe quelle région d'intérêt), le reste de la thèse est consacrée à l'analyse des cartes de jacobien obtenues à partir de tels algorithmes. Trois approches de type RNR ont été évaluées [98, 10, 157] pour mesurer l'atrophie de l'hippocampe sur les images IRM réelles. Les images ont été dégradées par un bruit gaussien (SNR = 35 dB) et aucun pré-traitement n'a été appliqué.

Résultats et discussion

Différents types de biais d'estimation ont pu être observés pour les trois méthodes évaluées dans cette expérience. Pour l'algorithme de Noblet *et coll.*, le biais reste presque constant quelque soit la valeur de l'atrophie simulée (biais de l'ordre de $(-4, 22\%, -3, 38\%)$). D'autre part, le biais observé pour la méthode ANTS augmente avec la valeur de l'atrophie simulée $(0, 72\%, 5, 39\%)$. Enfin, le biais de la méthode de Vemuri *et coll.* est inversement proportionnel à l'atrophie simulée $(-8, 8608\%, 0, 4788\%)$. Un biais négatif indique une sur-estimation de l'atrophie et vice versa. Dans le cas de la méthode de recalage non-rigide développée par Noblet *et coll.*, le biais dans les estimations peut être dû à l'utilisation simultanée des contraintes de préservation de la topologie et de la régularisation du champ de déformation. Nous croyons que l'approche par ensemble de niveaux de Vemuri *et coll.* est sujette à des erreurs provoquées par le bruit dans les images car le champ de déformation est ajusté par une fonction liée à la différence des intensités dans les deux images. Pour cette raison, il est possible que la direction de la déformation soit correcte, mais que la magnitude soit erronée. Pour les petites atrophies, les valeurs les plus basses des biais sont observés pour ANTS, possiblement en raison de l'utilisation d'une transformation symétrique. Ce résultat est en accord avec les études existantes [58, 174] qui ont identifié l'utilisation d'une transformation symétrique comme un moyen d'obtenir des estimations atrophie non biaisées. Il ressort de nos expériences que les estimations de grandes atrophies sont biaisés malgré l'utilisation d'une transformation symétrique et d'un modèle de déformation fluide (dont l'intérêt est pourtant de mesurer de grandes déformations). Une explication possible concerne l'utilisation d'une régularisation gaussienne qui pénalise les grandes déformations plus que les petites.

Les tests ont été effectués dans une approche de type "leave one out" afin de s'assurer de la robustesse de notre approche à de nouvelles anatomies. Les résultats

ont souligné que l’algorithme ANTS permet d’obtenir une incertitude plus faible que les méthodes de Noblet *et coll.* et de Vemuri *et coll.*. Les longueurs maximales des intervalles de confiance sont égales à 14,32%, 4,82% et 8,94% pour les méthodes de Noblet *et coll.*, ANTS et Vemuri *et coll.*, respectivement, pour la plus grande atrophie simulée dans la base d’apprentissage (−14,87%). Nous avons pu observer que ces méthodes semblent mieux se comporter dans le cas de petites atrophies.

La probabilité de couverture (“coverage probability”) désirée était de 95%, tandis que des probabilités de couverture de 89,81 – 93,52%, 92,59 – 100% et 90,74 – 98,18% ont été obtenues avec les méthodes de Noblet *et coll.*, ANTS et de Vemuri *et coll.*. De cette étude, on peut conclure que la méthode ANTS fournit des intervalles de confiance significatifs par rapport à ceux obtenus avec les deux autres méthodes.

En termes de temps de calcul, il a fallu 10 – 12 minutes, 1 – 2 minutes et 3 – 4 minutes avec les algorithmes de Noblet *et coll.*, d’ANTS et de Vemuri *et coll.*, respectivement, pour traiter une paire d’images de taille $64 \times 64 \times 64$ voxels. Cela signifie que la création d’une base de données d’une centaine d’images est une question d’un jour pour l’algorithme de Noblet *et coll.* et de quelques heures pour les deux autres méthodes. Toutefois, il ne faut pas oublier que ce coût de calcul est un investissement en temps. Cette base de données peut en effet être réutilisée dans les calculs futurs de l’incertitude de l’estimation de l’atrophie cérébrale.

En conclusion, ce cadre proposé ici n’est pas seulement flexible en termes de son applicabilité à toute méthode d’estimation d’atrophie et de l’incorporation d’artefacts des images IRM, mais apporte aussi un avantage supplémentaire de correction automatique des biais existants dans les estimations d’atrophie.

5 Recalage bayésien

L’estimation de l’incertitude est un point important quand il s’agit d’utiliser les mesures d’atrophie obtenues à partir de recalage déformable pour des décisions médicales. Il est dans ce cas intéressant de développer une méthode qui est capable de fournir conjointement des estimations ponctuelles de l’atrophie et l’incertitude de ces estimations. À cet égard, une approche bayésienne de recalage non-linéaire est développée pour quantifier les changements volumétriques longitudinaux dans le cerveau et prendre en compte les artefacts d’images d’IRM (bruit et inhomogénéité en intensité) simultanément.

La première étape consiste à sélectionner des modèles mathématiques pour la transformation non-linéaire du recalage, l’inhomogénéité en intensité et le bruit présents dans les images IRM. La transformation non-linéaire est paramétrée à l’aide de B-splines cubiques. Parce qu’elles peuvent modéliser n’importe quel objet en 3D, indépendamment de son modèle physique, les B-splines forment une classe popu-

laire de transformations. Afin de produire des solutions réalisables physiquement, il est souhaitable d'introduire des contraintes de régularisation. Parmi de nombreux termes de régularisation, le lissage de type plaques minces ("thin plate splines") est choisi. Cette quantité représente l'énergie de flexion d'une plaque mince de métal et possède la propriété de pénaliser les grandes déformations. Pour restreindre davantage le champ de déformation aux solutions topologiquement réalisables, seules les solutions qui conduisent à un déterminant positif de la matrice jacobienne sont autorisées. L'inhomogénéité de l'intensité varie lentement dans l'image IRM. Ainsi, elle peut être exprimée comme une combinaison linéaire de fonctions lisses. Les polynômes de Legendre représentent l'une des nombreuses fonctions de base capable de représenter cette inhomogénéité en intensité dans les images d'IRM. Enfin, une distribution gaussienne est utilisée pour définir la vraisemblance (*i.e.* le critère de similarité) car les images sont supposées être corrompues par un bruit gaussien. Des a priori uniformes sont utilisés pour les paramètres de la transformation et les paramètres du biais en intensité. La procédure consiste alors à maximiser la probabilité a posteriori d'obtenir des estimations des différentes inconnues.

En raison de la nature complexe de la loi a posteriori, il n'est ni possible de la calculer analytiquement, ni de tirer des échantillons suivant celle-ci directement. Dans de telles situations, les méthodes Markov Chain Monte Carlo (MCMC) sont généralement employées pour générer des échantillons. Dans notre cas, des échantillons des paramètres du biais en intensité et des paramètres de transformation peuvent être générés par la distribution a posteriori en utilisant des méthodes MCMC et une approche par maximum a posteriori (MAP) permet d'estimer le mode de cette loi.

Résultats et discussion

Quatre niveaux d'atrophies ($-14,6\%$, $-9,9\%$, $-4,7\%$, $-1,6\%$) ont été simulés pour les sujets 2 et 7 de la base d'images IBSR. La méthode présentée précédemment est comparée avec quatre algorithmes de recalage non-rigide : Noblet *et coll.* [98], ANTS [10], Vemuri *et coll.* [157] et une méthode utilisant sur les B-splines pour modéliser le champ de déformation, mise en œuvre dans ITK (Insight Toolkit). L'approche bayésienne permet d'obtenir de meilleurs résultats que les méthodes de Noblet *et coll.*, de Vemuri *et coll.* et l'approche disponible dans ITK (correspondant à une optimisation déterministe). L'approche s'est révélée également compétitive par rapport à ANTS pour six expériences sur huit en utilisant 648 points de contrôle pour les B-splines. Les différences de performances sont dues à différents modèles de transformation, de régularisation et autres contraintes employées telles que la préservation de la topologie par exemple. Tandis que les méthodes ANTS et Noblet *et coll.* sont des approches multi-résolution, Vemuri *et coll.* et notre algorithme ne sont pas mises en œuvre dans un cadre multi-résolution. Pour les mêmes raisons, les cartes de jacobien obtenues sont très différentes. Notre approche fournit notamment des cartes plus lisses que celles obtenues avec les autres méthodes, notamment

en raison du nombre de paramètres utilisés pour caractériser le champ de déformation. La distribution du jacobien obtenue à partir de l’algorithme MCMC, était mono-modale pour la plupart de nos tests. Pour l’approche MCMC, la longueur des intervalles de confiance a été d’environ 2% de l’atrophie en considérant un bruit gaussien de 35 dB et ces intervalles contenaient dans nos expériences l’estimation ponctuelle de l’atrophie par maximum a posteriori (MAP). Les erreurs observées avec cette méthode sont de l’ordre de 0,22 – 1,61%.

Pour conclure, l’approche bayésienne a fourni des résultats très intéressants tant en présence de bruit ou d’inhomogénéité en intensité dans les images. La modélisation bayésienne et les techniques d’échantillonnage MCMC nous a permis d’élaborer un cadre pour l’estimation de l’atrophie cérébrale. Choisir les bons modèles pour les différentes quantités d’intérêt est une étape clé. Cependant, les méthodes d’échantillonnage MCMC sont coûteuses en temps de calcul, notamment lorsque le nombre de paramètres à échantillonner augmente. La motivation concernant l’utilisation d’un tel cadre mathématique est de pouvoir fournir des estimations des incertitudes des mesures d’atrophie. Nos résultats ont justifié l’intérêt de développer une approche bayésienne et ce cadre semble être très prometteur pour l’avenir.

6 Conclusions

Le travail effectué dans le cadre de cette thèse est consacré à l’estimation de l’atrophie cérébrale longitudinale à partir d’images IRM.

La simulation d’atrophie cérébrale est un des moyens de générer une vérité terrain nécessaire pour l’évaluation de méthodes d’estimation d’atrophie. Les performances de trois algorithmes disponibles à la communauté (SIENA, SIENAX et BSI-UCD) et l’algorithme de recalage non-rigide de Noblet *et coll.* ont été évaluées dans ce cadre. Les expériences menées ont montré que SIENA fournit les estimations les plus fiables des trois algorithmes disponibles (SIENA, SIENAX, BSI-UCD). Cependant, l’approche développée par Noblet *et coll.* a permis d’obtenir de meilleurs que SIENA dans une seconde série d’expériences. Par ailleurs, l’inhomogénéité en intensité et le bruit contribuent majoritairement à la hausse des erreurs dans l’estimation de l’atrophie par rapport aux artefacts d’interpolation et aux distorsions géométriques.

Cette étude nous a permis d’identifier que les étapes essentielles pour SIENA et BSI-UCD sont le recalage et la segmentation du cerveau (liée à la précision de l’extraction du cerveau). Pour SIENAX, le goulot d’étranglement concerne principalement l’étape d’extraction du cerveau. L’algorithme de Noblet *et coll.* dépend à la fois la précision du recalage non-rigide et de l’étape d’extraction de la zone d’intérêt. Le recalage est notamment fonction de nombreux aspects tels que le modèle (le critère de similarité, la régularisation) et les paramètres algorithmiques

(multi-résolution et optimisation).

Cette étude a permis de montrer que les erreurs d'estimation obtenues dans ces expériences étaient comparables au taux annuel d'atrophie (0,5 à 2.8%) observée pour diverses pathologies. Il s'agit d'une des raisons qui nous ont motivé à élaborer un cadre permettant l'estimation des incertitudes pour la mesure de l'atrophie cérébrale.

À cette fin, un nouveau cadre pour la construction d'intervalles de confiance des méthodes d'estimation de l'atrophie a été proposé. La caractéristique la plus remarquable de cette approche est qu'elle a été capable de corriger les erreurs qui découlent de biais dans les estimations de l'atrophie ainsi que d'inclure les artefacts présents dans les images IRM dans l'estimation de l'atrophie. L'utilisation d'une base d'apprentissage permet l'étude de différentes zones du cerveau, d'inclure plusieurs types d'artefacts selon les besoins de l'étude. L'efficacité de ce cadre a été démontrée en comparant trois algorithmes de recalage déformable (Noblet *et coll.*, ANTS and Vemuri *et coll.*) pour la création d'intervalle de confiance de mesures de l'atrophie dans l'hippocampe. Cette analyse confirme que la correction de ce biais est nécessaire pour renforcer la précision des méthodes d'estimation de l'atrophie cérébrale. L'origine de ce biais peut être expliquée par l'utilisation de différents modèles de transformation ou de régularisation par exemple. Cette étude a montré également que l'algorithme ANTS fournit les intervalles de confiance les plus significatifs par rapport aux méthodes de Noblet *et coll.* et Vemuri *et coll.*

Dans la dernière partie de cette thèse, un nouvel algorithme d'estimation de l'atrophie a été développé. Nous avons construit un modèle bayésien complet facilitant l'estimation des intervalles de confiance des mesures de l'atrophie dans une région d'intérêt. La formulation bayésienne se composait de la modélisation de la transformation non-rigide et des inhomogénéités en intensité. Les résultats obtenus sur des images dégradées par un bruit gaussien ont montré que l'approche bayésienne fournit dans la plupart des cas de meilleurs résultats que les méthodes d'ANTS, de Noblet *et coll.* et de Vemuri *et coll.* L'étude menée a également montré que l'estimation des variations d'intensités et de l'atrophie sont liées [109]. Dans nos expériences, nous avons montré qu'il était préférable d'estimer le biais en intensité et l'atrophie séparément. Bien que cette approche soit très coûteuse en temps de calcul, l'approche développée dans cette thèse permet d'obtenir de façon rigoureuse des intervalles de confiance de l'atrophie mesurée à partir d'images IRM.

Les contributions principales de cette thèse sont les suivantes:

- Développement d'un cadre qui a permis la comparaison des approches existantes d'estimation de l'atrophie du cerveau par la création d'une vérité terrain simulée et l'évaluation des méthodes largement utilisées (effet d'inhomogénéité en intensité, bruit, distorsions géométriques, artefacts d'interpolation et lé-

sions).

- Estimation de l'incertitude dans l'atrophie estimée pour tout algorithme dans une région d'intérêt sous la forme d'intervalles de confiance; le point clé concerne la correction de biais des estimations de l'atrophie.
- Construction d'un modèle bayésien complet pour l'estimation simultanée de l'atrophie, des biais en intensité des images IRM et l'estimation des intervalles de confiance sur l'atrophie estimée.

General Introduction

Several diseases affecting the central nervous system, which include Multiple Sclerosis (MS), Alzheimer's and schizophrenia, are characterised by decreasing brain volume. This process is known as brain atrophy which is a general term representing the decrease in size or wasting away of the brain due to loss of neurons and connections between them. Clinical studies suggest that brain matter loss begins early in the course of diseases such as MS. In addition, there has been increasing amount of evidence linking brain atrophy to a variety of neurological and cognitive impairments. Thus, brain atrophy has emerged as a conspicuous biomarker of disease progression. Consequently, estimation of brain volumetric change has been given a fair amount of attention in the recent past by the research community. Automated tools, capable of detecting changes in the brain accurately, are the need of the day as they facilitate disease diagnosis and prognosis.

Magnetic Resonance Imaging (MRI) techniques are an important non-invasive means of examining changes in the brain because they provide good contrast between different tissues of the brain. Developing image processing programs for manipulating MR images, therefore, offer ways of measuring brain volume loss in the course of a disease.

The purpose of this thesis is to investigate image processing methods for estimating brain atrophy endured by MS patients. Longitudinally acquired 3D MR images of a patient are compared with an aim of obtaining an estimate of loss in brain volume in a period of time. Such analyses are referred to as longitudinal brain atrophy estimations.

Three major parts define the work produced in this thesis.

The first part deals with the development of a framework for evaluating existing approaches of longitudinal brain atrophy estimation using simulated ground-truths. Acknowledging the fact that MR acquisitions are not free of artefacts, one expects that they would interfere with the normal functioning of an image processing method. For this reason, extensive validation of existing approaches of estimating brain atrophy is necessary to understand the effect of various MR artefacts. Nevertheless, validation of methods is never an easy task, in the real scenario, due to the absence of gold standards for comparison.

In order to create gold standards, an approach for realistic simulation of brain tissue loss that relies on the estimation of a topology preserving B-spline based deformation fields is utilised. Using these gold standards, examination 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 performed in detail. The results show that, in general, bias-field inhomogeneity and noise lead to larger errors in the estimated atrophy than geometrical distortions and interpolation artefacts. Further experiments are performed in order to include a non-rigid registration (NRR) algorithm of Noblet *et. al* in the comparison. Experiments on 18 different anatomical models of the brain after simulating whole brain atrophies in the range of 0.2 – 0.6% indicated that, in the presence of bias-field inhomogeneity and noise, mean errors of 0.4472 ± 0.3074 , 3.4483 ± 1.7303 , 1.5255 ± 0.6828 and 0.3256 ± 0.1415 may be expected in the atrophy estimated by SIENA, SIENAX, BSI-UCD, and Non-rigid registration (NRR) algorithm of Noblet *et. al*, respectively. This step is instrumental in pinpointing the sources of errors arising from MR artefacts and disease evolution (MS lesions).

The second part of the thesis is dedicated to statistical analysis of maps of volumetric change in an individual brain so as to gain an insight into the uncertainties associated with the estimated atrophy. A generic framework for constructing confidence intervals for atrophy estimated in any region of interest in the brain is developed. All uncertainty estimations are performed separately *i.e.* they do not form a part of the atrophy estimation method. Hence, this framework can be used for *any given method* such as SIENA, BSI or an NRR algorithm.

This framework works in two steps. First a learning database is created with the help of a simulated ground truth. This step is carried out to learn the mean and variance of each estimator of atrophy. Next this database is used for constructing confidence intervals for an estimated atrophy. Although, we are interested in evaluating longitudinal atrophy in one patient, this database was built independent of patient’s anatomy. Due to the computational complexity involved in constructing this database, it is stored and reused for future uncertainty estimations in atrophy for any patient. Confidence intervals are developed for estimated hippocampal atrophy that a patient has undergone over time. This framework is tested for three NRR algorithms developed by Noblet *et al.*, Avants *et al.* (ANTS) and Vemuri *et al.*). Results show that ANTS algorithm produces compact confidence intervals in comparison to the other two methods. Maximum lengths of confidence intervals are 14.32%, 4.82% and 8.935% for Noblet *et. al*, ANTS and Vemuri *et al.*, respectively for the highest atrophy in the database (−14.87%).

Finally, in the quest of creating a novel atrophy estimation approach, that has the potential of quantifying longitudinal brain atrophy from serial images of a patient along with the related uncertainties, a Bayesian registration framework is devised.

Since a non-rigid registration method is implemented, atrophy in structures of interest could be evaluated. Additionally, Bayesian registration also allows us to obtain probability distributions of all the quantities of interest as a by-product of the estimations. In order to deal with images degraded with bias-field inhomogeneity and noise, the estimation of the deformation field, bias-field inhomogeneity and noise is achieved simultaneously. In our formulation, the deformation field is modeled using B-splines, bias-field inhomogeneity with Legendre polynomials and Gaussian noise is considered. Unfortunately, it is not possible to analytically analyse the posterior probability distribution for the deformation field or for other unknowns, nor is it feasible to draw samples directly from it. A widely used Markov Chain Monte Carlo (MCMC) method called Metropolis-Hastings algorithm allows us to perform the sampling. Tests are carried out on real images with simulated longitudinal hippocampal atrophy and comparisons are made with other NRR methods (Noblet *et al.*, ANTS and Vemuri *et al.*). While the observed errors with the new method are in the range of 0.22 – 1.61%, the confidence intervals could localise the atrophy within 2%. Besides, our approach also outperformed other deterministic algorithms in most of the cases. Thus, it can be concluded that our method holds great promise for atrophy estimations along with the uncertainty in the coming years.

Thesis plan

This dissertation is organised in four chapters.

The first part, chapter 2, is a brief introduction to the structure of the brain, MS disease, brain atrophy in MS, MRI methods and image processing tools studied and used in this work. Section 2.2 provides a description of MS and its subtypes followed by diagnosis and treatment of MS. In section 2.3, brain atrophy mechanisms in MS, types of brain atrophy, treatment procedures and importance of brain atrophy as a marker of disease progression is discussed. The basic principle of MRI, MR anatomical scans, MR artefacts (bias field inhomogeneity, noise, geometrical distortions) and their correction strategies are detailed in section 2.4. A detailed literature survey of existing methods of estimating atrophy using MRI is presented in section 2.5. Brain extraction algorithms and their performances are summarised in section 2.6. Current issues in brain atrophy estimation are addressed in section 2.7.

It is important for the user to know the potential reasons of failure and further quantification of errors for these methods. Hence, chapter 3 is dedicated to the validation of existing freely available atrophy estimation methods using a ground truth. A survey of existing validation studies and performance comparison of atrophy estimation methods is presented in section 3.1. The proposed framework is detailed in two parts: first, an atrophy simulation algorithm that allows us to create a ground truth is detailed in section 3.2.1 which is followed by simulation of artefacts for the

validation framework are explained in section 3.2.2. Results of atrophy simulation on BrainWeb images are presented in section 3.3.1 and evaluation experiments for SIENA, SIENAX, BSI and JI are shown in section 3.3.2. The findings from this study are listed in section 3.5.

Chapter 4 deals with the estimation of uncertainties in atrophy estimation. Existing uncertainty estimation methods in image registration are discussed in section 4.2 including Bayesian registration and MCMC methods in section 4.2.3. The proposed generic framework for estimating uncertainty in longitudinal brain atrophy by means of constructing confidence intervals *for any given method* is discussed in section 4.4. It consists of learning uncertainties from example simulations using a ground-truth (section 4.4.1) and construction of confidence intervals (section 4.4.2). The constructed confidence intervals are shown for three NRR methods [98], [10] and [157] on real images in section 4.5 which is mainly dedicated to quantification of bias in atrophy estimations and the evaluation of the confidence intervals.

Finally, a novel Bayesian registration method is described in chapter 5 to simultaneously compute bias field inhomogeneity, noise and brain atrophy. The proposed Bayesian model is described in section 5.1 in three parts: structural modelling (section 5.1.1), definition of relations between the transformation, bias field inhomogeneity, noise (section 5.1.2) and, probability modelling and MCMC sampling procedure (section 5.1.3). Experimental results on estimation of bias field inhomogeneity and atrophy from our framework are compared with popular bias field correction methods (N4ITK [155] and differential bias correction of Lewis *et al.* [80]) and four different non-rigid registration approaches of Noblet *et al.* [98], Vemuri *et al.* [157], ANTS [10] and another cubic B-spline based registration method in section 5.3. Results are discussed in section 5.4.

The overall conclusions, findings from this thesis and future perspectives are presented in chapter 6.

Background

Multiple sclerosis (MS) is a central nervous system disease, affecting between 2 to 150 persons per 100,000 in the world. In Europe, United States, Russia, Canada, Australia and New Zealand, this disease is rampant. The rarity of MS among Samis, Turkmen, Uzbeks, Kazakhs, Kyrgyzis, native Siberians, North and South Amerindians, Chinese, Japanese, African blacks and New Zealand Maoris, as well as the high risk among Sardinians, Parsis and Palestinians, indicate that different racial and ethnic groups have different susceptibility to this disease and explains the geographical unevenness in the presence of MS [123].

MS is a highly researched disease with decrease in brain volume being one of its characteristics [92]. With the development of robust image processing methods, MRI techniques have become indispensable in studying MS disease progression by measurement of brain atrophy, among other things. MRI comes with the advantage of providing high tissue contrast however, it is prone to artefacts that modify the acquired image. For example, MR noise is responsible for modifying the appearance of the true signal in the image while others such as geometrical distortion mimic the pathology or anatomy. Presence of artefacts impedes the correct functioning of image processing methods.

This chapter is a brief introduction to the structure of the brain, MS disease, MRI methods and image processing tools studied and used in this thesis. section 2.2 provides a description of MS disease and brain atrophy mechanisms in MS. Next, MRI artefacts and correction strategies are detailed in section 2.4. A literature survey of existing methods of estimating atrophy using MRI is presented in section 2.5.

2.1 The brain

Let us examine the structure of the brain cursorily in order to be better equipped to understand the subsequent sections. Neurons are basic building blocks of the nervous system which (see Fig. 2.1) consist of four main parts:

- the nucleus, the mitochondria and other organelles required by the cell (called the cell body).
- the dendrites that establish connections and assist in receiving chemical signals from other neurons.

- the axon or nerve fibre which connects different parts of the body to the brain, and passes electro-chemical signals.
- the axon terminal where the axon connects with other neurons or muscles to pass on signals (synapses).

2.1.1 White matter

White Matter (WM) consists mostly of myelinated axons. The function of an axon is to receive and interpret messages in the brain. In order to speed up transmissions of electric signals through the axons, an insulating layer known as myelin sheath surrounds them (see Fig. 2.1). It is due to the white colour of myelin that this part of the brain is called “white matter”.

WM is located in the inner layer of the cortex, the optic nerves, the central and lower areas of the brain (the brainstem) and surrounding the grey matter in the spinal cord.

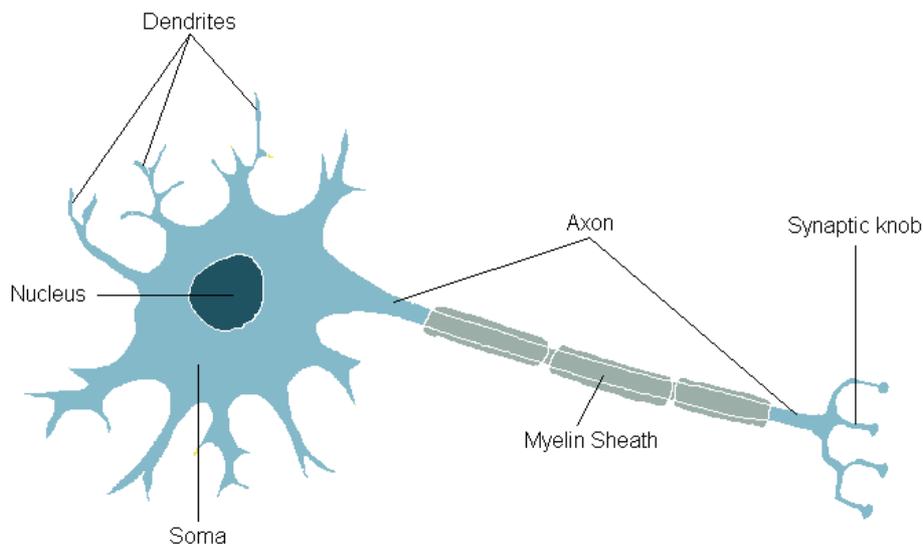


Figure 2.1: Structure of a neuron (Source: <http://www.ainenn.org/>). Figure shows the cell nucleus and the main body (soma). Neurons have extensions called dendrites whose main function is to receive chemical messages. Also shown are axons that pass electro-chemical signals to other neurons. The axons are covered with a myelin sheath that acts as an insulator and accelerates communication of signals. At the axon terminal are located the synaptic knobs, where the electro-chemical signals are converted to chemical signals and passed on to other neurons.

2.1.2 Grey matter

Grey matter (GM) contains neural cell bodies and their dendrites as opposed to white matter which contains mostly myelinated axons. In living tissue, grey matter has grey-brown colour and includes brain areas responsible for muscle control, sensory perception such as seeing and hearing, memory, emotions, and speech.

It is distributed at the surface of the cerebral hemispheres (cerebral cortex) and of the cerebellum (cerebellar cortex), as well as in the depth of the cerebrum (thalamus; hypothalamus; sub-thalamus, basal ganglia - putamen, globus pallidus, nucleus accumbens; septal nuclei), deep cerebellar nuclei (dentate nucleus, globose nucleus, emboliform nucleus, fastigial nucleus), brainstem (substantia nigra, red nucleus, olivary nuclei, cranial nerve nuclei) and spinal white matter (anterior horn, lateral horn, posterior horn).

2.1.3 Cerebrospinal fluid

It is a colourless fluid that surrounds the brain and the spinal cord. Cerebrospinal fluid (CSF) is produced by an organ, called the choroid plexus, in the ventricles and by the surface of the roof of the third and fourth ventricles. The major functions of CSF are to protect the brain from injury and to maintain chemical stability.

Some internal parts of the human brain are shown in Figure 2.2.

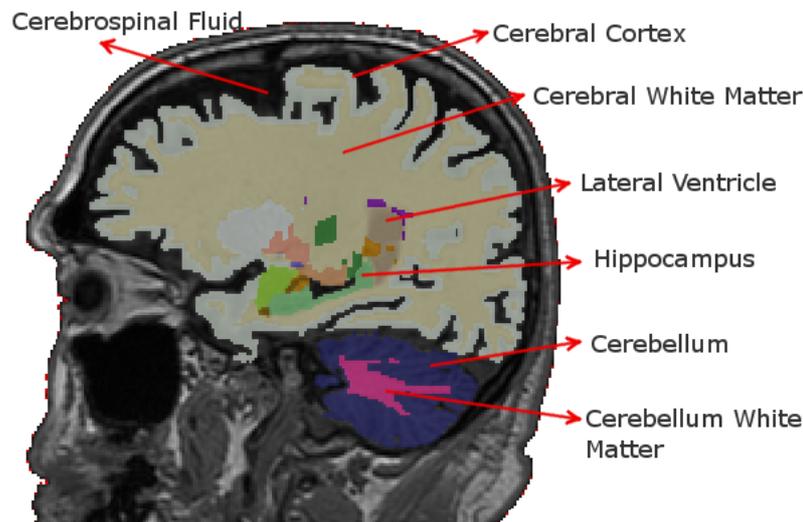


Figure 2.2: A sagittal view of the human brain. The cerebral and cerebellum WM, CSF and GM (cerebral cortex, hippocampus and cerebellum) are shown. This figure has been obtained by segmenting a T1 weighted MR image from the Alzheimer's Disease Neuroimaging Initiative (ADNI) <http://adni.loni.ucla.edu/> database with FreeSurfer's segmentation algorithm <http://surfer.nmr.mgh.harvard.edu/>.

2.2 Multiple sclerosis

MS is known to be an inflammatory demyelinating disease constituted mainly by multi-focal areas of WM lesions. MS falls into the category of autoimmune diseases where a damaged immune system is forced to attack its own tissues, on account of its inability to distinguish between virus proteins and body's own myelin (a substance responsible for insulating nerve endings) and thus produces antibodies that attack. This condition is known as autoimmunity meaning that the body has developed an allergy towards itself. Readers seeking a thorough treatment of MS are referred to [113].

2.2.1 Causes

The causes of MS have not, until now, been clearly understood. Researchers all over the world are working towards finding an explanation to this complicated phenomenon. Many scientists believe that the root cause of the symptoms experienced by MS patients can be explained through understanding the process of demyelination or loss of myelin. When the myelin sheath is lost, nerves are coerced to transmit signals tardily. Loss of myelin may even result in complete loss of transmission in severe cases. This process leads to patches of scarring or 'sclerosis', occurring where neurons have lost myelin. Such patches of inflammation are known as lesions or plaques and are abundant in the WM namely in centrum semiovale, corpus callosum, optic chiasm, optic nerves, brain stem and spinal cord [115]. MS derives its name from these areas of scarring. Fig. 2.3 illustrates an MS patient with lesions.

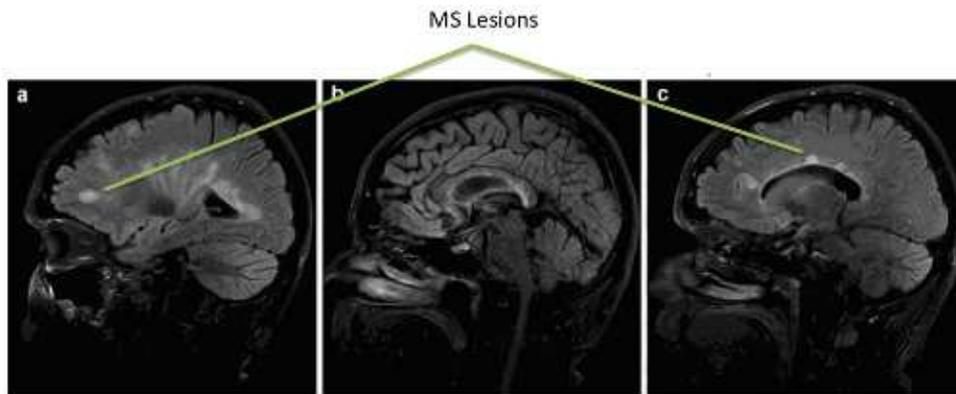


Figure 2.3: (a-c) Sagittal fluid-attenuated inversion recovery (FLAIR) images obtained for a 27 year old woman with MS with demyelinating lesions in the periventricular white matter and corpus callosum. There is extensive involvement of callosum, which is a hallmark of MS. (Source: [115] pp. 172).

Even when the patches of scarring caused by demyelination have healed and remyelination has occurred, the response time of the nerve endings tends to remain slower. The process of demyelination along with an in-vivo MR image of an MS patient illustrating confluent demyelination are shown in Fig. 2.4.

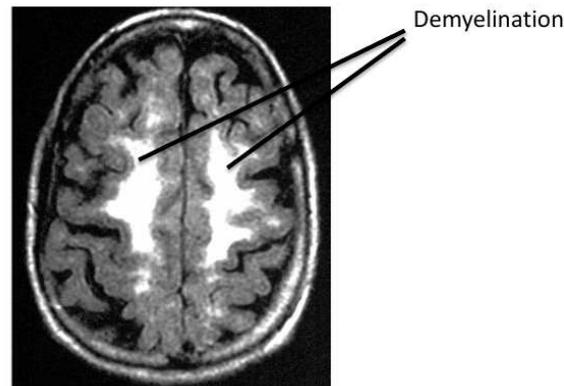


Figure 2.4: Axial T1-weighted FLAIR MR image showing confluent demyelination. (Source: [107]).

2.2.2 Symptoms

Most patients experience their first symptoms of MS when they are as young as 20 years of age. The initial symptom of MS is often blurred or double vision, red-green color distortion, or complete blindness in one eye. The course of patients with MS disease is largely unpredictable, with patients suffering from just any neurological symptom from muscle weakness, difficulty in coordination, impaired walking to transitory abnormal sensory feelings such as numbness and prickling. In worst cases, MS can produce partial or complete paralysis. Some patients may also experience pain. Aphasia, tremors, and dizziness are also frequently seen. Cognitive impairments such as difficulties with concentration, attention, memory, and poor judgement are prevalent in almost half of the MS patients. Depression is another common feature of MS. With the lapse of time, MS patients may suffer a variety of symptoms of which some are reversible and others not [127].

2.2.3 Multiple sclerosis subtypes

Based on factors including frequency of clinical relapses, time to disease progression, and lesion development on MRI, four types of MS can be distinguished [85]:

Relapsing/remitting MS (RRMS)

This type of MS is characterised by relapses and in this period new symptoms may appear and old ones may resurface or worsen. The relapses are followed by remission, during which the patient fully or partially recoups from the adverse effects during the relapse. Relapses can last for days, weeks or months and recovery can be slow and gradual or almost instantaneous. Almost all patients presenting with MS are first diagnosed with relapsing/remitting disease.

Secondary progressive MS (SPMS)

SPMS is the next phase of the disease for persons who have suffered from relapsing/remitting MS for some years. This is characterised by a gradual worsening of the disease between relapses. A few relapses may be experienced in the beginning of SPMS but after a while these will result into a general progression. There may be relapses followed by a periods of remissions, but once a person has entered this phase, full recovery is generally not possible.

Primary progressive (PPMS)

This type of MS is characterised by a gradual progression of the disease from its onset with no remissions at all. There may be periods when there is no disease activity or levelling off of disease activity. As with SPMS, there may be good and bad days or weeks. The disease is seen mostly affecting the spinal cord but often migrates into the brain, but is less likely to damage brain areas like in relapsing/remitting or secondary progressive.

Progressive relapsing MS (PRMS)

In this form of MS, progression of disease is seen from its onset, emphasised by relapses. The patient may recover significantly immediately following a relapse. At the same time, there is a gradual worsening of symptoms between the relapses.

Figure 2.5 illustrates how disability is related to time in different forms of MS.

2.2.4 Diagnosis and treatment

There is no single test that confirms the diagnosis of MS as the occurrence of symptoms may be random, affecting multiple areas of the body. In 2001, an international panel in association with the National Multiple Sclerosis Society (NMSS) of America recommended a revised diagnostic criteria for MS called the McDonald criteria [90]. These criteria facilitated the diagnosis of MS in patients with signs and symptoms suggestive of the disease such as mono-symptomatic disease, disease with a typical relapsing-remitting course or insidious progression but no clear attacks and remissions. Since 2001, this criteria has been revised two times in 2005 [90] and in 2011 [111] where the use of imaging for demonstration of distribution of central nervous system lesions in space and time has been simplified and in some cases distribution of lesions can be established by a single scan. The aim of these revisions was to simplify the criteria while preserving their diagnostic sensitivity and specificity in order to allow a early diagnosis, uniform and widespread use.

Evidence of chronic inflammation of the central nervous system is sought through a procedure known as lumbar puncture, where CSF is collected and analysed. Additionally, nerve conduction studies of optic, sensory and motor nerves can reveal

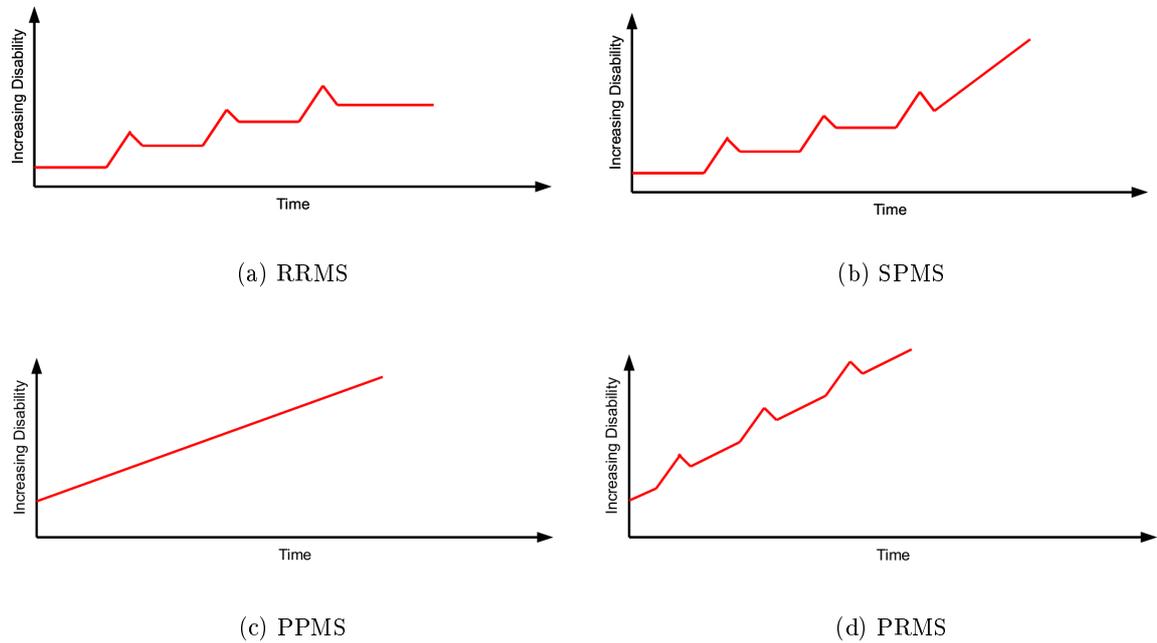


Figure 2.5: Figure showing the increase in disability with time for different types of MS. RRMS: unpredictable attacks followed by periods of remission, SPMS: initial RRMS suddenly begins to exhibit increased disability with no remissions, PPMS: steady increase in disability without attacks and PRMS: steadily increasing disability since onset with superimposed attacks. (Source: [113]).

evidence about presence of MS as the process of demyelination results in sluggish nerve conduction. Several laboratory tests may be necessary to exclude other diseases such as sarcoidosis, vasculitis and Lyme disease that exhibit symptoms similar to MS. Brain and spinal cord atrophy, conceded by MRI, is also used as another diagnostic criteria for MS [84, 93, 12].

Treatment of MS may be handled in many ways. Disease modifying drugs can help reduce the severity or frequency of relapses by altering or suppressing the activity of the body's immune system. Common drug treatments include interferon beta-1a, interferon beta-1b, glatiramer acetate and fingolimo for relapsing forms of MS and mitoxantrone for rapidly relapsing-remitting, primary-relapsing and secondary-progressive forms of MS [110, 24]. Some symptoms of MS can be managed with the help of specialists such as physiotherapists and psychologists who can help with mobility, coordination and memory or concentration problems.

2.3 Brain atrophy

Brain atrophy is a general term representing the decrease in size or wasting away of brain matter due to loss of neurons and connections between them. It is a char-

acteristic of many diseases affecting the central nervous system such as Alzheimer's (AD) [25, 138], MS [136] and Schizophrenia [17], these diseases being very prevalent in the world and evolving over decades. The extent of neurological dysfunctions and cognitive impairments in the disease course determine the quality of life of patients [74]. As a result, rapid advances in research in areas related to these diseases are imperative, given the duration of the disease and constant assistance needed by patients in order to perform simple chores. Although brain atrophy is a typical feature of various neurodegenerative diseases, the focus in this thesis is on studying longitudinal brain atrophy in MS patients. Figure 2.6 shows MRI scans of an individual taken at different instances of time indicating brain atrophy progression. Notice the prominent enlargement of ventricles as atrophy becomes more and more severe.

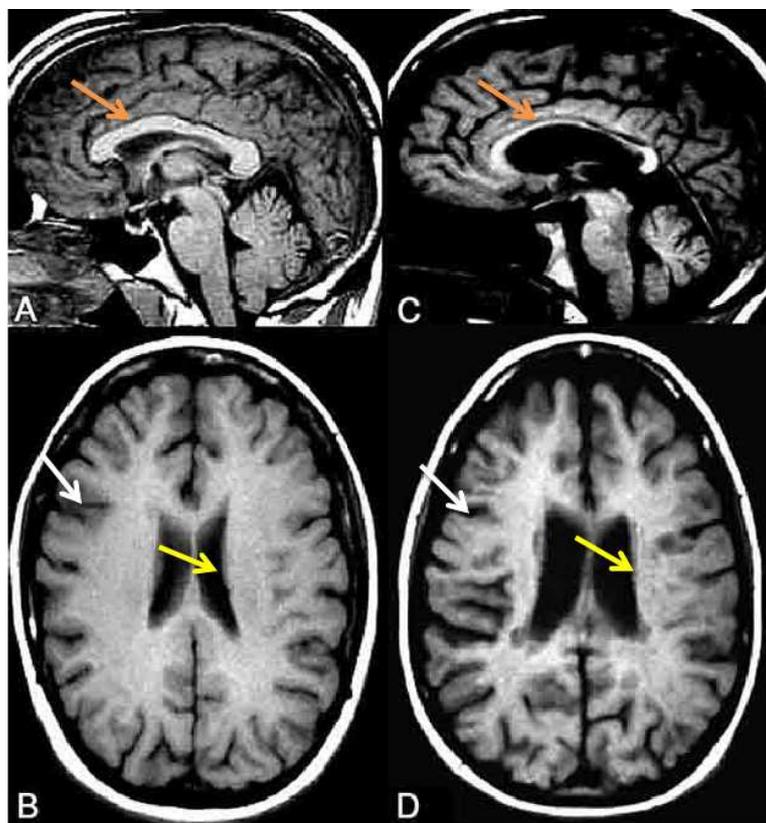


Figure 2.6: Brain Atrophy in MS shown on T1-weighted MR images are shown for normal controls in their fifth decade (a-b) compared with age-matched patients with MS (c-d). Note thinning and decreased volume of corpus callosum (marked with orange arrows). Also shown are cortical atrophy (marked with white arrows) and ventricular enlargement (marked with yellow arrows) in MS. (Source: [178])

2.3.1 Brain atrophy in MS

The emergence of brain atrophy in MS is related to a combination of multiple factors. The exact mechanisms underlying central nervous system atrophy in MS patients are largely unknown, evidence exists that atrophy may arise due to effects of dynamic inflammation within the central nervous system, including demyelination, axonal injury, neuronal loss, Wallerian degeneration, and possibly iron deposition [93]. 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 MS.

2.3.2 Types of brain atrophy

Brain atrophy may manifest locally in a part of the brain or globally affecting the entire brain. In local or regional atrophy, the damage is concentrated on a particular area of the brain, which may become functionally impaired. Generalised brain atrophy involves the whole brain, and may be associated with a whole range of problems. Parts of GM of the brain are generally atrophied early in the disease course [180, 41]. In MS, local cortical thinning is observed beside an overall reduction of the cortical thickness with disease progression [125]. Cortical atrophy has been shown to be related to lesion load and disability occurring even in patients with mild disability [26]. Another prominent feature of brain atrophy in MS is ventricular enlargement in Relapsing-Remitting MS. On the other hand, cortical atrophy seems to be more important in the progressive forms of MS [104]. Hippocampus, a critical brain area for memory retention, has been shown to undergo atrophy that surpasses the global brain atrophy in MS [135]. Dietemann *et al.* have shown that corpus callosum atrophy appears earlier than brain atrophy in the course of MS and severity of clinical symptoms is more pronounced in patients with severe corpus callosum atrophy [36]. Global brain atrophy measurements have also been widely studied and accepted not only in MS but in other diseases too. Brain atrophy is observed in early stages of PPMS and affects both GM and WM. WM atrophy appears more closely related to clinical outcome and WM local damage than GM atrophy [126]. 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 [138].

2.3.3 Treatment of brain atrophy

Until now, there is no complete cure for brain atrophy. The treatment for brain atrophy concentrates on ways of preventing further damage if possible through drugs. For example, Zivadinov *et al.* [177] have explored the effect of steroids on brain atrophy. They show that chronic use of high dose intravenous methylprednisolone (IVMP) in patients with MS may limit brain atrophy progression over long-term via different immunological mechanisms, including down regulation of adhesion molecule expression on endothelial cells, decreased cytokine and matrix metalloproteinase secretion, decreased auto reactive T-cell-mediated inflammation and T-cell apoptosis

induction, blood-brain barrier closure, demyelination inhibition and remyelination promotion. In another study, clinical trials dedicated to analysing brain atrophy progression as a result of undertaking interferon beta-1a therapy have observed slowing of whole brain, GM atrophy and T_1 -hypointense lesion (also called T_1 black hole and is an area of severe tissue damage detected by MRI) volume accumulation in Relapsing-Remitting MS [179]. Aside from that, therapy focused on aiding the patients in coping with the decreased brain function is performed. Patients regularly undergo neurological exams with an aim of monitoring their progress. In case of a seditious decline, patients may be put in a special care facility if their caretakers are unable to provide support due to excessive brain damage.

2.3.4 Brain atrophy as a biomarker

Clinical studies suggest that brain matter loss begins early in the disease course of MS and is seen in all stages of the disease [92, 42]. There has been increasing amount of evidence linking brain atrophy to a variety of neurological and cognitive impairments. In a recent study, it was shown that decline in ambulatory function is related to atrophy of central brain regions exclusively, whereas decline in neurologically more complex tasks for coordinated hand function is related to atrophy of both central and peripheral brain regions [65]. Moreover, there is an ongoing debate about the relationship between degenerative and inflammatory processes in MS, which may be crucial in deciding the future course of disease treatments [102]. Studies have highlighted the importance of the irreversible tissue destruction in determining disease progression more than lesion growth assessments [92, 15, 4, 136, 40]. Whole brain and regional atrophy estimates are sensitive measures of neuro-degenerative component of multiple sclerosis and have the potential of evaluating future anti-inflammatory, remyelinating or neuro-protective therapies in clinical trials. Regional atrophy detection is particularly useful in understanding specific neuro-physical dysfunctions or other specific clinical findings [35]. Another dimension of research where the atrophy measures prove to be valuable is when studying different forms of MS. They enable us to study any correlations that exist between regional atrophy and the disease subtype [104]. These destructive processes hold promise and are capable of providing novel therapeutic targets to help MS patients. In the light of these factors, brain atrophy in conjunction with MRI methods has emerged as an important biomarker of disease progression and severity [165]. Thus, the last decade has witnessed the use of MRI in the development of a number of sophisticated image processing tools for quantifying the loss in brain volume.

2.4 Magnetic resonance imaging

MRI is a non-invasive medical exam that helps the physicians in diagnosing and treating medical conditions. Since the eighties, MRI has presented itself as a powerful imaging technique as a way of visualising detailed structures in-vivo. MRI makes use of electromagnetic waves, which means that there is no exposure to radiation

such as in X-rays or computed tomography (CT) scans. In addition, it provides good tissue contrast enabling the detection or appearance of changes such as tumours or cancerous tissues. The basic principle of MRI is that a magnetic field, much stronger than that of the earth, forces hydrogen atoms in the body to line up in a certain way. When radio waves are sent towards these lined-up hydrogen atoms, they rebound, and a computer records the signal. Different types of tissues send back different signals and accordingly tissue contrast appears in an MR image. Interested readers are referred to [52] for an in-depth understanding of key fundamental and operational principles of MRI.

The basic parameters of image acquisition with MRI are echo time (T_E) and repetition time (T_R), which are optimised to provide image contrast based on the chemical sensitivity of MRI. The echo time (T_E) represents the time between the start of the Radio frequency (RF) pulse and the maximum in the signal. The second parameter, repetition time (T_R), is the amount of time that exists between successive pulse sequences applied to the same slice. Standard MRI scans are described below:

T_1 -weighted MRI These scans are the most commonly run scan that use a gradient echo (GRE) sequence, with short T_E and short T_R . This gives rise to contrast between fat and water: with water darker and fat brighter. Due to the short repetition time (T_R) this scan can be run very fast allowing the collection of high resolution 3D datasets. The main property of the T_1 -weighted scans is that they allow us to obtain a GM/WM contrast. In these images, WM appears brighter than the GM. See Fig. 2.7(a).

T_2 -weighted MRI As in the T_1 -weighted scan, fat is differentiated from water. However, unlike the T_1 -weighted images, fat shows brighter and water lighter. As a result, brain WM therefore shows as darker than the GM, which facilitates imaging oedema. T_2 -weighted scans use a spin echo sequence, with long T_E and long T_R . They have been in use for a long time since the spin echo sequence is less susceptible to inhomogeneities in the magnetic field. See Fig. 2.7(b).

Proton density weighted MRI Unlike T_1 or T_2 weighted scans, spin density or proton density weighted scans gain contrast only from differences in the amount of available spins (hydrogen nuclei in water). It uses a spin echo or sometimes a gradient echo sequence, with short T_E and long T_R . See Fig. 2.7(c).

Fluid attenuated inversion recovery (FLAIR) It is an inversion-recovery pulse sequence used in order to null signal from fluids. For example, it can be used in brain imaging to suppress CSF so as to bring out the periventricular hyperintense lesions, such as MS plaques. By carefully choosing the inversion time TI (the

time between the inversion and excitation pulses), the signal from any particular tissue can be suppressed. See Fig. 2.7(d).

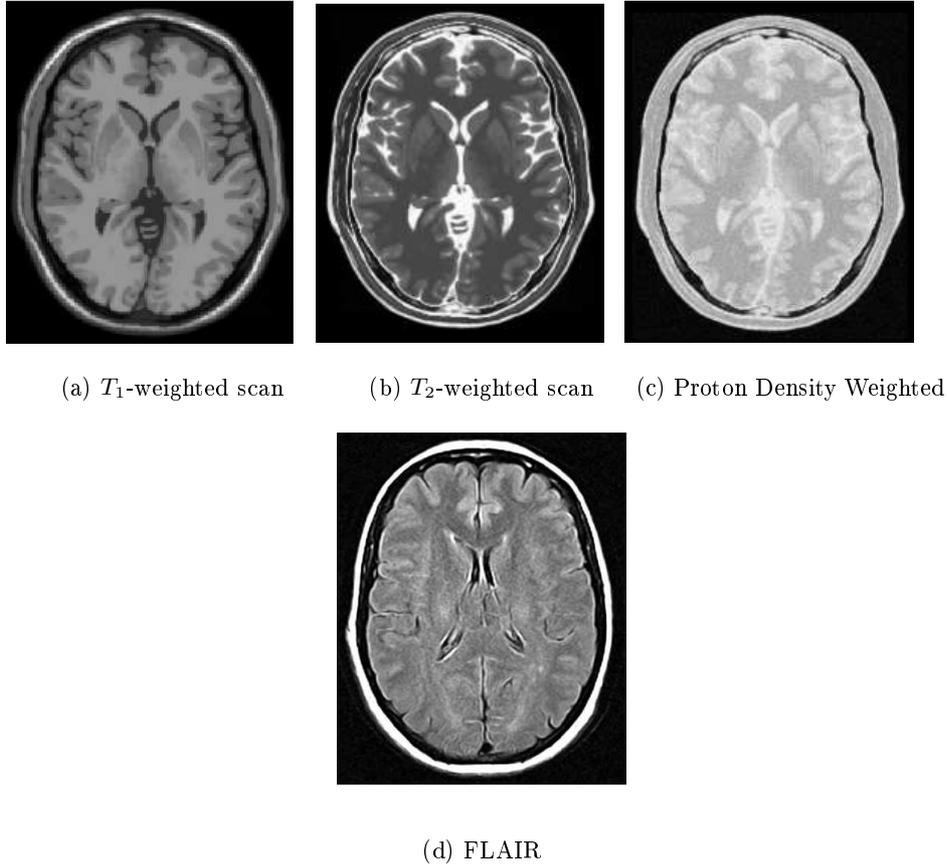


Figure 2.7: Anatomical MR scans (Source: (a-c) BrainWeb Image Database [7] and (d) <http://www.gehealthcare.com/euen/mri/images/neuro4.jpg>)

Other type of scans include Susceptibility weighted imaging (SWI) that exploit the susceptibility differences between tissues to provide contrast and specialised scans such as diffusion weighted MRI are currently in great use. The local characteristics of water diffusion allows acquisition of in-vivo images of biological tissues. Diffusion weighted MRI can provide information about damage to parts of the nervous system.

2.4.1 MRI artefacts

The acquired image is altered due to unwanted effects of the MRI, known as artefacts. Since our goal is to manipulate anatomical MR images for extracting information, the role of image artefacts must be clearly understood in order to be able to

develop robust algorithms. Artefacts affect the accuracy of image processing methods such as registration and segmentation. Therefore, they are discussed in more detail in the following paragraphs.

Noise is seen as an irregular granular pattern on the images degrading image information. The main source of noise in the image is the patient's body (RF emission due to thermal motion) as well as the whole measurement chain of the MR scanner (coils, electronics etc.) also contributes to the noise. This noise corrupts the signal coming from the transverse magnetisation variations of the intentionally excited spins (on the selected slice plane).

Noise in MRI is additive *i. e.* noisy image F is $F(x) = G(x) + N(x)$ for each voxel x and the noise free image G . It follows a Rician distribution. The magnitude of the signal is expressed as [1]:

$$M = \sqrt{(A + n_1)^2 + n_2^2} \quad (2.1)$$

where M is the magnitude image, A is the original noiseless signal level and n_1 and n_2 are uncorrelated Gaussian noise variables with zero mean and equal variance σ_n^2 . The probability density function (PDF) of such an image, given by the Rician distribution as follows:

$$p_M(M|A, \sigma_n) = \frac{M}{\sigma_n^2} e^{-\frac{M^2 + A^2}{2\sigma_n^2}} I_0\left(\frac{AM}{\sigma_n^2}\right) u(M) \quad (2.2)$$

where $I_0(\cdot)$ is the second order Bessel function of the first kind and $u(\cdot)$ the Heaviside step function. However, when the signal to noise ratio (SNR) is higher than 2 dB, this noise can be assumed to be distributed according to a Gaussian distribution [49].

A real T1-weighted MR image and the distribution of the noise in its background are illustrated in Figure 2.8(a-b).

RF inhomogeneity or bias-field inhomogeneity Slowly changing unwanted intensity variations arise as a result of inhomogeneity in the uniform magnetic field, inhomogeneity in the applied RF pulse sequence, RF field coil(s) and loading of the coils by the patient. A real T1-weighted MR image having areas of high bias field inhomogeneity is illustrated in Figure 2.8(a).

Generally, bias field is modelled as having a multiplicative effect on the images. The image (F) degraded by the bias field (B) is simulated as $F(x) = G(x).B(x)$ for each voxel x and the bias free image G .

Partial volume artefacts Partial volume effect is caused when a voxel contains a mixture of tissues and therefore possesses a signal that is an average of the signals corresponding to these tissues. Smaller voxel size (higher resolution) is recommended for taking care of this artefact but this is at the cost of a poorer signal to noise ratio in the image and higher acquisition time.

Geometrical distortions They result from errors in gradient field strength and non-linearity of gradient fields in the MR scanner. As a consequence, image deformations other than the actual anatomical changes in the brain and are introduced in the image.

2.4.2 Correction of artefacts

Estimation and eventual correction of artefacts has been an active area of research for many years. Most of the work has been dedicated to estimating noise, bias-field inhomogeneity and geometrical distortions in MR images.

Noise artefact

We briefly list the important methodologies adopted for correction of noise in MRI.

Parameter estimation of Rician distribution [1] The most natural way of correction of noise is by estimating the parameter of the Rician distribution. Complications come from the fact that moments of the distribution in Eq. 2.2 are difficult to calculate. However, the even-order moments can be expressed in simple polynomial forms. For instance, the second order moment is:

$$\mu_2 = E \{M^2\} = A^2 + 2\sigma_n^2 \quad (2.3)$$

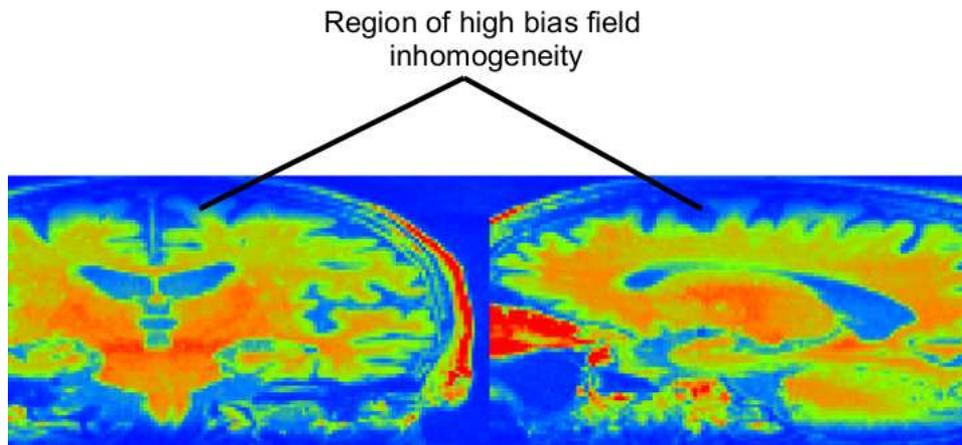
In the background of the image, the signal to noise ratio is zero due to lack of water-proton density in the air, the Rician probability distribution function simplifies to a Rayleigh distribution as follows:

$$p_M(M|\sigma_n) = \frac{M}{\sigma_n^2} e^{-\frac{M^2}{2\sigma_n^2}} u(M) \quad (2.4)$$

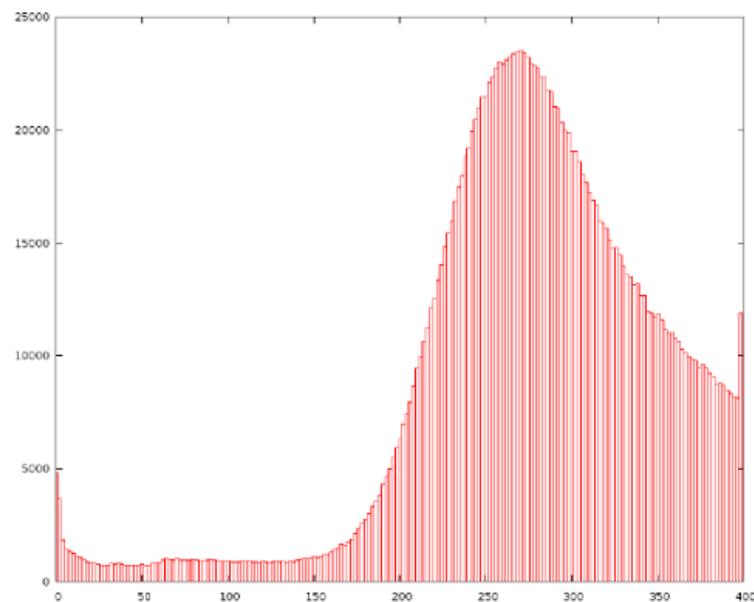
Another method that the authors call the “conventional approach” focuses on the relation between noise and signal of the second order moment in a Rician distribution from Eq. 2.3 and the noise free signal \hat{A} is estimated as

$$\hat{A} = \sqrt{\langle M^2 \rangle - 2\sigma_n^2} \quad (2.5)$$

where $\langle M^2 \rangle$ is the sample second order moment. The estimation of noise from a



(a) A T1-weighted image with visible bias field inhomogeneity



(b) MR noise distribution

Figure 2.8: Sagittal and coronal slices of a T1-weighted image. Notice the prominent bias field inhomogeneity at the top of the head. Noise in the background of an MR image follows a Rayleigh distribution [1]. The distribution of the noise in the background is also shown and resembles a Rayleigh distribution.

single image is performed on the background where the signal is assumed to be zero. An unbiased estimator based on methods of moments is:

$$\hat{\sigma}_n^2 = \frac{1}{2N} \sum_{i=1}^N M_i^2 \quad (2.6)$$

where N is the number of points in the background. This estimator is also the maximum likelihood (ML) estimate of the parameter for the Rayleigh distribution.

Linear minimum mean squared error (LMMSE) estimator The authors in [1] have provided a closed form solution to the LMMSE estimator that filters noise in images that follows a Rician model using even order moments of the Rician distributions. Closed form solutions are less expensive estimation methods as compared to optimisation based methods. The final expression of the LMMSE estimator is:

$$\hat{A}_{ij}^2 = \langle M_{ij}^2 \rangle - 2\sigma_n^2 + K_{ij}(M_{ij}^2 - \langle M_{ij}^2 \rangle) \quad (2.7)$$

with K_{ij} ,

$$K_{ij} = 1 - \frac{4\sigma_n^2 \left(\langle M_{ij}^2 \rangle - 2\sigma_n^2 \right)}{\langle M_{ij}^4 \rangle - \langle M_{ij}^2 \rangle^2} \quad (2.8)$$

where A_{ij} is the unknown intensity value at pixel ij , M_{ij} the observation vector and $\langle \cdot \rangle$ is the LMMSE operator.

The ‘‘conventional’’ approaches (Eq. 2.5), on first order moments Eq. 2.6 and LMMSE (Eq. 2.7) have the disadvantage that a segmentation of the background must be available and estimation is done considering that the signal is always be zero in the background. These requirements make the methods sensitive to errors and artefacts. In order to overcome these limitations, approaches that depend on some sample local statistics of the image may be used. Using the second order moment and the mean, two estimators are developed for images having a background. Finally, another estimator based on variance that can be used on any kind of image following a Rician distribution with no constraint on the background. Using the local images statistics also makes the estimators less dependent on the size of estimation window, uniformity and outliers. For more details and comparison of all these approaches refer to [1].

Non local means (NLM) filter based Another important class of denoising methods is based on NLM filters [32, 167]. In [32], NLM filters are used for correcting Gaussian noise. According to the basic principle of NLM filters, the denoised value $df(x_i)$ at a voxel x_i is as follows:

$$df(x_i) = \sum_{x_j \in \Omega} w(x_i, x_j) f(x_j) \quad (2.9)$$

where $f(x_j)$ is the intensity at voxel x_j and $w(x_i, x_j)$ is the weight assigned to $f(x_j)$ in the restoration of voxel x_i . The function of the weight is to provide a measure of similarities of the local neighbourhoods N_i and N_j of voxels x_i and x_j under the assumptions that $w(x_i, x_j) \in [0, 1]$ and $\sum_{x_j \in \Omega} w(x_i, x_j) = 1$. NLM filter definition states that each voxel can be linked to all other voxels, however, the number of voxels taken into account can be limited to a smaller subset of voxels (search volume V) based on their similarity to each other defined by the weight function, centred at the current voxel x_i .

The weight is calculated as:

$$w(x_i, x_j) = \frac{1}{Z_i} e^{-\frac{\|f(x_i) - f(x_j)\|^2}{h^2}} \quad (2.10)$$

where Z_i is the normalisation constant to make sure that $\sum_{x_j \in \Omega} w(x_i, x_j) = 1$ and h is the smoothing parameter controlling the decay of the exponential function. The parameter h can be calculated automatically.

In a similar way, NLM filters are used for correcting Rician noise in [167] where the corrected intensity $dfR(x_i)$ at voxel x_i is given by the second order moment equation of a Rician distribution (Eq. 2.3):

$$dfR(x_i) = \sqrt{\left(\sum_{x_i \in V} w_i x_i^2 \right) - 2\sigma_n^2} \quad (2.11)$$

where σ_n^2 is the Rician noise parameter. For conventional MRI, the Rician NLM filter outperforms the previous version, the Gaussian NLM filter. Interested reader may refer to [32, 167] for elaborated explanations.

Curious readers may find a survey of some denoising methods in MRI in [61].

Bias-field inhomogeneity

Work dealing with bias-field inhomogeneity correction has been published since the early eighties [38]. Sources related to the properties of the MRI device such as failure of the RF coil, non-uniform B1 field (RF field strength), non-uniform sensitivity of the receive only coil (spaces between wire in the coil, uneven distribution of wire), or presence of non-ferromagnetic material in the imaged object [38] can be corrected for by shimming techniques [82, 28], special imaging sequences and by employing a different sets of coils, or by calibrating the MRI device by a phantom or a mathematical model [97, 154, 156, 166]. Apart from these hardware related inhomogeneities, there are others related to the shape, position and orientation of the object inside the magnet, and to the specific magnetic permeability and dielectric properties of

the imaged object [158]. High magnetic field scanners are more susceptible to artefacts related to the imaged object than low magnetic field scanners.

Automated pre-processing methods also deal with these spurious smoothly varying image intensities. We present a short survey of such methods here.

Approaches that use one image One of the most widely used algorithms for multiplicative bias-field correction is an iterative approach by the name of non-parametric non-uniform intensity normalisation (N3) [137]. N3 makes it possible to estimate the bias-field and the distribution of the true tissue intensities simultaneously. The most important advantage of this algorithm is that no model of the tissue classes present is required. Recently an improved version of N3 (N4ITK) has been published where a fast and robust B-spline approximation routine and a modified hierarchical optimisation scheme has been incorporated for better bias field estimation [155].

Retrospective correction of bias-field inhomogeneity has been presented in a model-based correction method of Meyer *et al.* [91]. Unlike methods that consider only multiplicative bias field, Meyer *et al.* model inhomogeneity as having a multiplicative and an additive component represented by a combination of smoothly varying basis functions.

Styner *et al.* [148] have proposed PABIC (Parametric Bias Correction) where the low-frequency bias field is modelled using Legendre Polynomials. The estimation of bias-field is formulated as a non-linear energy minimisation problem using an evolution strategy. Optimisation is carried out in a way such that the information of the corrected image is minimised while the global intensity statistic is preserved.

In 3-D, the bias-field estimate \hat{B} is derived as follows:

$$\hat{B}(x, p) = \sum_{i=0}^l \sum_{j=0}^{l-i} \sum_{k=0}^{l-i-j} p_{ijk} P_i(x) P_j(x) P_k(x) \quad (2.12)$$

where $P_{i(\cdot)}$ denotes a Legendre polynomial of degree i . The image coordinates \underline{x} are scaled in the range of $[-1, 1]$. For Legendre polynomials of degree l , the size m of the parameter vector p is given by:

$$m = (l + 1) \frac{(l + 2)}{2} \frac{(l + 3)}{3} \quad (2.13)$$

Their results indicate that Legendre polynomials of maximum degree of three can model bias inhomogeneity presented in MRIs with head coils. One of the limitations of this method is that it models distributions of intensities of various tissue classes (GM, WM, CSF) with a Gaussian distribution. The parameters (mean and stan-

dard deviation) of this Gaussian need to be provided that are usually obtained from a segmentation of tissue classes. As a result, this method is susceptible to errors in segmentation.

Manjón *et al.* have devised another automatic non-parametric method of bias correction [88], which is a coarse to fine approach where bias fields are modeled with different frequency ranges. The bias field is modelled as a linear combination of k equidistant low frequency cubic B-Spline basis functions:

$$\beta(\alpha) = \sum_{i=1}^k \alpha_i \phi_i \quad (2.14)$$

where ϕ are low frequency basis functions satisfying the following condition,

$$\sum_{i=1}^k \phi_i(x_m) = 1, \forall m \in [1, M] \quad (2.15)$$

The highlight of this method is the use of a combination of intensity and gradient image features for more robust homogeneity measurement using an entropy-based cost function.

A combination of non-parametric non-uniformity normalisation (N3) algorithm and fuzzy-C-means (FCM)-based inhomogeneity correction algorithm has been developed into a new inhomogeneity correction method in [83].

Hadjidemetriou *et al.* [53] propose a inhomogeneity correction algorithm based on non-parametric statistics of high order local intensity co-occurrences. These statistics are restored with a non-stationary Wiener filter. The use of co-occurrence statistics improve robustness to whole head images for non-uniformities present in high field acquisitions. Results demonstrate that it outperforms N3 algorithm both in terms of performance and lower time requirements.

Approaches that use a template or longitudinal data Studholme *et al.* [146] take a template based approach to bias field correction by means of considering the underlying tissue structure. The intensity distorted MR images are warped to a reference template using a fine-scale entropy based registration method. Thus, the relative bias field between template and subject MR images can be computed from the ratio of their low-pass filtered intensity values.

A completely different approach dedicated to correcting differential bias in longitudinal MR images is discussed in [80] where no assumptions are made about signal distribution, bias field or signal homogeneity. The basis of this method is that the difference of registered serial scans has small scale structure whereas bias field by

nature is a low frequency signal. Thus, differential bias field could be estimated by applying an appropriately sized filter to the difference image.

Mathematically, for serially acquired images f_i , $i = 1, 2$, we have

$$\begin{aligned} \log(f_1(x)) - \log(f_2(x)) &= \log(u_1(x)) - \log(u_2(x)) + \log(b_1(x)) - \log(b_2(x)) \\ &+ \log\left[1 + \frac{n_1(x)}{u_1(x)b_1(x)}\right] + \log\left[1 + \frac{n_2(x)}{u_2(x)b_2(x)}\right] \end{aligned} \quad (2.16)$$

where $u_i(x)$ are the true images, $b_i(x)$ and the $n_i(x)$ are the bias fields and noise in the two images, $\log\left[1 + \frac{n_i(x)}{u_i(x)b_i(x)}\right]$ is the original additive noise of the system for $i = 1, 2$ at voxel x . The difference $\log(b_1(x)) - \log(b_2(x))$ gives the differential bias field in the two images.

Simultaneous estimation of bias field and segmentation/registration The problem of inhomogeneity correction has also been studied in conjunction with segmentation or registration in order to reduce registration or segmentation errors. Such approaches try to simultaneously rectify intensity inhomogeneity and perform tissue segmentation [81, 19], registration [94] or both segmentation and registration [6].

Ashburner *et al.* [6] present a probabilistic framework that combines image registration, tissue classification, and bias correction within the same generative model called Unified segmentation. The tissue classification is modelled as a mixture of Gaussians. Extra parameters are used in the mixture of Gaussian formulation that account for the smooth intensity variations caused by bias field inhomogeneity. This model parametrises bias as the exponential of a linear combination of a small number of low frequency basis functions. In order to include registration to a standard space in the same generative model, the tissue probability maps are allowed to be deformed according to a parametric model that models the deformations by a linear combination of about a thousand cosine transform bases. The objective of this work is to explain how tissue segmentation, registration and bias correction can be included in the same generative model and this framework can be extended to a more sophisticated implementation. This algorithm is also a part of Statistical Parametric Mapping (SPM) 8 [5, 6] and is used for brain volume estimation by adding segmented GM and WM volumes.

Modat *et al.* [94] perform differential bias field correction within a non-rigid registration framework. The spatial transformation and differential bias correction parameters are optimised simultaneously using normalised mutual information as a metric. The bias field is parametrised using a second regular lattice of control points overlaid on the reference image.

A review of bias field estimation procedures is presented in [158]. We would like

to mention that due to rapidly evolving MRI technology, the problem of inhomogeneity correction is not a completely solved problem.

Geometrical distortions

Until now, only few studies have attempted the correction of geometrical distortions arising from gradient non-linearity. Janke *et al.* [64] have presented the problem of distortion correction by expressing the gradient field as expansion of spherical harmonics as the basis function. A phantom is used in order to determine the coefficients using an iterative procedure.

The field B_V generated by a gradient field ($V = X, Y, \text{ or } Z$) can be written in spherical coordinates as:

$$B_{V(n,m)}(r, \phi, \theta) = r^n [a_{V(n,m)} \cos(m\phi) + b_{V(n,m)} \sin(m\phi)] P_{(n,m)}(\cos\theta) \quad (2.17)$$

where $B_{V(n,m)}(r, \phi, \theta)$ is a spherical harmonics expansion of order n and degree m of each component of the gradient field, $a_{V(n,m)}$ and $b_{V(n,m)}$ are constants and r is the radial distance from the magnet isocenter. The corresponding Legendre functions are $P_{(n,m)}(\cos)$. Eq. 2.17 is only an approximation of the true gradient field B_V .

Jovicich *et al.* [67] have verified the accuracy of the method of Janke *et al.* using phantom data and have ascertained the improved quality intensity reproducibility of brain data. They have also shown that the amount of distortion measured within a typical field of view (FOV) required for head imaging is sufficiently significant necessitating distortion correction in images especially those used for longitudinal studies.

The work of Wang *et al.* [160] has shown that the main source of geometrical distortion is gradient non-linearity. Maximum absolute geometric errors ranged between 10 and 25 mm within a volume of $240 \times 240 \times 240 \text{ mm}^3$ are observed, when imaging with the new generation of gradient systems that employ shorter coils, more than the old generation of scanners. The vendor's correction method can successfully reduce the geometric distortion measured but only within the plane in which these 2D correction methods are applied. Distortion along the axis normal to the plane was, as expected, virtually unchanged. If the distortion along the normal of the correction plane is not small, 2D and vendor's correction are inefficient. In the sequel to this work, Wang *et al.* [161] have implemented a 3D method based on the phantom-mapped geometric distortion data, with or without vendor's correction. When the vendor's 2D correction applied, the method corrects for both the "residual" geometric distortion left in the plane in which the correction method is applied (the axial plane) and the uncorrected geometric distortion along the axis normal to the plane.

Caramanos *et al.* [23] have established that inconsistent positioning of sub-

jects in the MRI along the magnet’s long axis (*i.e.* Z) and the gradient distortion effects associated with such Z-shifts can affect the accuracy and precision of MRI-derived measures of whole-brain atrophy. The error caused by geometric distortions increases in magnitude with the increase in Z-distances between the brains to be compared or between the magnet isocenter and the center of the pair of brains to be compared. These distortions can be corrected by accurate subject positioning when acquiring images or with the use of gradient-distortion correction fields post image acquisition. A novel phantom made using DUPLO plastic bricks (a toy for children) is used for learning gradient distortions which are modelled using spherical-harmonics. These corrections are applied to SIENA measurements for neutralising observed Z-shift-associated gradient distortion effects.

Finally, this set-up is also used for simulating gradient distortion effects for further testing. Synthetic Z-shifts of 50 mm are created by moving the scanner bed out by 50 mm. This resulted in a mean (range) MRI-measured Z-shift of -49.2 mm (-50.9 to 48.4) relative to each subjects’ CM (typical canthomeatal alignment) scan. Canthomeatal line is the reference line for correct head positioning. CM alignment results in an individual’s cerebrum being centred several centimetres further into the magnet’s isocenter. Other acquisitions are performed by trying to reposition the subject in relative to the CM alignment (accurate-as-possible repositioning). In spite of trying to reposition the image in the same place, an average MRI-measured Z-shift of 4.3 mm (-9.0 to 21.1) is observed, indicating that precise and accurate positioning and repositioning is difficult to achieve. Results with SIENA show that Z-shifts of 50mm in PBVC values with a significantly higher mean absolute error of about 0.40% (versus 0.17% for the accurate-as-possible repositioning), paired-samples t-test with 8 degrees of freedom $t = -2.96$, $p = 0.018$ and a maximum absolute error of 0.81%.

The key point for geometrical distortion correction is that when the amount of distortion measured within a typical field of view required for head imaging is sufficiently large then, not correcting for these distortions would render images to be of limited use for stereotaxy or longitudinal studies for obtaining reliable atrophy measurements.

2.5 Brain atrophy estimation with MRI

MRI acquisitions coupled with sophisticated image processing techniques have the power of providing ways of measuring brain volume loss. Several methods manipulating MR images have been developed for estimating brain atrophy in the past, each employing different methodologies and focused on different aspects of atrophy.

2.5.1 Classification of existing atrophy estimation approaches

The currently available image processing methods for brain atrophy estimation have been differentiated based on criteria determined by dependence on registration or segmentation [106], type of analysis they aid in conducting and whether they output a global or a local measurement of atrophy.

Use of registration or segmentation

Atrophy estimation and image registration Image registration is the process of geometrically aligning two (or more) images (generally called reference and floating images) in order to facilitate their comparison. The process of registration produces a transformation that maps the floating image into the coordinate system of the reference image. Registration is a step of primary importance when evaluating atrophy evolution in serially acquired scans of the same individual (longitudinal studies), in comparing an individual to an atlas (cross-sectional studies) or using different imaging modalities. From the point of view of atrophy estimation, the type of registration algorithm chosen depends on the type of alignment desired: rigid, affine or deformable.

Registration methods that estimate rigid (a linear transformation model that includes rotations, translations) or affine (a linear transformation that includes rotation, scaling and shear followed by a translation) transformations align the images globally as they compute a common set of transformation parameters for the each voxel. Atrophy estimation methods that do not perform a voxel based or regional analysis may use rigid or affine transformations for comparisons. For instance, whole brain atrophy measurements based on the movement of boundaries over serial scans of the same subject, matching a subject to an atlas for normalising for head size or correcting for geometrical distortions using the skull.

Deformable registration methods are also called non-linear or non-rigid registration approaches. Such registration approaches result in minimising local differences between the target and the floating images. A variety of transformation models exist and their use depends on the material being modelled. These could be physical models such as linear elasticity, fluid flow transformations or free form deformations such as B-splines. Linear elastic transformations are based on the assumption of a linear stress-strain relationship. However, many biological materials have a non-linear stress-strain relationship. Further approximations make it accurate only for small deformations [56]. Fluid flow transformations make it possible to model large deformations such as those arising in inter-subject registration. They are based on idealised physical properties of liquids and must satisfy physical laws such as conservation of mass, energy, and linear and angular momentum. Physical models preserve topology (i.e. do not allow tearing or folding) in the estimated deformation. Free form deformations (FFD) model the transformations using polynomial basis functions such as B-splines that allow the modelling of any 3-D object in space rather

than a surface. A disadvantage of employing FFDs is non-conservation of topology. Nevertheless, in a recent study a topology preserving B-spline transformation model is presented, where topology preservation is achieved by adding explicit constraints on the Jacobian [98].

Usually, a combination of affine and non-linear registrations are applied to images for correcting both global and local differences. Local registration methods provides us with a voxel-wise atrophy map between the floating and the target image, thus local atrophy evolution can be studied. An atrophy map can be obtained from the Jacobian determinant of the local transformation. It gives us the volume expansion or shrink factor at each voxel. Mathematically, the Jacobian determinant (or simply the Jacobian) $J(x, y, z)$ of a transformation $T(x, y, z) = (t_x, t_y, t_z)$ at a voxel (x, y, z) can be written as:

$$J(x, y, z) = \begin{vmatrix} \frac{\partial t_x}{\partial x} & \frac{\partial t_x}{\partial y} & \frac{\partial t_x}{\partial z} \\ \frac{\partial t_y}{\partial x} & \frac{\partial t_y}{\partial y} & \frac{\partial t_y}{\partial z} \\ \frac{\partial t_z}{\partial x} & \frac{\partial t_z}{\partial y} & \frac{\partial t_z}{\partial z} \end{vmatrix} \quad (2.18)$$

Registration-based methods include atrophy estimation by brain edge motion analysis (“Boundary Shift Integral” (BSI) [44], “Structural Image Evaluation, using Normalisation, of Atrophy” (SIENA) [140, 143]), Voxel-Based Morphometry (VBM) (“Statistical Parametric Mapping” (SPM) [5]), Template-Driven Segmentation [51] and Local Jacobian Analysis [18, 21, 20, 171, 170, 59]. Details of some of these methods will be provided in section 2.5.2.

A review of registration methods in medical imaging can be found in [86, 176, 56].

Atrophy estimation and image segmentation Image segmentation refers to the partitioning of an image into multiple classes. It often results in a simpler representation of the image due to grouping of those areas that are similar according to some criterion. For example, a brain image is often segmented into GM, WM and CSF.

In atrophy estimations, usually the total brain volume is computed by totalling the volume of GM and WM. Segmentation is also performed when evaluating atrophies in regions of interest (ROI) such as the ventricles or the hippocampus.

Segmentation based methods rely on different brain atrophy measurements like the “Brain Parenchymal Fraction” (BPF) [57, 132] and “Brain to IntraCranial Capacity Ratio” (BICCR) [31]. Details of some of these methods will be provided in section 2.5.2.

The reader may find a survey of existing automated and semi-automated image segmentation methods in [108, 11].

Longitudinal or cross-sectional analysis

When using brain atrophy as a marker of disease evolution, two types of analyses are conducted: longitudinal and cross-sectional (illustrated in a general manner in Fig. 2.9). 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 [44] and SIENA [140]. 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 [143] is a popular method utilised by cross-sectional studies.

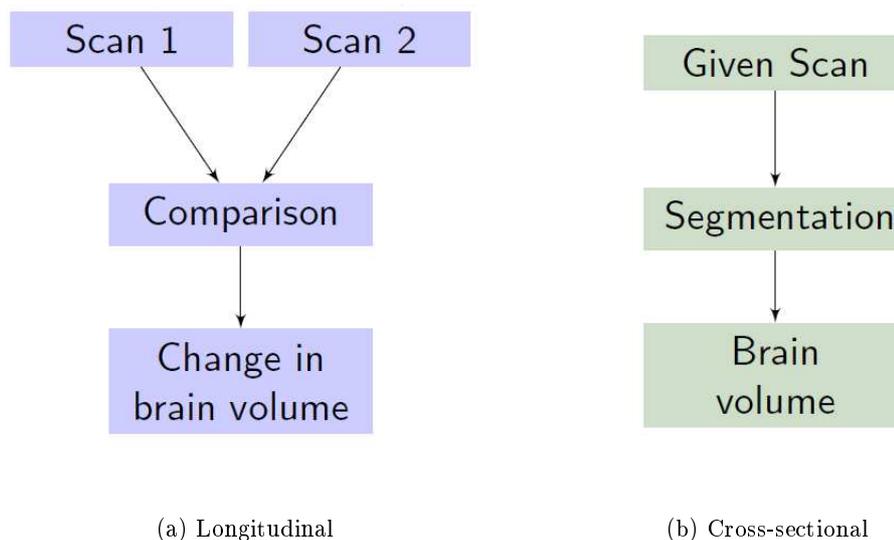


Figure 2.9: Types of analyses. (a) Longitudinal analysis: Serial scans of the same subject are analysed. (b) Cross-sectional: Analysis using one scan of a subject.

Locally or globally estimated atrophy

A global atrophy estimation method, such as BSI [44] or SIENA [140], gives an estimate of overall decrease or rate of decrease in the brain volume. Global measures are used in cases when one just wants to know whether the brain has undergone an atrophy globally. In MS, whole brain atrophies are sometimes utilised as biomarkers of disease progression [132]. Local measurements (Tensor Based Morphometry [18, 21, 20, 171, 59], SPM [5]) 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.

2.5.2 Atrophy estimation approaches

In this section, we briefly describe some widely used measures of brain atrophy. At this stage, the goal is to make the reader familiar with the existing methods for measuring atrophy. An evaluation of some of these methods is presented in chapter 3. We classify the methods on the basis on the analysis (longitudinal or cross-sectional) that can be performed with them.

Methods for longitudinal analysis

SIENA is a widely used automated tool that estimates longitudinal brain atrophy [140, 143]. This algorithm begins by extracting the brain and skull from the baseline and repeat scans using the Brain Extraction Tool (BET) [139]. The brains are then registered using an affine transformation while constraining the scaling with skull images. The affine transformation, thus obtained, is first divided into two halfway transformations that are then applied to the brain scans to resample 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 [175]. At this step, bias field correction is also applied. Next, a 1D array is filled with intensity values from a profile perpendicular to the edge of the baseline image. Using tri-linear interpolation, these values are sampled at sub-voxel positions as the array's elements will not in general fall exactly at voxel grid positions. The number of elements in this array is predetermined (± 3 mm from the edge), but may also change from region to region depending on the presence of another edge. Another 1D array is similarly constructed with intensity values of the second scan (that is registered into a common space as scan 1) from the same voxel positions as in the baseline image. Edge motion is now estimated by finding the relative shift, between the arrays, which produces the maximum correlation. The direction of the edge determines whether it has been affected with an atrophy or hypertrophy. By fitting a quadratic through the correlation values at the peak and its two neighbours, the position of optimal displacement is estimated to sub-voxel accuracy. The estimation of edge motion is pictorially described in Fig. 2.10. In order to make the algorithm robust, correlation is computed between smoothed derivatives of the 1D profiles instead of using the intensity values directly.

SIENA provides whole-brain atrophy as a single number called the percentage brain volume change (PBVC) as mean perpendicular brain edge motion, which is calculated according to the following formulae:

The mean surface motion l is given by,

$$l = \frac{v \sum m}{aN} \quad (2.19)$$

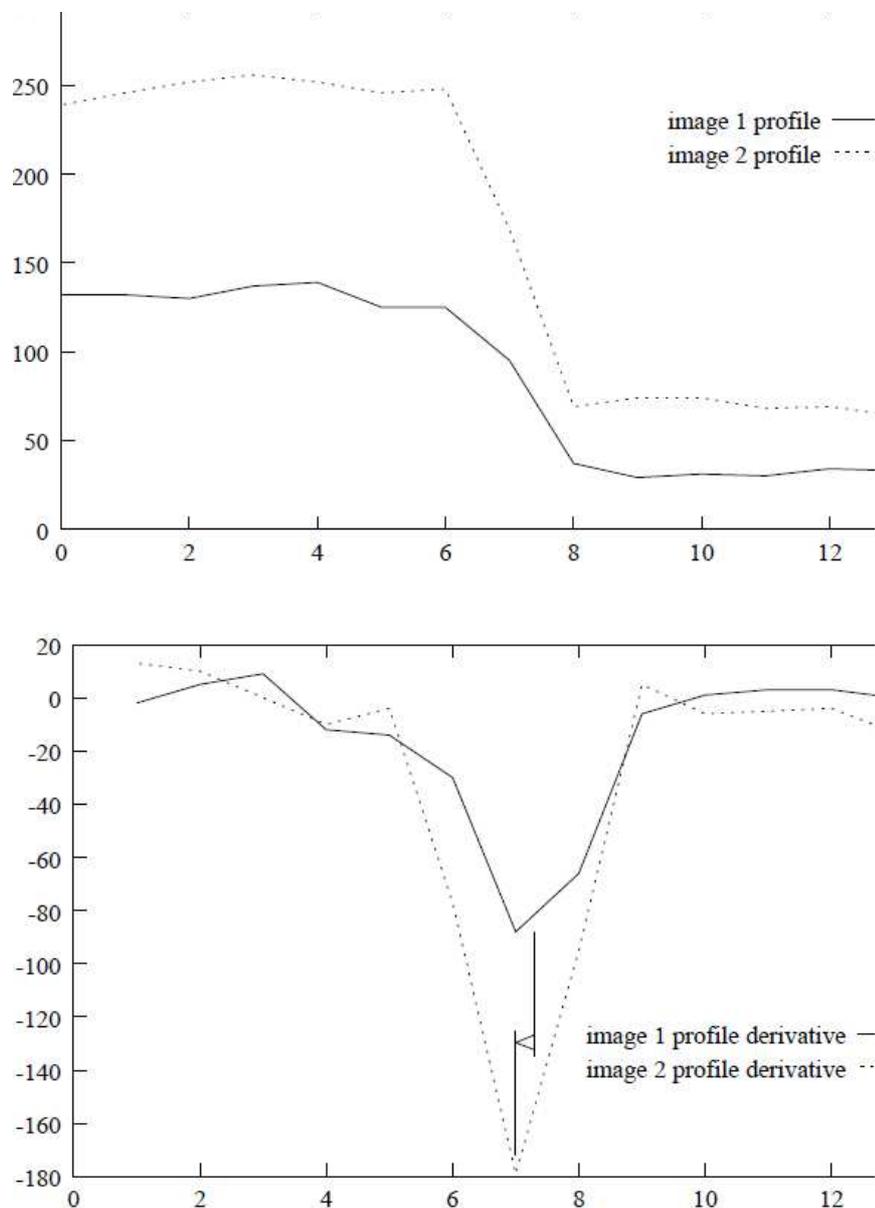


Figure 2.10: SIENA: edge motion detection. The x-axis and y-axis represent the motion in millimetres and image intensity (top image) or image intensity derivatives (bottom image). The intensity profiles of an edge in serially acquired images 1 and 2 and their derivatives. By fitting a quadratic through the correlation values at the peak and its two neighbours, the position of optimal displacement is estimated to sub-voxel accuracy. (Source: [141]).

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

$$\text{PBVC} = \frac{100lA}{V} = \frac{100lfV}{V} = 100lf \quad (2.20)$$

where A is the actual brain area (different from aN), V is the brain volume and f is the ratio of actual area to volume. A and V need not be known and f can be directly obtained by scaling a single image by a known amount and then comparing with the unscaled version using the complete SIENA algorithm. The correct PBVC is known from the scaling that was applied and from the measurement of l , f can be calculated. It varies across scanners, slice thicknesses and pulse sequence, but normally lies between 0.1 and 0.2mm^{-1} . This method (called self-calibration) helps reduce bias (systematic error) in the reported estimates of PBVC.

SIENA is a fully automated method that makes longitudinal analysis possible. Smith *et al.* [141] report the accuracy of SIENA (median brain volume change error) to be 0.15% based on scan-rescan atrophy estimations. In addition, the error introduced by skull-based registration is comparable to total error, meaning that the skull-based registration on average may be a major contributor to the overall error. The key step of this algorithm is to estimate the displacement of the edges between the scans; which depends on the segmentation and registration algorithm. Errors in segmentation and registration of the edge points thus affect the quality of the results obtained using SIENA. Figure 2.11 depicts the edge motion of boundaries as estimated by SIENA.

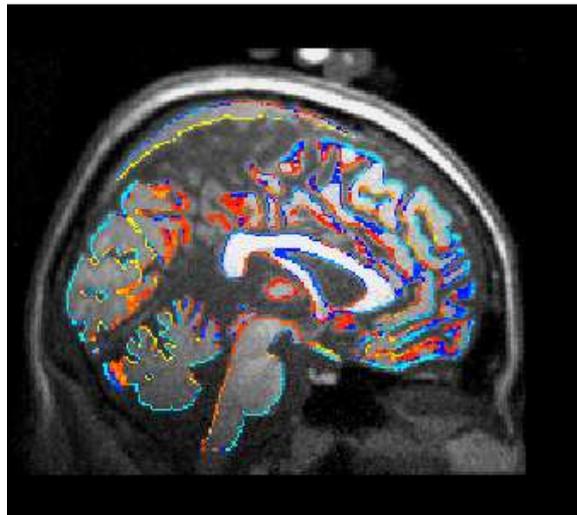


Figure 2.11: Brain edge movement as estimated by SIENA. A hot-cold map represents the movement of edges (hot colours indicate atrophy while cold colours represent growth).

The authors in [13] extend SIENA for voxel-wise analysis of atrophy across subjects. The edge displacement map between the time points is first estimated for each subject which is warped to align with a standard-space edge image and then carries out many voxel-wise cross-subject statistical analysis. For example, this can be used for locating brain edge points that are significantly atrophic for the group of subjects as a whole or for correlating factors such as age or disease progression to atrophy.

BSI Another popular tool for estimating cerebral atrophy from two time points is boundary shift integral [44]. In this approach, the repeat scan and the baseline scan are registered using an affine transformation. In the original BSI implementation [44], 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 normalised 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 BSI 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 according to Eq. 2.21.

$$\text{BSI} = \frac{K}{I_1 - I_2} \sum_{x,y,z \in E} (\text{clip}(i_{base}(x, y, z), I_1, I_2) - \text{clip}(i_{reg}(x, y, z), I_1, I_2)) \quad (2.21)$$

where K is the voxel volume, E is the set of voxels in the boundary region and $i_{base}(x, y, z)$ and $i_{reg}(x, y, z)$ are the normalised voxel intensities of the registered serial scans. $[I_2, I_1]$ is the intensity window in which BSI is calculated. The clip function for an intensity a is defined as:

$$\text{clip}(a, I_1, I_2) = \begin{cases} I_2 & a < I_2 \\ a & I_1 < a < I_2 \\ I_1 & a > I_1 \end{cases}$$

This computation is pictorially shown in Figure 2.12.

BSI is a popularly used semi-automated tool for estimating longitudinal atrophy. Freebrough *et al.* [44] find that brain and ventricular BSIs correlated tightly with correlation coefficients of $r = 1.000$ and $r = 0.999$, respectively, with simulated volumetric changes. Atrophy is introduced in a given structure by spatially scaling the scan about the central point of the structure by equal amounts in each dimension. The comparison of intensity values directly makes BSI measurements vulnerable to intensity variations in images caused by bias field inhomogeneities and noise. For the same reason, normalisation of intensities between the serial scans forms an important step.

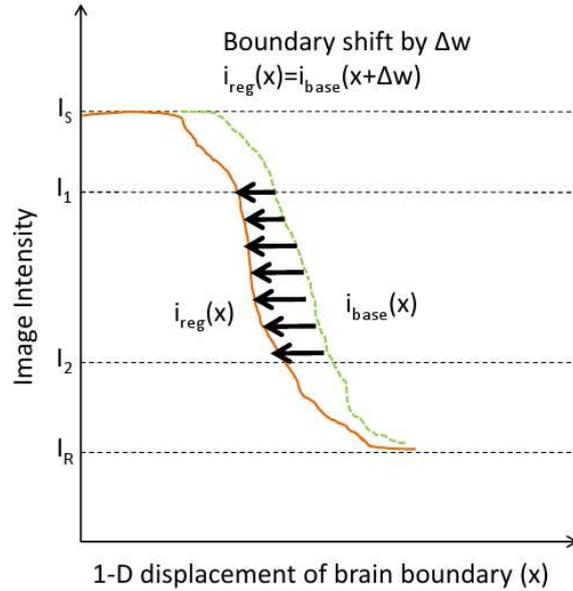


Figure 2.12: Boundary shift integral calculation. A 1-D representation of a boundary shift between a baseline scan, $i_{base}(x)$, and a registered serial scan, $i_{reg}(x)$. This shift Δw is the area A divided by the length of the intensity window $I_2 - I_1$.

Leung *et al.* [79] seek an improvement on the classic BSI algorithm. Their method (KN-BSI) performs tissue-specific intensity normalisation and parameter selection in order to obtain better performance. They introduce some new steps. The baseline and repeat brain regions are dilated by 3 voxels to include enough voxels in CSF to estimate the mean CSF intensity. k-means clustering is performed with 3 clusters (CSF, GM and WM) on the bias-corrected baseline and repeat scans inside the dilated regions. Linear regression coefficients are calculated using the mean intensities of CSF, GM, WM, and the interior brain region and are further applied to the images for normalising their intensities. From results of k-mean clustering, the intensity windows, which define the limits of BSI, are calculated. The estimated KN-BSI atrophy rates, on the ADNI database, are 0.09% higher than classic BSI rates in controls and 0.07% higher in ADs. The standard deviation of the KN-BSI rates is 22% lower in controls and 13% lower in ADs, compared to classic BSI algorithm. Results confirm that KN-BSI is more robust and reduces sample sizes by 32% (to detect an atrophy of 25% at 5% significance) needed in clinical trials as compared to classic BSI.

Cortical longitudinal atrophy detection algorithm (CLADA) This method addresses the problem of measuring changes in cortical thickness over time. CLADA [96] creates a subject-specific cortical model which is longitudinally deformed to match images from individual time points. In the first step, all the images are corrected for intensity variations [148, 80], registered to the baseline using a symmetric

affine registration algorithm followed by voxel-wise image averaging for producing a bias free image with improved image quality for each subject. The resulting subject specific average image is segmented into GM, WM, and CSF and a cortical GM-WM boundary is generated through a combination of segmentation and atlas-based classification [33]. Two explicit surfaces for the inner cortical surface (ICS *i.e.* the surface between WM and GM) and the outer cortical surface (OCS *i.e.* the surface between CSF and GM) are constructed. Finally, this subject-specific cortical model is deformed to fit the images from individual time points. Transforming the cortical model is preferred to transforming the image for preventing interpolation artefacts. Cortical thickness is determined for each surface vertex in the follow-up image as the distance between corresponding vertices on the ICS and OCS surfaces. The global cortical thickness is estimated as the surface-area-weighted average. Results on MS patients show a higher rate of cortical thinning in MS patients compared to healthy controls over 2 years and show that CLADA can measure cortical atrophy in longitudinal studies reliably. The annual global cortical thinning is found to be significantly higher in the MS group -0.025 mm/year ($-0.71 \pm 0.91\%$) as compared to the control group (-0.011 mm/year, or $-0.30 \pm 0.52\%$ /year; $p = 0.018$), indicating a higher rate of cortical thinning in the MS patients over the two years of follow-up. Also, the estimation of bias using the scan-rescan datasets was not significantly different from zero. Using scan-rescan tests, the estimated change is 0.45% for images with 1 mm^3 isotropic voxels and 0.77% for images with $1 \times 1 \times 5 \text{ mm}^3$ voxels. The mean absolute accuracy error is 0.43 mm (determined by comparison of CLADA measurements to cortical thickness measured directly in post-mortem tissue).

Sparse unbiased analysis of anatomical variance (SUA AV) Longitudinal analysis of cortical and WM atrophy rates are studied with penalised statistical methods by Avants *et al.* [9]. In this approach, there is a group template that contains cortical priors and labels that can be used in a parcellation scheme. The priors of the group template are warped to the “Single Subject Template” (SST) in order to initialise the segmentation. An SST is created from each subject’s serial data by rigidly aligning the first time point image to the group template. The probability maps are then deformed to the individual space and initialise a prior constrained segmentation of the baseline and the repeat scan. Thickness maps and cortical parcellation are then calculated for each time point. Parcellation is used for verifying the findings of the voxel-wise analysis. Test-retest data shows that this approach produces no change when the retest data contains the same image content as the test data and produces normally distributed low variance estimates of thickness change centred at zero when test-retest data is collected within a short interval when no change is actually expected. The standard deviation of changes measured is 0.1 mm across these images and is a measure of reproducibility for images of resolution approximately 1 mm^3 . The key highlight of this method is the use of subject specific priors that provide unbiased, prior-based segmentation and

measurement of cortical thickness.

Iterative principal component analysis (IPCA) In the atrophy estimation method of Chen *et al.* [27] iterative principal component analysis (IPCA) is carried out in order to compute changes in brain volume from sequential MR images for characterising whole-brain atrophy rates in patients with Alzheimer’s disease (AD). The idea is to first register the baseline and the repeat scan and identify those pairs of intensities voxel-wise that are sufficiently distant from the iteratively determined PCA major axis. An outlier distance threshold, depending on the required sensitivity and specificity in the detection of small volume changes while calculating global intensity changes is determined. The IPCA is evaluated with coregistered MRIs that are acquired 30-min apart in a normal subject, simulated regions-of-atrophy (ROA) of different sizes, ranging from 0.04% to 5.5% of whole-brain volume, into the follow-up image, two different levels of noise (50% or 100% of standard deviation of the intensities) in the ROA. ROAs are defined by identifying GM from fuzzy segmentations of the brain and, alternatively as the group of voxels representing the outer and inner GM boundaries or a manually defined brain region. The mean errors are found to be in the range of 0.125 – 4.6%.

Inverse problem approach An inverse problem approach is employed by Schweiger *et al.* [129] for studying atrophy of various structures of interest in the brain. A pre-labelled brain and serially acquired images of the same patient are used for generating finite element meshes from the image data. In this formulation, volume change is modelled using an unknown coefficient of expansion for each structure. It is an inverse problem because the coefficients of expansion for each structure are recovered from the observed volume changes in the mesh. For the purpose of reconstruction, the search space is restricted by assuming that the expansion coefficients be homogeneous or piecewise homogeneous within each of a set of anatomical structures of the brain. Thus, the problem is to find one expansion coefficient per segmented structure. Let m be the number of regions, and n the number of mesh nodes. The expansion coefficients $\alpha = \alpha(i) \in \mathbb{R}^m$, are computed through an iterative Levenberg-Marquardt solver as follows:

$$\alpha_{k+1} = \alpha_k + (J_k^T J_k + \eta I)^{-1} (J_k^T (a_k - a_{tgt})) \quad (2.22)$$

where a_k and a_{tgt} are the nodal positions of the deformed and the target mesh, respectively, and $J \in \mathbb{R}^{n \times m}$ is the Jacobian matrix $J_{ij} = \frac{\partial u_i}{\partial \alpha_j}$, and η is a trust region control parameter. Because m is small, J can be calculated by explicit perturbation of the region coefficients.

Atrophy estimation on 9 AD patients for a low-dimensional solution space of 15 regions of interest show that the reconstruction results capture the trend of atrophy in GM, WM, CSF and background but the absolute values of atrophy are not in accord. The salient feature of this algorithm is that it directly estimates the un-

known volume changes anatomically as opposed to providing a voxel-wise atrophy measure. By increasing the number of regions, that is to say, the dimension of α the accuracy could be improved. Also, the regions can be chosen according to the disease being investigated.

Methods for cross-sectional analysis

SIENAX [143] 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 [142]. 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 normalisation at the final stage of the algorithm. Next, tissue type segmentation [175] is performed on the original (unregistered) extracted brain. The segmentation method includes estimation of partial volume effects for edge voxels, giving higher volumetric accuracy than a hard segmentation. The total brain volume is derived from this segmentation result. To normalise for head size, the brain volume is multiplied with the volumetric scaling factor derived earlier.

Although SIENAX has been primarily developed for cross-sectional studies, through comparison of brain volumes of serial scans of a patient, it can be employed for studying longitudinal atrophy also. Brain volume can be estimated for each of the two given scans and then the percentage decrease in brain volume from time point one to two can be calculated. Tissue classification is performed on the brain making brain extraction an important step in its analysis. 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. In order to calculate the brain volume, number of voxels inside this segmentation are counted and multiplied by voxel volume.

Smith *et al.* [141] report an accuracy of 0.5 – 1% brain volume accuracy for SIENAX.

Methods for longitudinal and/or cross-sectional analysis

Voxel [5] and tensor based morphometry [45, 18, 21, 103, 20, 171, 109, 59, 162] VBM [5] is dedicated to analysis of local differences in the brain using a statistical approach. The procedure involves normalising high-resolution images from all the subjects under study into the same space, followed by segmentation of the GM from the spatially normalised images and smoothing the gray-matter

segments. The general linear model (GLM) provides a framework for application of statistical tests for identifying regions of gray matter concentration that are significantly related to parameters of the study such as group comparisons and identifying regions of grey matter that are significantly related to covariates such as age or disease severity. GLM is a linear statistical model [89],

$$\mathbf{Y} = \mathbf{XB} + \mathbf{U}, \quad (2.23)$$

where \mathbf{Y} is a matrix with series of multivariate measurements such as different subject brain scans, \mathbf{X} contains experimental design variables, \mathbf{B} is a matrix containing parameters that are usually to be estimated and \mathbf{U} is a matrix containing errors or noise. The errors are usually assumed to follow a multivariate normal distribution.

The smoothed gray-matter images are compared with voxel-wise parametric statistical tests, although non-parametric testing can also be applied such as in [122]. The underlying philosophy of this approach is that macroscopic volume and shape differences can be measured by means of employing spatial normalisation. However, mesoscopic volume differences cannot be accounted for, even after spatial normalisation has been performed. Due to the partial volume effect, smoothing reflects these small differences in the form of intensity differences.

In a study, receiver operating characteristic curves for a z-score in the bilateral medial temporal areas including the entorhinal cortex showed a high discrimination accuracy of 87.8% between controls and AD patients [54]. Another study [62] that performed z-score based analysis showed the sensitivity and specificity between AD and Vascular dementia are 92% and 80%, respectively.

Methods based on non-rigid registration constitute another important set. Maps of volumetric change can be obtained in the form of scalar Jacobian maps of the deformation field, representing the changes between serially acquired brain images. Analysis of maps such as Jacobian maps on a voxel-wise or regional basis is called tensor based morphometry (TBM). Usually TBM is performed on anatomical images, Pagani *et al.* [103] have extended TBM for diffusion tensor MRI by using diffusion anisotropy images. The use of diffusion anisotropy images makes the analysis sensitive to changes in the major white matter fibre bundles. Instead of obtaining a scalar map such as a Jacobian map, if statistics are performed on the deformation field directly, it is known as deformation based morphometry (DBM) or deformation field morphometry (DFM).

In TBM/DBM, analysis can be done in various ways. For example, Gaser *et al.* [45] divide subjects into two groups based on their mean ventricular/brain ratios and compute statistical maps of displacement vectors and their spatial derivatives in the GLM framework. Wang *et al.* [162] define a new multivariate surface morphometry statistic as a 4×1 feature vector containing the logged deformation tensor (3×1) and

the radial distance (a scalar). Radial distance is the distance from each parametric surface point to the center of 3D positions of the iso curves in the parameter domain. Finally, they use the Mahalanobis distance for measuring the difference between the mean vectors of two different groups of subjects. Jacobian integration (JI) is used in many studies which consists of averaging the Jacobian in a ROI [18, 20, 109] for estimating regional atrophy.

No explicit accuracy estimates have been mentioned by Gaser *et al.* and Wang *et al.* while the accuracy of JI has been reported to be in the range of 0.2 – 0.6% in [18, 20, 109].

Voxel or tensor based morphometries provide us with the freedom of obtaining brain atrophy in any area of interest such as the whole brain or voxel-wise. If a ROI based analysis is preferred, a segmentation of the region of interest must be made available in which atrophy is desired. Thus, the segmentation of the ROI will be a critical determinant of their performance. Since these techniques are based on registration, an accurate and precise registration method will vitally increase the accuracy of these methods. Assumptions related to parametric statistical tests should be satisfied, if not non-parametric tests are preferred. Also the sample size determines the performance of VBM / TBM. These techniques can be used in MS for both longitudinal and cross-sectional analysis such as to identify areas of significant volumetric change and lesion growth.

A schematic representation of VBM and TBM can be found in Fig. 2.13.

Other measures

Brain parenchymal fraction (BPF) Atrophy can be estimated as a change in the absolute volume of brain parenchyma, or as a change in a normalised index of brain volume such as the brain parenchymal fraction, BPF [57, 132] (Eq. 2.24). Brain parenchyma is the brain itself excluding its blood vessels, coverings or its support structures. In other words, brain parenchyma is brain's GM and WM. Bias field correction is applied before BPF calculation.

$$BPF = \frac{\text{parenchyma volume}}{\text{parenchyma volume} + CSF} \quad (2.24)$$

Using this approach, an assessment of the degree of atrophy of the brain can be obtained in cross-sectional studies by a single measurement, since the normalisation of the parenchymal volume takes account of absolute brain size. Secondly, any variation in the calibration of the MRI scanner gradient strengths should have little effect on the BPF.

Recently Souplet *et al.* [145] have presented an improved version of BPF for longitudinal studies. An Expectation-Maximisation (EM) segmentation framework

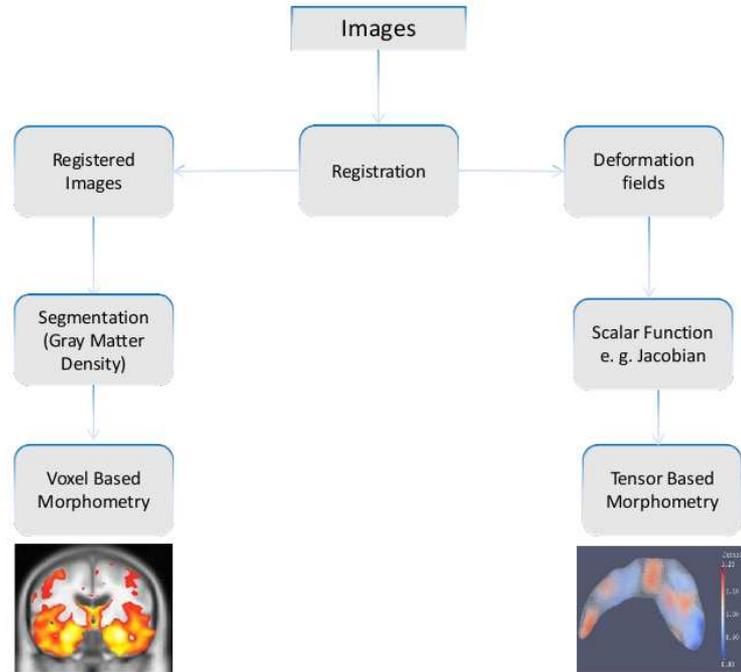


Figure 2.13: Voxel and tensor based morphometry (Source: <http://www.fil.ion.ucl.ac.uk/spm/>).

is proposed taking into account the partial volume effect. For better applicability to longitudinally acquired data, the segmentation parameters are computed from all the time points together and then used for segmenting each time point individually. Experiments on BrainWeb images (with 3% noise, 20% INU (intensity non-uniformity), slice thicknesses: 1 and 3 mm and moderate MS lesions model) exhibited errors in whole brain atrophies of +1% and +5% for 1 mm and 3 mm slice thicknesses, respectively.

Brain to intra-cranial capacity ratio The second ratio, BICCR Eq. 2.25, is calculated from the estimates of intracranial, brain parenchymal and CSF volumes [31]. Lesion volumes are also accounted for.

$$BICCR = \frac{GM + WM + L}{GM + WM + L + CSF} \quad (2.25)$$

where L refers to lesion volume. Each brain voxel is classified as brain tissue, CSF or background. Intensity inhomogeneity and noise correcting pre-processing is applied before this ratio is calculated. As opposed to BPF, normalisation for individual brain size is sought by linearly registering the brain into the standard Talairach space. The main difference between the BICCR and BPF metrics is the inclusion of extra-cerebral CSF in the denominator. Scan-rescan tests on 4 normal controls

show that mean error is 0.21%.

Note that BPF and BICCR are measures of atrophy derived from segmentation and are not methods that provide atrophy measurements exclusively. Some aspects of brain atrophy estimation algorithms discussed in this section with the roles of registration and segmentation steps are summarised in Table 3.6.

2.6 Brain extraction

Accurate dissociation of brain tissue (brain extraction) from non-brain tissue, in the whole head image is a crucial pre-processing step for most of the atrophy measurement tools. Brain extraction is also known as “skull stripping”. Many segmentation algorithms require a brain mask in order to perform tissue classification. Let us briefly go over some brain extraction tools.

FMRI software library’s (FSL) Brain extraction tool (BET) [139] is most extensively used for this purpose. At first, the intensity histogram is analysed to find lower and upper intensity values in an image and a rough brain/non-brain intensity threshold based on the image contrast in the image modality used. In the next step, the centre-of-gravity of the head is calculated in addition to the rough size of the head in the image. Inside the brain, a triangular tessellation of the surface spheres is initialised and deformed vertex by vertex in direction of the brain boundary. This algorithm can also segment the outer surface of the skull. An example of a brain extracted by BET is illustrated in Figure 2.14.

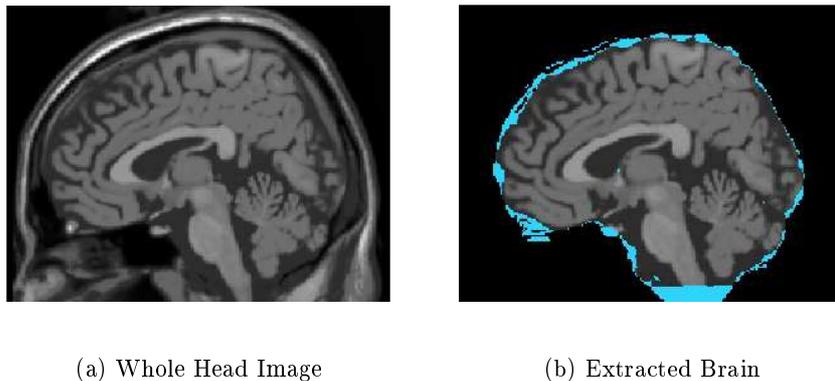


Figure 2.14: Brain Extraction by BET on the simulated BrainWeb image. The areas that BET fails to classify as brain (with respect to the ground truth) are shown in blue. It can be seen that the boundary of the brain is under-estimated by the BET algorithm.

BrainSuite’s brain surface extractor (BSE) [134] Shattuck *et al.*’s brain extraction works in three steps: First, the whole brain MRI is processed with an

Method	Registration	Segmentation	Type of Measurement	Type of Study	Freely Available Implementations	Accuracy
VBM	To register multi-patient scans with an atlas	GM Segmentation to perform voxel-wise statistics	Local	Long./Cross.	Yes ^a	-
TBM	To register serial scans	Segmentation of region of interest	Global/Local	Long./Cross	-	Jl 0.2 – 0.6%
SIENA	To register serial scans	To determine brain boundaries	Global	Long.	Yes ^b	≈ 0.15%
BSI	To register serial scans	To determine brain boundaries	Global	Long.	Yes ^{c d}	≈ 0.2%
CLADA	To register serial scans	Segment subject specific atlas	Local (cortical)	Long.	-	0.45%
SUAAV	To construct SST	Segmentation of WM, cortex	Local (cortical)	Long.	-	0.1 mm
IPCA	To register serial scans	-	Global	Long.	-	0.125 – 4%
SIENAX	To normalise for head size with an atlas	To calculate brain volume	Global	Cross.	Yes ^e	0.5 – 1%
BPF	-	To calculate BPF	Global	Cross.	-	+1%
BICCR	To normalise for head size with an atlas	To calculate BICCR	Global	Cross.	-	0.21%

Table 2.1: Summary of brain atrophy estimation algorithms describing the roles of registration and segmentation steps, type of measurement they provide, type of studies they are utilised in (Long.: longitudinal, Cross.: cross-sectional analysis) and their availability. Note that for BSI implementation details may be different from those described in the original BSI paper [44].

^a<http://www.fil.ion.ucl.ac.uk/spm/>

^b<http://www.fmrib.ox.ac.uk/fsl/siena>

^c<http://ideallab.ucdavis.edu/software/bbsi.php>

^d<http://sourceforge.net/projects/bsintegral/>

^e<http://www.fmrib.ox.ac.uk/fsl/sienax>

anisotropic diffusion filter to smooth out non-essential gradients. Next, a Marr-Hildreth edge detector is applied to the filtered image to extract important anatomical boundaries. Finally, a sequence of morphological and connected component operations identifies and refines the objects defined by these boundaries.

FreeSurfer’s brain extraction [33] This procedure involves deforming a tessellated ellipsoidal template into the shape of the inner surface of the skull of the subject. The deformation process is driven by two kinds of forces, an MRI-based force, that drives the template outwards from the brain and a curvature reducing force, constraining the deformed template to be smooth.

FreeSurfer’s MRI Hybrid Watershed (HWA) [130] Segonne *et al.* [130] have presented a brain extraction that combines watershed algorithms and deformable surface models. It has three steps. First, a single white matter voxel is extracted in a T1-weighted MRI image, for creating a global minimum in the white matter before applying a watershed algorithm. Next, the watershed algorithm builds an initial estimate of the brain volume based on the three-dimensional connectivity of the white matter. A statistical atlas validates and corrects the segmentation, and the MRI intensity values are locally re-estimated at the boundary of the brain. Finally, a high-resolution surface deformation is performed that matches the outer boundary of the brain.

Analysis of functional neuroimages (AFNI) 3d intracranial [163] In this algorithm, the grey, white matter and background voxel intensities are modelled with Gaussian distributions which allows the estimation of lower and upper bounds of grey and white matter intensities. Slice by slice, voxels are classified into the brain if their intensities are within the bounds calculated in the previous step. The set of voxels classified within the brain is a disconnected set (contains holes). By determining the connected components these holes are filled.

Brain extraction meta-algorithm [116] Rex *et al.* [116] have developed brain extraction meta-algorithm (BEMA) that executes several extraction algorithms parallelly (BET [139], BrainSuite’s Brain Surface Extractor [134], Analysis of Functional NeuroImages’ (AFNI) 3d Intracranial [163], FreeSurfer’s brain extraction [33]) and then combines their results in an efficient fashion so as to obtain improved results over any of the individual methods. By performing voxel-wise analysis on training data, BEMA attempts to calculate an optimal boolean combination of results of these four brain extraction algorithms that produces the most accurate output for a voxel.

Other extraction tools include Brainvisa’s¹ [87], SPM’s brain extraction algorithm [5], Minneapolis Consensus Strip (McStrip) [114] and EM based brain extrac-

¹<http://brainvisa.info>

tion [37]. Souplet *et al.* [144] combine the brain extraction results from 5 different algorithm ([87, 139, 163, 130, 37]) using the measures of specificity and sensitivity under the STAPLE framework [164].

Evaluation of some brain extraction algorithms has been performed in [76, 16, 39]. Lee *et al.* [76] found that automated methods (BET (kappa SI: 0.946) and BSE (kappa SI: 0.905)) produce erroneous results, but, at the same time, said that these errors can be corrected by adjusting their parameters. When parameter tuning does not work, manual intervention may be needed to correct the brain mask. Boesen *et al.* [16] demonstrated that McStrip (that is based on warping to a template, intensity thresholding, and edge detection) is better than SPM's brain extraction, BET, and BSE. The Dice similarity indices are McStrip: 0.994, SPM's method: 0.986, BET: 0.976 and BSE: 0.977. Fennema-Notestine *et al.* [39] have concluded that HWA and BSE are more robust across diagnostic groups than 3dIntracranial and BET. Mean Jaccard similarity indices obtained by comparison to manual segmentations are 3dIntracranial: 0.802, BET: 0.787, BSE: 0.863 and HWA: 0.855. Their results indicate that BSE could reach the surface of the brain. However, in some cases some brain tissue may have been removed. 3dIntracranial and BET often could not remove large chunks of non-brain tissue and/or removed some desirable brain regions. Besides, HWA worked well in difficult face and neck regions, carefully preserving the brain, although further stripping of other non-brain regions may have been needed.

Brain extraction is one of the first steps in many atrophy estimation algorithms (such as SIENA, SIENAX, BSI) and determines the region over which atrophy estimations are performed. Hence, their ability to remove non-brain tissue and leave brain tissue determines the accuracy of atrophy estimations. Because brain extraction is based on manipulating image intensities, it is affected by artefacts such as noise and bias field inhomogeneity.

2.7 Key issues in brain atrophy estimation

Brain atrophy is a feature of many neuro-degenerative diseases and is seen fairly early in the disease course. In addition, the phenomenon of atrophy is known to be related to cognitive decline and motor functions. Thus, interesting information can be derived through comparison of atrophies in groups, in serial scans of an individual or by relating atrophies regions in the brain to covariates such as age, disease severity or cognitive decline. However, the changes in the brain volume may be subtle and this requires the use of sophisticated measurement methods that can gauge minute differences in the brain. Methods that manipulate MRI images to extract atrophy information have been developed. Each of them are based on different interpretations of brain atrophy and target different applications. For example, movement of boundaries over time for estimating whole brain atrophy volume change

or non-rigidly aligning serial scans for tracking voxel-wise longitudinal atrophy or segmentation of the brain for cross-sectional studies.

Longitudinal atrophy estimation methods SIENA,[141], BSI [44], TBM [45, 18, 21, 20, 171, 109, 59] and; cross-sectional methods SIENAX [141] and VBM [5] are widely used by the research community.

The critical steps for BSI and SIENA are related to the calculation of movement at boundaries. Due to the use of image intensities directly for calculation of BSI, it is susceptible to MR artefacts such as noise and bias field inhomogeneity. SIENA may be less affected by this as it uses smoothed derivatives of intensities instead of the intensities directly for estimating edge movement. Both SIENA and BSI need aligned serial scans in order to be able to compare corresponding boundaries in the two scans, affine registration is used in both cases. We expect that mis-aligned edges are a source of error. Brain extraction is also a critical step as it is on this area, the movement is computed. Performance of SIENAX is dependent on the accuracies of the tissue classification and the brain extraction algorithms. Inaccurate brain extraction will affect SIENAX more than SIENA, where information from the brain extraction of the two serial scans is utilised for identifying brain tissue. Statistical tests are applied in VBM methods in order to perform voxel based analysis. The validity of assumptions of using parametric tests is very important. For TBM, the important steps are non-rigid alignment of scans. Within the non-rigid registration, the transformation model, regularisation and other constraints must be carefully chosen to suit the requirements. Segmentation of ROI is required in both VBM and TBM methods.

Table 3.6 listed the accuracy of some of these methods to be between 0.15 – 1% while medical research has shown that the annual whole brain atrophies in MS are in the interval of 0.5 – 0.8% [4]. This fact makes the existing atrophy estimation algorithms susceptible to errors comparable to the annual atrophy rates and calls for development of more accurate methods. In addition, a conscientious validation of these methods is an important task in order to ascertain errors in estimations as well as determining optimal values of parameters of these methods. All the methods discussed above provide a point estimate of atrophy, however, given the many sources of error, it would be worthier to provide an interval or reliability estimates. Such reliability estimates, for instance confidence intervals, must represent the reproducibility and bias in atrophy estimations. From the point of view of the end user, accuracy, level of automation, computational complexity, availability and ease of interpretation may be important considerations.

Evaluation of Existing Brain Atrophy Estimation Algorithms

Validation of brain atrophy estimation algorithms is a crucial task as brain atrophy is being increasingly employed as a biomarker of disease progression. Nonetheless, it is challenging to evaluate methods when no atrophy ground truths exist. In this chapter, a framework for assessing three popular freely available brain atrophy estimation methods (BSI-UCD [44], SIENA [143] and SIENAX [141]) is developed. A topology preserving non-rigid registration approach allows us to simulate realistic atrophies in the brain, which serve as the ground truth. The influence of bias field inhomogeneity, noise, geometrical distortions and interpolation artefacts on whole brain atrophy measurements is analysed. Effect of lesion load is also investigated. In another set of experiments, the consequence of anatomical variability on atrophy estimation by simulating atrophies on a cohort of 18 Brainweb images is examined. Lastly, experiments are conducted to compare a deformable registration algorithm (Noblet *et al.* [98]) to SIENA, SIENAX and BSI-UCD.

3.1 Studies on evaluation of brain atrophy estimation algorithms

Until now, few studies have aimed to evaluate existing brain atrophy estimation algorithms. They can be distinguished on the basis of the evaluation criteria (scan-rescan, consistency, patient-control separation and evaluation using a ground truth) that are adopted for determining the accuracy of methods. The most commonly used criteria are enlisted below, followed by a survey of existing works on validation of atrophy estimation methods.

3.1.1 Evaluation criteria

Same day scan-rescan : This measure is based on the fact that tests on two scans of the same subject obtained on the same day should yield zero atrophy. Use of this protocol can help in gauging the influence of non-destructive biological factors and MRI sequence variations on brain atrophy estimation. Let $\{A_{ij}, i \neq j\}$ be the set of brain atrophy estimations between exam i and exam j for a patient taken on the same day. 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 (zero atrophy metric) *i. e.* $E(A_{ij}) \sim 0$ and $Var(A_{ij})$ is the uncertainty.

Consistency : This is also a way of measuring the variability in atrophy estimations arising from non-destructive biologic factors and MRI sequence variations. 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. Mathematically, it involves analysis of the set $\{A_{ir}, i = 1 \dots n\}$ where A_{ir} is the brain atrophy estimation between the i th of the n baseline examinations and 1 year repeat exam r for a patient. We expect that $Var(A_{ir}) \sim 0$.

Another approach is incremental atrophy summation comparison which is often used for validating a method in the absence of a ground truth. This criteria must be interpreted cautiously as a consistent method may not necessarily be accurate. For a patient, $A_{ij} + A_{jl} \sim A_{il}$ is desired from a consistent method where $i = 1 \dots n - 1, j = i + 1 \dots n, l = j + 1 \dots n$ and n is the total number of scans acquired. The variance in these cases represents the uncertainty in the measurements. Smith *et al.* [142] adopt this technique for calculating the overall error in the estimated percentage brain volume change (PBVC) in terms of mean absolute difference. They calculate the mean absolute difference (error) using the following expression: $(A_{1n} - (A_{12} + A_{23} \dots + A_{(n-1)n}))$.

Patient-control separation : A reliable atrophy estimation method is expected to be able to reliably discriminate between the two groups such as that of patients and controls in a statistical analysis. Hence, the degree of patient-control group separation is another way of deciding the efficiency of a measurement. Various metrics can be used for measuring this. For example, Gunter *et al.* [50] use the difference between the group mean rates of atrophy ($mean_{patient}, mean_{control}$) divided by the pooled variance, $\frac{|mean_{patient} - mean_{control}|}{\sqrt{SD_{patient}^2 + SD_{control}^2}}$ where standard deviation (SD) of estimated atrophies of patients and controls. This approach of evaluation of an atrophy estimation measurement makes it possible to perform comparisons when no ground truth but patients and controls are available. 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.

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 biomechanical models [21, 20], Jacobian maps [68, 109] and manual brain segmentations [34]. 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 and makes it possible to provide a numerical estimate of the accuracy of a method in terms of the measured error. We expect $A_{ij} - GT_{ij} \sim 0, i = 1 \dots n - 1, j = i + 1 \dots n$, where GT_{ij} is the ground truth atrophy specified between exam i and exam j for a patient. The focus will be on this criterion in our study.

3.1.2 Existing studies on validation of atrophy estimation methods

In this section, we present some of the previous works on the evaluation of atrophy estimation methods, without the use of any simulated ground truth, and the various criteria that are adopted to perform the comparisons.

Notations used in this section: p represents the p-value and r the correlation coefficient between the specified quantities.

Using patient-control separation, same day scan-rescan and consistency criteria

The separation of controls and Alzheimer’s disease (AD) patients is used as a gold standard by Gunter *et al.* [50] for evaluating BSI and Gradient Matching Method (GMM, derived from SIENA). The yearly atrophy rate results show that, in terms of control/patient separation, GMM method is superior to BSI method for the whole brain (BSI: 0.46 ± 0.19 , GMM: 1.30 ± 0.22) as well as ventricular measurements (BSI: 0.81 ± 0.20 , GMM: 0.95 ± 0.20). Next, they evaluate major steps in image pre-processing. According to their findings, neither bias field corrections (control/patient separation with no bias field correction for BSI: 0.42 ± 0.18 , GMM: 1.28 ± 0.22) nor a hardware change between the scan pairs affected the group separation. The use of an alignment mask (including skull and scalp apart from the brain) for selecting the region over which alignment is to be optimised, left the whole brain measures of BSI largely unchanged while that of GMM worsened with respect to the case when default brain mask is used. They conclude that, in general, ventricular measure offers a better group separation than the whole brain measure.

The authors, Paling *et al.* [105], present a validation of BSI [44] by applying it to atrophy measurements in late-onset dementia. The validation is carried out to evaluate the effect of calibration of BSI parameters and image scaling to adjust intensity on overall accuracy of atrophy measurements. Accuracy is estimated by the analysis of differences in atrophy estimation values for same-day repeat scan pairs of 15 young and elderly controls. The overall accuracy is in the range of 2 – 4 ml for this study. The results of comparison of registration techniques with 6 and 9 DOF, respectively show that the use of an automated image registration with 9 DOF can efficiently correct scanner voxel size variations without being biased due to the presence of cerebral atrophy. Also, the patient group demonstrated considerable

atrophy ($p < 0.001$) as compared to the controls.

Boyes *et al.* [18] show the correspondence between two different approaches of measuring atrophy, JI (B-spline free form deformation (FFD) based registration), and BSI for Alzheimer's disease. They compare the performance of the methods on the basis of a number of criteria such as same day scan-rescan, consistency and patient-control group separation. Also, a simplistic atrophy is simulated by decreasing the voxel size in the x , y and z directions (scaled atrophy). Results suggest that the mean atrophy for the same-day scan pairs with BSI and JI was close to zero (BSI: $p = 0.56$, JI: $p = 0.10$), while the mean absolute atrophy is very similar (BSI 0.28%, JI 0.25%) for the two techniques. JI ($p > 0.6$) also proves to be a better discriminator between patient-control groups than BSI ($p < 0.01$). The authors infer that JI (absolute error 0.23 ± 0.28) technique is more accurate as compared to BSI (absolute error 0.78 ± 0.46) in terms of scaled atrophy.

Smith *et al.* [142] compare the longitudinal atrophies estimated by SIENA and BSI with each other and also with the cross-sectional atrophy estimated using SIENAX using Percentage Brain Volume Change (PBVC) as a measure. The evaluation is based on correlation of PBVC values, same day scan-rescan analyses, comparison of incremental atrophy summation with first-last differencing (when scans at several time points are available, the atrophy estimated between every two consecutive scans when summed should be close to the atrophy estimated between first and last time points) and patient-control group separation. They show that BSI and SIENA estimates correlate highly ($r = 0.87$, $p < 0.0001$) with each other and exhibit an atrophy estimation error rate of 0.2%. The same day scan-rescan results yield the mean(), mean(abs()) and median(abs()) values for SIENA and BSI (ratio adjusted by 1.20 to match SIENA) respectively as: 0.0006, 0.27, 0.16 and 0.1118, 0.22, 0.17. For first-last time point atrophy estimation results, median absolute differences (controls/patients) of 0.20%/0.16% in the case of SIENA and 0.20%/0.18% in the case of BSI are observed. In this case, SIENAX results also correlate well with SIENA ($r = 0.71$) and ($p < 0.0001$). BSI and SIENA show very similar sensitivity in discriminating the AD-control groups. It is also indicated that SIENAX correlates well with SIENA but is less sensitive, in terms of discrimination between patient and control groups.

Use of a gold standard for evaluating atrophy estimation methods

Creation of gold standard 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 [68], 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 minimising the sum of squared differences between the Jacobian of the transformation and the desired level of atrophy at each voxel. An additional penalisation term is also considered in order to prevent *corner Jacobians* from being negative to ensure that the estimated deformation field preserves topology. However, the penalisation 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)” [109]. 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 regularisation 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 approach proposed by Camara *et al.* [21] 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 onto the considered image using a fluid registration method. Atrophy is then simulated by a FEM (finite element model) 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 for deriving 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.

Evaluation using a gold standard Camara *et al.* [21, 20] evaluate global (BSI and SIENA) and local (JI) brain atrophy estimation methods using a simulated gold standard. To generate the gold standard data, atrophy is simulated using a cohort of scans of 27 probable Alzheimer’s disease and 19 age-matched controls. They find that SIENA and BSI results correlate well with gold standard data (Pearson coefficient of 0.962 and 0.969, respectively). Absolute differences in Percentage Brain Volume Change (PBVC) are reported to be $0.23\% \pm 0.26\%$ and $0.22\% \pm 0.28\%$ for BSI and SIENA respectively. They also use FFD-based and fluid techniques for JI and compare the PBVC errors with respect to the gold standard for different regions of the brain. Results show that mean absolute differences from the gold standard in PBVC are, for the whole brain: FFD = 0.31% and fluid = 0.58%, the lateral ventricles: FFD = 0.79%; fluid = 1.45%; the left hippocampus: FFD = 0.82%; fluid

= 1.42% and the right hippocampus: FFD = 0.95%; fluid = 1.62%. The largest errors for both local techniques occurred in the sulcal CSF (FFD = 2.27%; fluid = 3.55%) regions. For small structures such as the hippocampus, the errors are found to be larger, specially for controls (controls/patients) as compared to the errors in the estimated whole brain atrophy. When using JI, better results are obtained when FFD based registration is used, as opposed to fluid registration. Camara *et al.* identify brain extraction as a crucial step in global techniques whereas number of multi-resolution levels and stopping criteria determined the accuracy of JI approaches.

Pieperhoff *et al.* present a non-rigid registration based approach that quantifies longitudinal atrophy in [109]. They study the effect of noise and bias field inhomogeneity on the results of volume change. To analyse the differences between MR images that were longitudinally acquired, they test the specificity (to what degree are only actual volume changes taken into account) and sensitivity (ability of the method to detect small structural changes) of deformable field morphometry (DFM). They also present a method of simulating atrophy that provides a ground truth for their evaluation. Their experiments suggest that noise and RF inhomogeneity, when not corrected, lead to different types of errors in brain volume change estimation. Noise is responsible for scattered clusters of small volume changes while RF inhomogeneity causes stronger volume changes wide-spread on the brain, unrelated to brain anatomy. Experiments on Brainweb images [30] degraded with 5% additive noise; 40% INU or both showed that spurious changes of up to $\pm 4\%$ can be seen if these artefacts are not corrected for when using DFM. DFM is capable of detecting and estimating simulated atrophies in the brain but provides underestimated results. Same day scan-rescan analysis on healthy subjects reveals that SIENA and BSI report an error of 0.15% and 0.3% respectively, which is comparable to 0.29% for the proposed DFM approach.

The approach of de Bresser *et al.* [34] makes use of manual segmentations of the brain as a gold standard. SIENA [141], Unified Segmentation (US) [6] and k nearest neighbour (kNN) [2] approaches are employed for measuring brain volume change. Adopting precision (repeatability) and accuracy (ground truth) criteria they establish that US and kNN show a good precision, accuracy and comparability for brain volume measurements. The coefficient of repeatability (brain volume/volume change) is found to be larger for US (2.84%) than for kNN (0.31%) and SIENA (0.92%). When brain volume changes are compared, SIENA showed a good ($r = 0.82$; $p = 0.004$), kNN a moderate ($r = 0.60$; $p = 0.067$) and US a weak ($r = 0.50$; $p = 0.138$) correlation with the manual segmentations. To conclude, SIENA outperformed kNN and US in terms of whole brain volume change. kNN can be used if volume change measurements of brain structures are studied separately.

Recapitulation of existing evaluation studies

The most utilised criteria to perform validation of atrophy estimation techniques are: same day scan-rescan, consistency, group separation, incremental atrophy summation with first-last differencing and comparison with a ground truth. The point to be noted is that each criterion has its own limitations. For instance, with same day scan-rescan, consistency and group separation, there is no absolute standard for comparison whereas it is difficult to verify if a simulated ground truth represents atrophies seen in the real cases. It may also be mentioned that, in the context of atrophy simulation, 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 modelling 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.* [68] or Pieperhoff *et al.* [109].

This literature survey has highlighted that SIENA, BSI and JI are the three most studied approaches of atrophy estimation. Various validation studies report the errors in whole brain atrophy estimations for SIENA and BSI to be between 0.15 – 0.2% and 0.2 – 0.3%, respectively. In addition, SIENA overestimates while BSI underestimates atrophy. Besides, JI is validated for different transformation models. Errors in estimations on whole brain have been reported to be for different models as FFD: 0.25 – 0.31%, fluid flow: 0.58% and bio-mechanical: 0.29%. Note that it may be difficult to directly compare the results of these studies as they have been performed on different image databases. From investigations dealing with MR artefacts, it can be concluded that noise and bias field inhomogeneity, if not corrected, lead to different types of errors in the brain. Up to $\pm 4\%$ of spurious changes can be measured with DFM. Brain extraction is a crucial step in global atrophy estimation algorithms whereas multi-resolution levels and stopping criteria determine the performance of JI techniques.

The reader is referred to Table 3.1 for a summary of existing works on evaluation of existing atrophy estimation approaches detailed in section 3.1.2.

This remainder of the chapter is organised as follows. The evaluation framework consisting of atrophy simulation (section 3.2.1) and simulation of artefacts (section 3.2.2) is detailed next. Results of atrophy simulation on Brainweb images are presented in section 3.3.1 and evaluation experiments are discussed in section 3.3.2. The findings from this study are listed in section 3.5.

Approach	Pathology studied	Method(s) evaluated	Criteria used	Conclusions
Gunter <i>et al.</i> [50]	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 <i>et al.</i> [105]	AD (late-onset dementia)	BSI	Same-day scan-rescan, Patient/control separation (Zero-atrophy metric)	*BSI accuracy lies between 2 – 4ml (1σ), *Group separation ($p < 0.0001$)
Boyes <i>et al.</i> [18]	AD	BSI, JI	Same-day scan-rescan, Consistency, Patient/control separation, Scaled Atrophy	*Both measures are well correlated, *JI better estimates simulated tissue loss (absolute error 0.23 ± 0.28)
Smith <i>et al.</i> [142]	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 <i>et al.</i> [20]	AD	BSI, SIENA, JI	Mean absolute difference of PBVC values with simulated atrophies	*SIENA-BSI correlate well, *FFD performed better than fluid registration *Whole brain atrophy errors FFD = 0.31%; fluid = 0.58%
Pieperhoff <i>et al.</i> [109]	CBD	DFM	Specificity, Sensitivity (using simulated atrophy)	*Noise/Bias field inhomogeneity lead to different errors in volume change measurements *Error of DFM (0.29%) comparable to SIENA, BSI
de Bresser <i>et al.</i> [34]	-	US, kNN, SIENA	Manual segmentations as ground truth	*SIENA outperformed kNN, US for whole brain volume change kNN recommended for studying atrophy of brain structures separately.

Table 3.1: 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, **TBM:** Tensor Based Morphometry, **US:** Unified Segmentation, **kNN:** k Nearest Neighbour

3.2 Evaluation framework

The evaluation framework consists of two steps: generation of a test database through simulation of different atrophies and image artefacts and testing brain atrophy measurement algorithms (SIENA, SIENAX, BSI and JI (Noblet *et al.* [98]) on these images. A pictorial summary of the validation framework is shown in Fig. 3.1. Let us study the algorithm of atrophy simulation and addition of image artefacts in more detail.

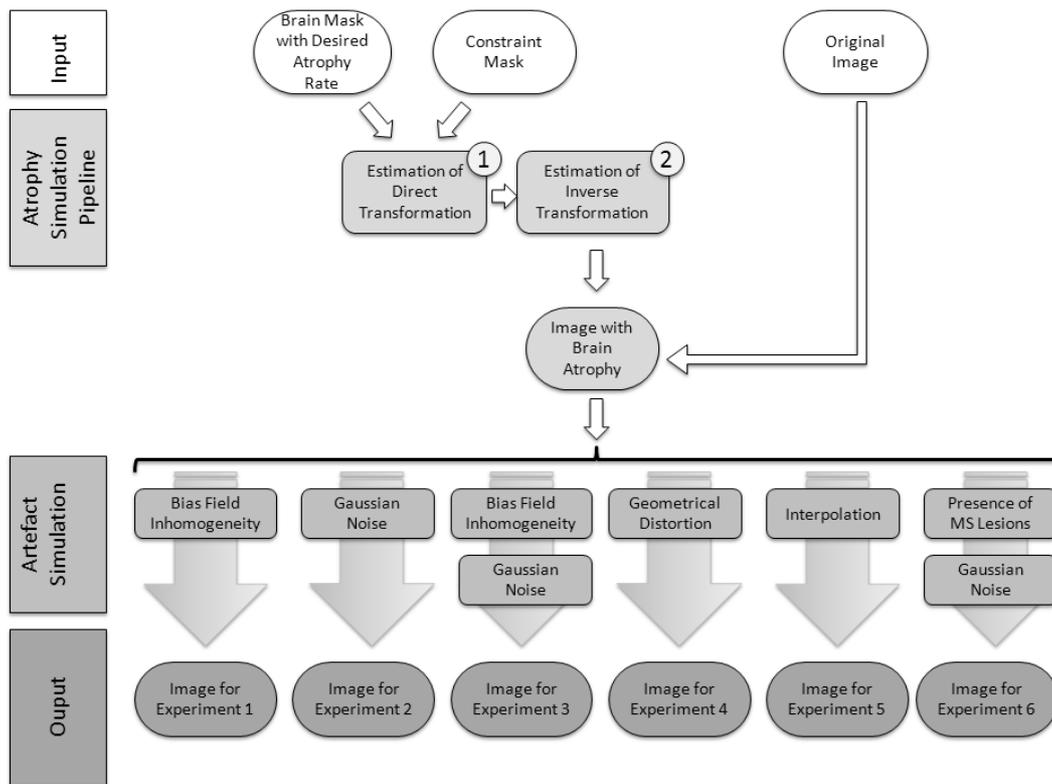


Figure 3.1: Proposed atrophy simulation framework and the evaluation framework showing the addition of various artefacts to the simulated images in order to carry out the validation procedure. This figure shows the various steps involved in the simulation of atrophy (Section 3.2.1) followed by the addition of various artefacts (Section 3.2.2). Here "Output" refers to the images that we simulate for the 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.

3.2.1 Simulation of atrophy: ground truth generation

The atrophy simulation approach estimates a deformation field that preserves topology so that the Jacobian at each voxel, is as close as possible to the desired local level of atrophy. Topology preservation ensures that connected structures remain connected and that the neighbourhood 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.

Contrary to Karacali *et al.* [68] who consider the sum of squared differences between the Jacobian of the transformation and the desired level of atrophy, the logarithm of the Jacobian is considered so that dilations ($1 < J < +\infty$) and contractions ($0 < J < 1$) have a similar influence on the objective function. The advantage of considering the logarithm of the Jacobian has already been highlighted in [78]. 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 current approach is presented in Table 3.2.

First, the B-spline based deformation model that is considered is described followed by the associated optimisation problem.

The multi-resolution deformation model

Let $\mathbf{s} \triangleq [x, y, z]^t \in \Omega \subset \mathbb{R}^3$. The mapping $\mathbf{h} : \Omega \mapsto \Omega$ writes $\mathbf{h}(\mathbf{s}) = \mathbf{s} + \mathbf{u}(\mathbf{s})$, where \mathbf{u} is the displacement vector field. Consider a decomposition of the displacement vector field \mathbf{u} over a sequence of nested subspaces $V_1 \subset \dots \subset V_l \subset V_{l+1} \subset \dots$, defining a multi-resolution approximation of \mathbf{u} [95]. 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

$$\begin{aligned} \Phi_{i,j,k}^l(x, y, z) &= \Phi_i^l(x) \Phi_j^l(y) \Phi_k^l(z) \\ &= 2^{3l/2} \Phi(2^l x - i) \Phi(2^l y - j) \Phi(2^l z - k). \end{aligned}$$

Study	Approach for Atrophy Simulation	Topology Preservation	Skull Invariance	Evaluated Methods	Simulated artefacts
Camara <i>et al.</i> [21]	Bio-mechanical framework	Not guaranteed	Yes	*SIENAX	*Rician noise *Rotational motion *Pulsatile motion
Camara <i>et al.</i> [20]	Bio-mechanical framework	Not guaranteed	Yes	*SIENA *BSI *B-spline and fluid based *non-rigid registrations	*Rician noise
Karacali <i>et al.</i> [68]	Jacobian-based framework	Consider a necessary but not sufficient condition	No	-	-
Pieperhoff <i>et al.</i> [109]	Jacobian-based framework	Consider a necessary but not sufficient condition	No	*SIENA *BSI *Elastic registration	*Noise *Bias field inhomogeneity
Present Method	Jacobian-based framework	Yes	Yes	*SIENA *BSI *SIENAX *B-spline FFD	*Gaussian noise *Bias field inhomogeneity *Geometrical distortions *Interpolation artefacts *Presence of lesions

Table 3.2: A survey of work on evaluation of brain atrophy methods using a simulated ground truth.

At scale l , *i.e.*, in space V_l , the displacement field \mathbf{u}^l is parametrised by the vector $\mathbf{a}^l \triangleq \{a_{x;i,j,k}^l, a_{y;i,j,k}^l, a_{z;i,j,k}^l\}$ as:

$$\mathbf{u}^l(x, y, z) = \begin{bmatrix} u_x^l(x, y, z) \\ u_y^l(x, y, z) \\ u_z^l(x, y, z) \end{bmatrix} = \begin{bmatrix} \sum_{i,j,k} a_{x;i,j,k}^l \Phi_{i,j,k}^l(x, y, z) \\ \sum_{i,j,k} a_{y;i,j,k}^l \Phi_{i,j,k}^l(x, y, z) \\ \sum_{i,j,k} a_{z;i,j,k}^l \Phi_{i,j,k}^l(x, y, z) \end{bmatrix}. \quad (3.1)$$

First degree polynomial B-spline scaling functions Φ are considered in the current implementation [98]. The reader is referred to [98, 99] for additional details on the deformation model.

Optimisation problem

Let $\Omega_J \subset \Omega$ be the area where the desired simulated atrophy level $J(\mathbf{s})$ (the value of the Jacobian at each voxel $\mathbf{s} \in \Omega_J$) is user-specified. The estimation of the corresponding deformation field \mathbf{u} is performed according to the following objective function:

$$E_{\mathbf{u},J,\lambda} = \int_{\Omega_J} |\log(J\mathbf{u}(\mathbf{s})) - \log(J(\mathbf{s}))|^2 d\mathbf{s} + \lambda C \int_{\Omega} E_{Reg}(\mathbf{u}(\mathbf{s})) d\mathbf{s}, \quad (3.2)$$

where $J\mathbf{u}$ stands for the Jacobian of \mathbf{u} , E_{reg} is a regularisation term that ensures that the estimated transformation is smooth, λ is the weight of the regularisation term and C is a scaling factor computed at the beginning of each scale so that the data term and regularisation term are comparable [99]. Among the many regularisation terms proposed in the literature, the membrane energy is chosen.

$$E_{Reg}(\mathbf{u}(\mathbf{s})) = \left(\frac{\partial u_x(\mathbf{s})}{\partial x}\right)^2 + \left(\frac{\partial u_x(\mathbf{s})}{\partial y}\right)^2 + \left(\frac{\partial u_x(\mathbf{s})}{\partial z}\right)^2 + \left(\frac{\partial u_y(\mathbf{s})}{\partial x}\right)^2 + \left(\frac{\partial u_y(\mathbf{s})}{\partial y}\right)^2 \\ + \left(\frac{\partial u_y(\mathbf{s})}{\partial z}\right)^2 + \left(\frac{\partial u_z(\mathbf{s})}{\partial x}\right)^2 + \left(\frac{\partial u_z(\mathbf{s})}{\partial y}\right)^2 + \left(\frac{\partial u_z(\mathbf{s})}{\partial z}\right)^2. \quad (3.3)$$

Contrary to Karacali *et al.* [68] who consider an additional term that prevents the violation of the topology preservation condition, we directly solve the following constrained optimisation problem, which ensures exact topology preservation in the underlying continuous domain:

$$\hat{\mathbf{u}} = \arg \min_{\forall \mathbf{s} \in \Omega, 0 < J_{\mathbf{u}}(\mathbf{s}) < +\infty} E_{\mathbf{u},J,\lambda}. \quad (3.4)$$

The procedure for solving this optimisation problem is quite involved and is detailed in [98]. The estimation is done in a hierarchical way until the desired final scale l_f , the optimisation procedure at scale l being initialised with the solution

at scale $l - 1$. Instead of considering a gradient descent algorithm as in [98], the Levenberg-Marquardt optimisation procedure [112] is made use of in order to improve the convergence rate.

To simulate realistic atrophy, it is desirable to enforce that the skull remains invariant under the simulated transformation. This constraint has been considered previously in [21] but not in [68]. In the present framework, the optimisation is performed only on those B-spline parameters that do not affect the skull, while setting the other parameters to zero. By doing this, the deformation field vanishes on the skull. Therefore, it is allowed to remain in the same position.

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 using the algorithm described in [100]. In this inversion algorithm of Noblet *et al.*, again a topology preserving parametric B-spline-based representation of the deformation field is considered. The inversion method consists of solving a system of non-linear equations using interval analysis techniques. This procedure provides the flexibility of controlling the accuracy of inversion to the user.

3.2.2 Simulation of artefacts

In order to conduct a realistic evaluation of atrophy estimation methods, artefacts (MR, method or disease related) must be incorporated in the images after having simulated atrophy in the brain. Effect of bias field inhomogeneity, noise, geometrical distortions, interpolation noise and MS lesions on the estimated atrophy are considered in this work and are simulated/used in the following manner:

Bias field inhomogeneity The baseline image as well as the atrophied images are degraded using two different multiplicative intensity non-uniformity (INU) fields (20% INU) available with Brainweb.

Noise In this work, a Gaussian noise is added to all brain scans, such that a signal to noise ratio (SNR) of 15 dB is achieved. Gaussian noise provides a good approximation to the noise present in MRI [49].

Geometrical distortions 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:

$$x' = Tx \tag{3.5}$$

Holden *et al.* [55] suggest that geometrical distortions can be modeled as a fourth degree polynomial as follows:

$$x' = P_0 + P_1x + P_2x^2 + P_3x^3 + P_4x^4 \quad (3.6)$$

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 relative geometrical distortions as well as any rotation and/or translation due to head movement between the two scans. We use “Automated Image Registration (AIR) [168]” algorithm (version 5.2.5) in order 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 3.2. 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 [22] who demonstrate a phantom based geometrical distortion field which is determined based on $T1$ weighted acquisitions. The range of displacements in [22] 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, tumours, 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, 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

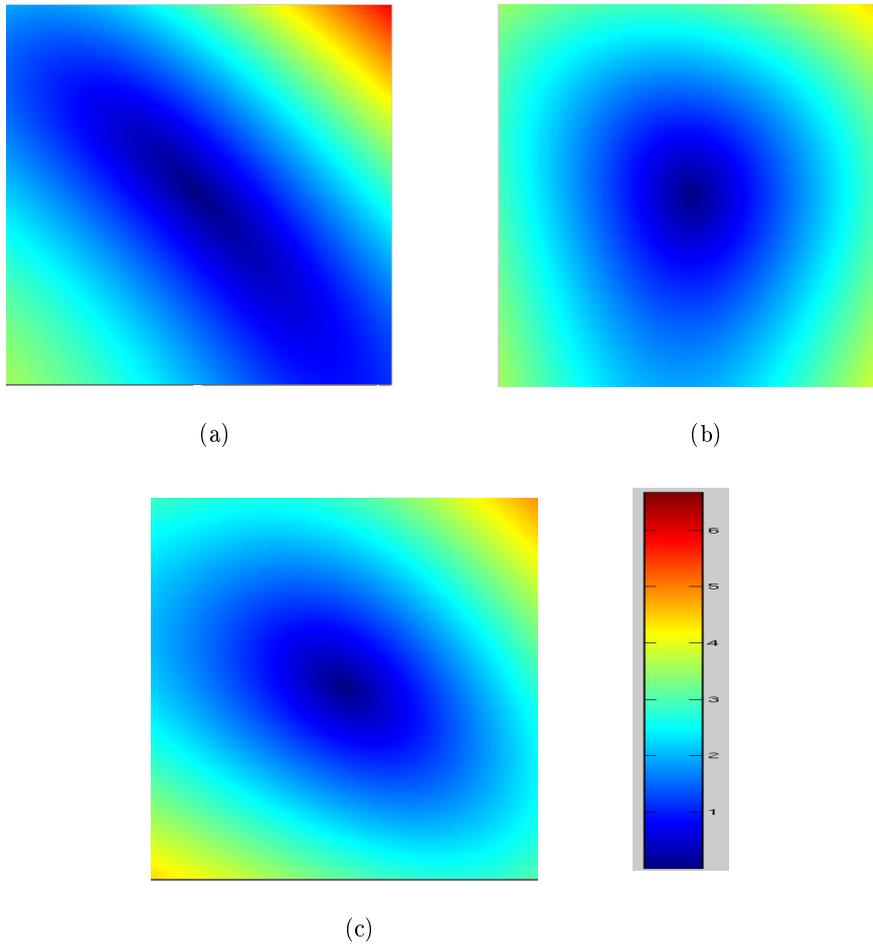


Figure 3.2: The modulus (in mm) of the simulated geometrical distortion field (a) Coronal (b) Sagittal (c) Axial views. The simulated deformations vary in the range of 0 – 6.5 mm.

are performed on normal as well as images with lesions that are degraded with 3% of noise. 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.

3.3 Results

Results of experiments on simulation of atrophy and evaluation of SIENA, SIENAX, BSI-UCD (BSI implementation by Imaging of Dementia and Ageing lab, UC, Davis) and JI (Non-rigid registration of Noblet *et al.*) are detailed in this section. The implementation details and the parameters used for experiments for SIENA, SIENAX, BSI-UCD and JI are explained in appendix B.1.

3.3.1 Results on simulation of atrophy and discussion

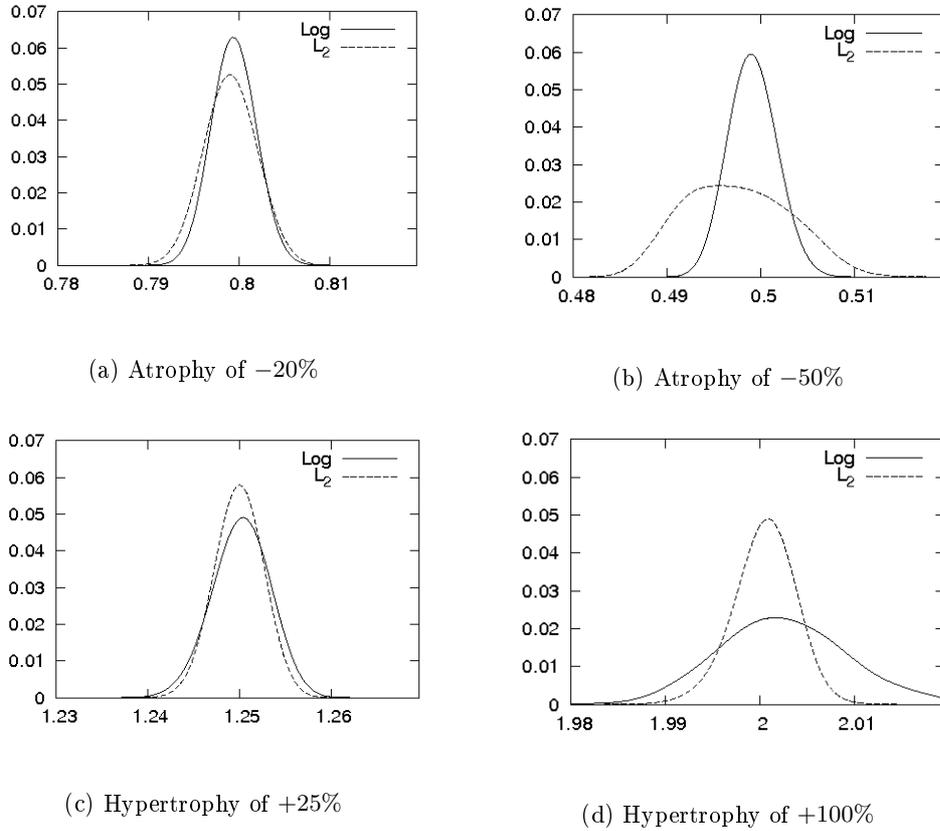


Figure 3.3: Distributions of Jacobian values obtained when simulating uniform atrophies [(a) 20% ($J = 0.8$) and (b) 50% ($J = 0.5$)] and hypertrophies [(c) 25% ($J = 1.25$) and (d) 100% ($J = 2$)].

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. 3.3. We can see that the use of the logarithm for simulating a large atrophy (Fig. 3.3(b)) yields a smaller dispersion of Jacobian values as compared to the L_2 norm. This dispersion increases when simulating large hypertrophies (Fig. 3.3(d)). When simulating small atrophies or hypertrophies, the use of the logarithm or the sum of squared differences tends to yield similar results (Fig. 3.3(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.

Table 3.3: 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

A quantitative analysis of the ability of the proposed algorithm to simulate the desired atrophy has also been done (see Table 3.3). 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 control 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 neighbourhood 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. 3.4 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 high atrophy, contrary to the method proposed in [68], which requires the estimation of a large atrophy in an incremental way. For example, it is 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. 3.5. Fig. 3.5(c) illustrates the simulated changes in the brain and also that there is no movement around the skull.



Figure 3.4: 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.

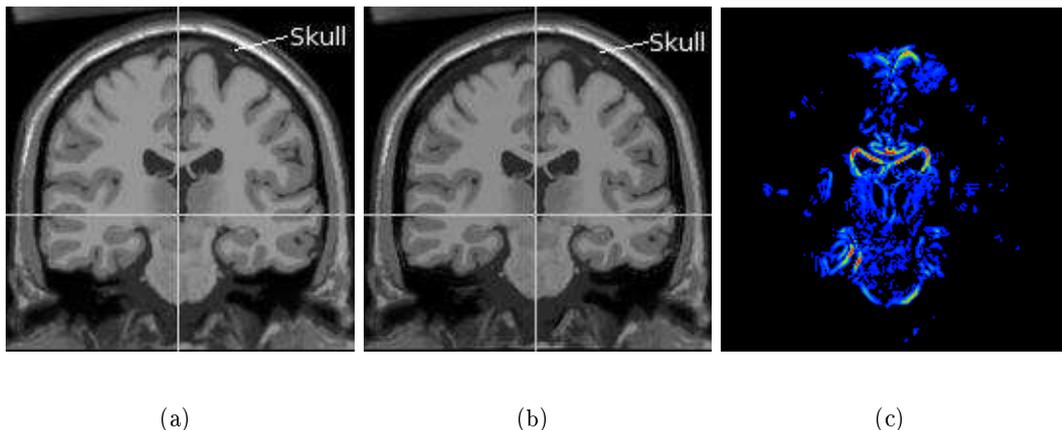


Figure 3.5: (a) Original Brainweb image (b) Simulated image with 10% of globally uniform atrophy (c) Difference between images (a) and (b). Note that there is no deformation on the skull between (a) and (b).

Experiments have also been conducted for the simulation of non-uniform atrophies. To this end, we utilise the “Internet Brain Segmentation Repository (IBSR)”¹. 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 the atrophies in real cases, we refer to [151] and [20], which mention atrophy rates observed in various parts of the brains of Alzheimer’s disease patients. We illustrate the desired and the obtained Jacobians with the cor-

¹<http://www.cma.mgh.harvard.edu/ibsr/>

responding error map in Fig. 3.6. It can be seen in this figure that the simulated Jacobian is close to the true Jacobian with errors in the range of 0–1.2% of atrophy. Fig. 3.6(c) depicts that these errors in the Jacobian are spread evenly all over the brain. Table 3.4 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.

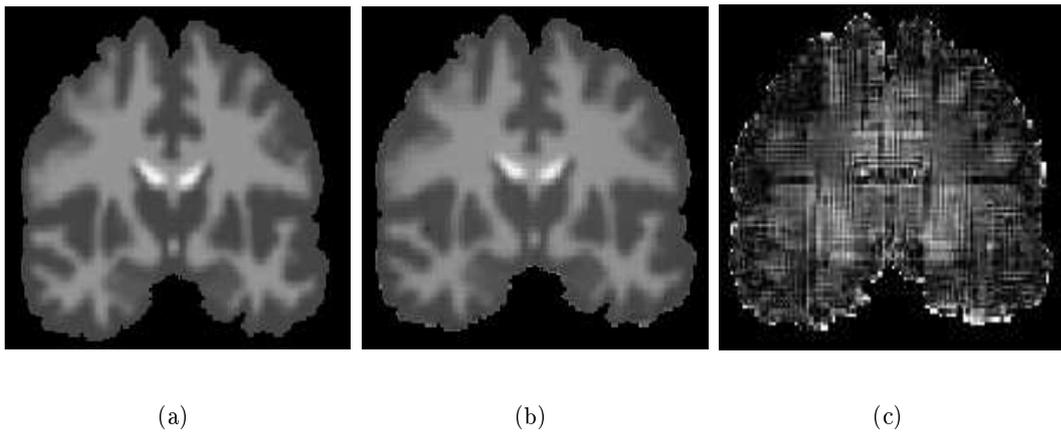


Figure 3.6: Simulation of non-uniform atrophies (a) Desired Jacobian map (b) Jacobian map obtained by using our algorithm (c) Error map. The gray levels (from black to white) depict the values of the Jacobian (from low to high) in the range of 0.94 – 1.08 for the maps shown in (a) and (b) and for the error map (c) in the range of 0 – 1.2% of atrophy.

Table 3.4: Table showing the desired and the obtained mean Jacobian values 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 Volume Change	Achieved Volume Change
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%

The proposed simulation framework is also versatile and can be used for simulating a more complicated pattern of atrophy. For instance, it can be used for simultaneously simulating a global brain atrophy and a change in a given pathological area such as MS lesion evolution or tumour growth. The Brainweb MS moderate lesion database is used for this experiment. The simulation is done by first simulat-

ing 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. 3.7, we present a simulation of 10% of global brain atrophy and 100% of MS lesion volume increase. Notice the increase in the sizes of lesions from Fig. 3.7 (a) to (b). The simulated atrophy can be clearly seen in the difference image (Fig. 3.7 (c)), especially at the boundaries along with the change in lesion sizes.

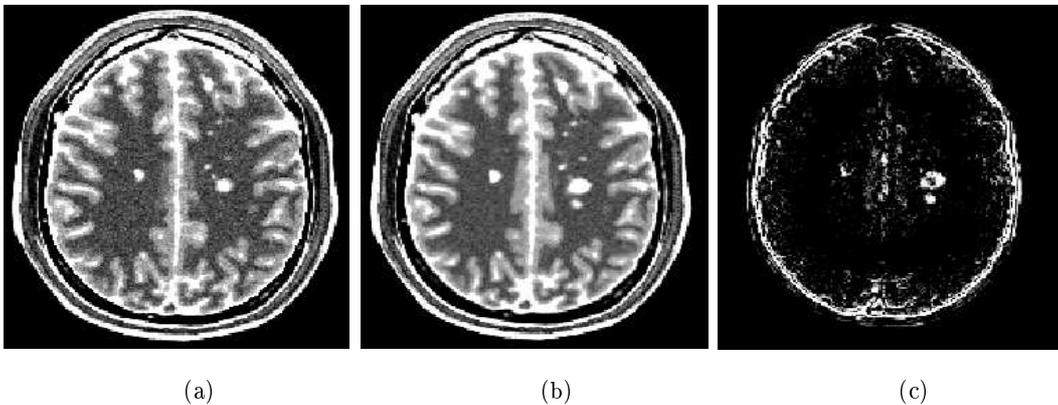


Figure 3.7: Simulation of atrophy in a brain with an increase in the MS lesion size: (a) Original Brainweb image with moderate lesions (b) Image with 10% of global atrophy and 100% of MS lesion increase (c) Difference between images (a) and (b).

3.3.2 Evaluation results of SIENA, SIENAX, BSI-UCD, JI and discussion

Our evaluation is divided into two parts. The first part consists of the simulation of a number of atrophies in a single normal brain image from Brainweb database [71] 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 [8]) in order to examine the effect of anatomical variability on these methods. Also, few results of comparison are shown for JI (Noblet *et al.* [98]). 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.

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. 3.8(a) shows the percentage brain volume change (PBVC) between any two brain scan pairs, $B1$ and $B2$, such that the simulated atrophy in brain $B1$ is less than the simulated atrophy in $B2$ (for the simulated atrophy ranges of 0-1% and 1-10%). For example, if the simulated atrophy in $B2$ is 10%, then it is compared with all simulated atrophies of 1 – 9% (in $B1$). Fig. 3.8(a) compares the methods under consideration with respect to the ground truth, for the noise-free case.

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 15 dB 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. 3.8 (b-d) show the PBVC estimation for SIENA, SIENAX and BSI-UCD for observations degraded with bias field only, noise only, both bias field and noise, respectively. The absolute errors in the estimated PBVC for the three methods are depicted using boxplots in Fig. 3.9 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. 3.8 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. 3.8(a)). Fig. 3.9(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. 3.8(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. 3.9(b) shows that SIENAX (maximum error 1%) outperforms SIENA and BSI-UCD (maximum error 3.7%).

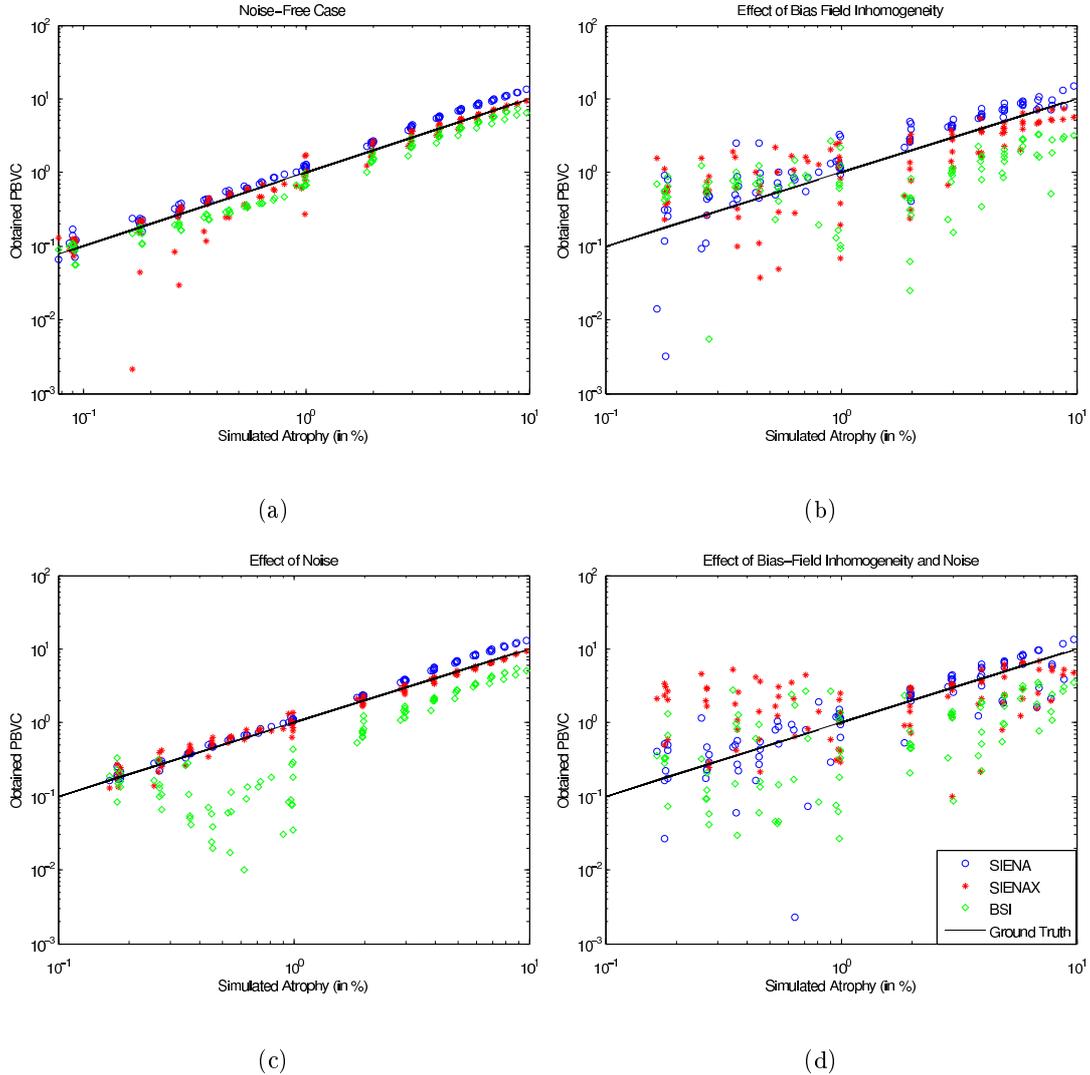


Figure 3.8: PBVC comparison of SIENA, SIENAX and BSI-UCD with the ground truth for (a) noise-free images and those degraded by (b) bias field inhomogeneity only (c) noise only and (d) bias field inhomogeneity and noise. (Refer to section 3.3.2.)

As can be seen from Fig. 3.8, the introduction of bias field inhomogeneity Fig. 3.8(b) and both noise and bias field inhomogeneity Fig. 3.8(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. 3.9(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

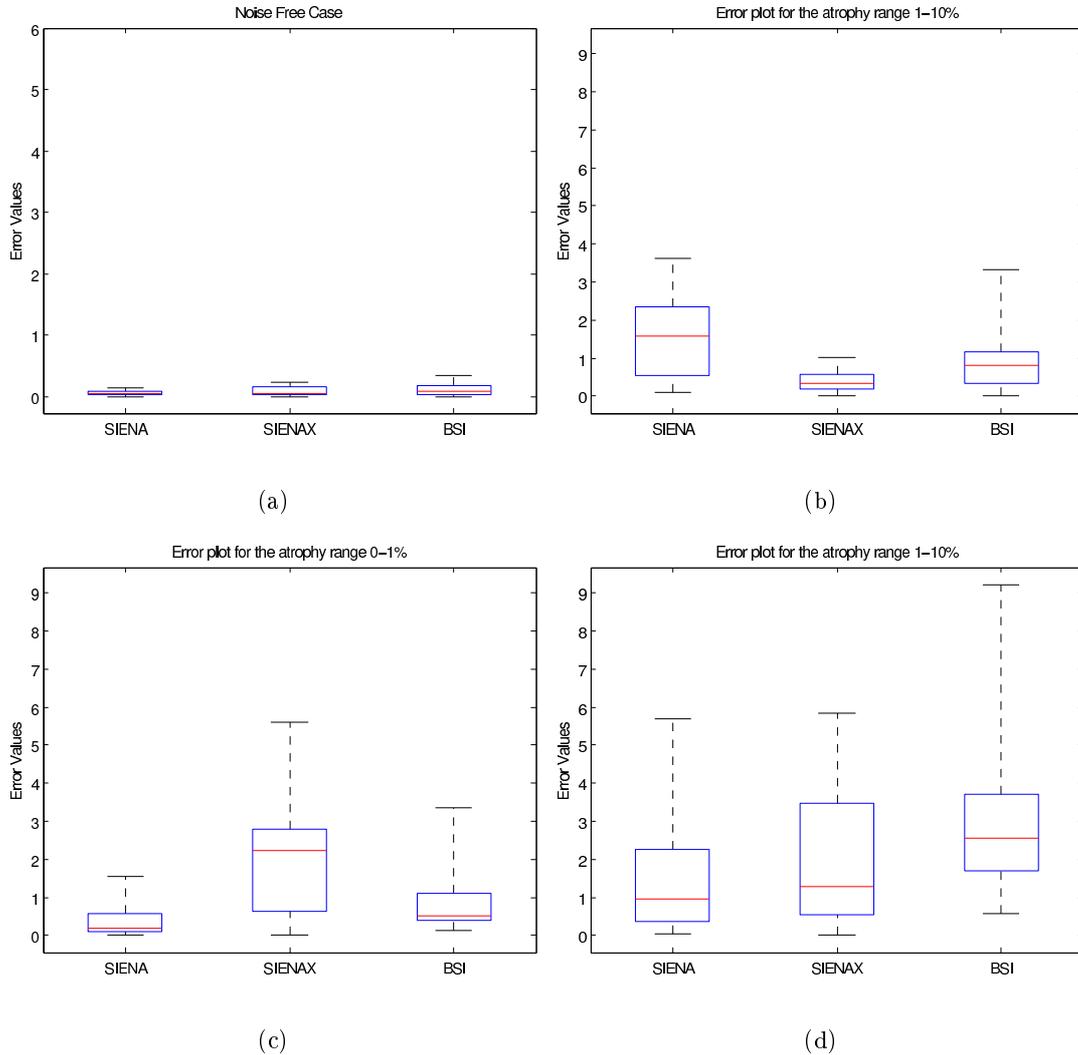


Figure 3.9: Boxplot showing absolute errors in the estimated PBVC by SIENA, SIENAX and BSI-UCD with respect to the ground truth for (a-b) noise-free images and (c-d) those degraded by bias field inhomogeneity and noise. The ends of the boxes show the lower quartile and upper quartile values. The red line denotes the median. The lines extending from each end of the boxes show the maximum and minimum values of the data.

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 scans

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 scans. 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 scans 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 [175] 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 3.6). 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 ($p = 3.0795e - 07$) and BSI ($p = 1.5435e - 27$) 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.

Effect of geometrical distortions Geometrical distortions are simulated as described in section 3.2.2. Fig. 3.10 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 normalise the brain volume with respect to the MNI brain atlas. As mentioned earlier, in our study only unnormalised 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.

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. 3.11. We see that SIENA and SIENAX remain generally unaffected, when compared to the noise-free case shown in Fig. 3.8(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. 3.8(a).

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 3.5. 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 3.5 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 3.5 that lesions can lead to a significant non-desired

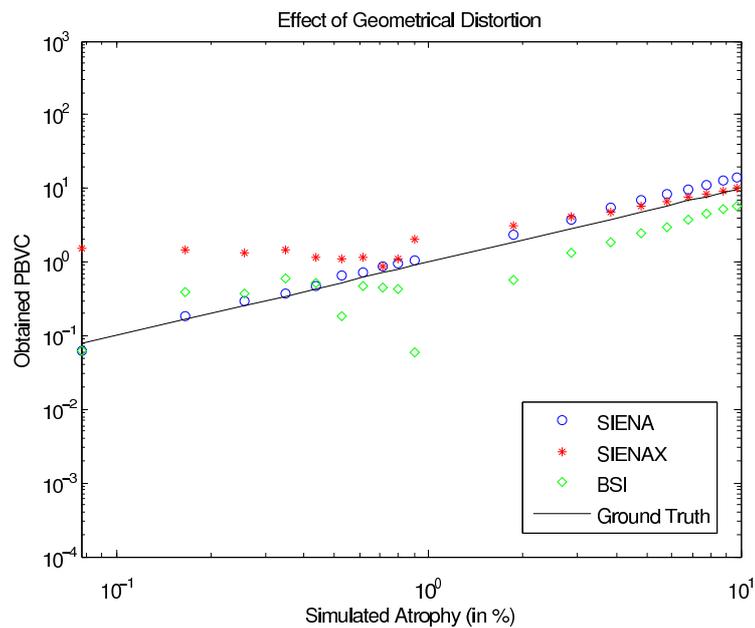


Figure 3.10: PBVC comparison for SIENA, SIENAX and BSI-UCD, respectively with the ground truth when the geometrically distorted Brainweb image is compared with simulated images for different atrophy rates.

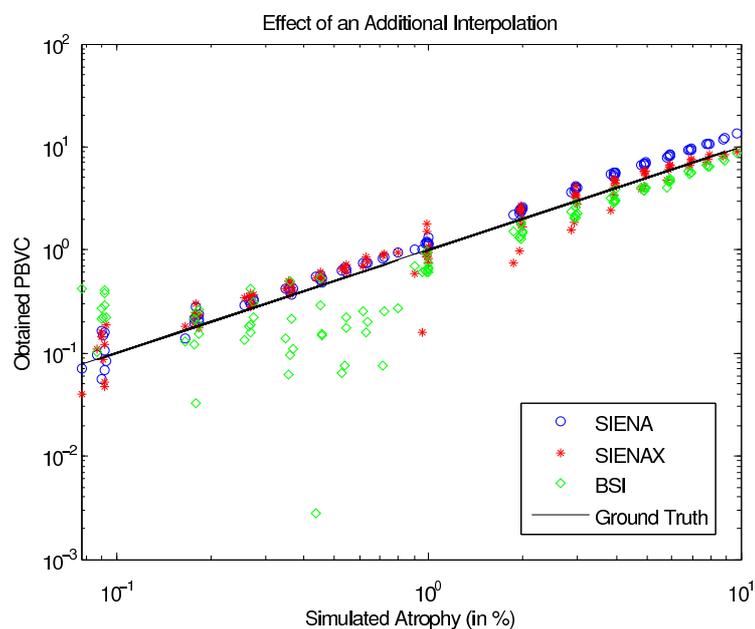


Figure 3.11: PBVC comparison of SIENA, SIENAX and BSI-UCD, respectively, with the ground truth when images are rotated by 1° to add an extra interpolation (B-Spline 5th order) step.

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 lower for both SIENA and SIENAX as depicted in Table 3.5. 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 tests are performed with images degraded by noise, which is mainly responsible for the poor performance of BSI-UCD (see Table 3.5). Our experiments on images degraded with noise have already illustrated the sensitivity of BSI-UCD towards presence of noise (see Table 3.6).

Table 3.5: 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 3.6 summarises the results corresponding to various sources of error discussed in this section.

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 [8]. 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 adopted for separating 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

Table 3.6: 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 3.3.2 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

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 neurodegenerative 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 3.3.2.

Fig. 3.12(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. Since non-uniform atrophy simulations represent the most general case, these tests are repeated using a local atrophy estimation method, Jacobian Integration (Non-rigid registration (Noblet *et al.* [98]) along with the three other methods. 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 $E_{ki} = | PBVC_{GTki} - PBVC_{Estimatedki} |$ where $k = 1 \dots 18$ and $i = 0.2\%, 0.5\%, 1\%, 1.5\%$; A, B, C .

$$\mu_i = \overline{E_{ki}} \quad (3.7)$$

$$\sigma_i = \sqrt{\frac{1}{n} \sum_{k=0}^n (E_{ki} - \mu_i)^2} \quad (3.8)$$

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. 3.12(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. 3.12(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. 3.12(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 the experiments (Fig. 3.12(c)). When the same tests are performed using JI, it is observed that JI surpasses all the other three methods in both noise-free and noisy cases Fig. 3.12(c) and (d). It may be clarified here that the simulation algorithm uses the transformation model, the regularisation term and topology preservation constraints in the same manner as Noblet *et al.*'s algorithm that is used here for performing JI. Nevertheless, the transformation of the simulated atrophy is inverted before it is applied to the images so that the model parameters do not remain the same as the original transformation. Hence, when the simulated transformation is recovered with Noblet *et al.*'s algorithm the estimation does not favour JI.

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 be a reason for 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

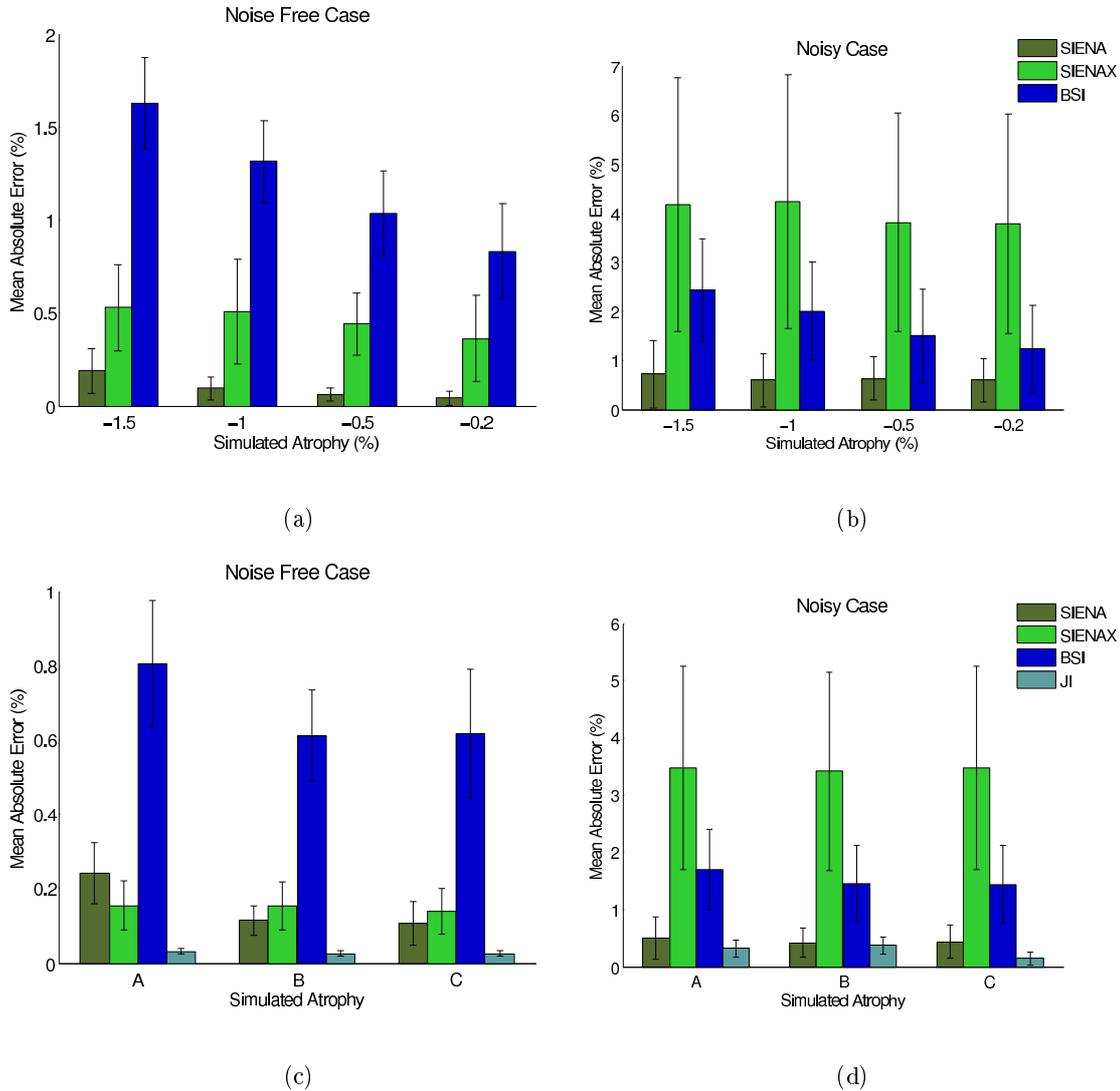


Figure 3.12: Bar plot showing the mean absolute error and its standard deviation in the estimated PBVC (in percentage) with respect to the ground truth for SIENA, SIENAX and BSI-UCD for the Uniform atrophy case using (a) Noise-free images (b) With images degraded with bias field inhomogeneity and noise 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.

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 3.3.2, 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 derivatives 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 Tables 3.7 and 3.8. 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, $p < 0.001$ for SIENA and SIENAX while $p < 0.001$ for BSI. In terms of overall mean also, JI proves to be the best performer among this group of four methods.

Table 3.7: Overall mean and standard deviation of absolute error in PBVC of SIENA, SIENAX and BSI-UCD for uniformly simulated atrophy. Noisy observations are images degraded with noise and bias field inhomogeneity.

Method	Uniform Atrophy	
	Noise-free Observations	Noisy Observations
SIENA	0.0944 ± 0.0656	0.6382 ± 0.5306
SIENAX	0.4492 ± 0.2210	4.0073 ± 2.4087
BSI-UCD	1.1881 ± 0.2418	1.7948 ± 0.9743

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. 3.13 shows the voxel by voxel mean error for the

Table 3.8: Overall mean and standard deviation of absolute error in PBVC of SIENA, SIENAX, BSI-UCD and JI for non-uniformly simulated atrophy. Noisy observations are images degraded with noise and bias field inhomogeneity.

Method	Non-Uniform Atrophy	
	Noise-free Observations	Noisy Observations
SIENA	0.1545 ± 0.0948	0.4472 ± 0.3074
SIENAX	0.1510 ± 0.0629	3.4483 ± 1.7303
BSI-UCD	0.6715 ± 0.1891	1.5255 ± 0.6828
JI (Noblet <i>et al.</i>)	0.0366 ± 0.0100	0.3256 ± 0.1415

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. 3.13(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.

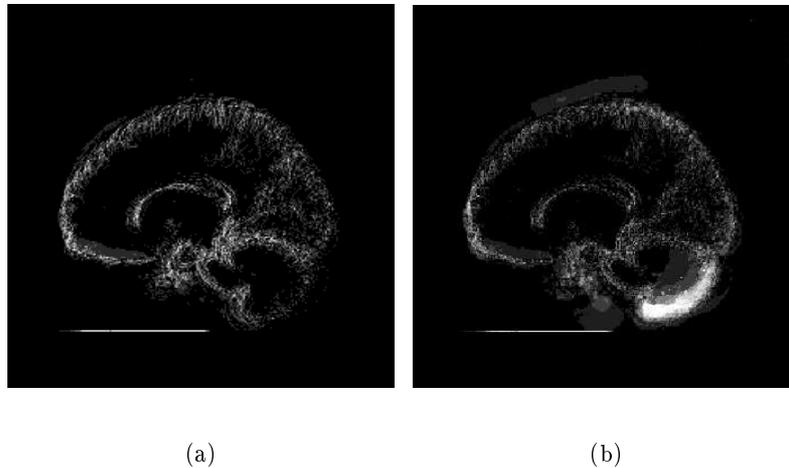


Figure 3.13: Mean error in the segmentation of brain for (a) the noise-free case (b) when using images degraded by bias field inhomogeneity and noise for the 18 baseline images.

We also perform some additional experiments on BSI-UCD, in order to explain the high error that is observed. Studies on BSI (such as [20]) have stated the mean absolute error to be around 0.2% which is lower than what can be seen in our analysis (See Tables 3.7 and 3.8). Three experiments are performed with BSI-UCD using

noise-free images in 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 free of intensity inhomogeneity 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, a test is carried out by replacing the registration of the BSI-UCD with an in-house mutual information based affine registration algorithm. These results are demonstrated in Fig. 3.14. 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 [20]. 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.

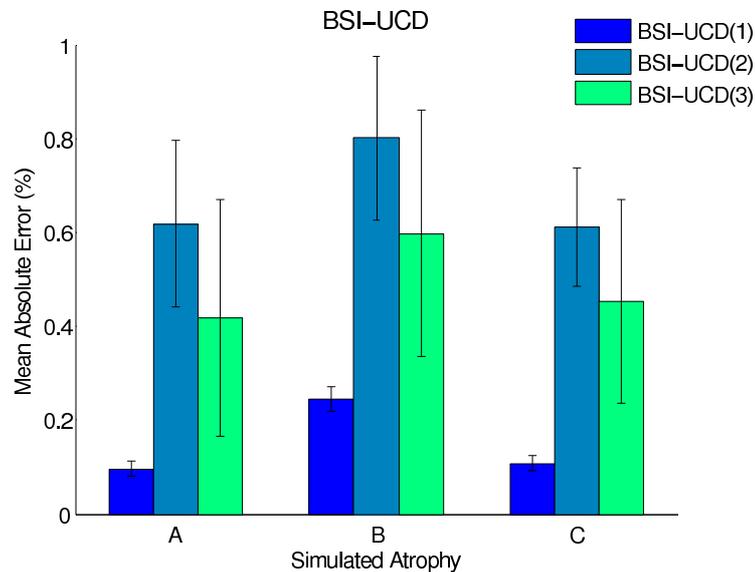


Figure 3.14: Bar plot showing the mean absolute error and the standard deviation in the estimated PBVC (in percentage) when the BSI-UCD algorithm is run with no pre-processing (BSI-UCD(1)), normal BSI-UCD algorithm (BSI-UCD(2)) and by replacing the registration algorithm of BSI-UCD with an in-house affine registration algorithm (BSI-UCD(3)).

In Jacobian integration, there are two primary steps: a non-rigid registration (NRR) step in order to obtain the Jacobian and the integration of this Jacobian over the area of interest. In order to quantify the effect of using an erroneous mask for integrating the Jacobian, another experiment is performed by integrating the same Jacobian map over different masks. Brain masks are created as follows: by manipulating the ground truth mask by performing dilations, erosions and translations on it and from the segmentation algorithm of Bricq *et al.* [19]. The list of the masks employed are listed below:

- M: Ground truth mask.
- MD1: Ground truth mask dilated 1 iteration using a $3 \times 3 \times 3$ cubic structuring element.
- MD2: Ground truth mask dilated 2 iterations using a $3 \times 3 \times 3$ cubic structuring element.
- ME1: Ground truth mask eroded 1 iteration using a $3 \times 3 \times 3$ cubic structuring element.
- ME2: Ground truth mask eroded 2 iterations using a $3 \times 3 \times 3$ cubic structuring element.
- MT: Ground truth mask translated by 1° each in X, Y and Z dimensions.
- SHMC: Mask obtained by brain extraction (BET) followed by the segmentation algorithm. The algorithm also makes use of N3 bias correction [137] for noisy images.

Table 3.9 shows the mean and standard deviation of absolute error in PBVC of NRR algorithm of Noblet *et al.* for non-uniformly simulated atrophy with respect to the ground truth when different masks are used for averaging the Jacobian. It can be seen that erosion and dilation lead to higher errors as compared to translation of the true brain mask. On the whole, using Bricq *et al.*'s algorithm can cause a difference of up to 0.1% in the estimated global atrophy (with respect to the case when ground truth mask is used). Statistical tests show that the results obtained using masks from Bricq *et al.*'s algorithm are significantly different to the case when true mask is used on the same deformation ($p \approx 10^{-13}$). These tests highlight the noteworthy role of the mask used for averaging the Jacobian in order to estimate atrophy.

3.4 Comparison with other studies

In this section, we compare the results of evaluation with other works. Camara *et al.* [20], who utilise the same criterion as in our study, report the mean absolute errors in the estimated atrophy as 0.22% and 0.23% for SIENA and BSI, respectively. These values do not agree with the mean absolute errors that we obtain in

Table 3.9: Overall mean and standard deviation of absolute error in PBVC of NRR of Noblet *et al.* for non-uniformly simulated atrophy with respect to the ground truth when different masks are used for averaging the Jacobian.

Mask	Noise-free Observations
M	0.0366 ± 0.0011
MD1	0.2563 ± 0.1148
MD2	0.3351 ± 0.1363
ME1	0.2807 ± 0.1641
ME2	0.1954 ± 0.1922
MT	0.08548 ± 0.0320
SHMC	0.0721 ± 0.0278

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). Their 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 utilised 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. In case of JI, Noblet *et al.*'s [98] method cannot be directly compared to Camara *et al.*'s [21] FFD as these two algorithms are based on different models. Note that the mean absolute errors can only be compared to those studies who use simulated ground truths. The employment of other criteria of evaluation by Smith *et al.* [142] and Boyes *et al.* [18] 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. 3.8). The least squares fitted scaling factors are $GT(\text{Ground Truth})=SIENA*0.89$; $GT=BSI-UCD*1.88$; $SIENA=BSI-UCD*2.15$ (for noise-free observations). A similar trend has been observed in Smith *et al.* [142] ($SIENA=BSI*1.20$) and Camara *et al.* [20] ($GT=SIENA*0.90$; $GT=BSI*1.18$; $SIENA=BSI*1.29$). 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: $r = 0.87, p < 0.001$, Camara: $r = 0.97$). Our results indicate a better correlation between the atrophy measured by SIENA and SIENAX ($r = 0.89; p < 0.001$) as compared to Smith *et al.* ($r = 0.71; p < 0.001$).

3.5 Findings

We have evaluated the performance of three popular methods for the estimation of cerebral atrophy using gold standards. To create the gold standards, a topology preserving scheme is used for simulating atrophy using a B-spline based deformation model. Additional constraints ensured skull invariance in the simulated image. We have shown that the 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.

The performance of three freely available algorithms (SIENA, SIENAX and BSI-UCD) and one JI algorithm by simulating atrophy in Brainweb images was assessed. Our analysis consisted of two steps: the simulation of atrophies in a single Brainweb image in order 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 higher as compared to SIENA. Experiments with non-uniformly simulated atrophies, where Jacobian integration (based on the non-rigid registration method of Noblet *et al.* [98]) was tested as well, also supported these results. In terms of mean error, JI outperformed SIENA in both noise free and noisy cases. The observed errors were also higher than the overall errors (mean absolute differences) obtained by Camara *et al.* [21] for SIENA and BSI.

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 in the group

of three widely used methods SIENA, BSI and SIENAX. However, experiments including JI showed that it performed better than SIENA. 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 SIENA, SIENAX and BSI. In the JI algorithm, the similarity criterion was based on the intensity values and was susceptible to intensity inhomogeneities. A good bias field correction algorithm is essential in order to improve the performance of the methods. The set of tests that we performed also indicated that SIENA, BSI-UCD and JI 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 scans 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 localise the brain edges and thereby improve the accuracy of SIENA. The critical steps for BSI-UCD are the registration of the two brain scans 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, [14]). JI depends both on the non-rigid registration of the serial images and the segmentation of brain for extracting the region of interest. The registration in turn is a function of many aspects such as the model (similarity criterion, regularisation function) and algorithmic settings (such as levels in the multi-resolution framework

and optimisation parameters). Pre-processing techniques play a critical role in its correct functioning. 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.

Uncertainty Estimation for Longitudinal Brain Atrophy Measurement Algorithms with a Ground Truth

Analysis of statistical properties of maps of volumetric changes of the brain, such as Jacobian maps, is a crucial step in studies based on voxel or tensor based morphometry. However, the accuracy of maps of volumetric change is questionable when they are obtained from images corrupted with artefacts. Bias field inhomogeneities and noise in the images were identified as major determinants causing errors in atrophy estimations in chapter 3. In addition, factors such as inaccuracies of registration and segmentation procedures used during an atrophy estimation algorithm may also result in biased estimates of atrophy, even in the absence of any image artefacts. Another critical consideration that is generally overlooked is model error (*i.e.*, errors in the mathematical model representing brain atrophy). For instance, in non-linear registration approaches free-form, linear elastic or viscous fluid flow transformations can be chosen as models. Jacobian maps obtained from these non-linear registration models may be biased [121]. Validation studies [142, 20, 109] including ours (chapter 3) have confirmed the existence of bias in atrophy estimations. Due to these reasons, it becomes difficult to separate real anatomical changes from spurious ones. It is thus important to make such maps of change more reliable by providing estimates of uncertainties in atrophy estimations. This will aid end users in decision making.

In this chapter, the goal is to develop a generic framework for estimating uncertainties in longitudinal brain atrophy by means of constructing confidence intervals *for any atrophy estimation method*. A simulated ground truth acts as an independent source of learning errors in atrophy measurements. Example simulations on multiple subjects allow us to estimate uncertainties that are stored for future measurements in a “learning database”. The novelty of the framework lies in its ability to incorporate both errors arising from MRI artefacts and method specific bias in the constructed confidence intervals. In the light of superior performance of the NRR method of Noblet *et al.* [98] over SIENA, SIENAX and BSI methods in section 3.3.2, coupled with the flexibility of manipulation (atrophy measurement in any region of interest) that NRRs offer, the rest of the thesis is dedicated to the analysis of Jacobian maps obtained from such algorithms. Three non-rigid registration (NRR)

approaches are chosen [98, 10, 157] to demonstrate how confidence intervals can be constructed for atrophies estimated by them. The proposed framework is tested for hippocampal atrophies in real MR brain images. Confidence intervals are quantitatively evaluated on the basis of coverage probability and length. Results show that the ANTS [10] method provides meaningful confidence intervals as compared to the NRR methods of Noblet *et al.* [98] and Vemuri *et al.* [157].

4.1 Uncertainty

Uncertainty is defined as the margin of error in a measured value and depends on the accuracy and precision of the system of measurement. Accuracy refers to the distance of an estimated value from its true value whereas precision is the variability in repeated measurements of the same quantity under identical conditions [149]. Precision can be increased by using a large sample size. However, if the system has a systematic error, a large sample size will not improve the accuracy. An independent source is required to measure accuracy.

Let θ^* be a scalar measured quantity then uncertainty on θ^* can be expressed as an interval (θ_l, θ_h) that is likely to contain the true value (or the ground truth) θ of this quantity with a predefined probability. The distance of mean estimated value θ^* and the true value θ (accuracy) is what we will refer to as the bias (b) in estimations in the rest of the thesis. Fig. 4.1 depicts the definition of precision and bias.

$$e = (\theta_l, \theta_h) \quad (4.1)$$

$$b = \theta - \theta^* \quad (4.2)$$

4.2 Uncertainty estimation in image registration

The discussion in the previous chapters has revealed how and why registration algorithms play a crucial role in brain atrophy estimation. Hence, this section is devoted to approaches of measuring uncertainty in registration algorithms. The existing approaches are classified depending on whether they employ a ground truth, manipulate the similarity criterion or the deformation field or are based on Bayesian registration for obtaining uncertainties.

4.2.1 Using a ground truth

The most popular way of measuring uncertainty is to simulate known transformations and compare the recovered results with the true transformation. Let T be the transformation estimated by a registration approach and let θ^* be the simulated

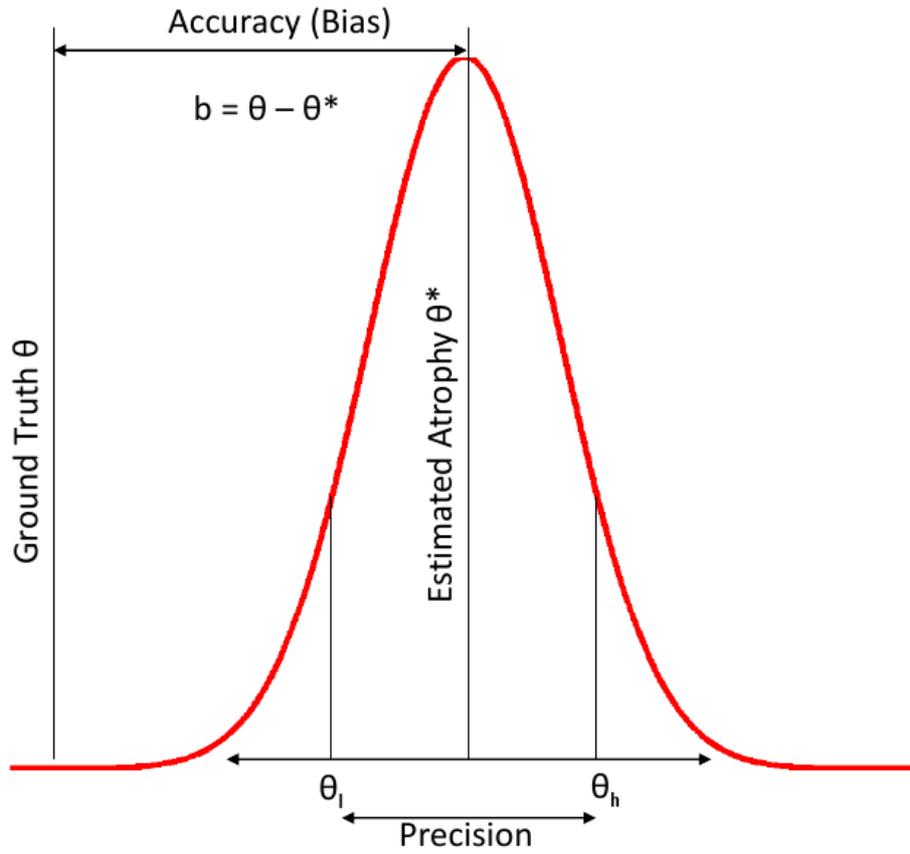


Figure 4.1: Assuming a Gaussian distribution for atrophy, the mean estimated atrophy θ^* and precision (variance) are obtained from repeated measurements of atrophy under identical conditions. In this depiction, the confidence interval $e = (\theta_l, \theta_h)$ does not contain the true value (ground truth). The distance between the mean estimated atrophy θ^* and the true atrophy θ is the bias b in atrophy measurements.

deformation (ground truth). The residual registration error at a voxel x can be calculated as:

$$\varepsilon = \frac{1}{N} \sum_{x \in \Omega} \|(x + \theta^*(x)) - T(x)\| \quad (4.3)$$

where N is the total number of voxels and Ω is the domain of the images.

Applications of FEM to modelling interrelations between different tissue types for creating artificial deformations have been demonstrated for breast tissue by Schnabel *et al.* [128], for prostate by Lee *et al.* [75] and for brain by Camara *et al.* [20].

Schnabel *et al.* [128] report average registration errors for non-rigid registration method of Rueckert *et al.* [124] as 0.4 mm and 0.85 mm when registering post-contrast breast images to a FEM deformed version of themselves and pre-contrast breast images to a FEM deformed version of post-contrast images.

Lee *et al.* [75] compare three registration algorithms: the Demon's method [150], level set method of Vemuri *et al.* [157] and a physically-based viscous fluid flow method [29]. They use the fact that the estimated deformation $\hat{\theta}$ and the simulated deformation θ^* must be inverse mappings of each others. Hence, their composition must produce values close to zero. They define absolute ε_{abs} and relative ε_{rel} measures of registration error as:

$$\varepsilon_{abs}(x) = \text{length}(\hat{\theta}(\theta^*(x))) \quad (4.4)$$

$$\varepsilon_{rel}(x) = \frac{\text{length}(\hat{\theta}(\theta^*(x)))}{\text{length}(\theta(x))} \quad (4.5)$$

where $\text{length}(\theta(x))$ represents the norm of the 3-D vector at voxel x . They conclude that Demon's method produces the best results on prostate images (mean $\varepsilon_{abs} = 0.41$ cm, mean $\varepsilon_{rel} = 1.5$ cm) followed by fluid flow registration method (mean $\varepsilon_{abs} = 0.37$ cm, mean $\varepsilon_{rel} = 3.93$ cm) and level set motion based method (mean $\varepsilon_{abs} = 0.97$ cm, mean $\varepsilon_{rel} = 15.58$ cm).

The contributions of Camara *et al.* [20] have already been discussed in section 3.1.2.

4.2.2 Manipulating the similarity criterion or deformation field

Robinson *et al.* [120] and Yetik *et al.* [172] present registration accuracy derived from the Cramér-Rao bound (CRB) in terms of covariance of deformation field parameters. The Cramér-Rao bound states that if θ is an unknown set of parameters then the covariance (C) of any unbiased estimator $\hat{\theta}$ of θ is bounded by the inverse

of the Fisher information $I(\theta)$:

$$C(\hat{\theta}) \geq \frac{1}{I(\theta)} \quad (4.6)$$

Let $f(x) = u'(x) + n_f(x)$ and $g(x) = u(x) + n_g(x)$ where f, g are images to be registered, $n_f(x), n_g(x)$ are zero mean i.i.d. noise with variance σ^2 . u is the true image and $u'(x) = u(x + \theta^*)$, where θ^* is the true deformation between the two images f and g . The Fischer information matrix I is given by [73]

$$I_{ij}(\theta) = \frac{1}{\sigma^2} \sum_{x \in \Omega} \frac{\partial u' \partial u'}{\partial \theta_i \partial \theta_j} \quad (4.7)$$

where Ω is the domain of the images f and g . Computation of the Fischer matrix is done using the ML estimates of u' and σ^2 [73].

The Cramér-Rao bound (Eq. 4.6) gives us a lower bound on the covariance of any unbiased estimator $\hat{\theta}$ of θ^* . The requirement of an unbiased estimator of deformation may be too limiting in practise where we deal with biased estimates frequently.

Fast Registration Accuracy Estimation (FRAE) is presented by Kybic *et al.* [72] which assumes that the similarity criterion and the error in registration parameters are normally distributed. Statistical analysis of the similarity criterion leads to a closed form solution to finding parameter covariance. Let J be a similarity criterion which can be written as a sum of pixel contributions as follows

$$J(\theta) = \sum_{x \in \Omega} s(\theta(x)) \quad (4.8)$$

where $s(\theta(x))$ is the similarity criterion at voxel x at which the estimated deformation is θ . The covariance under this framework can be estimated as,

$$C_{\theta}^{FRAE} = \frac{4\gamma}{F^{-1}((1 - \alpha)^2, d)} \mathbf{H}^{-1} \quad (4.9)$$

where F^{-1} is the inverse cumulative χ^2_d distribution function and \mathbf{H}^{-1} is the inverse of the Hessian of the parameter of interest θ . γ is the variance of the similarity criterion around the noiseless value at 95% level. FRAE can be applied to atrophy estimations performed with the help of NRR methods based on similarity functions that can be written as a sum of pixel contributions. However, some additional computational overhead is required. This approach has the advantage of being fast.

In another work, Kybic *et al.* [73] make use of bootstrap resampling in order to estimate uncertainty. Bootstrap methods are versatile as they allow the user to obtain an estimate of any parameter of interest such as the mean or standard deviation from an approximate distribution. The basic principle of bootstrap methods

is stated here. Let $X = \{x_1, x_2, \dots, x_N\}$ be N i.i.d samples of a random variable \mathcal{X} . The bootstrap procedure consists of drawing N samples from X with replacement, generating a bootstrap sample X^b . This process is repeated BR times and $b = 1, 2, \dots, BR$.

Kybic *et al.* [73] estimate uncertainty by applying bootstrap resampling on the image voxels for simulating the behaviour of the cost function. The bootstrap resampling method is applied to voxel coordinates Ω . A set of BR bootstrap resamples $\Omega^{(b)}$ by sampling on Ω with replacement to obtain bootstrap cost J^b , which is minimised to generate the bootstrap estimated deformation field, $\hat{\theta}^{(b)}$. Here, $b = 1, 2, \dots, BR$. The main advantage of the bootstrap approach is that it is applicable to any cost function that can be written as a sum of voxel contributions such as sum of squared or absolute differences but can be generalised to accommodate mutual information. The covariance matrix is calculated from bootstrap samples in order to facilitate a comparison to CRB (Eq. 4.6) and FRAE (Eq. 4.9).

$$C_{\hat{\theta}}^{boot} = \frac{1}{BR} \sum_{b=1}^{BR} (\hat{\theta}^{(b)} - \mu_{\theta}^{boot})(\hat{\theta}^{(b)} - \mu_{\theta}^{boot})^{\top} \quad (4.10)$$

and

$$\mu_{\theta}^{boot} = \frac{1}{BR} \sum_{b=1}^{BR} \hat{\theta}^{(b)} \quad (4.11)$$

Kybic *et al.* compare CRB, FRAE and bootstrap method by simulating 2-D translations in the range of $[-2, 2]$ and adding uncorrelated zero mean i.i.d. Gaussian noise, correlated Gaussian noise and salt and pepper noise to achieve SNRs from -10 dB to 70 dB using Lena and ultrasound images. CRB provides optimistic estimates of errors at low SNRs but its estimates are useful for Gaussian noise and high SNRs. For medium to high SNRs, FRAE overestimates the error but fails at lower SNRs (10 to 20 dB) and tests with salt and pepper noise. The bootstrap based method outperforms the other two methods for correlated, uncorrelated Gaussian and salt and pepper noise. The ratio of the estimated to the true translation is less than 2 . The overall conclusion from this study is that FRAE is superior to Cramér-Rao bound but is a worse performer than the bootstrap method. A drawback is that the uncertainties from these three methods do not include bias in estimations. For the bootstrap method, the image generating process needs to be assumed ergodic so that its behaviour across realisations can be learnt from their behaviour in space [73].

Hub *et al.* [60] identify ambiguity in homogeneous regions and misaligned edges due to inefficient optimisation as the major causes of errors in B-spline image registration. Their uncertainty estimation is based on characterising the sensitivity of the similarity measure to moderate and randomly performed variations in the B-spline coefficients. Let $\{c_1, c_2, \dots, c_{3N}\}$ be the B-spline coefficients obtained from the registration algorithm. Also let $\{q_1, q_2, \dots, q_{3N}\}$ be a set of random variables

distributed uniformly in the range of $(-10, 10)$ mm. Typical registration errors in lung imaging are expected in this range. Generation of test deformations is carried out by replacing each estimated c_n B-spline coefficient by $c_n + q_n$. As the next step, the new deformation field is calculated from the perturbed coefficients. The voxel wise uncertainty is defined as the maximum deviation deformation obtained after random variations in the coefficients and the initial deformation (*i.e.* before random variations). This procedure is repeated with K sets of random variations r_n and for each voxel and dimension the largest spatial deviation is stored for those K' random variations where the local similarity metric (SSD) is less than or equal to the initial local similarity metric. Finally, the voxel wise uncertainty $e_i(x)$ is given by the following expression:

$$e_i(x) = \theta_{max,k}(x) = \max_{k=1,\dots,K} \{|\theta_{k,i}(x) - \theta_i(x)|\} \quad (4.12)$$

where $\theta_{k,i}(x)$ is one of the K' deformations and $\theta_i(x)$ is the initial deformation at voxel x . $i = 1, 2, 3$ denotes the dimension.

Results of experiments on CT lung images exhibited maximum average local registration error of 5.7 ± 4.6 mm in the caudal-cranial directions. This method is based on the premise that randomly performed variations may locally improve the registration quality although the global registration could be worse as the global metric may increase.

Jalobeanu *et al.* [63] estimate the uncertainties as the inverse covariance matrix (or precision matrix) corresponding to a Gaussian approximation of the *a posteriori* probability distribution function of the deformation field around the minimum. The covariance matrix is obtained by computing the second derivatives of the cost function at the minimum. A major highlight of this method is the correction of uncertainties for systematic biases. First, they apply a soft thresholding to inverse covariance terms, and the threshold value is computed by the average second derivative for pure noise, computed from simulations. Second, resampling errors are taken into account in cases where raw images are not available. Their simulations on satellite images of Mars confirm that interpolation methods such as bi-cubic resampling are a source of systematic shifts (maximum 0.25 pixel). The distribution of these positioning errors is well-approximated by a Gaussian with a standard deviation of 0.1 pixel. In order to account for this, a diagonal matrix (with diagonals equal to 0.1) is added to the estimated covariance matrix.

4.2.3 Bayesian registration

Bayesian registration forms another category of methods of uncertainty and parameter estimation. A Bayesian registration framework consists of choosing a model for the deformation field and characterising its posterior distribution according to Bayes' rule. Some of the recent publications on Bayesian registration are reviewed

here.

A Bayesian framework driven by local image regions is investigated by Toews *et al.* [153] for use in image-guided surgical applications. They aim at multi-modal registration between a pre-operative MR to ultrasound brain images acquired during surgery. The idea is to incorporate local region information for coping with missing tissue between pre and intra-operative images and non-stationary image intensity characteristics. Their joint posterior probability is written as:

$$p(X, T|F, G) \propto p(F, G|X, T)p(T|X)p(X) \quad (4.13)$$

where F and G are the observed pre and intra-operative images. $p(X)$ represents the prior probability over regions defined according to relative sizes or overlap of regions, X and T is the transformation to be estimated. The likelihood $p(F, G|X, T)$ is represented in terms of mutual information (MI). The informativeness of a region is quantified in terms of an MI based criterion as related to spatial location S and intensity I :

$$MI(I|S) = H(I) - H(I|S) \quad (4.14)$$

$$H(I|S) = \sum_{i=1}^n p(S_i)H(I|S_i) \quad (4.15)$$

where $H(\cdot)$ is the entropy function and S represents a spatial location at sub-region image scale defined over a set of n discrete spatial labels. Based on the informativeness of a region, it is selected for the registration. The transformation T is recovered in the form of a global translation between the ultrasound image and the MR slice corresponding to the position of the ultrasound probe via efficient numerical integration.

The results presented in the article indicate that extracting information from local regions leads to an error reduction in comparison to the case when this information is not taken into account. Mean errors in estimated maximum a posteriori (MAP) transformation in (pixels) with respect to the ground truth is 2.5 ± 1.7 and 3.9 ± 2.9 when the local region information is incorporated and when it is disregarded.

Markov chain Monte Carlo sampling based methods

In the context of Bayesian registration, Markov chain Monte Carlo (MCMC) sampling approaches may be used when trying to estimate quantities such as MAP and compute distributions of parameters of interest. An MCMC approach consists of simulating samples from a probability distribution by constructing an ergodic Markov chain with a stationary distribution. The state of this chain after “sufficiently” large number of iterations is used as a sample from the desired distribution. Generally the samples before this are discarded (known as *burn-in*). By doing this,

the effect of the initial state of the chain is forgotten. One of the important factors is the determination of the number of iterations needed by the chain to reach its stationary state.

Metropolis-Hastings (MH) and Gibbs sampler are two widely utilised MCMC approaches based on random walks. In these algorithms, the Markov chain moves in small random steps around the equilibrium value. New candidates are generated according to a distribution known as the proposal distribution. By comparing the probabilities of the current state and the new proposal, the new proposal is either accepted or rejected. While these algorithms are easy to implement and analyse, they take a long time to explore the whole parameter space. More details on MH method and Gibbs sampler can be found in appendix C.1.

Risholm *et al.* [119] describe an elastic Bayesian registration framework that allows the estimation of the deformation field as well as the uncertainty with an application to neurosurgical image guidance. Their implementation is based on the use of FEMs which are numerical techniques for finding approximate solutions to partial differential equations and integrals. Its computations are performed on a mesh that covers the region of interest. In this work, the brain tissue is assumed to be a linear elastic material and a FEM derives a non-uniform tetrahedral mesh conforming to the brain boundaries delineated by grey, white matter and cerebrospinal fluid. When dealing with linear elastic finite element methods, the strain energy is defined as follows:

$$E_r = \frac{1}{2} \int_{\Omega} \varepsilon^T \sigma dx \quad (4.16)$$

where the strain vector ε for a displacement vector $\theta = [\theta_x, \theta_y, \theta_z]$ is defined as follows:

$$\varepsilon = \left[\frac{\partial \theta_x}{\partial x}, \frac{\partial \theta_y}{\partial y}, \frac{\partial \theta_z}{\partial z}, \frac{\partial \theta_x}{\partial y} + \frac{\partial \theta_y}{\partial x}, \frac{\partial \theta_x}{\partial z} + \frac{\partial \theta_z}{\partial x}, \frac{\partial \theta_y}{\partial z} + \frac{\partial \theta_z}{\partial y} \right] \quad (4.17)$$

where σ is the stress vector. Hooke's law describes the relation between stress and strain vectors as $\sigma = C\varepsilon$, where C is the material matrix characterised by Lamé constants, λ and μ . Next, the likelihood term and the prior on the deformation field are modelled as Boltzmann distributions:

$$p(g|\theta) = \frac{1}{Z_s} \exp\left(\frac{-E_s(\theta; f, g)}{T_s}\right) \quad (4.18)$$

$$p(\theta|\lambda, \mu) = \frac{1}{Z_r} \exp\left(\frac{-E_r(\theta; \lambda, \mu)}{T_r}\right) \quad (4.19)$$

where E_s is the similarity criterion (SSD), E_r is the regulariser (Eq. 4.16), $Z_s; Z_r$ are normalising constants and $T_s; T_r$ are temperature parameters of the distribution. f and g are the pre and intra-operative images. Lamé parameters λ and μ are modelled as following a Beta distribution.

They use Metropolis-Hastings algorithm for drawing samples from the posterior distribution of the deformation field and the elastic parameters. If the segmentation of the skull is available, their sampler can be restricted to values inside the intracranial space. The proposal distributions for the deformation field and the Lamé parameters are chosen to be three independent normals given by:

$$N_{\theta}(\theta^*|\theta^i, \sigma_{\theta})N_{\lambda}(\lambda^*|\lambda^i, \sigma_{\lambda})N_{\mu}(\mu^*|\mu^i, \sigma_{\mu}) \quad (4.20)$$

where θ^* , μ^* and λ^* are the new candidates proposed at iteration $(i + 1)$, θ^i , μ^i and λ^i are the values of these parameters at iteration i and σ_{θ} , σ_{λ} and σ_{μ} are the proposal variances.

Convergence is established when the scale reduction [48] for all parameters is less than 1.2. In their experiments, deformations are sampled at 349 boundary nodes. For the remaining nodes, deformation field is determined deterministically. The set of tests with λ and μ sampled uniformly in all tissue types such that $\lambda(GM) = \lambda(WM) = \lambda(CSF)$ and $\mu(GM) = \mu(WM) = \mu(CSF)$ result in a maximum registration error of 2.40 mm. Other tests, where these parameters are sampled non-uniformly, lead to a maximum maximum registration error of 2.43 mm. The registration error is computed as the difference from the ground truth at the boundary nodes. An advantage of their method is that the movement of the boundary nodes is computed by the algorithm and need not be specified.

The same authors also present ways of summarising uncertainty using a robust statistic (inter-quartile range (IQR)) and their visualisation [118]. The objective of this work is to enable sharing of this information with neuro-surgeons for aiding surgical decisions. IQR (difference of the first and third quartile) represents the dispersion of a distribution and can be useful even when dealing with non-Gaussian distributions. While IQR of the marginal distributions is used as a measure of uncertainty, it is visualised as IQR ellipsoids or as maximum IQR scalar maps. An ellipsoid is fitted to the IQRs of samples of deformation at each node and each dimension in the finite element mesh. The dispersion of the marginal distribution on deformations can be visualised with IQR ellipsoids three dimensionally. A scalar measure of the extent of the distribution can be generated by visualising the maximum of the IQRs in each dimension at one node. Fig. 4.2 is an example illustration of IQR maps.

Richard *et al.* [117] present a finite element based Bayesian registration framework. Like Risholm *et al.* [119], they also model the brain as a linear elastic material. SSD serves as the similarity criterion and strain energy of a linear elastic material is the regulariser. Sampling is accomplished by setting up a Metropolis-Hastings sampler. In order to make the Markov chain converge quickly, the authors introduce a scaling parameter and penalisation. It is a coarse to fine strategy where the sampler accepts large deformations in the early iterations and fine displacements are

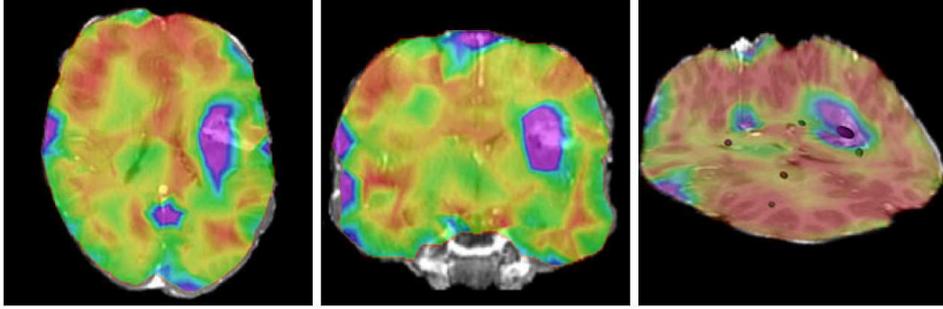


Figure 4.2: Inter-quartile visualisations overlaid on the post-operative image. Images on the left and centre show axial and coronal maximum IQR maps. Pink areas have maximum IQR of 7.5 mm which decreases down to 1.7 mm for red. Large IQRs (pink areas) are seen around the tumour because of the high intensity differences due to blood and oedema; lack of one to one correspondence between pre and post-operative images of the brain. The image of the right is a composite visualisation of intersecting axial, coronal and sagittal planes overlaid with the maximum IQR map and a few IQR ellipsoids. (Source: [118])

corrected in the later iterations. The scaling parameters at iteration i are defined as:

$$\delta_i = \delta_{max}\tau_i + \delta_{min}(1 - \tau_i) \quad (4.21)$$

where δ_{min} and δ_{max} are lower and upper bounds on δ_t variations and $\tau_i = (0.985)^i$. $\delta_{min} = 1$ is the precision to which the deformation field must be computed and $\delta_{max} = 30$ is the maximum expected deformation. This correction approach appears interesting, nonetheless, it may not always be a good strategy to accept large deformations in the beginning and exponentially decreasing the size of deformations accepted as the algorithm progresses. First, it will not allow the sampler to explore the whole parameter space. If the sampler strays away from the solution (which is possible because lower probability states are also accepted), it may become difficult to come back to the right direction as the proposal variance will become smaller and smaller with each passing iteration.

The penalisation term enforces regularisation constraints such that large deformations are accepted initially if they produce a significant correction, gradually loosening the constraint as the sampling progresses. It is chosen at iteration i as:

$$\beta_i = \frac{1}{1 - 0.999\tau_i} \quad (4.22)$$

The penalisation is expressed on the regularisation term as $E_r^{\beta_i}$.

The average registration error between the ground truth and the estimated deformation field at the mesh nodes when the regularisation function is not penalised

is 1.16 ± 2.67 pixels compared to 1.06 ± 2.35 pixels when it is, indicating that use of penalisation does not improve the results significantly. Tests are conducted on Lena image corrupted with a Gaussian noise of $\sigma = 10$.

Bayesian registration methods and atrophy estimation

Bayesian registration methods allow us to characterise posterior distributions of the deformation field from which atrophy information can be extracted. The intractability of the posterior distribution calls for sophisticated sampling methods such as MCMC approaches or numerical integration techniques for computing MAP and gather other information such as uncertainty in registration.

The basis of Bayesian registration methods is the model chosen for quantities of interest. The underlying interest in adopting bio-mechanical models for the brain is that they are connected to the physiological behaviour and material parameters characterising an organ. This choice is in contrast to the use of free form deformation models, whose parameters do not have any physiological basis and are applicable to any 3D object. Risholm *et al.* [118] mention that most successful methods for intra-operative image registration are based on bio-mechanical models. One disadvantage of the linear elastic model of [117, 118] is that it can only handle small deformations efficiently. In the context of longitudinal atrophy estimation, it is not clear if a bio-mechanical model may be more relevant or accurate than its non bio-mechanical counterparts.

Finally, the three methods that are detailed here are developed for image guided surgeries. No effort has been made to apply a Bayesian framework to the problem of atrophy estimation. The main difference between these applications is that, in brain tumour operations, the brain loses a fairly considerable part (a large deformation) as opposed to the occurrence of brain atrophy (usually characterised by smaller deformations).

4.2.4 Key issues in uncertainty estimation

Uncertainty estimation in registration algorithms can be performed with a ground truth [128, 75, 20] and other approaches that involve manipulation of the similarity criterion or the deformation field [73, 60] and Bayesian registration [118, 117].

A ground truth provides one of the easiest ways of estimating errors. Care should be taken to use physically plausible simulations and, if possible, they should be based on some physical evidence. For example, Schnabel *et al.* [128] state that due to confinement on the breast tissue in the scanner RF coil, tissue motion of 10 mm is expected in most of cases. They use these values as boundary conditions for the FEM solver that simulates known deformations for them. Arbitrary simulations may not provide a realistic estimate of uncertainty. Also, the selected model of

simulation of deformations must be different from the one that is tested, in order to avoid biased estimation of error. Other methods that involve manipulation of the similarity criterion or deformation field need to be specifically included within the optimisation of the cost function, but are attractive due to the non-requirement of a ground truth. Besides, a pre-requisite is that the deformation field or the similarity criterion follows a Gaussian distribution for these computations to be valid. As for the Bayesian registration methods, the models for parameters of interest such as the transformation must be selected cautiously. Remember that Bayesian registration methods coupled with MCMC are time consuming approaches.

All these uncertainty methods, though developed and tested for other applications, can be employed or extended for atrophy estimations. The mean registration errors are reported to be 5.7 mm for lung applications [60], 0.1 pixel in satellite applications [63] and, 2.5 pixels [153]; 2.43 mm in [118] in image guided surgical applications for the brain. It should be mentioned here that these errors are high from the point of view of atrophy, where often deformations of this order (1 – 6 mm) are seen and must be quantified to increase sensitivity to volume change measurements.

Also, except for the approach of Jalobeanu *et al.* [63], uncertainty estimation approaches have not attempted to simultaneously examine the bias in estimated registration parameters. Some of the available studies dedicated to bias estimation in atrophy approaches are discussed here.

4.3 Bias in estimated atrophy

The existence of bias in atrophy estimations has been brought to light in many evaluation efforts including ours (Smith *et al.* [142], Camara *et al.* [20], Pieperhoff *et al.* [109], Sharma *et al.* [133]).

The effect of non-rigid registration deformation model and deformation constraints is studied by Rohlfing *et al.* [121] with deformation based morphometry. Three non-rigid registration methods are evaluated: based on a B-spline based transformation model, Demons algorithm and a curvature PDE based registration algorithm. It is revealed that all three methods produce identical overlap of underlying structures but the deformation field and the Jacobian are highly different. They conclude that the statistics performed on Jacobian in DBM are biased due to the underlying models and constraints of the non-rigid registration. Another important finding of this work is the average Jacobian in homogeneous regions is close to the correct volume ratios. Hence, results from DBM are reliable when performed on a segmented region rather than voxel by voxel.

Further, correction of bias has also been attempted, mainly by introducing regularisation terms in non-rigid registration methods. Leow *et al.* [78] have presented

an unbiased non-linear image registration with a new regularisation term based on the logarithm of the Jacobian, which reduces the skewness of the Jacobian map, a requirement when dealing with tensor based morphometry.

According to Leow *et al.*, a registration method can be considered unbiased if it does not produce any spurious changes in the absence of changes. In other words, it must produce Jacobian maps that indicate zero change. The statistical distributions of log-Jacobian maps are studied by defining the Kullback-Leibler (KL) distance on density functions of materials arising in continuum-mechanical models. With this framework, unbiased image registration can be constructed by quantifying the symmetric KL-distance between the identity map and the resulting deformation. They minimise the following functional:

$$\arg \min_{\theta} C(F, F(\theta^{-1}), G, G(\theta)) + \text{wt}(KL(\text{pdf}_{\theta}, \text{pdf}_{id}) + KL(\text{pdf}_{id}, \text{pdf}_{\theta^{-1}})) \quad (4.23)$$

where C is a similarity criterion, F and G are the two images to be registered, θ and θ^{-1} are the recovered deformation field and its inverse, wt is the weight parameter. pdf_{id} and pdf_{θ} are probability density functions of the identity mapping and the deformation θ . The symmetric KL distance is defined as:

$$KL(\text{pdf}_{\theta}, \text{pdf}_{id}) + KL(\text{pdf}_{id}, \text{pdf}_{\theta^{-1}}) = \quad (4.24)$$

$$\int (|J_{\theta}(x)| - 1) \log |J_{\theta}(x)| dx = \quad (4.25)$$

$$\int (|J_{\theta^{-1}}(x)| - 1) \log |J_{\theta^{-1}}(x)| dx \quad (4.26)$$

A serial MRI example showing the Jacobian map estimated by Leow *et al.* [78] and another viscous fluid method is depicted in Fig. 4.3. The reduction of inverse consistency errors is demonstrated by comparing the deformations obtained by switching the order of the source and target images. Since this is a symmetric formulation, the deformation should not depend on the order of the input images. For a maximum displacement of 4.2, the error for inverse consistent mapping is 0.8616 ± 0.0115 whereas the error for the same experiment with inverse inconsistent mapping is higher, 0.9617 ± 0.0685 .

Yanovsky *et al.* [170] extended this unbiased method to an unbiased symmetric registration method. Symmetry comes from the regularisation term that is used (see Eq. 4.26). From tests conducted on baseline and repeat scans from the ADNI database, they conclude that unbiased methods both symmetric and asymmetric are less inclined towards producing biased estimates.

Yushkevich *et al.* [174] have established that DBM produces biased results (over-estimation of atrophy) when estimating hippocampal changes. A non-zero regression line fitted to 6 and 12 month rates of hippocampal atrophy of controls and

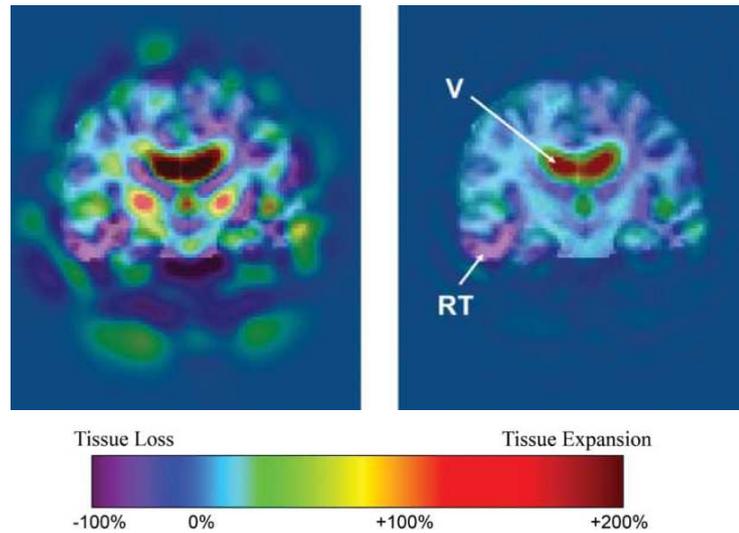


Figure 4.3: A 3-D serial MRI example showing Jacobian maps for a viscous fluid method (left) and (right) the method of Leow *et al.* [78]. Right temporal atrophy (RT) and ventricular enlargement (V) can be easily discerned in the Jacobian map generated using the method of Leow *et al.* [78]. The viscous fluid method generated a comparatively noisy map. (Source: [78]).

patients with mild cognitive impairment (MCI) from ADNI confirms the presence of an additive bias in estimations. Explicit estimation of this additive bias is performed by registering baseline and repeat scans of subjects acquired on the same day. They examined the effect of symmetric and asymmetric global and deformable transformations on atrophy estimations employing ANTS [10] and Rueckert *et al.*'s [124] free form deformation algorithms. A symmetric registration algorithm forces the computed deformations to be the same regardless of which image is the target *i.e.*, if the reference and target images change places. Their study concludes that effect of interpolation (when applying global and local transforms) affects the asymmetric DBM much more than the symmetric configuration. Also, asymmetry does not have significant effect on the power of MCI-control group difference comparison. Experiments on MCI and control groups with the intercept based method show that asymmetry in the application of the global transform results in a bias of 2–3% when ANTS is used, while asymmetry in the deformable registration has less effect on bias. Finally, Yushkevich *et al.* suggest that symmetric transforms be used in order to reduce the effect of bias in estimated atrophy.

Recently Hua *et al.* [58] have endorsed the presence of biased estimations with TBM with the hippocampus as the region of interest. They assume that any bias in atrophy estimates has a constant additive offset and a second component whose magnitude may depend on the true level of atrophy. Like Yushkevich *et al.* [174], Hua *et al.* find that asymmetric registration is the main reason of biased atrophy estimations. As a remedy, they reimplemented their registration method to be fully

inverse-consistent by further symmetrising the similarity metric MI and using a symmetric-KL regularisation term (Eq. 4.26). By enforcing inverse-consistency in their TBM method, the bias (offset) reduced from 1.4% to 0.28%. Transitivity error accounted for the remaining offset. By transitivity, it is meant that atrophy calculations between time points 1 and 2 and subsequent calculations between time points 2 and 3, should add up to direct calculations between times points 1 and 3. This error is computed through any conflicts in the direct and the indirect transforms. It is also mentioned that transitivity errors may lead to an additive or multiplicative bias present in all mappings, which may relate to the true change, or with the estimated amount of change.

Fox *et al.* [43] study recent works on TBM [58, 174] and suggest that accurate examination of longitudinal brain atrophy can be made possible by assessing methods through: simulation of atrophy by taking into account that simulated atrophies seldom represent the complex phenomena in the human brain perfectly and sophisticated simulations may be susceptible to biases of their own, symmetry and transitivity of measurements, manual measurements on large datasets could expose systematic biases in methods, reproducibility assessed with scan-rescan techniques and comparison with known disease biology to decide if the changes measured are probable.

In summary, bias in atrophy estimations from deformable registration algorithms is due to the asymmetric estimation as well as application of the global and the local transforms. Ensuring symmetry of the estimated transformations is a way of minimising the unwanted bias in atrophy measurements.

Most approaches for uncertainty estimation in image registration methods do not account for bias in atrophy estimates. What we describe in this chapter, is a framework that can be applied to any atrophy estimation approach as the uncertainty estimations are performed separately and compensates for biases in atrophy estimations. We will show later that depending on the NRR algorithm used, bias may be proportional to the true atrophy value. In such cases, bias cannot be learnt from scan-rescan images as is the case in [174].

In this chapter, a generic framework for estimating uncertainty in longitudinal brain atrophy by means of constructing confidence intervals is discussed. The proposed framework is described in section 4.4 consisting of learning uncertainties from example simulations using a ground-truth (section 4.4.1) and construction of confidence intervals (section 4.4.2). Results are shown on three NRR methods [98, 10, 157] on real images in section 4.5.

4.4 Proposed framework for uncertainty estimation

Let us suppose that we would like to test a set of longitudinally acquired images of a patient, $I(t_1)$ and $I(t_2)$, at times t_1 and t_2 , for existence of (longitudinal) atrophy. Using an atrophy estimation method, m , an atrophy change map $c_{I(t_1;t_2)}^m$ is obtained. A change map represents the volumetric changes that the brain has endured between times t_1 and t_2 .

Our aim is to define confidence intervals for $c_{I(t_1;t_2)}^m$. For a given method m , error distributions are learnt for different atrophies and example anatomies by means of creating a database. This will be referred to as the “learning database”. The construction of confidence intervals is accomplished via errors distributions in the learning database.

4.4.1 The learning database

The learning database is constructed with two objectives: to be able to capture the role of MRI artefacts and that of bias in measurements in the mis-estimation of atrophy.

Firstly, a ground truth is designed with an atrophy simulation algorithm capable of introducing synthetic local and global changes in the brain volume. The idea is to carry out simulations of varying magnitudes of volumetric changes in a region of interest in order to obtain a number of ground truths. For this purpose, the method described in chapter 3.2 is employed.

Let $\mathcal{A} = \{a_n\}$ be the set of all desired atrophies a_n to be included in the learning database, $n = 1, 2, \dots, N$, where N is the number of atrophies. An atrophy could be defined as each voxel of the region of interest shrinking uniformly or non-uniformly (based on some atrophy pattern discerned from real cases).

If the database is required to be patient independent, different patient anatomies must be taken into account in the process of its creation. Let us denote each image in the image database, comprising P patients, used in the construction of the learning database as $\mathcal{I}(p, t)$. This represents image for a patient p acquired at time instant t where $p = 1, 2, \dots, P$. For the database, we consider only the first time point of each of these patients. Atrophies are simulated in the image $\mathcal{I}(p, 1)$ (first time point) such that

$$\mathcal{I}(p, 1) \xrightarrow{a_n} \mathcal{I}(p, t_n) \quad (4.27)$$

where $a_n \in \mathcal{A}$ and $\mathcal{I}(p, t_n)$ is the n^{th} atrophied version of $\mathcal{I}(p, 1)$ for the p^{th} patient.

Intuitively, this process represents shrinking of the region of interest over time each time undergoing an atrophy $a_n \in \mathcal{A}$. In essence, this step allows us to generate images of patients with known changes in the brain, starting from the baseline, in

a progressive manner. In our implementation, $a_1 < a_2 < \dots < a_n \dots < a_N$ and the region of interest shrinks by $a_n\%$ *i.e.*, average atrophy in this region is $a_n\%$

A source of variability is MR artefacts (such as noise, bias-field inhomogeneity) that can be incorporated in this analysis. By example of MR noise addition, the inclusion of MR artefacts in this framework is explained. MRI noise is approximated with Gaussian noise, which is a valid assumption when $SNR \geq 2$ dB [49]. Let MRI noise, η , is a random variable distributed as $\mathcal{N}(0, \sigma^2)$. The value of σ is chosen depending on the image acquisition system and can be computed from noise estimation methods discussed in section 2.4.2. Note that several independent realisations of noise need to be added to the images in order to obtain the distribution of the estimated atrophy. Independently and identically distributed realisations of Gaussian noise are added to the baseline and the atrophied images in the following manner,

$$\mathcal{I}^k(p, 1) = \mathcal{I}(p, 1) + \eta_k \quad (4.28)$$

$$\mathcal{I}^k(p, t_n) = \mathcal{I}(p, t_n) + \eta_k \quad (4.29)$$

where

$$\begin{aligned} n &= 1, 2, \dots, N \\ k &= 1, 2, \dots, R. \end{aligned}$$

R is the total number of noise realisations. The ensemble of images after atrophy simulation and addition of noise will be referred to as \mathcal{L} in the forthcoming discussion:

$$\mathcal{L} = \left\{ \mathcal{I}^k(p, 1), \mathcal{I}^k(p, t_n); p = 1, 2, \dots, P; n = 1, 2, \dots, N; k = 1, 2, \dots, R \right\} \quad (4.30)$$

The next step is to estimate volumetric change maps by applying method m on the images in \mathcal{L} , which will enable us to estimate the error distribution of each atrophy in \mathcal{A} as the true atrophy is known. Estimation of volumetric change maps is denoted by $c_{\mathcal{I}^k(p;1;t_n)}^m$, obtained for the n^{th} simulated atrophy, k^{th} artefact realisation and p^{th} patient. If we assume that these estimations follow a Gaussian distribution, the ML parameters of the expected distribution (μ_n, σ_n^2) can be estimated from them for each atrophy as the sample mean,

$$\mu_n = \frac{1}{P \cdot R} \sum_{p=1}^P \sum_{k=1}^R c_{\mathcal{I}^k(p;1;t_n)}^m \quad (4.31)$$

and variance

$$\sigma_n^2 = \frac{1}{P \cdot R} \sum_{p=1}^P \sum_{k=1}^R (c_{\mathcal{I}^k(p;1;t_n)}^m - \mu_n)^2 \quad (4.32)$$

Since the R realisations for an atrophy a_n , forming the learning database, are assumed to follow a Gaussian distribution with parameters (μ_n, σ_n^2) , the errors \mathcal{E} also follow a Gaussian distribution with parameters,

$$e_n = (b_n, \sigma_n^2), \quad (4.33)$$

and

$$b_n = a_n - \mu_n, \quad (4.34)$$

Here, b_n is the bias in the estimated atrophy with respect to the ground truth and,

$$\mathcal{E} = \{e_1, e_2, \dots, e_n\}, n = 1, 2, \dots, N \quad (4.35)$$

The difference $(a_n - \mu_n)$ allows us to incorporate in the estimated error distributions, the method dependent bias in atrophy estimations arising from the errors in atrophy modelling and MRI artefacts. Despite the fact that building this database is computationally intensive, the framework is practically implementable because once the database is created, it can be reused for uncertainty estimations for any patient.

To summarise, the database \mathcal{E} contains the distribution of errors for varying magnitudes of atrophies. The procedure of learning database formation is represented pictorially in Fig. 4.4.

4.4.2 Confidence interval calculation

At this stage, the learning database is made use of in order to obtain confidence intervals. Let $c_{I(t_1;t_2)}^m$ be an atrophy change map estimated by method m on two longitudinally acquired MR images in a region of interest. We would like to construct confidence intervals for this atrophy change map.

$c_{I(t_1;t_2)}^m$ is classified into one of the most probable of the N atrophy classes of the learning database so as to be able to associate an error distribution with this atrophy change map. That is to say, the closest distribution that could have generated $c_{I(t_1;t_2)}^m$ is chosen and the corresponding error distribution is taken for estimating confidence intervals. Employing z -score as a distance measure, all distributions satisfying the following inequality are accepted at first, where α is the desired significance level:

$$|z| \leq z_{\frac{\alpha}{2}} \quad (4.36)$$

The selection of atrophies is shown in Fig. 4.5 and the z -score is calculated for each atrophy a_n having a distribution (μ_n, σ_n) as:

$$z_n = \frac{\mu_n - c_{I(t_1;t_2)}^m}{\sigma_n} \quad (4.37)$$

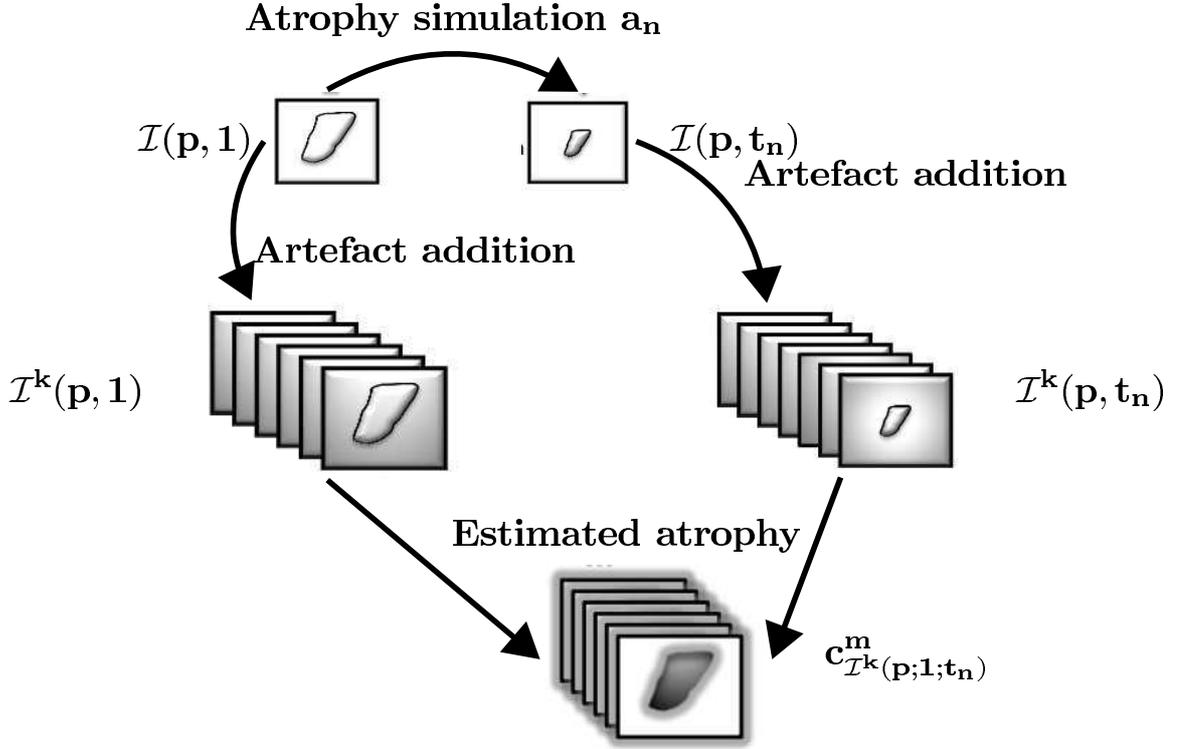


Figure 4.4: Learning database construction for atrophy a_n . Atrophy a_n is simulated in the region of interest in image $\mathcal{I}(p, 1)$ (baseline image) of patient p to obtain its atrophied version $\mathcal{I}(p, t_n)$. As a second step, artefacts are added to $\mathcal{I}(p, 1)$ and $\mathcal{I}(p, t_n)$ for building the learning database $\mathcal{L} = \{\mathcal{I}^k(p, 1), \mathcal{I}^k(p, t_n); p = 1, 2, \dots, P; n = 1, 2, \dots, N; k = 1, 2, \dots, R\}$. Atrophy estimations by a method m generate R change maps $c_{\mathcal{I}^k(p;1;t_n)}^m$. This process is repeated for all N atrophies represented in \mathcal{A} and each patient $p \in P$.

If no atrophies in \mathcal{A} satisfy the z -score criterion, the atrophy that we are looking for does not exist in the learning database. Consequently, the scope of the database must be expanded. The expansion of the database requires the addition of more ground truths and hence is a matter of more computational time.

Let $\mathcal{A}' \subset \mathcal{A}$ be the set of accepted atrophies. Assuming that \mathcal{A}' is not empty, the distribution with the minimum z -score, denoted by z_{min} , is selected from \mathcal{A}' and the corresponding error distribution parameters are used for estimating confidence intervals. Let $a_{n'}$ be the atrophy with z -score z_{min} then the confidence interval ϕ for $c_{I(t_1;t_2)}^m$ is given as:

$$\phi = c_{I(t_1;t_2)}^m + (a_{n'} - \mu_{n'} \pm z_{\frac{\alpha}{2}} \sigma_{n'}) \quad (4.38)$$

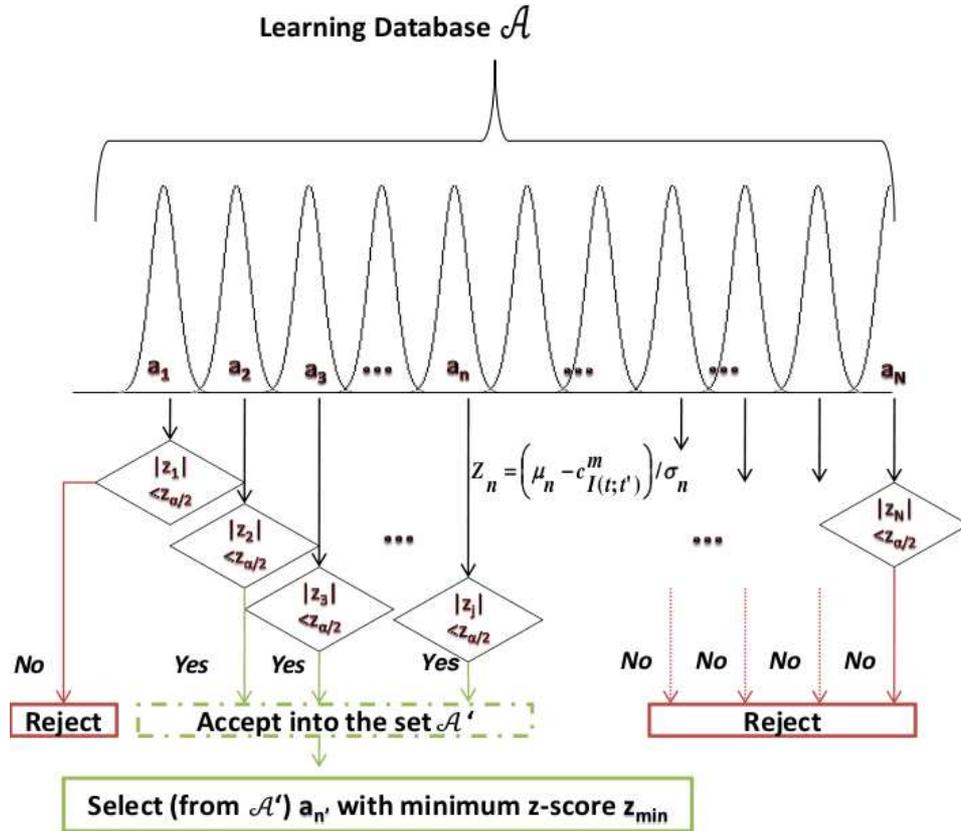


Figure 4.5: An example of selecting the closest atrophy distribution from the learning database. z -scores are computed for each atrophy a_n in \mathcal{A} present in the learning database as $z_n = \frac{\mu_n - c^m I(t_1; t_2)}{\sigma_n}$. From the database \mathcal{A} , all atrophies satisfying the relation $|z| \leq z_{\frac{\alpha}{2}}$ are selected in the database \mathcal{A}' at first. In the second step, the error distribution corresponding atrophy with the minimum z -score z_{\min} within the set \mathcal{A}' is chosen as the closest distribution.

4.5 Results

All experiments are carried out on three non-rigid registration algorithms (Noblet *et al.* [98], ANTS [10] and Vemuri *et al.* [157]). We demonstrate the construction of the learning database and the estimated confidence intervals, followed by a quantitative evaluation of the confidence intervals in terms of coverage probability and length of the interval. The implementation details and parameters used for experiments with Noblet *et al.* [98], ANTS [10] and Vemuri *et al.* [157] are explained in appendix B.1.

4.5.1 Learning database construction

The learning database is constructed with 18 anatomically different in-vivo 3D images from the IBSR database which contains real MR image data with manually guided segmentations. Our experiments have been performed on T1-weighted MR image data from 18 anatomies with a slice thickness of 1.5 mm.

This study will focus on the change in hippocampus' volume for MS patients. Anderson *et al.* report the annual mean decrease in hippocampal volume in MS patients as 9.4% on the right and 8.9% on the left [3]. Therefore, atrophies surrounding these are included in the learning database. This step is accomplished by simulating uniform volume changes in the range of 1 – 15% (with step size of 1%) in the hippocampus using the algorithm described in section 3.2.1. Variability is introduced by degrading the images by addition of Gaussian noise (SNR = 35 dB). The noisy image is verified visually to confirm that this choice of SNR results in a realistic image.

A total of 108 independent realisations (6 per patient) are simulated for each atrophy for the learning database of all the three NRR algorithms. An example of the image that has been used and simulated changes in the hippocampus are shown in fig. 4.6 (a-b).

4.5.2 Gaussian assumption

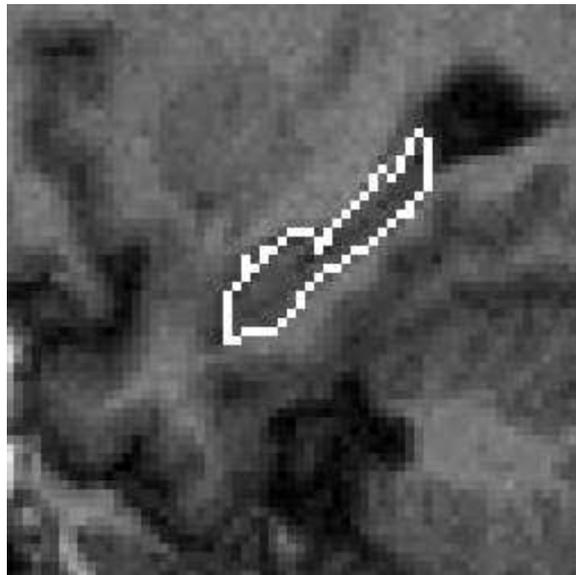
The error distribution of the estimated atrophies for the three methods [98, 10, 157] are learnt by examining the error distributions of the estimated Jacobians. Applying logarithm to the Jacobian (J) renders the distribution symmetric as hypertrophy ($J > 1$) and atrophy ($0 < J < 1$) are scaled evenly. Thus, all calculations are performed with the average values of the log-transformed Jacobian over the hippocampus, which we will refer to as hippocampal log-Jacobian and the corresponding atrophy as hippocampal atrophy in the future sections. They are computed as follows:

$$\log J_{\text{avg}} = \frac{1}{V} \sum_{x=0}^N \log(J(x)) \quad (4.39)$$

$$a_{\text{avg}} = (1 - e^{\log J_{\text{avg}}}) \cdot 100 \quad (4.40)$$

where $\log J_{\text{avg}}$ is the average Jacobian (hippocampal log-Jacobian), $\log(J(x))$ is the logarithm of the Jacobian at voxel x , a_{avg} is the average atrophy in percentage. N is the total number of voxels in the hippocampus.

All the log-transformed Jacobian distributions in the learning database are verified to follow a Gaussian distribution. A rule of thumb is to accept the null hypothesis that a distribution is Gaussian at a significance level of 95% if its skewness



(a)



(b)

Figure 4.6: Figure depicting (a) A sample image degraded with Gaussian noise ($SNR=35dB$) with hippocampus delineated and (b) simulated changes (as the difference between the baseline image and after simulating a hippocampal atrophy of 10.89% in the baseline image).

and kurtosis lie in the interval $[-2, 2]$. Table 4.1 lists the skewness and kurtosis values for all distributions indicating that Gaussian assumption indeed holds for the

three registration methods. Histograms depicting these distributions of errors in estimated log-Jacobians are shown in Fig. 4.7 for true atrophies of -14.87% , -5.94% and -1.67% for all the NRR methods.

True Atrophy (in %)	Skewness			Kurtosis		
	Noblet <i>et al.</i>	ANTS	Vemuri <i>et al.</i>	Noblet <i>et al.</i>	ANTS	Vemuri <i>et al.</i>
-14.87	-0.8330	0.6612	-0.3074	0.9344	0.3226	-0.7560
-13.93	-0.9556	-0.0428	-0.7523	1.0940	-0.4609	0.8238
-12.94	-1.0371	0.4143	-0.5932	1.1469	0.3803	0.9243
-11.90	-0.8488	-0.0169	0.1171	0.8425	-0.0480	-0.3756
-10.89	-0.9541	0.1617	-0.0521	0.8609	-0.1399	-0.9736
-10.23	-0.7933	0.6162	0.1180	0.3858	0.3735	-0.5756
-8.98	-0.9425	-0.1468	0.1955	0.6657	0.3792	0.5815
-7.98	-0.6529	-0.0446	0.2446	-1.1387	1.6377	-0.5038
-7.39	-0.8625	0.0555	-0.2704	0.2293	0.4361	-0.4690
-5.94	-1.1272	0.2257	-0.0839	1.5227	0.3059	-0.7118
-4.93	-0.8600	0.3880	-0.2884	0.9519	0.5673	-0.8076
-4.05	-0.9099	0.3757	-0.6436	0.7589	0.3387	-0.4809
-2.98	-0.8568	-0.0654	-0.2648	0.6254	-0.7390	-0.6419
-1.67	-0.9949	0.0199	-0.0691	0.7735	0.0812	-0.7539

Table 4.1: Skewness and kurtosis of log transformed Jacobian distributions. Skewness and kurtosis are within the range of $[-2, 2]$ for all the simulated atrophies.

4.5.3 Bias in atrophy estimations

Average atrophies (Eq. 4.40) and the bias (Eq. 4.34) in their estimations computed for atrophies in the learning database are illustrated in Fig. 4.8 (a-c), where each NRR algorithm shows a different pattern in bias. For Noblet *et al.*'s algorithm, the bias remains almost constant with the true atrophy value (bias in the range of $(-3.38\%, -4.22\%)$). On the other hand, ANTS method's bias increases with increasing ground truth atrophy (fig. 4.8) (bias in the range of $(+0.72\%, +5.39\%)$). Bias is inversely proportional to true atrophies for the estimations of Vemuri *et al.* (bias in the range of $(+0.48\%, -8.86\%)$). Negative bias indicates an over-estimation of atrophy and vice-versa. The key point here is that the bias may depend on atrophy values. Hence, the bias in atrophy estimations that are obtained by registering baseline and repeat scan may not always hold true.

4.5.4 Quantitative evaluation

The accuracy of the confidence intervals can be evaluated on the basis of two criteria: coverage probability and length of the confidence interval. For ensuring that the calculation of coverage probability and the length of the interval is not biased, a leave one out approach is followed. The patient for whom confidence intervals are constructed is not included in the training of the learning database. Considering a

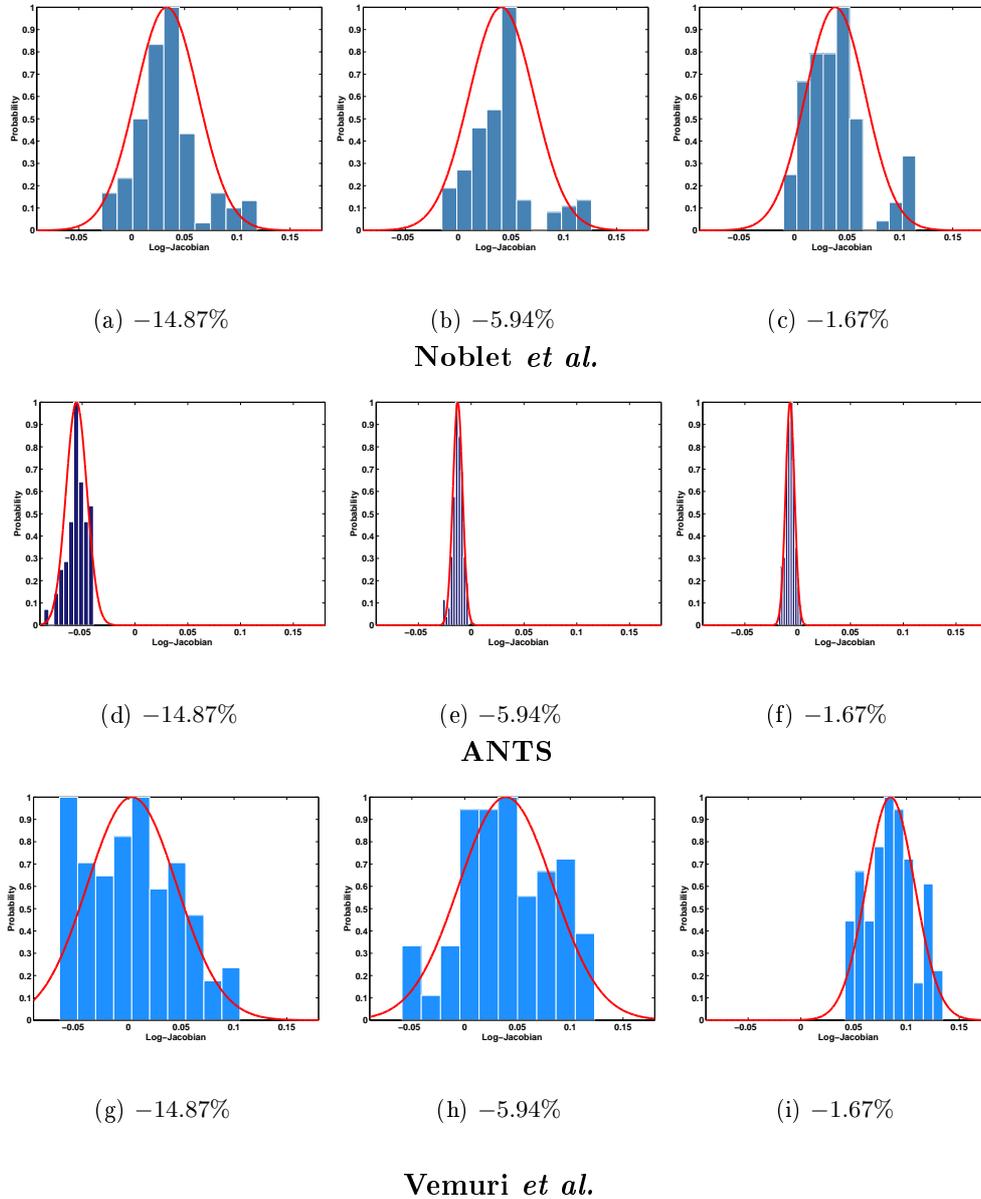
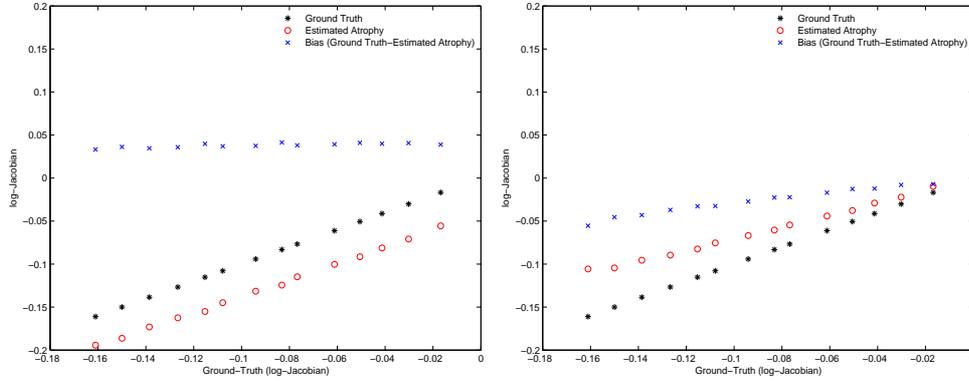


Figure 4.7: Histograms showing the distributions of error (in log-Jacobian scale) when the true atrophies are -14.87% , -5.94% and -1.67% for the three NRR methods. Note that the bin size has been set to 10 and the plots have been shown on the same scale. By doing so, the spreads of the distributions of the three methods under consideration can be compared. The plots allow us to conclude that ANTS is associated with the least amount of spread among the three methods which will translate later into shorter length confidence intervals.

database of P patients, this process is repeated P times and the coverage probability and the length are averaged at the end.

(a) Noblet *et al.*

(b) ANTS

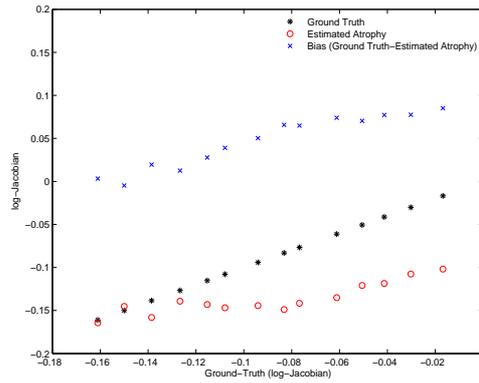
(c) Vemuri *et al.*

Figure 4.8: Bias b_n in the estimated atrophies used in building the learning database \mathcal{L} in terms of log-Jacobian for the methods being studied. Bias is calculated as the distance of the estimated atrophy from the ground truth atrophy.

Coverage probability (or actual coverage) refers to the number of times the confidence interval contains the true value. When confidence intervals are constructed, they are expected to contain the value of interest with a certain probability. This is known as nominal coverage, like in our case, it is usually set to 95%. Nominal and actual coverage are equal when the assumptions made in deriving confidence intervals are met. In our case, there are two assumptions: the error distributions follow a Gaussian distributions and the bias in atrophy estimations is additive in nature. The actual coverage can be lower or higher than the nominal coverage if the assumptions are not properly satisfied and can be used for evaluating the accuracy

of the constructed confidence intervals. In these set of experiments, the coverage probability is simply the number of test samples for which, the ground truth atrophy is contained in the interval. Coverage probability is calculated with 100 test samples per atrophy. The results are shown for log-transformed hippocampal Jacobian in Table 4.2 indicate that the learning database is capable of providing an estimate that contains the true value between 89.81 – 93.52%, 92.59 – 100% and 90.74 – 98.18% with the NRR methods of Noblet *et al.*, ANTS and Vemuri *et al.*, respectively. These results confirm that the Gaussian assumption is satisfied and bias correction is effective for ANTS and Vemuri *et al.* as they are able to achieve a coverage close to the desired coverage of 95% while Noblet *et al.*'s approach shows a poorer coverage in comparison to the other two methods. The coverage probability did not depend on the atrophy.

Length of the interval depends on how well the non-rigid registration algorithms cope with the sources of variations such as images artefacts, differing anatomies and model errors. Length of the intervals is proportional to dispersion of data (standard deviation). Compact confidence intervals are more informative as they localise the value of interest in a smaller range. Table 4.2 also displays the mean lengths of the confidence intervals in percentage of hippocampal atrophy. With NRR methods of Noblet *et al.*, ANTS and Vemuri *et al.*, the maximum lengths of confidence intervals are 14.32%, 4.82% and 8.94%, respectively. According to these results, ANTS method provides meaningful confidence intervals as opposed to algorithms of Noblet *et al.* and Vemuri *et al.* that provide large intervals. Results also allowed us to conclude that the length of the interval is dependent on the value of the atrophy. Higher values of atrophy resulted in larger confidence intervals in all the three methods.

Constructed confidence intervals The premise of the learning database is that the uniformly simulated atrophy in the hippocampus represents atrophies in a real scenario. If this assumption holds, the error distributions would be capable of finding reliable confidence intervals. In order to demonstrate the ability of the framework to estimate more general atrophies, 14 atrophies are simulated such that each voxel undergoes a different atrophy but the mean atrophy over the hippocampus still lies within the scope of the learning database. These simulations are performed for one patient of the IBSR database. Note that this patient is removed from the learning database before computing the intervals.

The confidence intervals estimated for these atrophies are shown in Table 4.3. It is indeed possible to estimate atrophies for both methods. These results are in agreement with those presented in the previous section confirming the superior performance of ANTS method in terms of lengths of intervals, over those of Noblet *et al.* and Vemuri *et al.*. For Noblet *et al.* and Vemuri *et al.*, the true atrophy is found within the interval for all the atrophies in the learning database. However, for

True Atrophy (in %)	Coverage Probability			Interval Length		
	Noblet <i>et al.</i>	ANTS	Vemuri <i>et al.</i>	Noblet <i>et al.</i>	ANTS	Vemuri <i>et al.</i>
-14.87	0.9259	0.9722	0.9815	8.9535	4.8200	14.3168
-12.94	0.9074	1.0000	0.9444	9.2716	3.5070	12.7654
-11.90	0.9167	0.9630	0.9444	9.5253	3.4588	11.1728
-10.89	0.9167	0.9907	0.9259	9.2262	3.0040	11.6553
-10.23	0.9259	0.9907	0.9537	9.5072	2.7028	15.5321
-8.98	0.9259	0.9907	0.9444	9.6157	2.7806	13.3435
-7.98	0.9074	0.9722	0.9074	9.4348	2.3324	11.3982
-7.39	0.9074	0.9259	0.9352	10.1654	2.7612	12.2597
-7.39	0.8889	0.9352	0.9352	10.2372	2.6639	10.4151
-5.94	0.9352	0.9444	0.9722	8.9444	1.9213	10.6057
-4.93	0.9352	0.9537	0.9722	9.2625	1.5774	8.8066
-4.05	0.9074	0.9259	0.9444	9.3170	1.3015	9.1981
-2.97	0.9352	0.9907	0.9630	8.9353	1.2422	8.4815
-1.67	0.8981	0.9352	0.9818	8.4605	1.4001	7.4677

Table 4.2: Average coverage probability using log-Jacobian values and lengths (% of volume loss with respect to the total hippocampal volume) of the constructed confidence intervals. All but one of the intervals contain the true value of atrophy. This interval is shown in red.

ANTS for 1 out of 14 atrophies the true value is outside the constructed confidence interval, which is highlighted in Table 4.3.

True Atrophy (in %)	Confidence Interval		
	Noblet <i>et al.</i>	ANTS	Vemuri <i>et al.</i>
-14.87	(-19.9402, -12.0319)	(-17.9408, -12.6115)	(-17.8160, -3.5021)
-13.93	(-18.7320, -10.1951)	(-16.4603, -12.4954)	(-14.2921, -0.9400)
-12.94	(-18.0697, -9.4958)	(-14.7293, -10.6740)	(-15.5948, -3.9525)
-11.90	(-16.9299, -8.1880)	(-12.8147, -9.1926)	(-19.9626, -8.8496)
-10.89	(-15.7578, -6.7546)	(-10.3662, -7.0424)	(-22.7274, -5.6283)
-10.23	(-15.3949, -5.7069)	(-12.7324, -9.3787)	(-13.7539, 3.0007)
-8.98	(-14.2713, -4.8983)	(-10.6501, -7.8014)	(-14.0809, -0.4525)
-7.98	(-14.3404, -3.6037)	(-9.9778, -6.5600)	(-13.8439, 1.9576)
-7.39	(-13.3304, -2.3820)	(-8.9090, -5.5201)	(-11.1788, 2.7105)
-5.94	(-11.2729, -2.7607)	(-7.6454, -5.1987)	(-11.4884, 2.5558)
-4.93	(-10.3265, -0.9738)	(-6.7062, -4.7089)	(-12.2917, -1.1914)
-4.05	(-9.8035, -0.0351)	(-5.0100, -3.3159)	(-8.6878, 4.2306)
-2.97	(-4.2373, 0.5599)	(-3.7908, -2.2303)	(-7.6977, 4.3176)
-1.67	(-7.7270, 1.4471)	(-3.2507, -1.5254)	(-6.5251, 4.8842)

Table 4.3: Constructed Confidence Intervals. For ease of interpretation, the upper and lower bounds are shown in terms of % atrophy.

4.6 Discussion and findings

In this chapter, we discussed a novel and generic framework for constructing confidence intervals by an atrophy estimation method. A learning database was built in order to obtain an estimate of error in the measured atrophy through simulation of different magnitudes of volumetric changes and addition of artefacts. The main feature of this database is that it is capable of correcting for errors that may arise due to the bias in atrophy estimations as well as of capturing the role of MRI artefacts in the mis-estimation of atrophy. This database is flexible as any atrophy (*e.g.* in a region of interest such as the hippocampus or the entire brain) and other image artefacts, depending on the user's requirement can be included. Atrophies that compose the learning database can be chosen on the basis of the disease and region of interest being studied.

This framework was employed for comparing the performance of three NRR algorithms, by Noblet *et al.* [98], ANTS [10] and by Vemuri *et al.* [157] for building confidence intervals for hippocampal atrophies. Images were degraded with a Gaussian noise (SNR = 35 dB) and no pre-processing was applied.

The bias in estimations revealed different patterns among the three methods. For Noblet *et al.*'s algorithm, the bias remains almost constant with the true atrophy value (bias in the range of $(-4.22\%, -3.38\%)$). On the other hand, ANTS method's bias increased with increasing ground truth atrophy (bias in the range of $(+0.72\%, +5.39\%)$). Bias was inversely proportional to true atrophies for the estimations of Vemuri *et al.* (bias in the range of $(-8.86\%, +0.48\%)$). Negative bias indicates an overestimation of atrophy and vice versa. In case of Noblet *et al.*, the bias in estimations may be because of the simultaneous use of the regulariser and constraints that were applied to preserve the topology. We believe that the level set based approach of Vemuri *et al.* is prone to errors with noisy images as the deformation vectors are scaled by a function of the difference of intensities in the two images. Due to this reason, it is possible that the direction of the deformation is correct but the magnitude is erroneous. For small atrophies, lowest bias values are observed for ANTS, possibly due to the configuration of ANTS that estimates a symmetric transformation used in this study. This result is in agreement with existing studies in bias estimation [58, 174] that identify the use of a symmetric transformation as a means of obtaining unbiased atrophy estimates. From our results, it appears that large atrophy estimations are biased, in spite of using symmetric transformation and fluid flow transformation model that allows us to measure large deformations. A possible explanation is the use of Gaussian regularisation that penalises large deformations more than small ones. As compared to Yushkevich *et al.* [174] that provide the bias estimation from scan rescan method to be between 2 – 3% for ANTS, the bias in our case for low atrophies is smaller (for an atrophy of -1.68% the bias is 0.7174%). Reasons for this include differences in the quality of the ADNI database that Yushkevich *et al.* use and IBSR database in our case. Secondly, in our case a

global transform was not estimated as the observations were already aligned with only deformable changes between them.

Tests were carried out with leave one out method in order to ascertain that new anatomies can be detected without being used in training of the learning database. Results have highlighted that the ANTS algorithm produced shortest intervals as compared to NRRs of Noblet *et al.* and Vemuri *et al.*. Maximum lengths of confidence intervals are 14.32%, 4.82% and 8.94% for methods of Noblet *et al.*, ANTS and Vemuri *et al.*, respectively for the highest atrophy in the database (-14.87%). Large confidence intervals were observed for large atrophies for all the 3 NRR approached. Hence, these methods localised smaller atrophies better than large ones.

The desired coverage probability was 95% whereas the achieved coverage probabilities were 89.81–93.52%, 92.59–100% and 90.74–98.18% with the NRR methods of Noblet *et al.*, ANTS and Vemuri *et al.*, respectively. Through this study, it can be concluded that the ANTS approach provides meaningful confidence intervals as compared to the approaches of NRRs of Noblet *et al.* and Vemuri *et al.*

All three NRR methods exhibited different bias patterns in their estimations. These patterns depended on the ground truth atrophy value in the database for ANTS and Vemuri *et al.* while bias in estimations was approximately constant for Noblet *et al.*. Such observations affirm the need for learning errors for each method and separately for atrophies depending on their magnitudes. In this work, only additive bias was considered. Other models such as multiplicative bias could be explored in the future.

Our tests were based on addition of noise. The effect of bias-field inhomogeneity and other artefacts such as geometrical distortion was not considered. However, it is possible to incorporate these artefacts within our framework. Also, the effect of different levels of noise was not tested. Note that adding more variable factors would lead to higher variability in results and a poorer localisation of atrophy. A limitation of this database is that it works for images acquired on the same scanner with similar noise characteristics. Huge differences in the quality of the images will not let this database function efficiently. Therefore, we recommend that in a real scenario, this database be built for a specific scanner. MR artefacts with parameters pertaining to the scanner can be included in this framework. For instance, Rician noise can be used instead of Gaussian noise that was used in our tests, whose parameters can be learnt from the acquired images. Further, depending on the type of coil used the bias field inhomogeneity characteristics and geometrical distortion characteristics can be a part of the database.

Apart from anatomies and MR artefacts, another dimension of training of this database comes from atrophies. A more pragmatic approach would be to learn realistic patterns of brain deformations in the region of interest from real subjects.

Simulations conforming to realistic voxel-wise atrophies can be created, thereby making the database practically usable.

When constructing such a framework, an inherent assumption is that a “good” segmentation of the region on interest is available. The fact that atrophy is averaged over this segmentation renders its accuracy indispensable to the correct functioning of this database and NRR algorithms used for measuring atrophy in general.

In terms of computational complexity, it took 5 – 6 minutes, 1 – 2 minutes and 3 – 4 minutes with the algorithms of Noblet *et al.*, ANTS and Vemuri *et al.*, respectively, per image pair on images of size $64 \times 64 \times 64$. This means that the creation of a database of 100 images is a matter of 1 day for the algorithm of Noblet *et al.* and few hours of computer time for the other two methods. However, remember that this computational cost is one time investment. This database can be reused in future calculations of uncertainty.

A future task is to compare this study to confidence intervals extracted from other approaches such as bootstrap method of Kybic *et al.* or Bayesian registration.

In conclusion, this uncertainty estimation framework is not only flexible in terms of its applicability to any atrophy estimation method and incorporation of MR artefacts, but also brings an added advantage of automatic correction of method dependent bias, existent in atrophy estimations.

Atrophy Estimation using Bayesian Non-Linear Registration

Uncertainty estimation is a point of paramount significance when it comes to putting image registration based measurements to use in real life decisions. The last chapter dealt with an uncertainty estimation framework whose salient feature was estimating uncertainties using a ground truth and could be applied to any atrophy estimation method. Similarly, it is also interesting to develop a method that is capable of providing uncertainties along with point estimates of atrophy. In this regard, a Bayesian non-linear registration model is presented in this chapter in order to quantify longitudinal volumetric changes in the brain and to estimate MR image artefacts (noise and bias field inhomogeneity), simultaneously. The non-linear transformation is modelled as 3D cubic B-splines, noise is assumed to be distributed according to a Gaussian law and bias field inhomogeneity is modelled with Legendre polynomials. Furthermore, Monte Carlo Markov Chain (MCMC) methods are employed for sampling from the desired probability distributions.

In this chapter, the proposed probability model is described followed by the sampling procedure in section 5.1. Experimental results on estimation of bias field inhomogeneity and atrophy from our framework are compared with popular bias field correction methods (N4ITK [155] and differential bias correction of Lewis *et al.* [80]) and four different non-rigid registration approaches of Noblet *et al.* [98], Vemuri *et al.* [157], ANTS [10] and another cubic B-spline based registration method in section 5.3. Results are discussed in section 5.4.

5.1 Modelling

In the construction of a full probability model, three distinct components can be identified [46]:

- Specifying model quantities and their conditional independence structure using a graphical representation.
- Specifying the parametric form of direct relationships between these quantities.
- Probability modelling *i.e.*, calculate the likelihood terms in the model and specifying prior distribution of parameters and hyper-parameters.

Following these steps, our framework is described below.

5.1.1 Structural modelling

Directed Acyclic Graphs (DAGs) are often used for representing tasks performed in an orderly fashion, depending on certain constraints. DAGs are directed as the link between two nodes is represented by an arrow and acyclic because starting from a node it is not possible to return to the same node by following the arrows.

Throughout this chapter, majuscule and minuscule letters indicate random variables and their realisations, respectively.

Figure 5.1 shows a DAG illustrating the problem. In our formulation, estimating atrophy is equivalent to estimating a non-linear mapping (\mathbf{T}), between serially acquired scans of an individual \mathbf{I}_1 and \mathbf{I}_2 , from which atrophy can be obtained through its Jacobian. Images are assumed to be degraded with a smooth multiplicative bias (\mathbf{B}) that gives rise to inhomogeneity in intensities and also with spatially invariant additive Gaussian noise (\mathbf{N}). Only differential bias field inhomogeneity and noise in the images are considered, therefore only image \mathbf{I}_2 is shown to be affected by these artefacts in the DAG. Θ ; Π are the parameters of the models chosen for the non-linear transformation and bias field inhomogeneity, respectively; they are assumed to be random variables by the design. β_n is the noise parameter. β_r , t_l , t_h and, p_l , p_h are hyper-parameters for Θ and Π , respectively.

Variables in the DAG

Let $\Omega = \{(x, y, z) | 0 \leq x < \mathbf{x}, 0 \leq y < \mathbf{y}, 0 \leq z < \mathbf{z}\}$ be the domain of the images \mathbf{I}_1 and \mathbf{I}_2 .

Non-Linear transformation The non-linear transformation \mathbf{T} is parametrised using uniform cubic B-splines. Because they can model any 3D object irrespective of its physical model, B-splines form a popular class of transformations. Cubic B-spline based registration has been developed before in [124, 173].

B-splines are defined by a control lattice ϕ superimposed on Ω . ϕ is a set of $\eta_x \times \eta_y \times \eta_z$ control points $\Theta_{i,j,k}$ with a uniform spacing δ . Then 3D field $\mathbf{T}(x, y, z)$ at a voxel (x, y, z) can be represented as a product of 1D cubic B-splines [77]:

$$\mathbf{T}(x, y, z) = \sum_{l=0}^3 \sum_{m=0}^3 \sum_{n=0}^3 S_l(u) S_m(v) S_n(w) \Theta_{i+l, j+m, k+n} \quad (5.1)$$

where $i = \lfloor \frac{x}{\eta_x} \rfloor - 1$, $j = \lfloor \frac{y}{\eta_y} \rfloor - 1$ and $k = \lfloor \frac{z}{\eta_z} \rfloor - 1$ and $u = \frac{x}{\eta_x} - \lfloor \frac{x}{\eta_x} \rfloor$, $v = \frac{y}{\eta_y} - \lfloor \frac{y}{\eta_y} \rfloor$ and $w = \frac{z}{\eta_z} - \lfloor \frac{z}{\eta_z} \rfloor$. S_l represents the l th basis function of the cubic B-spline defined as

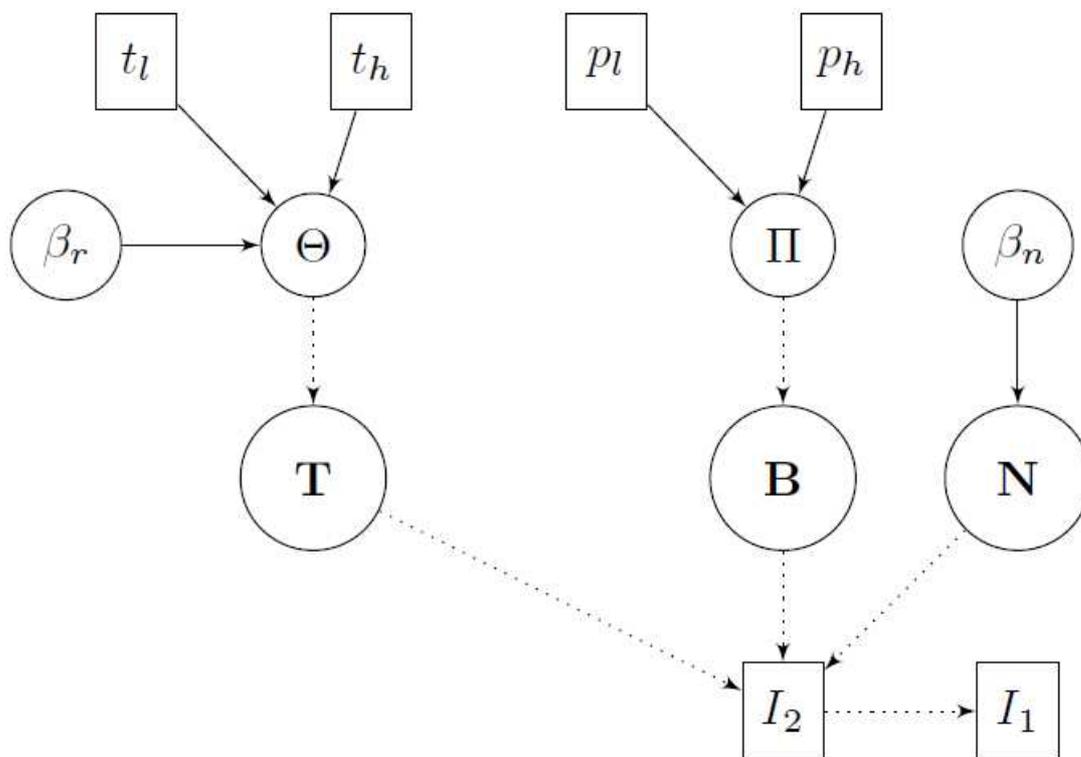


Figure 5.1: A DAG representing relations the deformable transformation \mathbf{T} , bias field inhomogeneity \mathbf{B} , noise \mathbf{N} and the observed images \mathbf{I}_1 , \mathbf{I}_2 . Conventional representations are used where square boxes represent quantities assumed fixed by the design and circles represent quantities that need to be estimated. Dotted and solid arrows represent deterministic and probabilistic relations, respectively.

$$\begin{aligned}
S_0(h) &= \frac{(1-h)^3}{6}, \\
S_1(h) &= \frac{(3h^3 - 6h^2 + 4)}{6}, \\
S_2(h) &= \frac{(-3h^3 + 3h^2 + 3h + 1)}{6}, \\
S_3(h) &= \frac{h^3}{6}
\end{aligned}$$

where $0 \leq h < 1$.

The basis functions determine the contribution of each control point to $\mathbf{T}(x, y, z)$ depending on their distances from voxel (x, y, z) . Spacing δ between the control points decides how densely control points are distributed on the image. It in turn is related to the changes that can be detected as it determines the shape of the approximating function. For example, if we choose $\eta_x = \eta_y = \eta_z = c'$ instead of $\eta_x = \eta_y = \eta_z = c$, the location of a control point $\Theta_{i+l, j+m, k+n}$ determined by i , j and k will be different, hence the approximating functions will differ in shape. The new location will be $i' = \lfloor \frac{x}{c'} \rfloor - 1$, $j' = \lfloor \frac{y}{c'} \rfloor - 1$ and $k' = \lfloor \frac{z}{c'} \rfloor - 1$ instead of $i = \lfloor \frac{x}{c} \rfloor - 1$, $j = \lfloor \frac{y}{c} \rfloor - 1$ and $k = \lfloor \frac{z}{c} \rfloor - 1$. Transformation is estimated at control points and then according to Eq. 5.1 the value of the transformation at each voxel (x, y, z) is computed. As δ becomes larger (the control lattice becomes coarser), more voxels affect the same control point to produce a smoother shape at the expense of compromising the accuracy. Processing large number of control points needs more computational time.

Due to such a parametrisation, $\mathbf{T}(x, y, z)$ is deterministically dependent on the transformation at control points Θ . The objective of this work is to estimate volume change in a region of interest through the Jacobian of the transformation \mathbf{T} . We will refer to Θ as transformation parameters in the upcoming sections.

Regularisation function In order to produce physically feasible solutions, it is desirable to constrain the deformable transformation to be smooth. Among many regularisation terms, the thin plate spline smoothing is chosen [159]:

$$\begin{aligned}
R_{smooth} = \frac{1}{V} \int_{x=0}^X \int_{y=0}^Y \int_{z=0}^Z & \left[\left(\frac{\partial^2 \mathbf{T}}{\partial x^2} \right)^2 + \left(\frac{\partial^2 \mathbf{T}}{\partial y^2} \right)^2 + \left(\frac{\partial^2 \mathbf{T}}{\partial z^2} \right)^2 \right. \\
& \left. + \left(\frac{\partial^2 \mathbf{T}}{\partial xy} \right)^2 + \left(\frac{\partial^2 \mathbf{T}}{\partial xz} \right)^2 + \left(\frac{\partial^2 \mathbf{T}}{\partial yz} \right)^2 \right] dx dy dz \quad (5.2)
\end{aligned}$$

where V denotes the volume of the image domain. This quantity is the bending energy of a thin-plate of metal and possesses a property of penalising large deformations. In applications of intra-subject registration, large deformations are not expected and are taken care of by this smoothing function. This regularisation function has been earlier used in conjunction with cubic B-splines in breast [124] and inter-subject brain [147] registrations.

Bias field \mathbf{B} changes slowly in the image leading to smooth intensity variations. Thus, it can be expressed as a linear combination of smooth basis functions. Legendre polynomials represent one of the many basis functions capable of parametrising bias field inhomogeneity in MR images [148].

In 3D, the bias field estimate $\mathbf{B}(x, y, z)$ given the maximum degree d of the Legendre polynomials and the parameter set Π is derived as follows:

$$\mathbf{B}(x, y, z) = \sum_{i=0}^d \sum_{j=0}^{d-i} \sum_{k=0}^{d-i-j} \Pi_{ijk} L_i(x) L_j(y) L_k(z) \quad (5.3)$$

where $L_i(\cdot)$ denotes a Legendre polynomial of degree i . The image voxel co-ordinates (x, y, z) are scaled in the range of $[-1, 1]$. For Legendre polynomials of maximum degree l , the size m of the parameter vector Π is given by:

$$m = (d+1) \frac{(d+2)}{2} \frac{(d+3)}{3} \quad (5.4)$$

The maximum degree d of Legendre polynomials is related to the accuracy and stability of the estimated bias field. Choosing a large degree will result in inefficient computation time, unstable coefficients and chances are that anatomical structures will be mistaken for intensity inhomogeneity [148]. On the other hand, low degree polynomials may not be able to remove bias completely. Experiments of Styner *et al.* [148] have established that a 3rd order Legendre polynomial is sufficient for modelling bias field inhomogeneity for MR scanners with head coils. We will refer to Π as bias field parameters in the upcoming discussion.

A brief description of Legendre polynomials can be found in Appendix A.1.

Noise in the image is modelled by a zero mean Gaussian distribution with standard deviation β_s . It is worth noting that approximating the Rician noise in MRI with Gaussian noise is valid when the signal to noise ratio in the image is greater than 2 dB [49].

5.1.2 Direct relations between variables in DAG

In a real scenario, during the acquisition interval, the patient's brain undergoes volumetric changes that are modelled as a non-linear transformation \mathbf{T} . According to our model, images \mathbf{I}_1 and \mathbf{I}_2 suffer multiplicative bias field inhomogeneity and additive MR noise. Let $(\mathbf{B}_1, \mathbf{N}_1)$ and $(\mathbf{B}_2, \mathbf{N}_2)$ be the bias field inhomogeneities and Gaussian noise (with standard deviations β_1 and β_2) in images \mathbf{I}_1 and \mathbf{I}_2 , respectively. The imaging process is modelled as:

$$\begin{aligned} \mathbf{B}_1(x, y, z)\mathbf{I}_1(x, y, z) + \mathbf{N}_1(x, y, z) &= \mathbf{B}_2(\mathbf{T}(x, y, z))\mathbf{I}_2(\mathbf{T}(x, y, z)) + \mathbf{N}_2(x, y, z) \\ \implies \mathbf{I}_1(x, y, z) &= \mathbf{B}(\mathbf{T}(x, y, z))\mathbf{I}_2(\mathbf{T}(x, y, z)) + \frac{\mathbf{N}(x, y, z)}{\mathbf{B}_1(x, y, z)} \end{aligned} \quad (5.5)$$

where $\mathbf{B}(\mathbf{T}(x, y, z)) = \frac{\mathbf{B}_2(\mathbf{T}(x, y, z))}{\mathbf{B}_1(x, y, z)}$ and $\mathbf{N}(x, y, z)$ is Gaussian noise distributed according to a Gaussian law with a standard deviation of $\beta_n = \sqrt{\beta_1^2 + \beta_2^2}$.

Eq. 5.5 can be interpreted as follows. If we consider a spatially invariant noise in images \mathbf{I}_1 and \mathbf{I}_2 , the differential correction will result in noise enhancement in areas where bias field is dark as compared to bright areas due to the factor $\frac{\mathbf{N}(x, y, z)}{\mathbf{B}_1(x, y, z)}$ (the differential noise is scaled by a factor equal to the bias field in image \mathbf{I}_1). For example, let $b_1(x, y, z)$ be a bias field in the first image with $0.9 < b_1(x, y, z) < 2$ and let original differential noise be $n(x, y, z)$, the ratio $\frac{n(x, y, z)}{b_1(x, y, z)}$ is higher when $b_1(x, y, z) = 0.9$ (dark area) than when $b_1(x, y, z) = 1.5$ (bright area). Thus, the noise $\frac{n(x, y, z)}{b_1(x, y, z)}$ cannot be represented by the same parameters as the original noise $n(x, y, z)$. Consequently, differential bias correction will achieve a common brightness level in the images but will result in noise enhancement in dark areas. This detail may not have a huge impact if the bias field is not strong, for instance, $0.9 < \mathbf{B}_1 < 1.1$, corresponding to 20% INU, which is the case in our experiments. So, the noise parameter can be safely assumed to be the same for the entire image. This assumption is important because the noise parameter determines the minimum energy that can be reached.

It is this differential bias field (\mathbf{B}) and noise (\mathbf{N}) whose estimation is carried out in this framework. When estimating longitudinal atrophies for a patient, it is sufficient to correct for differential \mathbf{B} and \mathbf{N} since this makes the intensities of the two images comparable to allow the use of a similarity criterion such as sum of squared differences (SSD).

Considering SSD as the similarity criterion, the energy in a 3D region of interest is given by:

$$E_s = \sum_{\Omega} (\mathbf{I}_1(x, y, z) - \mathbf{B}(\mathbf{T}(x, y, z))\mathbf{I}_2(\mathbf{T}(x, y, z)))^2 \quad (5.6)$$

Let $E_r = R_{smooth}$ (See Eq. 5.2) be the energy of the regularisation term. The total

energy can be written as:

$$E_t = E_s + \frac{E_r}{\beta_r} \quad (5.7)$$

β_r is the weight of the regulariser. The total energy E_t must be minimised in order to find the unknown parameters. To further constrain the deformation field to topologically feasible solutions [98], only those solutions that lead to a positive Jacobian determinant are allowed. At each iteration the Jacobian is checked at each voxel, and the solution is rejected if it does not satisfy the criterion of positivity.

5.1.3 Probability modelling

After having established a structural model, we move to probability modelling. Owing to the parametric modelling of random variables \mathbf{T} , \mathbf{B} and \mathbf{N} , the transformation parameters Θ , bias field parameters Π and standard deviation β_n of the noise are random variables and the estimation of these parameters directly leads us to estimates of \mathbf{T} , \mathbf{B} and \mathbf{N} .

This guides us to the following joint probability model.

$$\begin{aligned} P(\mathbf{I}_1, \mathbf{I}_2, \mathbf{T}, \mathbf{B}, \mathbf{N}) &= P(\mathbf{I}_1, \mathbf{I}_2, \Theta, \Pi, \beta_n) \\ &= P(\mathbf{I}_1, \mathbf{I}_2 | \Theta, \Pi, \beta_n) P(\Theta | \beta_r, t_l, t_h) P(\Pi | p_l, p_h) P(\beta_n) \end{aligned} \quad (5.8)$$

where I_1 and I_2 are the given images. β_r , t_l , t_h and p_l , p_h are hyper-parameters for Θ and Π , respectively. The modelling of the likelihood and prior functions in the joint probability distribution (Eq. 5.8) are detailed in the forthcoming discussion.

Likelihood

A Gaussian distribution is used for defining the likelihood (also known as the similarity criterion) because images are assumed to be corrupted by Gaussian noise:

$$P(\mathbf{I}_1, \mathbf{I}_2 | \Theta, \Pi, \beta_n) = \frac{1}{Z_s} e^{\frac{-E_s(\mathbf{I}_1, \mathbf{I}_2, \Theta, \Pi)}{2\beta_n^2}} \quad (5.9)$$

where Z_s is the normalising constant.

Prior distributions

The choice of the prior distributions is crucial in incorporating the knowledge of the system we possess, in the framework. In the present case, there are three quantities whose prior distributions are needed (Θ , Π and β_n):

- **Transformation parameters:** Two priors are chosen on the transformation parameters: a prior to regularise the deformation field and a uniform prior to restrict the values of the deformation field during sampling. The regulariser is chosen so as to make calculations simpler (the logarithm of the similarity

criterion (E_s) and regularisation energies (E_r) can be written as a sum). Further, a uniform prior is imposed on the transformation parameters Θ with hyper-parameters t_l and t_h . The range (t_l, t_h) is set according to the smallest and the largest deformation expected in the region of interest. In essence, t_l and t_h control the range of deformation field values from which samples are drawn in the MCMC framework.

Thus, we have the complete prior as:

$$P(\Theta|\beta_r, t_l, t_h) \propto e^{\frac{-E_r(\Theta)}{\beta_r}} U(t_l, t_h) \quad (5.10)$$

- **Bias field parameters:** A uniform prior is chosen for the bias field parameters Π with hyper-parameters p_l and p_h . This allows us to restrict the range in which the sampling is performed for the bias field parameters. If the largest and the smallest values of bias field inhomogeneity are known, an approximate range in which the Legendre polynomial parameters lie can be found analytically. The prior for the bias field parameters is given as,

$$P(\Pi|p_l, p_h) = U(p_l, p_h) \quad (5.11)$$

Remember that the parameters t_l , t_h , p_l and p_h , even if estimated approximately, can facilitate sampling by further constraining the MCMC sampler to probable solutions.

- **Noise parameter β_n :** The standard deviation β_n of the differential noise present in the images can be estimated from the noise in the background of MR images where the signal can be assumed to be zero. Alternative methods such as those discussed in section 2.4.2 can also be employed.

Segmentation of the background is obtained through Otsu's thresholding approach [101]. If the MRI noise can be approximated with a Gaussian distribution, the ML estimate of the noise standard deviation (β_n) is given by,

$$\beta_n = \sqrt{\frac{1}{N} \sum_{i=0}^N n_i^2} \quad (5.12)$$

where n_i is the value at the i th voxel of the background and N is the total number of voxels in the background. Note that β_n will not be sampled in the MCMC framework.

Posterior probability

The standard procedure involves maximising the a posteriori probability to obtain estimates of unknowns. Therefore, let us write the posterior probability which comes directly from the definition of conditional probability and Eq. 5.8 as:

$$\begin{aligned} P(\Theta, \Pi, \beta_n | \mathbf{I}_1, \mathbf{I}_2) &= \frac{P(\mathbf{I}_1, \mathbf{I}_2, \Theta, \Pi, \beta_n)}{P(\mathbf{I}_1, \mathbf{I}_2)} \\ &\propto P(\mathbf{I}_1, \mathbf{I}_2 | \Theta, \Pi, \beta_n) \\ &\quad P(\Theta | \beta_r, t_l, t_h) P(\Pi | p_l, p_h) P(\beta_n) \end{aligned} \quad (5.13)$$

where $P(\mathbf{I}_1, \mathbf{I}_2)$ is constant because \mathbf{I}_1 and \mathbf{I}_2 are given images. $P(\beta_n) = \text{constant}$ because it is determined deterministically. β_r, t_l, t_h and p_l, p_h are hyper-parameters for Θ and Π , respectively.

Expanding the posterior by substituting for all likelihood and priors distributions from Eqns. 5.9, 5.10, 5.11, and 5.13 we have

$$P(\Theta, \Pi, \beta_n | \mathbf{I}_1, \mathbf{I}_2) \propto e^{\frac{-E_s(\mathbf{I}_1, \mathbf{I}_2, \Theta, \Pi)}{2\beta_n^2}} e^{\frac{-E_r(\Theta)}{\beta_r}} U(t_l, t_h) U(p_l, p_h) \quad (5.14)$$

5.2 Maximum a posteriori estimation

Solving the maximum a posteriori (MAP) problem is tantamount to finding those parameters of the transformation field and the bias field that maximise the posterior probability in Eq. 5.14. Due to the complicated nature of our posterior it is neither possible to analytically compute it nor samples can be drawn from it directly. In such situations, MCMC methods are generally employed for generating samples. MCMC methods stochastically optimise the posterior distribution based on probability distributions.

In our case, samples of bias field parameters Π and transformation parameters Θ can be generated from the posterior distribution using MCMC methods. The energy is monitored iteration per iteration and the values of Π and Θ that lead to the minimum energy (or MAP) are the estimated values. Estimated atrophy is calculated from the estimated Θ .

5.2.1 Sampling schedule

The sampling is carried out with Metropolis-Hastings algorithm where candidate samples for the target distribution are drawn from a proposal distribution (from which sampling is easy). Gaussian proposal distributions are chosen for bias field (Π) and transformation (Θ) parameters. The next candidate state ($i+1$) is computed by means of generating samples as a function of the current state of the chain for bias field parameters π_i and transformation parameters θ_i according the following

rules:

$$\begin{aligned}\pi'^{(i+1)} &\sim \mathcal{N}(\pi^{(i)}, \sigma_\pi) \\ \theta'^{(i+1)} &\sim \mathcal{N}(\theta^{(i)}, \sigma_\theta)\end{aligned}\tag{5.15}$$

where σ_π and σ_θ are proposal distribution parameters. It is assumed that the two parameter sets are independent and thus can be sampled from two independent proposal distributions as shown in Eq. 5.15. The complete proposal distribution is

$$Q(\psi'; \psi^{(i)}) = \mathcal{N}(\pi^{(i)}, \sigma_\pi) \mathcal{N}(\theta^{(i)}, \sigma_\theta)\tag{5.16}$$

where $\psi = (\pi, \theta)$. These candidates are accepted with a probability

$$A(\psi'|\psi^{(i)}) = \min\left(1, \frac{P(\psi'|\mathbf{I}_1, \mathbf{I}_2)Q(\psi^{(i)}; \psi')}{P(\psi^{(i)}|\mathbf{I}_1, \mathbf{I}_2)Q(\psi'; \psi^{(i)})}\right)\tag{5.17}$$

$P(\cdot|\mathbf{I}_1, \mathbf{I}_2)$ is the posterior probability given by Eq. 5.14.

Note that the ratio $\frac{Q(\psi^{(i)}; \psi')}{Q(\psi'; \psi^{(i)})} = 1$ because a symmetric distribution (Gaussian) is chosen as a proposal distribution. Also, the knowledge of the normalising constant Z_s is not required for computing the acceptance probability.

Proposal variances An important factor is the variance of the proposal distributions that determines the time taken by the chain to reach its stable state or convergence. The distribution of the parameters studied via the Markov chain in its stable state is called their stationary distribution. Usually, the proposal variance is obtained by studying the acceptance rate (which is the number of samples accepted among the last N proposals). A large variance leads to low acceptance rates as most of the new samples will land in regions of low probability and hence will be rejected. Similarly, a small variance will result in more proposals being selected and the acceptance rate will be high. The optimal value of the proposal variance must be able to aid the Markov chain in reaching its stable state soon from the initial point while covering the entire possible parameter space.

Consider a Markov chain, $\{x_1, x_2, \dots, x_N\}$ with stationary distribution ψ , then the expectation $\bar{\mu}$ of x can be estimated by a delayed averaging,

$$\bar{\mu}_N = \frac{1}{N} \sum_{n=M+1}^{M+N} \mu(x_n),\tag{5.18}$$

The period until iteration M is the burn-in period. $\bar{\mu}_N$ is a consistent estimator if the chain is irreducible [152]. This computation will not hold if this is not the case and there is dependence, within the Markov chain x_n . Approximating $\mu(x_n)$ as a first order autoregressive process with autocorrelation ρ , the variance of $\bar{\mu}_N$ is given

by,

$$\text{var}(\bar{\mu}_N) = \frac{\sigma^2}{N} \frac{1 + \rho}{1 - \rho} \quad (5.19)$$

where σ is the proposal distribution. This means that when autocorrelation ρ is high, a large number of samples are needed, in order to achieve the desired accuracy in $\bar{\mu}_N$.

Gilks *et al.* [47] attempt to solve the problem of estimating the proposal variance for when the target distribution has an exchangeable form $\prod_{i=1}^M \psi(x_i)$ where x_i is the i th element of x . They have shown that for large M , it can be determined that the variance in Eq. 5.19 is minimised when σ is chosen such that 23.4% of the total candidates are accepted. It has been shown through empirical results that in most cases 15 – 50% of the proposals must be accepted for optimal performance [48]. Note that there is no optimal theoretical value of the proposal variance.

The optimal variance may be determined empirically, by monitoring the acceptance rates of candidates in nested windows of iterations. Gilks *et al.* [48] provide a way for setting a value for this quantity. Let R_i be the acceptance rate after i iterations, then the proposal variance or scaling for the next iteration σ_{i+1} can be calculated as follows:

$$\log(\sigma_{i+1}) = \log(\sigma_i) + \frac{(\text{logit}(R_i) - \text{logit}(r))}{r} \quad (5.20)$$

where r is the target acceptance rate and the function *logit* of a number p between 0 and 1 is $\text{logit}(p) = \log(p) - \log(1 - p)$. If $R_i < r$, σ will be reduced and, as a result, the acceptance rate will increase and vice versa.

In our implementation, “optimal” values for both σ_π and σ_θ are set according to the Eq. 5.20. This expression is not evaluated at each iteration as the acceptance rate may not vary substantially in one iteration. Acceptance rates are calculated using overlapping windows of candidates accepted in the last 20 iterations and variance values are updated at every 5th iteration. This is done in order to reduce the computational overhead of recalculating acceptance rate and σ at every iteration.

Convergence Convergence is established by following the evolution of the total energy E_t Eq. 5.7 graphically.

Algorithm 1 provides an overview of the sampling algorithm.

5.3 Results

Results of experiments conducted to demonstrate and compare the proposed framework with other bias field correction and non-rigid registration methods, are pre-

```

while no convergence do
  Sample  $\pi^{(i)} = \mathcal{N}(\pi^{(i-1)}, \sigma_\pi)$ .
  Sample  $\theta^{(i)} = \mathcal{N}(\theta^{(i-1)}, \sigma_\theta)$ .
  Compute total energy  $E_t = E_s + E_r$  from Eqns. 5.9 and 5.10.
  Compute MH criterion in Eq. 5.17.
  if Accept then
    |  $\pi^{(i)} = \pi^{(i)}$  and  $\theta^{(i)} = \theta^{(i)}$ 
  else
    |  $\pi^{(i)} = \pi^{(i-1)}$  and  $\theta^{(i)} = \theta^{(i-1)}$ 
  end
  Adjust  $\sigma_\pi$  and  $\sigma_\theta$  values according to Eq. 5.20.
end

```

Algorithm 1: Bayesian estimation of bias field and transformation parameters

sented in this section in three parts:

- performing only bias correction but no atrophy estimation
- performing only atrophy estimation and
- simultaneous estimation of bias field inhomogeneity and atrophy

5.3.1 Bias field correction

In order to study the efficiency of the bias field correction algorithm, images are degraded with two different synthetic bias field inhomogeneities of 20% INU from the Brainweb database and additive Gaussian noise (SNR=35 dB). Tests are performed on the whole brain simulated Brainweb image down-sampled to an image of size $64 \times 64 \times 64$. Neither atrophy is simulated nor atrophy estimation is performed. The similarity cost in the Eq. 5.9 is minimised where \mathbf{T} is set to identity (no deformation exists between the two images).

Tests are carried out with Legendre polynomials of maximum degree 2 to 5 in order to determine the optimum number of parameters for further tests. Next, the variance of the proposal distribution for bias field parameters is dynamically computed following Eq. 5.20. The hyper-parameters of the uniform prior are set to $p_l = -5$ and $p_h = 5$. Convergence is established by studying the evolution of total energy as the iterations progress. An example of change in total energy with progressing iterations is plotted in Fig. 5.2 (for the case when Legendre polynomials of maximum degree 3 are used for bias field parameter estimation). It can be seen from Fig. 5.2 that after 700 iterations the chain reaches a stable state.

The MCMC based differential bias estimation is compared with N4ITK¹ [155], differential bias correction of Lewis *et al.* [80] and PABIC of Styner *et al.* [148].

¹<http://www.insight-journal.org/browse/publication/640>

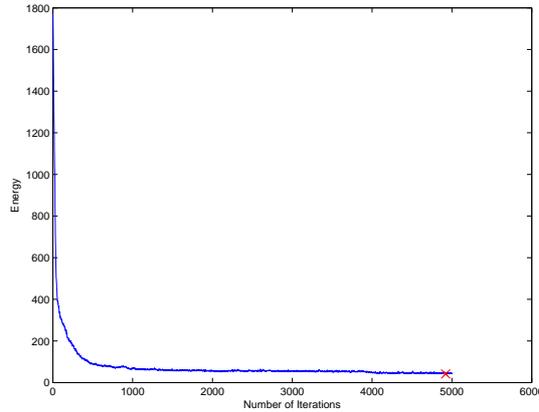


Figure 5.2: Evolution of energy with the sampling iterations when using a Legendre polynomial of degree 3 for the bias field B . The position of the MAP energy is marked with a red cross. Total energy converges after 700 iterations when the chain reaches a stable state.

While N4ITK is an improved version of the popular N3 algorithm that uses one image, Lewis *et al.*'s method estimates differential bias from a pair of serially acquired images only on the brain area (brain extraction must be performed prior to bias estimation). The PABIC approach also estimates bias field from a single image and models the classes in the image (*e.g.* GM, WM) using a Gaussian model. For algorithms that take one input image, the differential bias field is calculated as the ratio of the individual bias fields obtained by applying bias field estimation to each image separately. Examples of estimated fields by the four methods are in shown in Fig. 5.3. An explanation of the parameters of N4ITK, Lewis *et al.*'s and PABIC methods is available in appendix B.1.

The ground truth bias field that is added to the Brainweb image is shown in Fig. 5.3 (a) and the resulting bias fields from our approach in Fig. 5.3 (b-e) with maximum degree of Legendre polynomials in the range of 2 – 5, Lewis *et al.*'s method Fig. 5.3 (f), N4ITK Fig. 5.3 (g) and PABIC Fig. 5.3 (h) are also shown. The errors between the estimated fields and the ground truth are depicted in Fig. 5.4. Quantitative results are shown in Table 5.1. For all the four approaches, mean squared errors (MSE) in the recovered bias field and the bias corrected image with respect to the ground truth bias fields and Brainweb image are computed. MSE is computed according to the following equation:

$$MSE = \frac{1}{N} \sum_{(x,y,z) \in \Omega'} (G(x,y,z) - E(x,y,z))^2 \quad (5.21)$$

where Ω' is the region of the image considered and N is the total number of voxels in this region. G and E are expected and estimated quantities (anatomical image

or bias field), respectively.

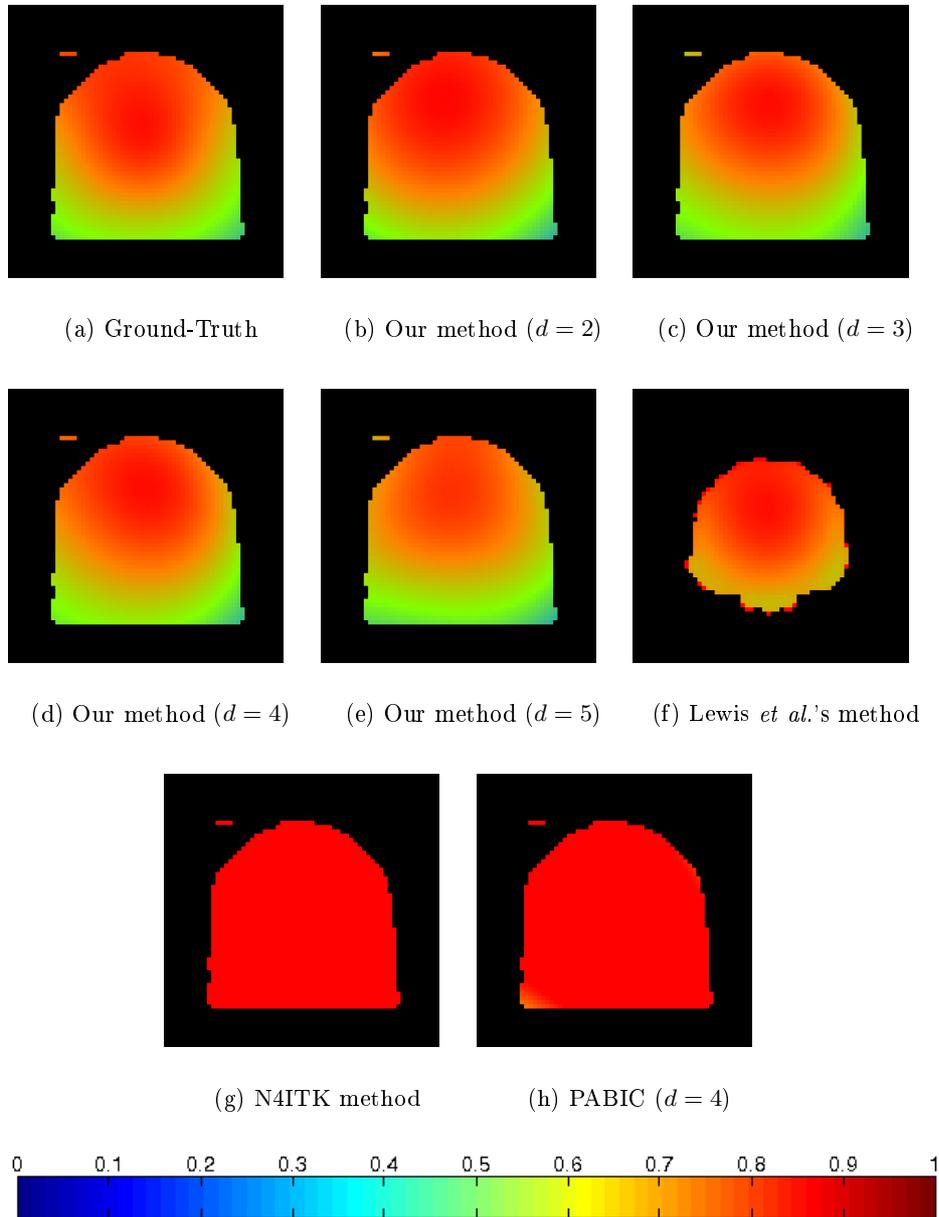


Figure 5.3: Comparison of estimated bias fields by our method, Lewis *et al.*'s method, N4ITK correction and PABIC approach. $d = 2, 3, 4, 5$ for our method represent the maximum degree of Legendre polynomial used in modelling of the bias field. For PABIC approach, the results obtained with $d = 4$ are illustrated because this results in the best estimations. See Table 5.1 for quantitative results.

The estimated fields in Fig. 5.3 and the errors in estimated bias fields in Fig. 5.4

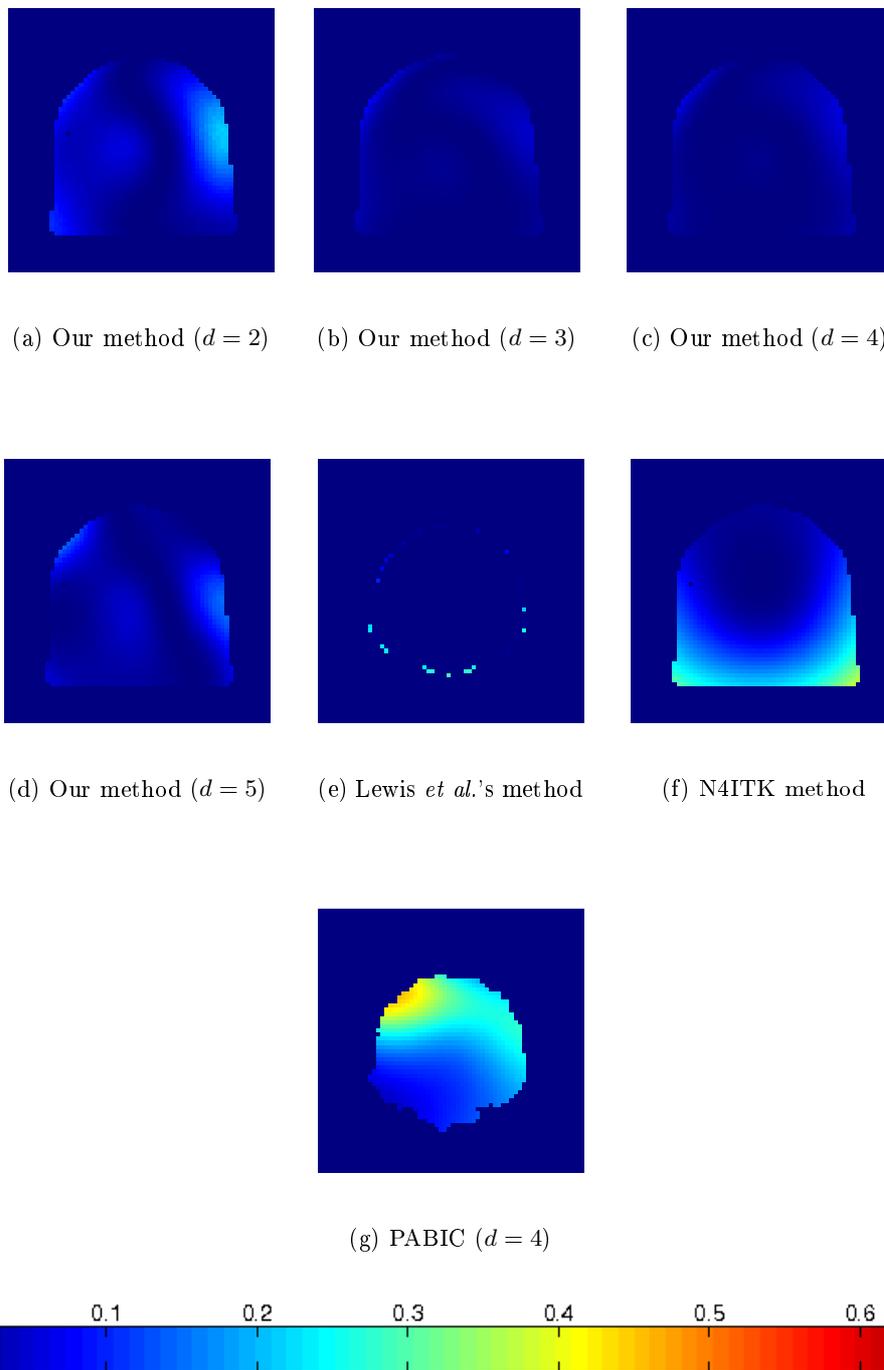


Figure 5.4: Error in the estimated bias fields and the ground truth for our method, Lewis *et al.*'s method, N4ITK correction and PABIC. $d = 2, 3, 4, 5$ for our method represent the degree of Legendre polynomial used in modelling the bias field. For PABIC, the results obtained with $d = 4$ are illustrated because this results in the best estimations. See Table 5.1 for quantitative results.

allow us to conclude that with Lewis *et al.*'s method and our approach the pattern of bias is well captured. Let us notice that the bias field is underestimated and patternless with N4ITK correction. The results of estimated fields and MSE show that the PABIC approach overestimates the bias field. Considering the accuracy of estimations for our method when Legendre polynomials of different degrees are used, it can be seen from Fig. 5.3 (b-e) that the estimated bias fields resemble the ground truth visually in all the four cases. If we look at the errors in Fig. 5.4, a Legendre polynomial of 3rd or 4th degree models the ground truth bias field better than a 2nd or 5th degree polynomial. The quantitative analysis (Table 5.3) confirms these observations. Legendre polynomial of maximum degree $d = 4$ leads to the lowest MSE values among experiments with other degree polynomials. Table 5.3 also highlights that Lewis *et al.*'s method performs better than our approach. Nevertheless, the performance of our method (with $d = 4$) is comparable to that of Lewis *et al.*'s method as opposed to N4ITK and PABIC that leave a considerable residual bias hence the larger MSE values. According to these results, PABIC approach is the worst performer out of the methods tested.

The maximum degree of the polynomial is related to the number of parameters with the Eq. 5.4. For maximum degrees between 2–5, the number of parameters are 10, 20, 35 and, 56, in the same order. Generation of 5000 samples is accomplished in 1, 1.8, 2.2 and 3.8 hours, respectively for maximum degrees of 2 – 5.

		MSE Bias Field		MSE Bias Corrected Image	
		A	B	A	B
Method	Bias				
		A	B	A	B
	Our method ($d = 2$)	0.001871	0.001037	40.60	23.02
	Our method ($d = 3$)	0.001643	0.0008509	31.93	19.06
	Our method ($d = 4$)	0.001513	0.0005330	21.89	14.28
	Our method ($d = 5$)	0.001351	0.0007008	20.19	10.99
	PABIC ($d = 2$)	0.385496	0.339491	6109.91	6333.93
	PABIC ($d = 3$)	0.290426	0.336078	6437.22	6347.76
	PABIC ($d = 4$)	0.260397	0.280823	7093.02	6605.07
	PABIC ($d = 5$)	0.268169	0.279130	6346.25	6510.79
	Lewis <i>et al.</i>	0.000980	0.000461	21.908	9.913
	N4ITK	0.051025	0.026226	1285.714	597.455

Table 5.1: MSE in the estimated bias fields and the bias corrected images with respect to the ground truth bias field and Brainweb image. A and B refer to the two simulated bias fields used in these experiments. $d = 2, 3, 4, 5$ represents the maximum degree of the Legendre polynomial used in the modelling of the bias field. The best results are shown in bold.

5.3.2 Atrophy estimation

Results are shown on two in-vivo 3D images (subject number 2 and 7) from the IBSR database (See appendix B.1 for details). The reason for choosing these two images is that they are different anatomically and in appearance. Uniform atrophies of around 2%, 5%, 10% and 15% are simulated in the hippocampus with the algorithm described in section 3.2.1. This provides us with a ground-truth for validation of our algorithm. A $32 \times 32 \times 32$ ROI containing the hippocampus is extracted from the whole brain. Computations are performed on this extracted ROI. Gaussian noise (SNR=35 dB) is added to the baseline as well as the simulated atrophied images.

The same tests are carried out with three other NRR algorithms (Noblet *et al.* [98], ANTS [10], Vemuri *et al.* [157]) and a cubic B-spline based non-rigid registration method implemented in ITK to inspect the performances of these approaches, compared to the proposed method. The cubic B-spline based registration will allow us to compare the implemented MCMC algorithm with a deterministic method that is based on the same transformation model. Regular step gradient descent optimisation is used. We will refer to this method as ITK cubic B-spline registration. The implementation details and values of parameters used for experiments with algorithms of Noblet *et al.* [98], ANTS [10], Vemuri *et al.* [157] and ITK cubic B-spline registration are given in appendix B.1.

The value of parameter β_n is estimated to be 2 from Eq. 5.12 with mean found to be close to zero. We use this value for our estimations. The value of β_r is fixed to 100 for experiments with images from the IBSR database while the range (t_l, t_h) is set to $(-20, 20)$ mm as the deformations are expected to lie within this interval. The variance of the proposal distribution for transformation parameters is dynamically computed following Eq. 5.20.

A minimum support of $3 \times 3 \times 3$ control points is required by cubic B-splines which encouraged us to set the size of the control point lattice to $5 \times 5 \times 5$ and $6 \times 6 \times 6$, amounting to $125 \times 3 = 375$ (lattice spacing = 6.5 mm) and $216 \times 3 = 648$ B-spline control points (lattice spacing = 5.5 mm), respectively. Image voxel size is 1 mm \times 1 mm \times 1 mm.

Evolution of energy Energy vs iteration plot for IBSR patient number 7 with simulated atrophy of -14.64% is shown in Fig. 5.5. This chain reaches convergence in 300 iterations. The number of iterations to convergence (or stability) depends on the starting point. When the chain is initialised with an identity transform, the number of iterations to stability averaged over all cases are 250 ± 100 .

Visual comparison of Jacobians An example of estimated Jacobians are illustrated in Fig. 5.6 for a simulated atrophy of -14.61% in the hippocampus of the IBSR subject 7. The difference of the Jacobians from the ground truth is depicted

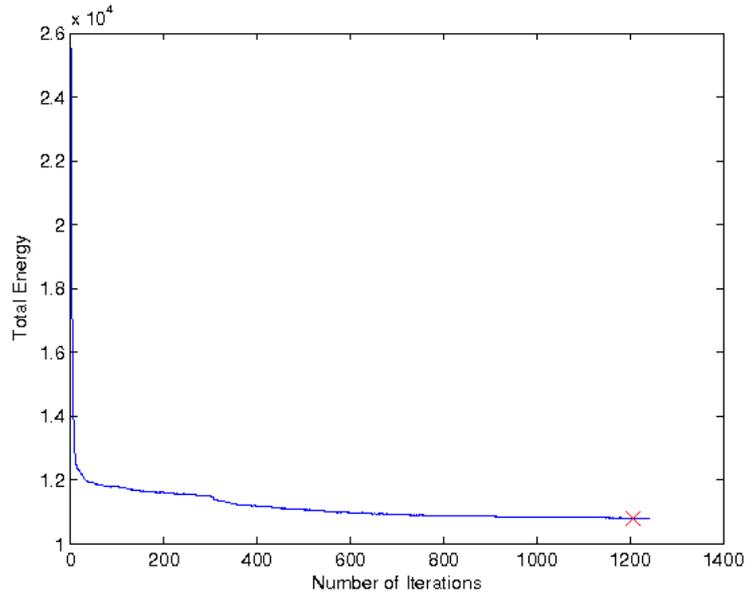


Figure 5.5: Energy evolution for IBSR patient number 7 with simulated atrophy of -14.64% . The chain reaches convergence in 300 iterations. The MAP energy is marked with a red cross.

in Fig. 5.7 for the Jacobian maps displayed in Fig. 5.6. It can be seen that the error is lower when the model is based on 648 control points than the one of 375 control points in Fig. 5.7. For the same experiment, Jacobian maps from NRR algorithms of Noblet *et al.*, ANTS, Vemuri *et al.* and ITK cubic B-spline registration (run with 648 control points) are shown in Fig. 5.8.

The atrophy simulation approach is a multi-resolution implementation. Simulations are performed on images of size $64 \times 64 \times 64$ in 5 resolutions where at the lowest level calculations are performed in blocks of 4 neighbouring voxels. As a result, the Jacobian appears blocky. From this an ROI of $32 \times 32 \times 32$ is extracted in order to speed up the MCMC iterations. The Jacobian maps from all approaches are drastically different. Our Jacobian maps (Fig. 5.6 (b-c)) and that of ITK cubic B-spline registration (Fig. 5.8 (d)) show some resemblance to the pattern of the ground truth map. The mean differences in Jacobians are -4.74% and -1.35% of atrophy in the ROI when the number of B-spline control points are 375 and 648, respectively for MCMC method and 5.01% for ITK cubic B-spline based method with 648 control points. No pattern can be observed in other approaches, however, the mean Jacobian in the ROI is close to the ground truth (See Table 5.2). This indicates that the main reason behind the methods not being able to recover the simulated atrophy patterns has to do with the amount or the type of regularisation and the resolution at which computations are performed. Note that cubic B-splines, by nature, are smooth, so our approach and ITK cubic B-spline based registration

method both produce visibly regularised Jacobian maps. ITK cubic B-spline based registration method did not contain any explicit regularisation term yet managed to produce an even Jacobian map.

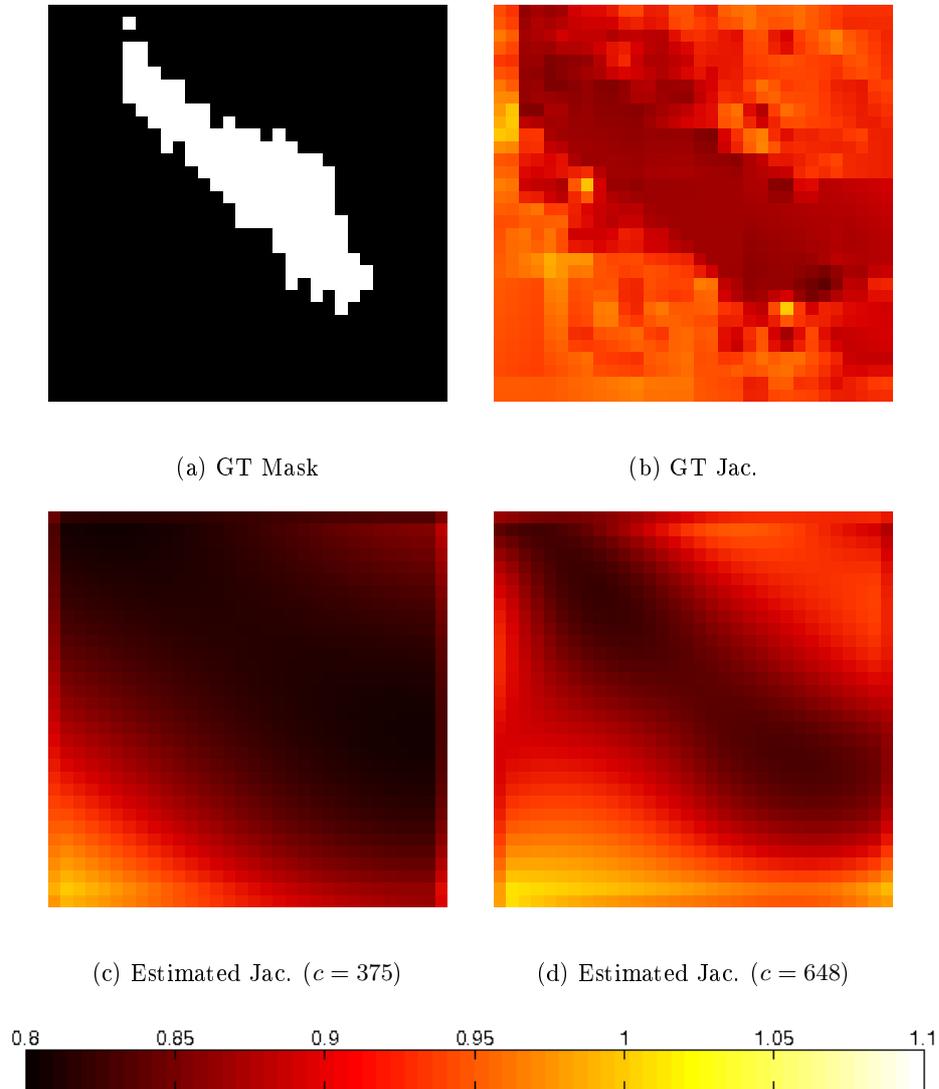


Figure 5.6: Ground truth mask of the hippocampus, the ground truth Jacobian along with the estimated Jacobians by the Bayesian framework for experiments on IBSR image number 7 for a simulated atrophy of -14.61% . $c = 375, 648$ refers to the number of B-spline control points.

Quantitative evaluation A comparison of the MAP atrophies in the hippocampus from our approach to other non-rigid registration methods Noblet *et al.* [98], Vemuri *et al.* [157] and ANTS [10] is provided in Table 5.2 for two subjects

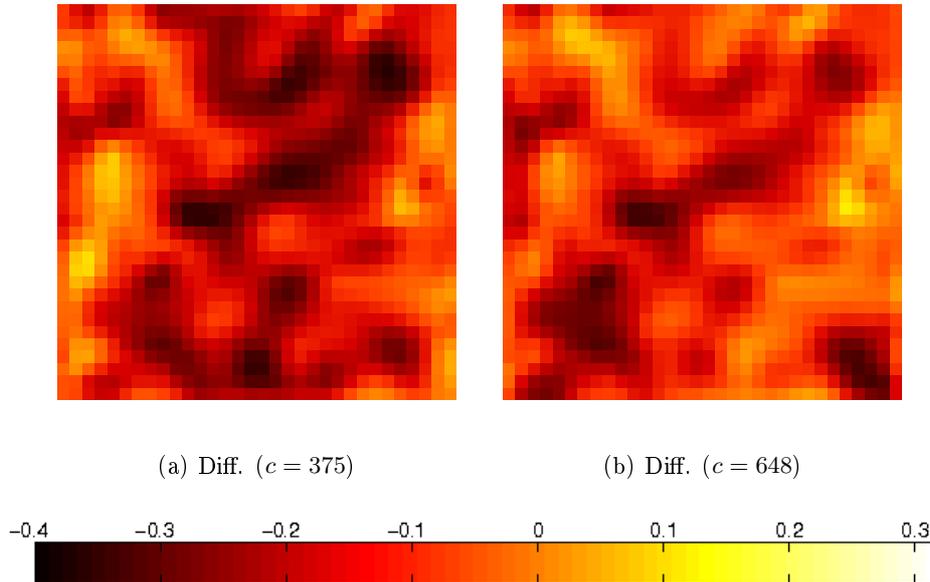


Figure 5.7: Difference of the estimated Jacobians shown in Fig. 5.6 from the ground truth Jacobian for the Bayesian framework for experiments on IBSR image number 7 for a simulated atrophy of -14.61% . $c = 375, 648$ refers to the number of B-spline control points.

from the IBSR database. Estimated mean atrophy is the estimated atrophy averaged over the hippocampus. 1200 samples are generated for both experiments using our framework.

Table 5.2 shows that, for our method, experiments using 648 B-spline control points are able to estimate large atrophies more accurately ($\approx 14.6\%$ and $\approx 9.9\%$) than simulations with 375 control points. The bias in the mean estimated atrophies when ground truths are $\approx 14.6\%$ and $\approx 9.9\%$ for subjects 2; 7 are $(-0.72, 0.76)$; $(0.61, 1.28)$ with 648 control points. The bias in estimations increases to $(6.43, 3.51)$; $(4.74, 2.85)$ for subjects 2 and 7 with 375 control points. On the other hand, tests with 375 control points estimate the smaller atrophies ($\approx 4.7\%$ and $\approx 1.6\%$) better than higher atrophies. The bias in estimations for smaller atrophies are for $c = 648$ and subject 2: $(-1.34, -1.66)$; subject 7: $(0.63, -1.37)$ and for $c = 375$ and subject 2: $(0.004, 0.17)$; subject 7: $(0.40, 0.36)$. It can be observed that better results are obtained when the number of control points is set to 648. Implementation of a NRR algorithm with the same transformation model (uniform cubic B-splines) in a deterministic setting shows that 648 B-spline control points lead to smaller errors as opposed to experiments with 375 control points in all cases. This is not observed in all MCMC based experiments. A possible reason is instability or slow mixing of the chain that was mistaken for convergence. In spite of this, the MCMC algorithm is a better performer than its deterministic counterpart in all the experiments.

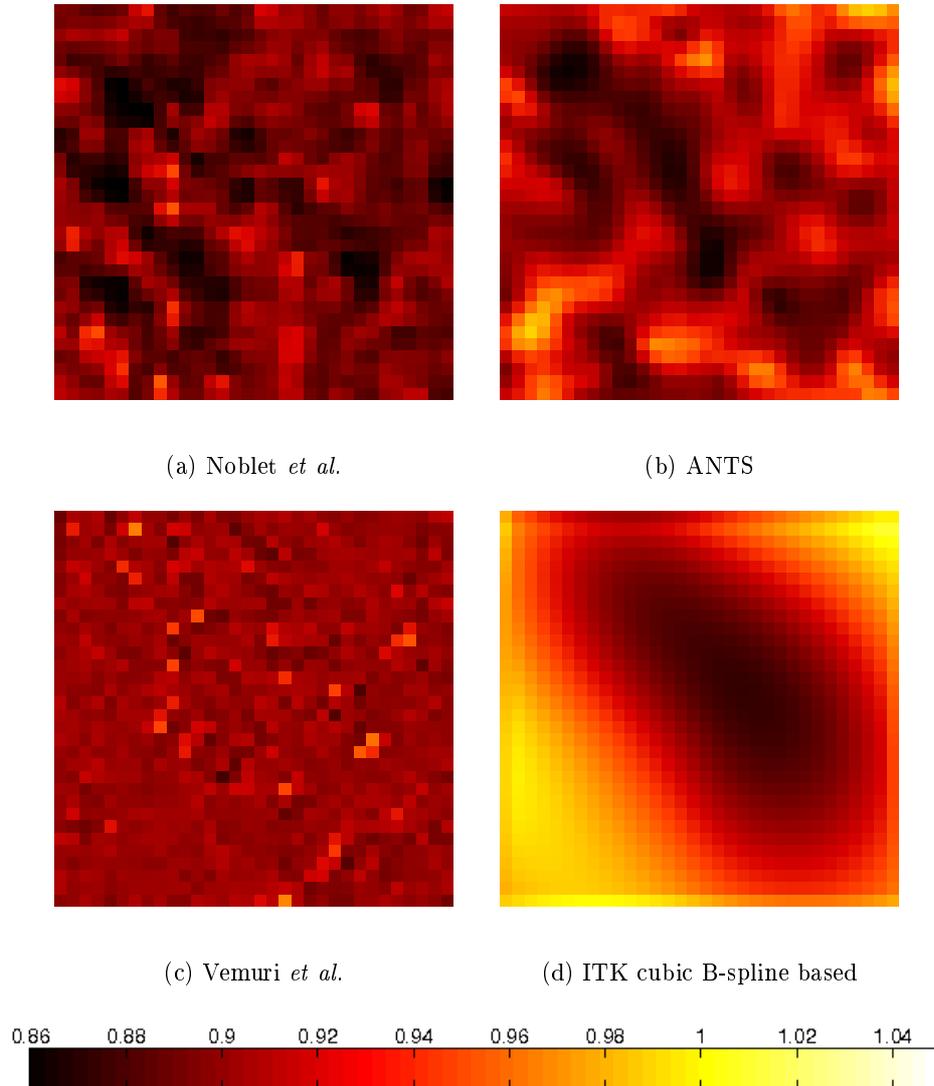


Figure 5.8: The estimated Jacobian maps for experiments on IBSR image number 7 for a simulated atrophy of -14.61% in the hippocampus using the NRR algorithms of Noblet *et al.*, ANTS, Vemuri *et al.* and ITK cubic B-spline based registration (with $c = 648$).

The NRR algorithms of Vemuri *et al.* and Noblet *et al.* overestimate whereas ANTS approach underestimates atrophy for both subjects. On the other hand, for our method no such trend is observed. In terms of bias in estimated mean atrophy, experiments with 648 control points outperform the other three NRR algorithms for large atrophies ($\approx 14.6\%$ and $\approx 9.9\%$) in all cases except one, where the algorithm of Vemuri *et al.* is better. For the same experiments with smaller atrophies ($\approx 4.7\%$ and $\approx 1.6\%$) ANTS is a better performer (in 3 out of 4 cases) than our method

but Vemuri *et al.* and Noblet *et al.* still fall behind. This trend is reversed in tests with 375 control points, where the performance of MCMC approach surpasses that of ANTS, Vemuri *et al.*'s and Noblet *et al.*'s approaches for smaller atrophies.

Confidence intervals The Bayesian registration approach also allows us to construct confidence intervals on the estimated atrophy. The estimated confidence intervals are shown in Figs. 5.9 (a-b) for both subjects. It can be seen that ground truths are not always contained in the confidence intervals but the estimated atrophy is always found within the confidence interval. In addition, the length of the intervals is approximately 2%.

Fig. 5.10 illustrates the distributions of the Jacobians estimated for -14.64% and -1.61% of hippocampal atrophy, for our experiments with different number of B-spline parameters. The location of the MAP with the corresponding 95% confidence intervals is also shown. The estimated atrophy falls at the mode of the distribution of the Jacobian.

5.3.3 Simultaneous estimation of bias field inhomogeneity and noise

A bias field of 20% INU and a Gaussian noise of 35 dB are added to the IBSR image of subject 2. In the same image, an atrophy of -14.6% is simulated. To conform to a real scenario, care is taken to extract the same $32 \times 32 \times 32$ ROI from the bias field as the image. This image is compared to the original IBSR image of subject 2 to which Gaussian noise with the same noise parameters is added. In order to perform bias correction, two existing bias correction methods are employed: N4ITK and Lewis *et al.*'s method, before atrophy estimation using the NRR registration approaches of Noblet *et al.*, ANTS, Vemuri *et al.*, ITK cubic B-spline and the proposed MCMC framework (without bias field correction). In addition, bias field parameters and atrophy are estimated simultaneously in the MCMC framework (648 control points and Legendre polynomials of maximum degree of 4 are used in these simulations).

The MSE between the estimated bias fields and the ground truth bias fields when bias field parameters and atrophy are simultaneously estimated in the MCMC scheme, with Lewis *et al.*'s method and with N4ITK are 0.009073, 0.0024387 and 0.032344. The results are in agreement with what we observed in earlier experiments (see Table 5.1). Like the previous tests, the differential bias correction of Lewis *et al.* is better than N4ITK correction and our method. The results of comparison of estimated atrophies in the hippocampus are presented in Table 5.3 and show that the bias correction methods of Lewis *et al.* and N4ITK are effective as there is no drastic performance difference from the results presented in Table 5.2 for the atrophy of -14.6% for all the methods. The maximum error with respect to the ground truth is seen for the NRR method of Noblet *et al.* at 2.51% when bias is corrected using N4ITK and lowest of -0.12% also for the same method but with Lewis *et al.*'s bias correction. Vemuri *et al.*'s algorithm produces similar results with

IBSR Subject 2

Method	Ground Truth	-14.64	-9.91	-4.75	-1.61
Our method ($c = 375$)		-8.21 (6.43)	-6.40 (3.51)	-4.746 (0.004)	-1.44 (0.17)
Our method ($c = 648$)		-15.36 (-0.72)	-9.15 (0.76)	-6.09 (-1.34)	-3.27 (-1.66)
Noblet <i>et al.</i>		-17.74 (-3.10)	-14.42 (-4.51)	-9.39 (-4.64)	-6.34 (-4.73)
ANTS		-10.42 (4.22)	-8.21 (1.70)	-3.85 (0.90)	-1.26 (0.35)
Vemuri <i>et al.</i>		-20.10 (-5.46)	-12.27 (-2.36)	-13.25 (-8.50)	-12.4 (-10.79)
ITK cubic B-spline ($c = 375$)		-5.80 (8.84)	-3.69 (6.22)	-2.58 (2.17)	-0.70 (0.91)
ITK cubic B-spline ($c = 648$)		-9.86 (4.78)	-7.44 (2.47)	-4.04 (0.71)	-2.36 (-0.75)

IBSR Subject 7

Method	Ground Truth	-14.61	-9.92	-4.91	-1.68
Our method ($c = 375$)		-8.21 (4.74)	-7.07 (2.85)	-4.52 (0.40)	-1.32 (0.36)
Our method ($c = 648$)		-13.10 (1.51)	-9.24 (0.68)	-4.28 (0.63)	-1.90 (-0.22)
Noblet <i>et al.</i>		-17.40 (-2.79)	-13.24 (-3.32)	-8.82 (-3.91)	-6.11 (-4.43)
ANTS		-11.9 (2.71)	-8.09 (1.83)	-3.99 (0.92)	-1.12 (0.56)
Vemuri <i>et al.</i>		-15.95 (-1.34)	-13.57 (-3.65)	-9.58 (-4.67)	-10.12 (-8.44)
ITK cubic B-spline ($c = 375$)		-7.74 (6.93)	-7.02 (2.89)	-2.74 (2.17)	-0.0013 (1.68)
ITK cubic B-spline ($c = 648$)		-9.66 (5.01)	-7.33 (2.59)	-4.02 (0.89)	-2.42 (-0.74)

Table 5.2: Comparison of MAP estimated mean atrophies (the quantity in brackets is difference from the ground truth *i.e.* the bias in the estimated atrophy) by our method to the ground truth, and NRR approaches of Noblet *et al.*, ANTS, Vemuri *et al.* and ITK cubic B-spline based non-rigid registration method. All values are shown in percentage of atrophy. Negative values show overestimation of atrophy and vice-versa. $c = 375, 648$ are the number of B-spline control points used in these experiments. The best results for each atrophy and subject are shown in bold.

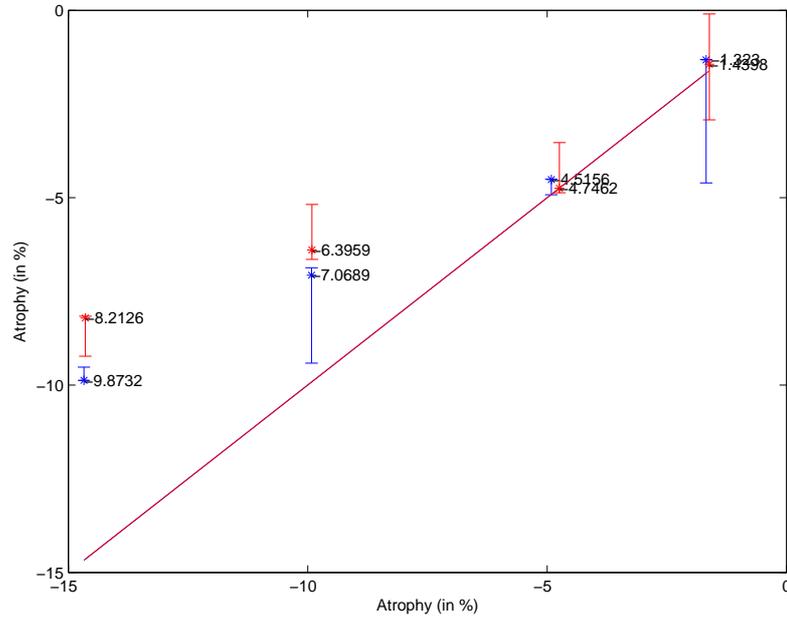
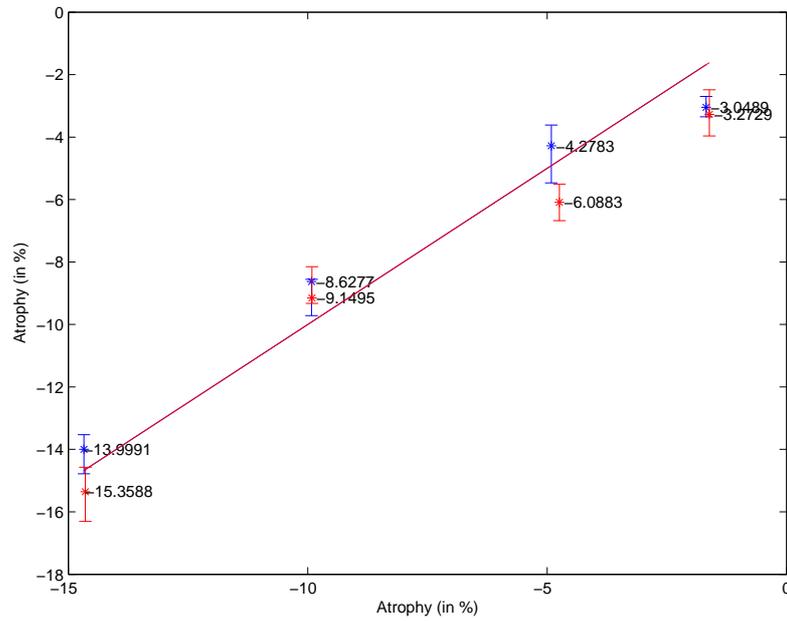
(a) $c = 375$ (b) $c = 648$

Figure 5.9: Estimated confidence intervals by the Bayesian registration method. All values are shown in percentage of atrophy. $c = 375, 648$ are the number of B-spline control points used in these experiments. Red and blue colours are used for subjects 2 and 7, respectively. The blue and red straight lines represent the ground truth atrophy and star (*) is the location of the MAP.

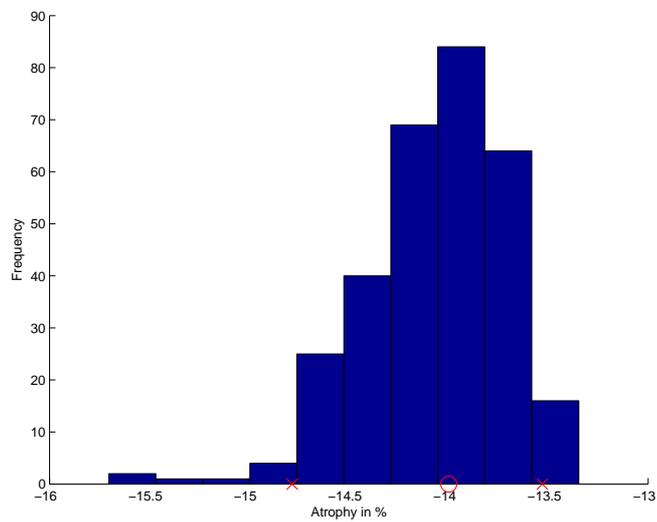
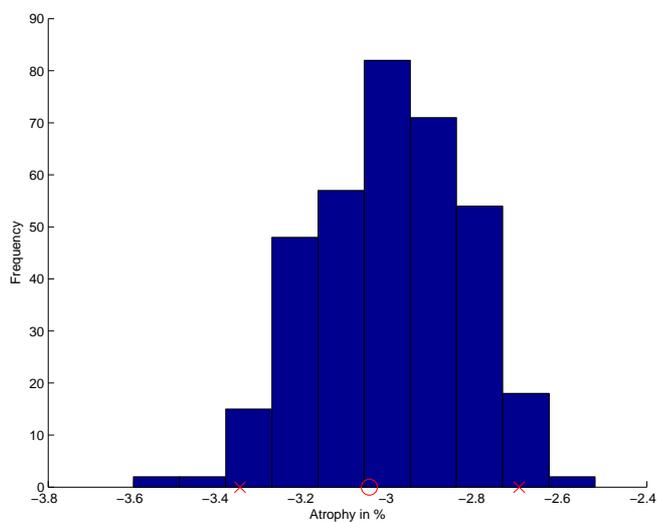
(a) $c = 375$, -14.64% (b) $c = 648$, -1.61%

Figure 5.10: The distributions of the Jacobians estimated for atrophies -14.64% and -1.61% with 648 of B-spline parameters. The location of the MAP is shown with a red circle and the 95% confidence intervals are marked with red crosses. Histograms are plotted with 10 bins.

IBSR Subject 2	
Bias field: 20% INU, Noise: 35 dB, GT: -14.6%	
Bias correction	Atrophy (in %)
Our approach	
Our approach	-19.18 CI (-16.64, -22.36)
N4ITK	-15.2 CI (-13.21, -14.92)
Lewis <i>et al.</i>	-15.18 CI (-14.43, -16.10)
Noblet <i>et al.</i>	
N4ITK	-17.11
Lewis <i>et al.</i>	-14.72
ANTS	
N4ITK	-10.56
Lewis <i>et al.</i>	-10.998
Vemuri <i>et al.</i>	
N4ITK	-15.72
Lewis <i>et al.</i>	-15.71
ITK cubic B-spline	
N4ITK	-9.68
Lewis <i>et al.</i>	-9.10

Table 5.3: Estimated atrophies with NRR algorithms of Noblet *et al.*, ANTS, Vemuri *et al.*, ITK cubic B-spline based as well as our approach with $c = 648$ B-spline control points. Differential bias correction of Lewis *et al.* and N4ITK bias field correction are employed. The proposed framework is used for estimating atrophy (no bias field corrections are performed) after bias corrections of Lewis *et al.* and N4ITK are applied to the images. An experiment is conducted where bias field correction is performed simultaneously with atrophy estimations. CI refers to the 95% confidence intervals. GT is the ground truth atrophy.

the two correction techniques. The results of simultaneous estimation of atrophy and bias field by making use of the MCMC approach are not on par with other methods (error with respect to the ground truth is 4.58%). The confidence intervals reveal that the estimated atrophy is not contained in them when the bias field correction is carried out with N4ITK. Further, the ground truth is not found in the intervals when bias field and atrophy are simultaneously evaluated using our approach. Nevertheless, note that the MCMC method manages to perform well (second to Noblet *et al.*'s NRR approach with Lewis *et al.*'s bias correction) with both Lewis *et al.*'s and N4ITK bias correction (respective errors with respect to the ground truths: -0.58%, -0.60%). The ground truth is contained inside the confidence intervals for these tests.

5.3.4 Consistency test

Further demonstration of our framework is performed with serially acquired images of an Alzheimer’s disease patient from the ADNI database. In the absence of a ground truth, one has to rely on other criteria of evaluation such as consistency between the measurements. ADNI database contains more than two longitudinally acquired images of a subject and provides a means to conduct consistency tests. An image (subject number 002_S_0295) is arbitrarily chosen for hippocampal atrophy estimation in our framework. Four serial scans are used in three consistency tests as follows: $a_{12} + a_{23} = a_{13}$, $a_{23} + a_{34} = a_{24}$ and $a_{12} + a_{23} + a_{34} = a_{14}$ where a_{ij} is the estimated atrophy between scans taken at instants i and j , where $i = 1, 2, 3$ and $j = 2, 3, 4$. The reader may refer to section 3.1.1 to find a description of the consistency criterion. No bias correction is applied as ADNI images have been already treated for intensity inhomogeneity with N3 algorithm. Each pair of images is registered with the affine registration available in ANTS. Besides, hippocampus’ segmentations are obtained from FreeSurfer’s segmentation routine. Experiments are performed on $32 \times 32 \times 32$ image regions containing the hippocampus. The results of these experiments are depicted in Fig. 5.11 where it can be observed that our method and ANTS are the most consistent due to the fact that the measurements are close to the $x = y$ line whereas Vemuri *et al.*’s NRR approach is the worst performer in this group of 5 methods.

5.4 Discussion and findings

In this chapter, a Bayesian non-rigid registration approach to measuring longitudinal brain atrophy was introduced. The most important contribution of this piece of work was that it allowed us to estimate bias field inhomogeneity and the deformation field simultaneously.

The Bayesian formulation consisted of modelling the non-rigid transformation as 3D cubic B-splines. Further, the bias field inhomogeneities were parametrised using Legendre polynomials. Gaussian likelihoods were chosen for the transformation and as a prior for the parameters of the cubic B-splines (which also acted as a regulariser). This was primarily because we considered Gaussian noise in the images. This assumption is valid when SNR in the image ≥ 2 dB and was verified in all our experiments. A different likelihood needs to be chosen if the noise is modelled by a Rician distribution, which is the actual distribution of MRI noise. Likewise, the noise standard deviation β_n sampling could be incorporated, in the present implementation its estimation was not in the MCMC framework. A uniform prior was imposed on the bias field. For the transformation field parameters, a Gaussian prior was chosen in conjunction with a uniform prior. The hyper-parameters (t_l , t_h , p_l and p_h) controlled the range in which the sampler could wander looking for probable values of the bias field and transformation field parameters, and were crucial in achieving convergence. β_r controlled the smoothness of the estimated transforma-

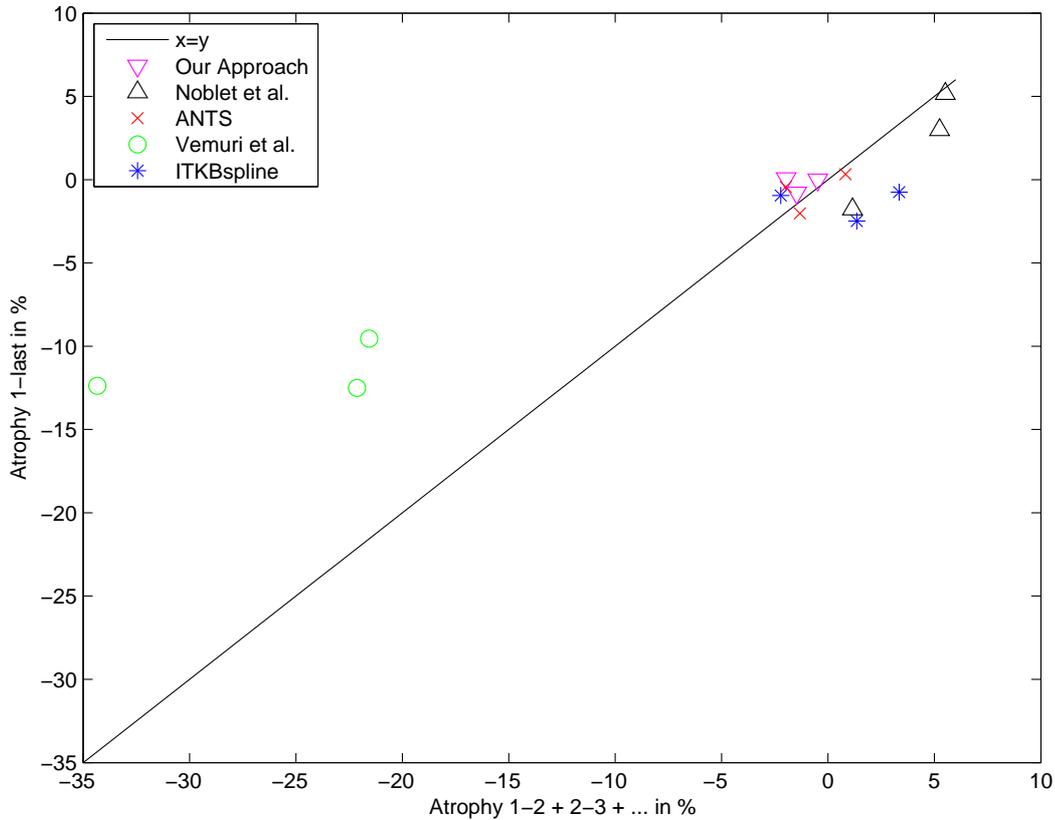


Figure 5.11: Consistency test. The x and y axes show summed and directly estimated atrophies between the first and the last time instants. If a method is perfectly consistent, its measurements must lie on the straight line $x = y$.

tion field.

The results were classified into three parts:

Estimation of bias field inhomogeneity Two simulated bias fields of 20% INU from the Brainweb simulated image database were recovered with the use of our method (with the maximum degree of Legendre polynomials in the range of $d = 2 - 5$), N4ITK [155] and differential bias correction of Lewis *et al.* [80]. In these experiments, only bias field inhomogeneity and noise artefacts were simulated. The images were free of any atrophy. The MSE between the MAP estimated fields and the ground truth field showed that Lewis *et al.*'s method could reproduce the bias field better than our method and N4ITK. However, the MSE achieved with our Legendre polynomial based modelling was not far from that of Lewis *et al.*'s method. We believe that the exceptional performance of differential bias correction method of Lewis *et al.* can be attributed to the fulfilment of its pre-requisites. The method requires registered and skull stripped images. Next, a median filter is ap-

plied for removing any anatomical structure from the difference image while being able to maintain sensitivity to adequately represent the differential bias field. As the two images in our experiments did not contain any anatomical changes, there was no residual anatomical structure in the difference image. Moreover, there was no registration error, as the images were already registered. Our approach being a parametric approach, depends on the validity of the assertion that bias field can be represented by Legendre polynomials as well as the registration error. Like Lewis *et al.*'s method, the absence of registration error and anatomical changes worked in favour of our algorithm too. At the same time, remember that for our approach, no knowledge of underlying tissue model or segmentation of the brain is needed.

The resulting bias fields from our tests showed that Legendre polynomial of degree 4 was optimal for the fields in the Brainweb database. The lowest MSE among experiments employing PABIC of Styner *et al.* [148], based on Legendre polynomial modelling, was also achieved for degree 4. However, it showed consistently higher MSE than the other methods. Possibly, this was due to the errors in segmentations as PABIC method takes the means and standard deviations of the tissue classes (WM, GM) as input. Values from K-means segmentation of the input image were provided. In order to confirm our assertions, we went back to the original contribution of Styner *et al.* [148] where they achieved an average MSE (in percentage of the contrast between white and grey matter) of $0.000904 - 0.322544$ over 20 bias fields that were generated from random Legendre parameter vectors (of 3rd degree) and applied to Brainweb images along with Gaussian noise between $0 - 9\%$ of intensity range. It may be mentioned here that the simulation of bias fields using Legendre polynomial parameters and recovering them with a Legendre polynomial based approach, might have driven PABIC to produce favourable results. As for N4ITK [155], the authors use default parameters in their experiments on the Brainweb database of 20 patients and comparisons are made by computing correlation coefficients. Due to differences in the experimental set-ups, our results could not be directly compared to neither PABIC nor N4ITK methods on the basis of the results produced in these articles.

A drawback of our MCMC based bias field inhomogeneity was the time requirement which is in hours. This is on the higher side when the other algorithms needed a few minutes. Recall that the main purpose of bias field inhomogeneity estimation was to aid in estimation of accurate confidence intervals on atrophies. To serve a larger perspective, confidence intervals for bias field parameters could be obtained for use in situations such as generation of bias field realisations to be incorporated in the uncertainty estimation database that was presented in chapter 4.

Estimation of atrophy Four atrophies (-14.6% , -9.9% , -4.7% , -1.6%) were simulated in the hippocampus of subject 2 and 7 from the IBSR database. The Bayesian method outperformed Noblet *et al.*'s and Vemuri *et al.*'s approach in al-

most all cases whereas it was better than ANTS in 6 out of 8 cases when 648 B-spline control points were used. The MCMC approach was better than ITK cubic B-spline based registration with a deterministic optimisation. In other experiments with 375 parameters, only smaller atrophies were well-estimated. From the visual Jacobian results, we could identify under-regularisation as a problem, which also explained the performance results of other algorithms. As discussed in 4.6, Vemuri *et al.*'s algorithm is particularly sensitive to noise and under-regularisation could be a reason of its failure. It is possible that transformation models and constraints could have been an influence. Similar observations about Jacobians were also made in [121].

The length of the confidence intervals was around 2% of atrophy with Gaussian noise of 35 dB, which is a good localisation range for hippocampal atrophy known to be around -9% in MS patients [3]. The estimated atrophy was contained in the intervals in all our tests whereas some intervals presented a slight bias with respect to the true atrophy. This may have arisen for two reasons: insufficiency of the lattice spacing or inefficiency of the cubic B-spline model.

Consistency tests on four serial scans from the ADNI database showed that our method as well as ANTS approach proved to be the most consistent among the 5 NRR methods tested. The differences in performances could be due to many factors such as the size of the image over which atrophy was estimated, due to the noise or residual artefacts that could not be taken care of by the ADNI pre-processing routines.

The issue of selection of number of control points is pivotal in determining the time complexity as well as the stability. Let us compare the lattice spacing in our experiments to other work based on cubic B-splines. Rueckert *et al.* [124] used a lattice spacing of 10 mm on breast images. Studholme *et al.* [147] studied registration of a group of subjects to a reference image in a multi-resolution setting with a lattice spacing of 2.4 mm at the finest level. Our experiments have shown that a lattice spacing of 5.5 mm could handle the simulated hippocampal atrophies (intra-subject registration) well. Nevertheless, if different atrophies are simulated in various ROIs at once, higher number of B-spline parameters would be required to achieve the same accuracy. Also, the regularisation will start to play a more important role as the lattice becomes finer. In our opinion, the trade off of time complexity vs accuracy is a crucial consideration. On an Intel Xeon processor at 2.5 GHz, 1000 samples were produced in 30 and 48 hours for 375 and 648 B-spline parameters. Contemporary publications such as Risholm *et al.* [119] sampled only the deformation at 349 boundary nodes with the MCMC algorithm leaving the deformation field calculation of the interior nodes to a deterministic algorithm. The total computation time for them was 14 hours (5 samples/second were generated). We suspect that an inefficient code may have contributed to the high computational time in our case.

Simultaneous estimation of bias field inhomogeneity and atrophy Results on IBSR image 2 with a simulated atrophy of -14.6% ; corrupted with a 20% INU bias field and noise of 35 dB; showed that the N4ITK method and differential bias correction approach of Lewis *et al.* are effective. After these corrections, our approach (without bias correction), Noblet *et al.*'s, ANTS and Vemuri *et al.*'s approach could estimate the atrophies in the hippocampus, with the errors with respect to the ground truth atrophy being less than 2%, barring one case of Noblet *et al.*'s method for which the error is 2.58%.

Let us recall certain facts to understand why there is a change in the performances of these approaches with respect to results on the Brainweb image. The better performance of N4ITK correction method in these tests could be due to differences in image quality between the simulated Brainweb and IBSR (contains real images) databases. Experiments with differential bias correction methods of Lewis *et al.* and ours, are also affected by the presence of atrophy in one of the images while N4ITK uses only one image for the bias estimation.

Simultaneous estimation of atrophy and bias field parameters using the MCMC framework did not perform as well as the previous case. The error with respect to the ground truth is 4.58%. As a result, the confidence intervals were skewed with respect to the location of the ground truth atrophy. Considering the MSE of the estimated bias field and the over-estimated atrophy, chances are that a part of the atrophy compensated for the bias field inhomogeneity. Although, a polynomial of maximum degree of 4 has only 56 parameters as compared to 648 parameters for the B-spline transformation, owing to the small size of the image ($32 \times 32 \times 32$), it is possible that during simultaneous estimation they interfere with each other's estimations, especially when the bias correction and atrophy estimation are performed at the same resolution. To remedy this situation, a multi-resolution framework could be built where the bias field estimation is performed only at lower resolutions. Also, further loss in interference between bias field and atrophy parameter sampling scheme could be sought by sampling atrophy more than one time in one iteration.

To conclude, the Bayesian approach works well, both in the presence of noise as well as both noise and bias field inhomogeneity in the images. Simultaneous estimation of bias field inhomogeneity and atrophy performed worse than other bias correction and non-rigid registration approaches.

Bayesian modelling and MCMC sampling techniques allowed us to develop a framework for estimating atrophy together with bias field inhomogeneity. Carefully choosing the right models for the quantities of interest is a key step [48]. Furthermore, MCMC sampling methods could be time consuming, particularly when the number of parameters to be sampled are large. The motivation behind the use of such an expensive framework is to be able to estimate uncertainties in volumetric measurements. Our results have justified the interest of developing a Bayesian ap-

proach and that this framework holds promise in the future. This is a step towards estimating uncertainties in atrophy, an area that was relatively untouched by the makers of algorithms dedicated to atrophy measurements.

General conclusion and future perspectives

The work performed in the course of this thesis is dedicated to the analysis of longitudinal atrophy estimation from serially acquired scans of a subject.

Simulation of atrophy is a way of generating desired ground truths for evaluation of methods. The performance of three freely available algorithms (SIENA, SIENAX and BSI-UCD) and NRR algorithm of Noblet *et al.* [98] by simulating atrophy in Brainweb images was assessed. By means of simulating whole brain atrophies in a single Brainweb image (to examine the robustness to bias field inhomogeneity, noise, geometrical distortions and interpolation artefacts) and statistical analysis of the results obtained on 18 different anatomical models of the brain, an extensive validation was carried out. 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 in the group of three widely used methods SIENA, BSI and SIENAX. Other experiments that included JI showed that it performed better than SIENA. Moreover, bias field inhomogeneity and noise contributed to higher errors in atrophies as compared to interpolation artefacts and geometrical distortions.

With the help of this study, we also identified that bottlenecks for SIENA are registration of the two given exams and segmentation of the boundary voxels (that is affected by the accuracy of the brain extraction). The critical steps for BSI-UCD are the registration of the two brain scans 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, [14]). JI depends both on the non-rigid registration of the serial images and the segmentation of brain for extracting the region of interest. The registration in turn is a function of many aspects such as the model (similarity criterion, regularisation function) and algorithmic settings (such as levels in the multi-resolution framework and optimisation parameters). Pre-processing techniques play a critical role in its correct functioning.

The key learning from this study is that errors that were observed here were

comparable to the whole brain annual atrophy rates (0.5 – 2.8%) that have been reported for various pathologies. One of the reasons that motivated us to develop a framework that allowed the estimation of uncertainties in estimated atrophies.

To this end, a novel and generic framework for constructing confidence intervals by an atrophy estimation method was devised. The most remarkable feature of this approach is that it was capable of correcting for errors that arose from bias in atrophy estimations as well as of capturing the role of MRI artefacts in the mis-estimation of atrophy. This database is flexible as any atrophy (*e.g.* in a region of interest such as the hippocampus or the entire brain) and other image artefacts, depending on the user's requirement can be included. Besides, this database can be created for images acquired from the same scanner and reused for future estimations. The efficiency of this framework was demonstrated by comparing the performance of three NRR algorithms, by Noblet *et. al* [98], ANTS [10] and by Vemuri *et al.* [157] for building confidence intervals for hippocampal atrophies. The bias in atrophy estimations revealed different patterns among the three methods. For Noblet *et al.*'s algorithm, the bias remained almost constant with the true atrophy value (max. value -4.22%). On the other hand, for the ANTS method, the bias in estimations increased with increasing ground truth atrophy (max. value $+5.39\%$). Bias in estimations was inversely proportional to true atrophies for the estimations of Vemuri *et al.* (max. value -8.86%). Smith *et al.* [142] also confirmed such behaviour in SIENA and BSI. This analysis confirms that the correction of this bias is necessary to strengthen the accuracy of atrophy estimation methods. The origin of this bias may be explained by the use of different transformation models; smoothness and other constraints that were a part of the NRR methods. Furthermore, through this study, it was concluded that ANTS provides meaningful confidence intervals as compared to those of NRR methods of Noblet *et. al* and Vemuri *et al.*.

In the final stage of the thesis, a novel atrophy estimation algorithm was developed. We built a complete Bayesian model that facilitated the estimation of confidence intervals for atrophies in an ROI along with the estimation of bias field inhomogeneity and noise, given two serially acquired images of an individual. The Bayesian formulation consisted of modelling the non-rigid transformation as 3D cubic B-splines and the bias field inhomogeneities were parametrised using Legendre polynomials. It considered Gaussian noise in the images. A combination of Gaussian and uniform priors was used for the transformation parameters whereas a uniform prior was selected for bias field parameters.

Results on images degraded with Gaussian noise have shown that the Bayesian approach outperformed ANTS, Noblet *et. al.*'s and Vemuri *et al.*'s NRR methods in most of the cases. The maximum and minimum errors with respect to the ground truth atrophy were 0.22% and -1.66% , respectively. It was also seen that the recovered Jacobian maps from these methods were very distinct. An implementation of a non-rigid registration method having the same transformation model

(cubic B-splines) but optimised by a deterministic optimisation was worse than our approach. Simultaneous estimation of bias field inhomogeneity and noise in the Bayesian framework overestimated the atrophy and underestimated the bias field. Accurate estimation of bias field is related to accuracy of the estimated atrophy [109]. Atrophy estimation in the Bayesian framework after correction of images for bias field inhomogeneity by methods of Lewis *et al.* [80] and N4ITK [155], produced better results than simultaneous estimation of bias field inhomogeneity and atrophy in the same framework. A possible reason for this kind of behaviour could be using the data from a small image patch for estimating bias field inhomogeneity and atrophy iteratively at the same resolution. The major drawback of this method comes from the time consuming nature of MCMC sampling, particularly when the number of parameters to be sampled are large. Nevertheless, an expensive framework like the one described in this thesis is required to be able to estimate uncertainties in volumetric measurements. Another limitation of this approach is that the constructed confidence intervals do not always contain the ground truth in all cases due to the resolution of the B-spline lattice or inefficiency of the models employed.

Note that the problem of determining uncertainties in longitudinal atrophies explicitly is an area that was until now mostly ignored by the makers of algorithms dedicated to atrophy measurements.

Summary of contributions

- Development of a framework that enabled the comparison of existing brain atrophy estimation approaches by creating a simulated ground truth and evaluation of widely used atrophy estimation approaches to assimilate the effect of bias field inhomogeneity, noise, geometrical distortions, interpolation artefacts and MS lesions on atrophy estimation.
- Uncertainty estimation for any brain atrophy estimation algorithm in a region of interest in the form of confidence intervals for atrophy; the highlight of uncertainty estimation is the correction of bias in atrophy estimations.
- Construction of a full Bayesian model for simultaneous estimation of atrophy, bias field inhomogeneity and noise as well as estimation of confidence intervals on the estimated atrophy.

Future perspectives

In this thesis, the atrophy simulated was simplistic and was not learnt from real cases. It is possible to construct an intricate model of deformations that parts of the brain undergo and use them to simulate atrophy such as in [169]. However, in doing so, the simulations run the risk of presenting biases of their own in estimations and should be taken into account. May be in some cases the constraint of topology

preservation can be relaxed as it may be too restrictive on the type of deformations that can be generated.

From our analyses, in chapters 4 and 5 it is clear that atrophy estimations are biased and must be corrected to deliver accurate atrophy measurements. One of the main reasons for biased estimations is transformation model error, smoothness and other constraints such as topology preservation. The Bayesian model can be extended to incorporate this error. For this, one or more independent sources of ascertaining atrophy is required, for example, a simulated ground truth or manual segmentations. This is not a simple task as the error may be dependent on image artefacts; magnitude of volumetric change and has to be learnt from a pair of images. Moreover, the accuracy of the ground truth must be accounted for (Is the simulated ground truth biased? How accurate are manual segmentations?).

Let G be the ground truth atrophy obtained from an independent source of measuring atrophy (such as manual segmentations), A be the estimated longitudinal atrophy, δ be the model error and ε_G and ε_θ are the measurement errors in the ground truth and the estimated atrophy. A model on the following lines can then be used and put in a Bayesian framework by finding appropriate models for δ and ε . [69]:

$$G + \varepsilon_G = A + \delta + \varepsilon_\theta \quad (6.1)$$

Here, the term model error (δ) accounts for the fact a mathematical model may not represent the complicated physical phenomenon of atrophy accurately. Measurement error (ε) is a measure of reproducibility of the same measurement under unchanged conditions (*e.g.* precision of manual segmentations). For the same reason, if a mathematical model is used for generating the ground truth it may not correspond to way atrophy actually happens in the brain. Therefore, we recommend the use of several independent sources in order to obtain the true atrophy and/or manual segmentations (that are not based on any mathematical model but are prone to human error). Bias correction within the Bayesian framework will ensure that the constructed confidence intervals contain the true value.

Furthermore, the Bayesian model can be implemented in a multi-resolution framework and bias correction can be performed only at lower resolutions to minimise the interaction between the bias field parameters and the transformation parameters. Improvements can be sought in the way MCMC sampling was performed in chapter 5. Because of the smoothness constraints, the deformation field is correlated in neighbourhoods and therefore if these parameters are sampled individually, it leads to slow mixing of the chain. To alleviate this problem, we can think of performing MCMC in blocks instead of sampling each parameter at each iteration. Different proposal distributions for different B-spline parameters can improve convergence rate. In addition, statistical voxel-wise analysis can be performed using the Jacobian distributions obtained from the Bayesian framework.

It was also found that different configurations of non-linear transformations, regulariser and constraints led to different types of Jacobian maps and deformation fields. An extensive evaluation of the Bayesian model using other non-linear transformations, regulariser and constraints such as diffeomorphism can be explored to determine the best configuration and learn the error distributions. It would be beneficial to perform this evaluation on a real as well as simulated images.

Finally, some advice..

We would like to bring an end to this thesis by reiterating that atrophy estimation is a subject that must be handled delicately and methods of measuring volumetric changes must be evaluated thoroughly before any conclusions are drawn from them. One of the major issues arises from conflicting results produced by different algorithms, on the same images, because the sensitivity of each method to image artefacts and adaptation to anatomy is subject to variation. Therefore, image processing specialists must resort to understanding the behaviour of these approaches, analysis of uncertainties, reasons for failure and determining the possibility of changes measured in the brain being real changes. It is through efforts in this direction that brain atrophy will become a more reliable biomarker of disease progression.

Appendix I

A.1 Legendre polynomials

The bias field $B(x)$ at a voxel x can be expressed as a linear combination of smooth basis functions. Legendre polynomials represent one of the many basis functions capable of parametrizing the smooth bias-field inhomogeneity in MR images. For Legendre polynomials of degree l , the size m of the parameter vector of the Legendre polynomial coefficients π is given by:

$$m = (l + 1) \frac{(l + 2)}{2} \frac{(l + 3)}{3} \quad (\text{A.1})$$

In 3-D, the bias-field estimate \hat{B} is derived as follows:

$$\hat{B}(x, p) = \sum_{i=0}^l \sum_{j=0}^{l-i} \sum_{k=0}^{l-i-j} \pi_{ijk} L_i(x) L_j(x) L_k(x) \quad (\text{A.2})$$

where $L_{i(\cdot)}$ denotes a Legendre polynomial of degree i . The image coordinates x are scaled in the range of $[-1, 1]$.

In mathematics, Legendre functions are solutions to Legendre's differential equation:

$$\frac{d}{dx} \left[(1 - x^2) \frac{d}{dx} L_i(x) \right] + i(i + 1) L_i(x) = 0. \quad (\text{A.3})$$

These solutions for $i = 0, 1, 2, \dots$ (with the normalization $L_i(1) = 1$) form a polynomial sequence of orthogonal polynomials called the Legendre polynomials. Each Legendre polynomial $L_i(x)$ is an i th-degree polynomial and can be expressed using Rodrigues' formula:

$$L_i(x) = \frac{1}{2^i i!} \frac{d^i}{dx^i} [(x^2 - 1)^i]. \quad (\text{A.4})$$

First few Legendre polynomials are:

i	$P_i(x)$
0	1
1	x
2	$\frac{1}{2}(3x^2 - 1)$
3	$\frac{1}{2}(5x^3 - 3x)$
4	$\frac{1}{8}(35x^4 - 30x^2 + 3)$
5	$\frac{1}{8}(63x^5 - 70x^3 + 15x)$
6	$\frac{1}{16}(231x^6 - 315x^4 + 105x^2 - 5)$
7	$\frac{1}{16}(429x^7 - 693x^5 + 315x^3 - 35x)$
8	$\frac{1}{128}(6435x^8 - 12012x^6 + 6930x^4 - 1260x^2 + 35)$
9	$\frac{1}{128}(12155x^9 - 25740x^7 + 18018x^5 - 4620x^3 + 315x)$
10	$\frac{1}{256}(46189x^{10} - 109395x^8 + 90090x^6 - 30030x^4 + 3465x^2 - 63)$

A.1.1 Orthogonality property

The Legendre polynomials $L_i(x)$, $i = 0, 1, 2, \dots$ are orthogonal on the interval from $[-1, 1]$ which is expressed by the following integral:

$$\int_{-1}^1 L_i(x)L_j(x)dx = \delta_{ij} \frac{2}{2i+1} \quad (\text{A.5})$$

where $\delta_{ij} = 1$ if $i = j$ and 0 otherwise.

Appendix II

B.1 Experimental set-up

This chapter describes the implementations of algorithms and parameters used in the experiments presented in this thesis.

B.1.1 Image database

Images from three databases are processed in our tests.

Brainweb simulated image database

Brainweb¹ is a simulated brain database that contains a set of realistic MRI data volumes produced by an MRI simulator [7]. These data provide the researchers with a ground truth to validate their algorithms. This database contains simulated brain MRI data for normal subjects and MS patients. For both cases, full 3-D data volumes have been simulated using three sequences (T1-weighted, T2-weighted, and proton-density weighted) and for varying slice thicknesses, noise levels, and levels of intensity non-uniformity. Furthermore, another simulated database containing 20 anatomical models for 20 normal subjects with fuzzy segmentations are also provided [8]. Brainweb database is used in experiments in chapters 3 and 5.

Internet brain segmentation repository

The IBSR² database contains real MR image data with manually guided segmentations. Our experiments have been performed on T1-weighted MR image data from 18 anatomies with expert segmentations of 43 individual structures (1.5mm slice thickness). IBSR database is used in experiments in chapters 3, 4 and 5.

Alzheimer's disease neuroimaging initiative (ADNI)

The ADNI³ database is a collection of images acquired using different protocols and imaging systems. One of the objectives of this database is to create a standardised repository of scans acquired from Alzheimer's disease patients longitudinally. The following image correction procedures are applied to raw ADNI images:

¹<http://mouldy.bic.mni.mcgill.ca/brainweb/>

²www.cma.mgh.harvard.edu

³<http://adni.loni.ucla.edu/about-data-samples/image-data/database>

1. Gradwarp is a correction procedure that corrects for image distortions due to gradient non-linearity that are dependent on the scanner.
2. B1 non-uniformity employs B1 calibration scans to correct for image intensity non-uniformity for image intensity inhomogeneity due to receiver coil non uniformity.
3. N3 intensity inhomogeneity correction is applied after Gradwarp and B1 correction for systems on which these two correction steps are performed.

Tests in chapter 5 are carried out on 4 longitudinally acquired T1-weighted MR images of a patient to which the aforementioned corrections had been applied. Image are acquired using a 1.5T scanner.

B.1.2 Bias field inhomogeneity correction

Following bias field correction strategies have been employed in the thesis:

N4ITK

The N4ITK is an improved version of N3 correction. Its implementation⁴ is given as a single class `itk::N3MRIBiasFieldCorrectionImageFilter` in Insight Toolkit (ITK). Here is a description of the parameters of the algorithm:

1. `m_MaskLabel` (default = 1): Mask of the area of the image over which bias field is computed.
2. `m_NumberOfHistogramBins` (default = 200): For intensity profile construction from the intensities of the uncorrected input image and a triangular parzen windowing scheme.
3. `m_WeinerFilterNoise` (default = 0.01): Field estimation is performed by deconvolution using a Wiener filter which has an additive noise term to prevent division by zero.
4. `m_BiasFieldFullWidthAtHalfMaximum` (default = 0.15): The full width at half maximum (FWHM) characterising the Gaussian modelling of the bias field.
5. `m_MaximumNumberOfIterations` (default = 50): Maximum number of iterations.
6. `m_ConvergenceThreshold` (default = 0.001): The standard deviation of the ratio between subsequent field estimates is used.
7. `m_SplineOrder` (default = 3): A smooth field estimate is produced using B-spline after each corrective iteration.

⁴<http://www.kitware.com/products/html/N3ITKImplementationForMRIBiasFieldCorrection.html>

8. `m_NumberOfControlPoints` (default = `m_SplineOrder+1`): Number of control points in the B-spline. Minimum is `m_SplineOrder+1`.
9. `m_NumberOfFittingLevels` (default = 1): Number of resolutions for fitting the B-spline.

The parameters are set to their default values in our experiments in chapter 5.

Differential bias field correction [80]

The implementation of differential bias field correction algorithm is done in ITK. This algorithm has two parameters:

1. Smoothing filter neighbourhood size: set to 5.
2. BET parameters: default parameters are used. When a ROI is extracted from the brain, brain extraction is not needed.

PABIC [148]

The implementation of PABIC is available in ITK applications. This algorithm has the following parameters:

1. `-input file` : Image to be corrected for bias field inhomogeneity.
2. `-class-mean` : Intensity means of the different tissue classes WM: 110.83; GM+CSF: 54.88
3. `-class-sigma` WM: 16.00; GM+CSF: 25.50
4. `-input-mask file`: Same as input file (bias field is estimated over the brain area).
5. `-degree int`: degree of Legendre polynomial used for the approximation of the bias field. These values are between 2 – 5.
6. `-use-log`: A multiplicative bias field is assumed. This parameter is set to “yes”.
7. `-grow double` (default 1.05), `-shrink double` (default $\text{grow}^{(-0.25)}$): optimisation parameters. Default parameters are used.
8. `-volume-max-iteration` (default 20): Number of iterations for 3D volume correction. It is set to 50.
9. `-init-step-size` (default 1.02): Initial step size for optimisation. This parameter is set to its default value.
10. `-coefficients`: Initial coefficients of the polynomial. No initialisation is provided.

11. `-schedule` (default 2 2 2 1 1 1): Multi-resolution schedule. Default schedule is used.

All three algorithms take a few seconds on images of size $64 \times 64 \times 64$.

B.1.3 Simulation of atrophy

The algorithm of atrophy simulation is described in section 3.2.1. The Brainweb simulated image database [7, 8] is used in evaluating brain atrophy estimations in chapter 3, where whole brain atrophies are studied. In experiments on uncertainty estimations (chapter 4, 5), atrophies are simulated in a region of interest (the hippocampus) with an aim of reducing the time complexity. Tests on hippocampal atrophy evolution are performed on 18 brains from IBSR database.

B.1.4 Brain atrophy estimation methods

SIENA, SIENAX and BSI-UCD

In our work, we used the SIENA and SIENAX implementations available as a part of the FMRIB Software Library (FSL)⁵ version 4.1 [141].

When the experiments were being performed with BSI, the original BSI implementation was not available, so we used the BSI implementation developed by Imaging of Dementia and Ageing lab, University of California, Davis (BSI-UCD)⁶. The calculation of the boundary shift integral is done in the same way as described in [44] but different pre-processing algorithms are used. The bias correction of BSI-UCD⁷ 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 package⁸. 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 this thesis. These bias correction and registration routines are used in all the experiments with BSI unless mentioned otherwise.

The experiments are carried out with 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 “unnormalised” brain volume estimates are used for all subsequent calculations and comparisons.

⁵<http://www.fmrib.ox.ac.uk/fsl/fsl/list.html>

⁶<http://neuroscience.ucdavis.edu/idealab/software/index.php>

⁷http://neuroscience.ucdavis.edu/idealab/software/bias_correction.php

⁸http://neuroscience.ucdavis.edu/idealab/software/linear_coreg.php

Jacobian Integration (JI)

JI has been performed on deformations from three non-rigid registration algorithms, namely Noblet *et al.* [98], Advanced Normalisation Tools (ANTS) [10] and Vemuri *et al.* [157], individually based on different modelling schemes:

Noblet *et al.*'s NRR method The implementation of Noblet *et al.* [98] is available as it is an in-house non-rigid registration method. They estimate 3-D topology preserving mappings where the deformation map is modeled as a hierarchical displacement field, decomposed on a multi-resolution B-spline basis. By restricting the Jacobian of the transformation topology preservation in the continuous domain is ensured. Interval analysis optimisation are employed. They use sum of squared differences of image intensities (SSD) as a matching criterion with elastic membrane regularisation 3.3. Parameters are set to the following values:

1. Regularisation: Membrane energy with a weighting factor of $\lambda = 1$.
2. Normalisation: Histogram normalisation by quantile of the Insight Toolkit (ITK)⁹ is used with number of quantiles=7.
3. Bias-field correction: Differential Bias Correction (Lewis *et al.* [80]) (used only for bias affected images).
4. Resolution: 7 (for images of size $256 \times 256 \times 256$) and 4 (for images of size $32 \times 32 \times 32$).

Same parameters are used for the algorithm of simulating atrophy.

ANTS method The method of **Avants *et al.* [10]** (available with Advanced Normalisation Tools, ANTS¹⁰) is also employed, as it has been ranked as one of the best among 14 NRRs [70]. Avants *et al.* have developed a symmetric image normalisation method (SyN) for maximising the cross-correlation within the space of diffeomorphic maps and use the Euler-Lagrange equations for optimisation. The following command is used ANTS 3 -m PR[fixedImage, movingImage, 1, 1] -o outputImage -i $30 \times 100 \times 100$ -r Gauss[1, 0] -t SyN [0.25] -number-of-affine-iterations 0. where

1. PR metric is a strict implementation of correlation taking the parameters [fixedImage,movingImage,weight,radius] where fixedImage and movingImage images are the two images to be registered, weight for this metric and the region radius for computing cross correlation.
2. -i $30 \times 100 \times 100$ refers to the number of iterations at each resolution. A level three pyramid with 100 iterations each at the first two levels and 30 iterations at the finest level is used.

⁹<http://www.itk.org/>

¹⁰<http://www.picsl.upenn.edu/ANTS/>

3. -r Gauss[1, 0] Gaussian regulariser with a standard deviation of 1 that operates only on the similarity gradient.
4. -t SyN [0.25] means that the estimated transformation is represented in terms of a symmetric (both forward and backward transforms are calculated) diffeomorphic model. 0.25 is the gradient step length.
5. Affine registration is not needed because images are already placed in the same co-ordinate system.

The method by Avants *et al.* [10] will be referred to as ANTS in the future.

Vemuri *et al.*'s NRR method In this deformable registration framework, a level-sets framework achieves image intensity morphing and a simple non-linear PDE representing motion as optical flow is used for the corresponding coordinate registration. An implementation of this algorithm is available in ITK as `itk::LevelSetMotionRegistrationFilter`, which takes only one parameter of the number of iterations to be executed. We set the maximum number of iterations to 200.

ITK cubic B-spline based NRR A cubic B-spline based NRR is implemented in ITK. SSD is chosen as the similarity metric, linear interpolation is used and the optimiser is set to regular step gradient descent.

B.1.5 Brain Segmentation

While the implementations of SIENA and SIENAX are completely automated, the implementation of BSI-UCD 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) [139] of the FSL library is used for performing 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. B.1(b)) such as the one obtained as an output of BET after the removal of the non-brain tissue and another mask defining the gray-white matter (the atrophy affected areas of the brain) boundaries (Fig. B.1(c)). For performing integration of the Jacobian obtained from the algorithm of Noblet *et al.* [98], the ground-truth mask of the whole brain (gray-white matter mask) is employed. Manual hippocampus segmentations are available in the IBSR for atrophy estimation in this region of interest by Noblet *et al.* [98], ANTS [10] and Vemuri *et al.* [157].

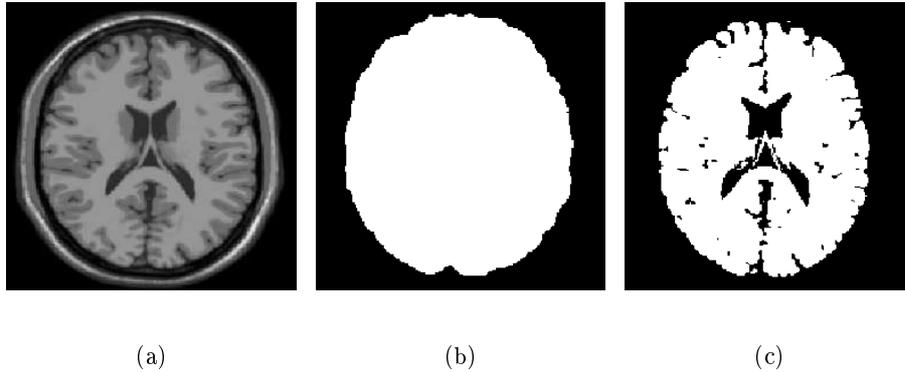


Figure B.1: (a) A T1-weighted image (b) Brain mask (extracted by BET) and (c) Gray-White matter mask of the brain computed from the T1-weighted image shown in (a).

A summary of the brain extraction, registration and bias field correction algorithms used for SIENA, SIENAX, BSI and NRR of Noblet *et al.* in this work is illustrated in the Table B.1. These four algorithms are evaluated in the first part (3) of the thesis. The similarity criteria and transformation models for deformable

Table B.1: Summary of the brain extraction, registration and bias field correction algorithms used for SIENA, SIENAX, BSI-UCD and NRR of Noblet *et al.*.

	SIENA	SIENAX	BSI-UCD	Noblet <i>et al.</i>
Brain extraction	BET	BET	BET	-
Registration type	Affine(FLIRT)	Affine(FLIRT)	Affine	B-spline based deformable
Similarity criterion	Correlation ratio	Correlation ratio	Cross-correlation	SSD
Final interpolation	Tri-linear	Tri-linear	Tri-linear	B-spline order 5
Bias-Field correction	With segmentation	With segmentation	Template based	DBC

BET: Brain Extraction Tool [139], **FLIRT**: [66], **Segmentation (EM based)**: [175], **DBC**: Differential Bias Correction[80], **SSD**: Sum of Squared Differences

registration methods tested in this work are enlisted in the Table B.2. Noblet *et al.*, ANTS and Vemuri *et al.* are used for experiments on uncertainty estimation in chapters 4.4. We used histogram normalisation by quantile (available with ITK) for all the methods in order to achieve intensity normalisation.

Table B.2: Deformable registration algorithms of Noblet *et al.*, ANTS, Vemuri *et al.* and ITK cubic B-spline based .

	Noblet <i>et al.</i>	ANTS	Vemuri <i>et al.</i>	ITK cubic B-spline
Similarity criterion	SSD	Cross-correlation based	SSD	SSD
Transformation model	Linear B-spline	Bi-directional diffeomorphism	Optical flow	Cubic B-spline

Time Complexity

- **Experiments on Evaluation of Brain Atrophy Estimations (SIENA,**

SIENAX, BSI-UCD and Noblet *et al.*) 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, BSI-UCD and JI (Noblet *et al.* [98]) are approximately 1 hour, 20 minutes, 1 hour and 1.5 hours, respectively. All the experiments are rendered on an Intel Dual Core 2.40 GHz processor with images of size $256 \times 256 \times 256$.

- **Experiments on Uncertainty Estimation in Brain Atrophy Estimations (Noblet *et al.*, Avants *et al.* and Vemuri *et al.*)** The time required for one experiment of simulation of atrophy using the proposed approach is 5 – 6 minutes. The computational times (including the time required for pre-processing) when performing one experiment with Noblet *et al.*, Avants *et al.* and Vemuri *et al.* on images of size $32 \times 32 \times 32$ are approximately 5 – 6 minutes, 1 – 2 minutes and 3 – 4 minutes, respectively. All the experiments are rendered on an Intel Dual Core 2.40 GHz processor.

Appendix III

C.1 MCMC algorithms

This appendix is dedicated to widely used MCMC methods: Metropolis-Hastings and Gibbs Sampler.

C.1.1 Metropolis-Hastings algorithm

The Metropolis-Hastings algorithm, developed by Metropolis, Rosenbluth, Rosenbluth, Teller, and Teller (1953) and generalised by Hastings (1970), is an MCMC method.

With MH, only a function proportional to the target probability distribution is required. The normalising constant need not be known, which is hard to calculate in most Bayesian applications. Gibbs sampler is a special case of MH applicable only in certain situations but possesses the advantage of being faster.

MH method generates a random walk using a proposal density and a criterion for accepting proposed moves. It creates a Markov Chain where the next state (s^{i+1}) depends on the current state of the sampler (s^i). A proposal density $Q(s'; s^i)$ generates a proposal for the next state (s') depending on the current state s^i .

An example of the proposal density is Gaussian $Q(s'; s^i) \sim \mathcal{N}(s', \sigma^2 I)$ which will generate samples with a mean equal to the current density and a variance of $\sigma^2 I$. The best results are obtained when the proposal density follows the shape of the target distribution $Q(s'; s^i) \approx P(s')$. Generally, the target distribution is unknown.

This proposal is “accepted” as the next value according to the following rules. Let

$$k = \frac{P(s')Q(s^i; s')}{P(s^i)Q(s'; s^i)} \quad (\text{C.1})$$

where is $\frac{P(s')}{P(s^i)}$ the likelihood ratio between the proposed sample s' , and the previous sample s^i and $\frac{Q(s^i; s')}{Q(s'; s^i)}$ is the ratio of the proposal density between the two states. This ratio is equal to 1 for a symmetric proposal density such as the Gaussian distribution. If the ratio k is:

```

if  $k \geq 1$  then
   $s^{i+1} = s'$ 
else
   $s^{i+1} = \begin{cases} s' & \text{with probability } k \\ s^i & \text{with probability } 1 - k. \end{cases}$ 

```

C.1.2 Gibbs sampler

When sampling for multivariate distributions, it is simpler to sample from a conditional distribution than to marginalise by integrating over a joint distribution. This is useful when the joint distribution is unknown or difficult to draw samples from directly but the conditional distribution for each variable is known (this is the case for Markov random field models). Suppose that we would like to draw samples of $\mathbf{s} = \{s_1, s_2, \dots, s_n\}$ from a joint distribution $p(s_1, s_2, \dots, s_n)$.

Let the t th sample be denoted by $\mathbf{s}^{(t)} = \{s_1^{(i)}, s_2^{(i)}, \dots, s_n^{(i)}\}$. Beginning with some initial value, for each variable, $s_j^{(i)}$, sample from the conditional distribution $p(s_j^{(i)} | s_1^{(i)}, \dots, s_{j-1}^{(i)}, s_{j+1}^{(i-1)}, \dots, s_n^{(i-1)})$. In the i th iteration, each variable $s_j^{(i)}$ is sampled once conditioned on all other variables, making use of the newly sampled values of variables $(s_{j-1}^{(i)}, s_{j-2}^{(i)}, \dots, s_n^{(i-1)})$ that have been sampled before $s_j^{(i)}$ in the i th iteration.

Interested readers can find an in-depth treatment of MCMC algorithms in [46].

Publications

- [1] Sharma, S., Noblet, V., Rousseau, F., Heitz, F., Rumbach, L., and Armspach, J.-P. (2008). Evaluation of brain atrophy estimation algorithms using simulated ground-truth data. In *Workshop MIAMS - MICCAI 2008*, <http://miams08.inria.fr>. (Not cited.)
- [2] Sharma, S., Noblet, V., Rousseau, F., Heitz, F., Rumbach, L., and Armspach, J.-P. (2009a). Influence of the presence of MS lesions on the measured atrophy. In *25th Congress of the European Committee for the Treatment and Research in Multiple Sclerosis (ECTRIMS)*. (Not cited.)
- [3] Sharma, S., Noblet, V., Rousseau, F., Heitz, F., Rumbach, L., and Armspach, J.-P. (2009b). Use of simulated atrophy for performance analysis of brain atrophy estimation approaches. In *Medical Image Computing and Computer Assisted Intervention*, volume 5762 of *Lecture Notes in Computer Science*, pages 566–574. Springer. (Not cited.)
- [4] Sharma, S., Noblet, V., Rousseau, F., Heitz, F., Rumbach, L., and Armspach, J.-P. (2010). Evaluation of brain atrophy estimation algorithms using simulated ground-truth data. *Medical Image Analysis*, 14(3):373–389. (Not cited.)

Bibliography

- [1] Aja-Fernandez, S., Alberola-Lopez, C., and Westin, C. F. (2008). Noise and signal estimation in magnitude MRI and Rician distributed images: A LMMSE approach. *IEEE transactions on image processing*, 17(8):1383–1398. (Cited on pages 29, 30, 31 and 32.)
- [2] Anbeek, P., Vincken, K. L., van Bochove, G. S., van Osch, M. J., and van der Grond, J. (2005). Probabilistic segmentation of brain tissue in MR imaging. *NeuroImage*, 27(4):795 – 804. (Cited on page 63.)
- [3] Anderson, V. M., Fisniku, L. K., Khaleeli, Z., Summers, M. M., Penny, S. A., Altmann, D. R., Thompson, A. J., Ron, M. A., and Miller, D. H. (2010). Hippocampal atrophy in relapsing-remitting and primary progressive MS: A comparative study. *Multiple Sclerosis*, 16(9):1083–90. (Cited on pages 117 and 156.)
- [4] Anderson, V. M., Fox, N. C., and Miller, D. H. (2006). MRI measures of brain atrophy in multiple sclerosis. *Journal of Magnetic Resonance Imaging*, 23:605–618. (Cited on pages 25, 26 and 57.)
- [5] Ashburner, J. and Friston, K. (2000). Voxel-based morphometry: the methods. *NeuroImage*, 11:805–821. (Cited on pages 2, 4, 36, 40, 41, 49, 55 and 57.)
- [6] Ashburner, J. and Friston, K. J. (2005). Unified segmentation. *NeuroImage*, 26(3):839 – 851. (Cited on pages 36 and 63.)
- [7] Aubert-Broche, B., Collins, D. L., and Evans, A. C. (2006a). A new improved version of the realistic digital brain phantom. *NeuroImage*, 32(1):138–145. (Cited on pages 28, 166 and 169.)
- [8] Aubert-Broche, B., Griffin, M., Pike, G. B., Evans, A. C., and Collins, D. L. (2006b). 20 new digital brain phantoms for creation of validation image data bases. *IEEE Transactions on Medical Imaging*, 25(11):1410–1416. (Cited on pages 77, 84, 166 and 169.)
- [9] Avants, B., Cook, P., McMillan, C., Grossman, M., Tustison, N., Zheng, Y., and Gee, J. (2010). Sparse unbiased analysis of anatomical variance in longitudinal imaging. In *MICCAI*, volume 6361 of *Lecture Notes in Computer Science*, pages 324–331. Springer Berlin / Heidelberg. (Cited on page 47.)
- [10] Avants, B. B., Epstein, C. L., Grossman, M., and Gee, J. C. (2008). Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical image analysis*, 12(1):26–41. (Cited on pages 7, 9, 16, 97, 110, 111, 116, 117, 124, 127, 143, 145, 160, 170 and 171.)

- [11] Babalola, K. O., Patenaude, B., Aljabar, P., Schnabel, J., Kennedy, D., Crum, W., Smith, S., Cootes, F., Jenkinson, M., and Rueckert, D. (2008). Comparison and evaluation of segmentation techniques for subcortical structures in brain MRI. In *MICCAI, Part I, MICCAI '08*, pages 409–416, Berlin, Heidelberg. Springer-Verlag. (Cited on page 40.)
- [12] Bakshi, R., Dandamudi, V. S. R., Neema, M., De, C., and Bermel, R. A. (2005). Measurement of brain and spinal cord atrophy by magnetic resonance imaging as a tool to monitor multiple sclerosis. *Journal of Neuroimaging*, 15:30S–45S. (Cited on page 23.)
- [13] Bartsch, A., Bendszus, N., De Stefano, N., Homola, G., and Smith, S. (2004). Extending SIENA for a multi-subject statistical analysis of sample-specific cerebral edge shifts: Substantiation of early brain regeneration through abstinence from alcoholism. In *Tenth Int. Conf. on Functional Mapping of the Human Brain*. (Cited on page 45.)
- [14] Battaglini, M., Smith, S. M., Brogi, S., and De Stephano, N. (2008). Enhanced brain extraction improves the accuracy of brain atrophy estimation. *NeuroImage*, 40:583–589. (Cited on pages 94 and 159.)
- [15] Bernal, R. A. and Bakshi, R. (2006). The measurement and clinical relevance of brain atrophy in multiple sclerosis. *Neurology*, 5:158–170. (Cited on page 26.)
- [16] Boesen, K., Rehm, K., Schaper, K., Stoltzner, S., Woods, R., Luders, E., and Rottenberg, D. (2004). Quantitative comparison of four brain extraction algorithms. *NeuroImage*, 22(3):1255 – 1261. (Cited on page 56.)
- [17] Boos, H. B. M., Cahn, W., van Haren, N. E. M., Derks, E. M., Brouwer, R. M., Schnack, H. G., Hulshoff Pol, H. E., and Kahn, R. S. (2011). Focal and global brain measurements in siblings of patients with schizophrenia. *Schizophrenia Bulletin*. (Cited on pages 1 and 24.)
- [18] Boyes, R. G., Rueckert, D., Aljabar, P., Whitwell, J., Schott, J. M., Hill, D. L. G., and Fox, N. C. (2006). Cerebral atrophy measurements using Jacobian integration: Comparison with the boundary shift integral. *NeuroImage*, 32:159–169. (Cited on pages 2, 4, 40, 41, 49, 51, 57, 61, 65 and 92.)
- [19] Bricq, S., Collet, C., and Armspach, J. P. (2008). Unifying framework for multimodal brain MRI segmentation based on hidden Markov chains. *Medical Image Analysis*, 12(6):639 – 652. Special issue on information processing in medical imaging 2007. (Cited on pages 36 and 91.)
- [20] Camara, O., Schnabel, J. A., Ridgway, G. R., Crum, W. R., Douiri, A., Scahill, R. I., Hill, D. L. G., and Fox, N. C. (2008). Accuracy assessment of global and local atrophy measurement techniques with realistic simulated longitudinal Alzheimer’s disease images. *NeuroImage*, 42(2):696–709. (Cited on pages 2, 4, 6, 40, 41, 49, 51, 57, 59, 62, 65, 68, 75, 89, 90, 91, 92, 96, 99, 107 and 108.)

- [21] Camara, O., Schweiger, M., Scahill, R., Crum, W., Sneller, B., Schnabel, J., Ridgway, G., Cash, D., Hill, D., and Fox, N. (2006). Phenomenological model of diffuse global and regional atrophy using finite-element methods. *IEEE Transactions on Medical Imaging*, 25(11):1417–1430. (Cited on pages 2, 4, 40, 41, 49, 57, 59, 62, 68, 70, 92 and 93.)
- [22] Caramanos, Z., Fonov, V. S., Francis, S. J., Narayanan, S., Collins, D. L., and Arnold, D. L. (2008). 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)*, pages 13–24. (Cited on page 71.)
- [23] Caramanos, Z., Fonov, V. S., Francis, S. J., Narayanan, S., Pike, G. B., Collins, D. L., and Arnold, D. L. (2010). Gradient distortions in MRI: Characterizing and correcting for their effects on SIENA-generated measures of brain volume change. *NeuroImage*, 49(2):1601 – 1611. (Cited on page 37.)
- [24] Carter, N. J. and Keating, G. M. (2010). Glatiramer acetate: A review of its use in relapsing-remitting multiple sclerosis and in delaying the onset of clinically definite multiple sclerosis. *Drugs*, 70:1545–1577. (Cited on page 23.)
- [25] Chan, D., Fox, N. C., Jenkins, R., Scahill, R. I., Crum, W. R., and Rossor, M. N. (2001). Rates of global and regional cerebral atrophy in AD and frontotemporal dementia. *Neurology*, 57:1756–1763. (Cited on pages 1 and 24.)
- [26] Charil, A., Dagher, A., Lerch, J. P., Zijdenbos, A. P., Worsley, K. J., and Evans, A. C. (2007). Focal cortical atrophy in multiple sclerosis: Relation to lesion load and disability. *NeuroImage*, 34(2):509–517. (Cited on page 25.)
- [27] Chen, K., Reiman, E. M., Alexander, G. E., Bandy, D., Renaut, R., Crum, W. R., Fox, N. C., and Rossor, M. N. (2004a). An automated algorithm for the computation of brain volume change from sequential MRIs using an iterative principal component analysis and its evaluation for the assessment of whole-brain atrophy rates in patients with probable Alzheimer’s disease. *NeuroImage*, 22(1):134 – 143. (Cited on page 48.)
- [28] Chen, Z., Li, S. S., Yang, J., Letizia, D., and Shen, J. (2004b). Measurement and automatic correction of high-order B0 inhomogeneity in the rat brain at 11.7 tesla. *Magnetic Resonance Imaging*, 22(6):835 – 842. (Cited on page 33.)
- [29] Christensen, G. E., Rabbitt, R. D., and Miller, M. I. (1996). Deformable templates using large deformation kinematics. *IEEE Transactions on Image Processing*, 5:1435–1447. (Cited on page 99.)
- [30] Cocosco, C. A., Kollokian, V., Kwan, R. K.-S., and Evans, A. C. (1997). Brainweb: Online interface to a 3D MRI simulated brain database. *NeuroImage*, 5(4):S425. (Cited on page 63.)

- [31] Collins, D. L., Montagnat, J., Zijdenbos, A. P., Evans, A. C., and Arnold, D. L. (2001). Automated estimation of brain volume in multiple sclerosis with BICCR. In *Lecture Notes In Computer Science*, volume 2082, pages 141–147. (Cited on pages 2, 40 and 52.)
- [32] Coupé, P., Yger, P., Prima, S., Hellier, P., Kervrann, C., and Barillot, C. (2008). An optimized blockwise nonlocal means denoising filter for 3D magnetic resonance images. *Medical Imaging, IEEE Transactions on*, 27(4):425–441. (Cited on pages 32 and 33.)
- [33] Dale, A. M., Fischl, B., and Sereno, M. I. (1999). Cortical surface-based analysis. I. segmentation and surface reconstruction. *Neuroimage*, 9(2):179–194. (Cited on pages 47 and 55.)
- [34] de Bresser, J., Portegies, M. P., Leemans, A., Biessels, G. J., Kappelle, L. J., and Viergever, M. A. (2011). A comparison of MR based segmentation methods for measuring brain atrophy progression. *NeuroImage*, 54(2):760 – 768. (Cited on pages 4, 59, 63 and 65.)
- [35] De Stephano, N., Battaglini, M., and Smith, S. M. (2007). Measuring brain atrophy in multiple sclerosis. *Journal of Neuroimaging*, 17:10S–15S. (Cited on page 26.)
- [36] Dietemann, J. L., Beigelman, C., Rumbach, L., Vouge, M., Tajahmady, T., Faubert, C., Jeung, M. Y., and Wackenheim, A. (1988). Multiple sclerosis and corpus callosum atrophy: Relationship of MRI findings to clinical data. *Neuroradiology*, 30:478–480. 10.1007/BF00339686. (Cited on page 25.)
- [37] Dugas-Phocion, G., Ballester, M. A. G., Malandain, G., Lebrun, C., and Ayaiche, N. (2004). Improved EM-based tissue segmentation and partial volume effect quantification in multi-sequence brain MRI. In *MICCAI*, pages 26–33. (Cited on page 56.)
- [38] Erasmus, L. J., Hurter, D., Naude, M., Kritzinger, H. G., and Acho, S. (2004). A short overview of MRI artefacts. *South African Journal of Radiology*, 8(2):13–17. (Cited on page 33.)
- [39] Fennema-Notestine, C., Ozyurt, I. B., Clark, C. P., Morris, S., Bischoff-Grethe, A., Bondi, M. W., Jernigan, T. L., Fischl, B., Segonne, F., Shattuck, D. W., Leahy, R. M., Rex, D. E., Toga, A. W., Zou, K. H., and Brown, G. G. (2006). Quantitative evaluation of automated skull-stripping methods applied to contemporary and legacy images: Effects of diagnosis, bias correction, and slice location. *Human Brain Mapping*, 27(2):99–113. (Cited on page 56.)
- [40] Fillipi, M. and Agosta, F. (2010). Imaging biomarkers in multiple sclerosis. *Journal of Magnetic Resonance Imaging*, 31:770–778. (Cited on page 26.)

- [41] Fisher, E., Lee, J.-C., Nakamura, K., and Rudick, R. A. (2008). Gray matter atrophy in multiple sclerosis: A longitudinal study. *Annals of Neurology*, 64(3):255–265. (Cited on page 25.)
- [42] Flilippi, M., Bozzali, M., Rovaris, M., Gonen, O., Kesavadas, C., Ghezzi, A., Martinelli, V., Grossmann, R. I., Scotti, G., Comi, G., and Falini, A. (2003). Evidence for widespread axonal damage at the earliest clinical stage of multiple sclerosis. *Brain*, 126:433–437. (Cited on page 26.)
- [43] Fox, N. C., Ridgway, G. R., and Schott, J. M. (2011). Algorithms, atrophy and Alzheimer’s disease: Cautionary tales for clinical trials. *NeuroImage*, 57(1):15 – 18. (Cited on page 111.)
- [44] Freeborough, P. A. and Fox, N. C. (1997). The boundary shift integral: An accurate and robust measure of cerebral volume changes from registered repeat MRI. *IEEE Transactions on Medical Imaging*, 16:623–629. (Cited on pages 2, 3, 4, 40, 41, 45, 54, 57, 58, 60 and 169.)
- [45] Gaser, C., Nenadic, I., Buchsbaum, B. R., Hazlett, E. A., and Buchsbaum, M. S. (2001). Deformation-based morphometry and its relation to conventional volumetry of brain lateral ventricles in MRI. *NeuroImage*, 13(6):1140 – 1145. (Cited on pages 4, 49, 50 and 57.)
- [46] Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (1995). *Markov Chain Monte Carlo in Practice: Interdisciplinary Statistics*. Chapman and Hall/CRC, 1 edition. (Cited on pages 127 and 175.)
- [47] Gilks, W. R., Roberts, G. O., and Sahu, S. K. (1996). Strategies for improving MCMC. *MCMC in practice*, pages 89–114. (Cited on page 137.)
- [48] Gilks, W. R., Roberts, G. O., and Sahu, S. K. (1998). Adaptive Markov chain Monte Carlo using regeneration. *Journal of the American Statistical Association*, 93:1045–1054. (Cited on pages 105, 137 and 157.)
- [49] Gudbjartsson, H. and Patz, S. (1995). The Rician distribution of noisy MRI data. *Magnetic Resonance in Medicine*, 34(6):910–914. (Cited on pages 29, 70, 113 and 131.)
- [50] Gunter, J. L., Shiung, M. M., Manduca, A., and Jack, C. R. (2003). Methodological considerations for measuring rates of brain atrophy. *Journal of Magnetic Resonance Imaging*, 18:16–24. (Cited on pages 4, 59, 60 and 65.)
- [51] Guttmann, C. R., Benson, R. and Warfield, S. K., Wei, X., Anderson, M., Hall, C., Abu-Hasaballah, K., Mugler, J., and Wolfson, L. (2000). White matter abnormalities in mobility-impaired older persons. *Neurology*, 54(6):1277–1283. (Cited on page 40.)
- [52] Haacke, E. (1999). *Magnetic resonance imaging: physical principles and sequence design*. J. Wiley & Sons. (Cited on page 27.)

- [53] Hadjidemetriou, S., Studholme, C., Mueller, S., Weiner, M., and Schuff, N. (2009). Restoration of MRI data for intensity non-uniformities using local high order intensity statistics. pages 36–48. (Cited on page 35.)
- [54] Hirata, Y., Matsuda, H., Nemoto, K., Ohnishi, T., Hirao, K., Yamashita, F., Asada, T., Iwabuchi, S., and Samejima, H. (2005). Voxel-based morphometry to discriminate early Alzheimer’s disease from controls. *Neuroscience Letters*, 382(3):269 – 274. (Cited on page 50.)
- [55] Holden, M. (2001). *Registration of 3D serial MR Brain Images*. PhD thesis, University of London. (Cited on page 71.)
- [56] Holden, M. (2008). A review of geometric transformations for nonrigid body registration. *IEEE Trans. Med. Imaging*, 27(1):111–128. (Cited on pages 39 and 40.)
- [57] Horsfield, M. A., Rovaris, M., Rocca, M. A., Rossi, P., Benedict, R. H., Filippi, M., and Bakshi, R. (2003). Whole-brain atrophy in multiple sclerosis measured by two segmentation processes from various MRI sequences. *Journal of Neurological Sciences*, 216(1):169–177. (Cited on pages 2, 40 and 51.)
- [58] Hua, X., Gutman, B., Boyle, C. P., Rajagopalan, P., Leow, A. D., Yanovsky, I., Kumar, A. R., Toga, A. W., Jack Jr., C. R., Schuff, N., Alexander, G. E., Chen, K., Reiman, E. M., Weiner, M. W., and Thompson, P. M. (2011). Accurate measurement of brain changes in longitudinal MRI scans using tensor-based morphometry. *NeuroImage*, 57(1):5 – 14. (Cited on pages 7, 110, 111 and 124.)
- [59] Hua, X., Lee, S., Yanovsky, I., Leow, A. D., Chou, Y.-Y., Ho, A. J., Gutman, B., Toga, A. W., Jack Jr, C. R., and Bernstein, M. A. (2009). Optimizing power to track brain degeneration in Alzheimer’s disease and mild cognitive impairment with tensor-based morphometry: An ADNI study of 515 subjects. *NeuroImage*, 48(4):668–681. (Cited on pages 2, 40, 41, 49 and 57.)
- [60] Hub, M., Kessler, M. L., and Karger, C. P. (2009). A stochastic approach to estimate the uncertainty involved in B-spline image registration. *IEEE Trans Med Imaging*, 28(11):1708–16. (Cited on pages 101, 107 and 108.)
- [61] Ishak, N., Gangeh, M., and Logeswaran, R. (2008). Comparison of denoising techniques applied on low-field MR brain images. In *Computer Graphics, Imaging and Visualisation, 2008. CGIV ’08. Fifth International Conference on*, pages 345–349. (Cited on page 33.)
- [62] Ishii, S., Shishido, F., Miyajima, M., Sakuma, K., Shigihara, T., Tameta, T., Miyazaki, M., and Kuroda, H. (2009). Comparison of Alzheimer’s disease with vascular dementia and non-dementia using specific voxel-based z score maps. *Annals of Nuclear Medicine*, 23:25–31. 10.1007/s12149-008-0210-8. (Cited on page 50.)

- [63] Jalobeanu, A. and Fitzenz, D. (2008). Inferring deformation fields from multirate satellite images. In *IEEE International Geoscience and Remote Sensing Symposium*, Boston MA, USA. (Cited on pages 102 and 108.)
- [64] Janke, A., Zhao, H., Cowin, G. J., Galloway, G. J., and Doddrell, D. M. (2004). Use of spherical harmonic deconvolution methods to compensate for nonlinear gradient effects on MRI images. *Magnetic Resonance in Medicine*, 52(1):115–122. (Cited on page 37.)
- [65] Jasperse, B., Vrenken, H., Sanz-Arigita, E., de Groot, V., Smith, S. M., Polman, C. H., and Barkhof, F. (2007). Regional brain atrophy development is related to specific aspects of clinical dysfunction in multiple sclerosis. *Neuroimage*, 38(3):529–37. (Cited on page 26.)
- [66] Jenkinson, M. and M., S. S. (2001). A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, 36:143–156. (Cited on page 172.)
- [67] Jovicich, J., Czanner, S., Greve, D., Haley, E., van der Kouwe, A., Gollub, R., Kennedy, D., Schmitt, F., Brown, G., MacFall, J., Fischl, B., and Dale, A. (2006). Reliability in multi-site structural mri studies: Effects of gradient non-linearity correction on phantom and human data. *NeuroImage*, 30(2):436 – 443. (Cited on page 37.)
- [68] Karacali, B. and Davatzikos, C. (2006). Simulation of tissue atrophy using a topology preserving transformation model. *IEEE Transactions on Medical Imaging*, 25(5):649–652. (Cited on pages 59, 62, 64, 67, 68, 69, 70 and 74.)
- [69] Kennedy, M. C. and O’Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):425–464. (Cited on page 162.)
- [70] Klein, A., Andersson, J., Ardekani, B. A., Ashburner, J., and et al. (2009). Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *NeuroImage*, 46(3):786–802. (Cited on page 170.)
- [71] Kwan, R. K.-S., Evans, A. C., and Pike, G. B. (1999). MRI simulation-based evaluation of image-processing and classification methods. *IEEE Transactions on Medical Imaging*, 18(11):1085–1097. (Cited on page 77.)
- [72] Kybic, J. (2008). Fast no ground truth image registration accuracy evaluation: Comparison of bootstrap and hessian approaches. In *ISBI*, pages 792–795. (Cited on page 100.)
- [73] Kybic, J. (2010). Bootstrap resampling for image registration uncertainty estimation without ground truth. *IEEE Transactions on Image Processing*, 19(1):64–73. (Cited on pages 100, 101 and 107.)

- [74] Lanz, M., H.-H. H. H. (2007). Brain atrophy and cognitive impairment in multiple sclerosis: a review. *Journal of Neurology*, 254:II43–II48(6). (Cited on page 24.)
- [75] Lee, H.-P., Lin, M. C., and Foskey, M. (2008). Physically-based validation of deformable medical image registration. In *MICCAI, Part II*, MICCAI '08, pages 830–838, Berlin, Heidelberg. Springer-Verlag. (Cited on pages 99 and 107.)
- [76] Lee, J.-M., Yoon, U., Nam, S. H., Kim, J.-H., Kim, I.-Y., and Kim, S. I. (2003). Evaluation of automated and semi-automated skull-stripping algorithms using similarity index and segmentation error. *Computers in Biology and Medicine*, 33(6):495 – 507. (Cited on page 56.)
- [77] Lee, S., Wolberg, G., and Shin, S. Y. (1997). Scattered data interpolation with multilevel B-splines. *IEEE Transactions on Visualization and Computer Graphics*, 3:228–244. (Cited on page 128.)
- [78] Leow, A. D., Yanovsky, I., Chiang, M.-C., Lee, A. D., Klunder, A. D., Lu, A., Becker, J. T., Davis, S. W., Toga, A. W., and Thompson, P. M. (2007). Statistical properties of Jacobian maps and the realization of unbiased large-deformation nonlinear image registration. *IEEE Transactions on Medical Imaging*, 26(6):822–832. (Cited on pages 67, 108, 109 and 110.)
- [79] Leung, K. K., Clarkson, M. J., Bartlett, J. W., Clegg, S., Jack Jr., C. R., Weiner, M. W., Fox, N. C., and Ourselin, S. (2010). Robust atrophy rate measurement in Alzheimer’s disease using multi-site serial MRI: Tissue-specific intensity normalization and parameter selection. *NeuroImage*, 50(2):516 – 523. (Cited on page 46.)
- [80] Lewis, E. B. and Fox, N. C. (2004). Correction of differential intensity inhomogeneity in longitudinal MR images. *NeuroImage*, 23(1):75 – 83. (Cited on pages 16, 35, 46, 127, 138, 154, 161, 168, 170 and 172.)
- [81] Li, C., Huang, R., Ding, Z., Gatenby, C., Metaxas, D., and Gore, J. (2008). A variational level set approach to segmentation and bias correction of images with intensity inhomogeneity. In *MICCAI, Part II*, MICCAI '08, pages 1083–1091, Berlin, Heidelberg. Springer-Verlag. (Cited on page 36.)
- [82] Liang, Z.-P. and Lauterbur, P. C. (1999). *Principles of Magnetic Resonance Imaging: A Signal Processing Perspective*. Wiley-IEEE Press. (Cited on page 33.)
- [83] Lin, M., Chan, S., Chen, J.-H., Chang, D., Nie, K., Chen, S.-T., Lin, C.-J., Shih, T.-C., Nalcioglu, O., and Su, M.-Y. (2011). A new bias field correction method combining N3 and FCM for improved segmentation of breast density on MRI. *Med Phys*, 38(1):5–14. (Cited on page 35.)
- [84] Lin, X., Tench, C. R., Evangelou, N., Jaspan, T., and Constantinescu, C. S. (2004). Measurement of spinal cord atrophy in multiple sclerosis. *Journal of*

- neuroimaging official journal of the American Society of Neuroimaging*, 14(3 Suppl):20S–26S. (Cited on page 23.)
- [85] Lublin, F. D. and Reingold, S. C. (1996). Defining the clinical course of multiple sclerosis: Results of an international survey. national multiple sclerosis society (USA) advisory committee on clinical trials of new agents in multiple sclerosis. *Neurology*, 46(4):907–911. (Cited on page 21.)
- [86] Maintz, J. and Viergever, M. (1998). A survey of medical image registration. *Medical Image Analysis*, 2(1):1–36. (Cited on page 40.)
- [87] Mangin, J.-F., Coulon, O., and Frouin, V. (1998). Robust brain segmentation using histogram scale-space analysis and mathematical morphology. In Wells, W. M., Colchester, A., and Delp, S., editors, *Proc. 1st MICCAI*, LNCS-1496, pages 1230–1241, MIT, Boston. Springer Verlag. (Cited on pages 55 and 56.)
- [88] Manjón, J. V., Lull, J. J., Carbonell-Caballero, J., García-Martí, G., Martí-Bonmatí, L., and Robles, M. (2007). A nonparametric MRI inhomogeneity correction method. *Medical Image Analysis*, 11(4):336 – 345. (Cited on page 35.)
- [89] Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979). *Multivariate analysis*. Academic Press. (Cited on pages 4 and 50.)
- [90] McDonald, W. I., Compston, A., Edan, G., Goodkin, D., Hartung, H.-P., Lublin, F. D., McFarland, H. F., Paty, D. W., Polman, C. H., Reingold, S. C., S.-W., M., Sibley, W., Thompson, A., Van Den N., S., Weinshenker, B. Y., and Wolinsky, J. S. (2001). Recommended diagnostic criteria for multiple sclerosis: Guidelines from the international panel on the diagnosis of multiple sclerosis. *Annals of Neurology*, 50(1):121–127. (Cited on page 22.)
- [91] Meyer, C., Bland, P., and Pipe, J. (1995). Retrospective correction of intensity inhomogeneities in MRI. *Medical Imaging, IEEE Transactions on*, 14(1):36 –41. (Cited on page 34.)
- [92] Miller, D. H., Barkhof, F., Frank, J. A., Parker, G. J. M., and Thompson, A. J. (2002). Measurement of atrophy in multiple sclerosis: pathological basis, methodological aspects and clinical relevance. *Brain*, 125:1676–1695. (Cited on pages 17 and 26.)
- [93] Minagar, A., Toledo, E. G., Alexander, J. S., and Kelley, R. E. (2004). Pathogenesis of brain and spinal cord atrophy in multiple sclerosis. *Journal of Neuroimaging*, 14:5S–10S. (Cited on pages 23 and 25.)
- [94] Modat, M., Ridgway, G. R., Hawkes, D. J., Fox, N. C., and Ourselin, S. (2010). Nonrigid registration with differential bias correction using normalised mutual information. In *Biomedical Imaging: From Nano to Macro, 2010 IEEE International Symposium on*, pages 356 –359. (Cited on page 36.)

- [95] Musse, O., Heitz, F., and Armspach, J.-P. (2003). Fast deformable matching of 3D images over multiscale nested subspaces. Application to atlas-based MRI segmentation. *Pattern Recognition*, 36(8):1881–1899. (Cited on page 67.)
- [96] Nakamura, K., Fox, R., and Fisher, E. (2011). CLADA: Cortical longitudinal atrophy detection algorithm. *NeuroImage*, 54(1):278 – 289. (Cited on page 46.)
- [97] Narayana, P. A., Brey, W. W., Kulkarni, M. V., and Sievenpiper, C. L. (1988). Compensation for surface coil sensitivity variation in magnetic resonance imaging. *Magnetic Resonance Imaging*, 6(3):271 – 274. (Cited on page 33.)
- [98] Noblet, V., Heinrich, C., Heitz, F., and Armspach, J.-P. (2005). 3D deformable image registration: A topology preservation scheme based on hierarchical deformation models and interval analysis optimization. *IEEE Transactions on Image Processing*, 14(5):553–566. (Cited on pages 5, 7, 9, 16, 40, 58, 66, 69, 70, 77, 85, 92, 93, 96, 97, 111, 116, 117, 124, 127, 133, 143, 145, 159, 160, 170, 171 and 173.)
- [99] Noblet, V., Heinrich, C., Heitz, F., and Armspach, J.-P. (2006). Retrospective evaluation of a topology preserving non-rigid registration method. *Medical Image Analysis*, 10(3):366–384. (Cited on page 69.)
- [100] Noblet, V., Heinrich, C., Heitz, F., and Armspach, J.-P. (2008). Accurate inversion of 3D transformation fields. *IEEE Transactions on Image Processing*, 17(10):1963–1968. (Cited on page 70.)
- [101] Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man and Cybernetics*, 9(1):62–66. (Cited on page 134.)
- [102] Owens, T. (2003). The enigma of multiple sclerosis: Inflammation and neurodegeneration cause heterogeneous dysfunction and damage. *Current opinion in neurology*, 16(3):259–265. (Cited on page 26.)
- [103] Pagani, E., Horsfield, M. A., Rocca, M. A., and Filippi, M. (2007). Assessing atrophy of the major white matter fiber bundles of the brain from diffusion tensor MRI data. *Magnetic Resonance in Medicine*, 58(3):527–534. (Cited on pages 49 and 50.)
- [104] Pagani, E., Rocca, M. A., Gallo, A., Rovaris, M., Martinelli, V., Comi, G., and Filippi, M. (2005). Regional brain atrophy evolves differently in patients with multiple sclerosis according to clinical phenotype. *AJNR Am J Neuroradiol*, 26(2):341–346. (Cited on pages 25 and 26.)
- [105] Paling, S. M., Williams, E. D., Barber, R., Burton, E. J., Crum, W. C., Fox, N., and O’Brien, J. T. (2003). The application of serial MRI analysis techniques to the study of cerebral atrophy in late-onset dementia. *Medical Image Analysis*, 8:69–79. (Cited on pages 4, 60 and 65.)

- [106] Pelletier, D., Garrison, K., and Henry, R. (2004). Measurement of whole-brain atrophy in multiple sclerosis. *Journal of NeuroImaging*, 14(3):11S–19S. (Cited on pages 2 and 39.)
- [107] Pershadsingh, H., Heneka, M., Saini, R., Amin, N., Broeske, D., and Feinstein, D. (2004). Effect of pioglitazone treatment in a patient with secondary multiple sclerosis. *Journal of Neuroinflammation*, 1(1):3. (Cited on page 21.)
- [108] Pham, D. L., Xu, C., and Prince, J. L. (2000). A survey of current methods in medical image segmentation. In *Annual Review of Biomedical Engineering*, volume 2, pages 315–338. (Cited on page 40.)
- [109] Pieperhoff, P., Sudmeyer, M., Homke, L., Zilles, K., Schnitzler, A., and Amunts, K. (2008). 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):269–287. (Cited on pages 4, 6, 11, 49, 51, 57, 59, 62, 63, 64, 65, 68, 96, 108 and 161.)
- [110] Plosker, G. L. (2011). Interferon-1b: A review of its use in multiple sclerosis. *CNS Drugs*, 25:67–88(22). (Cited on page 23.)
- [111] Polman, C. H., Reingold, S. C., Banwell, B., Clanet, M., Cohen, J. A., Filippi, M., Fujihara, K., Havrdova, E., Hutchinson, M., Kappos, L., Lublin, F. D., Montalban, X., O’Connor, P., Sandberg-Wollheim, M., Thompson, A. J., Waubant, E., Weinshenker, B., and Wolinsky, J. S. (2011). Diagnostic criteria for multiple sclerosis: 2010 revisions to the mcdonald criteria. *Annals of Neurology*, 69(2):292–302. (Cited on page 22.)
- [112] Press, W., Teukolsky, S., Vetterling, W., and Flannery, B. (1992). *Numerical recipes in C. The art of scientific computing*. Cambridge University Press, second edition. (Cited on page 70.)
- [113] Raine, C. S., McFarland, H., and Hohlfeld, R. (2008). *Multiple Sclerosis: A Comprehensive Text*. Saunders Ltd. (Cited on pages 20 and 23.)
- [114] Rehm, K., Schaper, K., Anderson, J., Woods, R., Stoltzner, S., and Rottenberg, D. (2004). Putting our heads together: A consensus approach to brain/non-brain segmentation in T1-weighted MR volumes. *NeuroImage*, 22(3):1262 – 1270. (Cited on page 55.)
- [115] Reimer, P., Parizel, P. M., Meaney, J. F. M., and Stichnoth, F. A. (2010). *Clinical MR Imaging: A Practical Approach: Third Edition*. Springer. (Cited on page 20.)
- [116] Rex, D. E., Shattuck, D. W., Woods, R. P., Narr, K. L., Luders, E., Rehm, K., Stolzner, S. E., Rottenberg, D. A., and Toga, A. W. (2004). A meta-algorithm for brain extraction in MRI. *Neuroimage*, 23(2):625–637. (Cited on page 55.)

- [117] Richard, F. J. and Samson, A. M. (2007). Metropolis-hasting techniques for finite-element-based registration. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 0:1–6. (Cited on pages 105 and 107.)
- [118] Risholm, P., Pieper, S., Samset, E., and Wells III, W. M. (2010a). Summarizing and visualizing registration uncertainty in non-rigid registration. 6362:554–561. (Cited on pages 105, 106, 107 and 108.)
- [119] Risholm, P., Samset, E., and Wells III, W. (2010b). Bayesian estimation of deformation and elastic parameters in non-rigid registration. *Biomedical Image Registration*, 6204:104–115. (Cited on pages 104, 105 and 156.)
- [120] Robinson, D. and Milanfar, P. (2004). Fundamental performance limits in image registration. *Image Processing, IEEE Transactions on*, 13(9):1185–1199. (Cited on page 99.)
- [121] Rohlfing, T. (2006). Transformation model and constraints cause bias in statistics on deformation fields. In *MICCAI, Part I*, volume 4190 of *Lecture Notes in Computer Science*, pages 207–214, Berlin/Heidelberg. Springer-Verlag. (Cited on pages 96, 108 and 156.)
- [122] Rorden, C., Bonilha, L., and Nicholas, T. E. E. (2007). Rank-order versus mean based statistics for neuroimaging. *Neuroimage*, 35(4):1531–1537. (Cited on pages 4 and 50.)
- [123] Rosati, G. (2001). The prevalence of multiple sclerosis in the world: An update. *Neurological Science*, 2:117–139. (Cited on page 17.)
- [124] Rueckert, D., Sonoda, L. I., Hayes, C., Hill, D. L. G., Leach, M. O., and Hawkes, D. J. (1999). Nonrigid registration using free-form deformations: Application to breast mr images. *IEEE Transactions on Medical Imaging*, 18(8):712–721. (Cited on pages 99, 110, 128, 131 and 156.)
- [125] Sailer, M., Fischl, B., Salat, D., Tempelmann, C., Schönfeld, M. A., Busa, E., Bodammer, N., Heinze, H. J., and Dale, A. (2003). Focal thinning of the cerebral cortex in multiple sclerosis. *Brain*, 126(8):1734–1744. (Cited on page 25.)
- [126] Sastre-Garriga, J., Ingle, G. T., Chard, D. T., Ramió-Torrentà, L., Miller, D. H., and Thompson, A. J. (2004). Grey and white matter atrophy in early clinical stages of primary progressive multiple sclerosis. *NeuroImage*, 22(1):353–359. (Cited on page 25.)
- [127] Schapiro, R. T. (2010). *Managing the Symptoms of Multiple Sclerosis: Fifth Edition*. ReadHowYouWant. (Cited on page 21.)
- [128] Schnabel, J. A., Tanner, C., Castellano-Smith, C. A., Degenhard, A., Leach, M. O., Hose, D. R., Hill, D. L. G., and Hawkes, D. J. (2003). Validation of non-rigid image registration using finite element methods: Application to breast MR images. *IEEE Trans. Med. Imaging*, 22(2):238–247. (Cited on pages 99 and 107.)

- [129] Schweiger, M., Camara-Rey, O., Crum, W., Lewis, E., Schnabel, J., Arridge, S., Hill, D., and Fox, N. (2005). An inverse problem approach to the estimation of volume change. In *MICCAI*, volume 3750 of *Lecture Notes in Computer Science*, pages 616–623. Springer Berlin / Heidelberg. (Cited on page 48.)
- [130] Segonne, F., Dale, A. M., Busa, E., Glessner, M., Salat, D., Hahn, H. K., and Fischl, B. (2004). A hybrid approach to the skull stripping problem in MRI. *NeuroImage*, 22(3):1060 – 1075. (Cited on pages 55 and 56.)
- [131] Sharma, J., Sanfilipo, M. P., Benedict, R. H., Weinstock-Guttman, B., Munschauer, F. E. r., and Bakshi, R. (2004a). Whole-brain atrophy in multiple sclerosis measured by automated versus semiautomated MR imaging segmentation. *American Journal of Neuroradiology*, 25(6):985–996. (Cited on page 2.)
- [132] Sharma, J., Sanfilipo, M. P., Benedict, R. H. B., Weinstock-Guttman, B., Munschauer, F. E., I., and Bakshi, R. (2004b). Whole-brain atrophy in multiple sclerosis measured by automated versus semiautomated MR imaging segmentation. *AJNR Am J Neuroradiol*, 25(6):985–996. (Cited on pages 2, 40, 41 and 51.)
- [133] Sharma, S., Noblet, V., Rousseau, F., Heitz, F., Rumbach, L., and Armspach, J.-P. (2010). Evaluation of brain atrophy estimation algorithms using simulated ground-truth data. *Medical Image Analysis*, 14(3):373–389. (Cited on page 108.)
- [134] Shattuck, D. W., Sandor-Leahy, S. R., Schaper, K. A., Rottenberg, D. A., and Leahy, R. M. (2001). Magnetic resonance image tissue classification using a partial volume model. *NeuroImage*, 13(5):856 – 876. (Cited on pages 53 and 55.)
- [135] Sicotte, N. L., Kern, K. C., Giesser, B. S., Arshanapalli, A., Schultz, A., Montag, M., Wang, H., and Bookheimer, S. Y. (2008). Regional hippocampal atrophy in multiple sclerosis. *Brain*, 131(Pt 4):1134–41. (Cited on page 25.)
- [136] Simon, J. H. (2006). Brain atrophy in multiple sclerosis: What we know and would like to know. *Multiple sclerosis (Houndmills, Basingstoke, England)*, 12(6):679–687. (Cited on pages 1, 24 and 26.)
- [137] Sled, J., Zijdenbos, A., and Evans, A. (1998). A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *Medical Imaging, IEEE Transactions on*, 17(1):87 –97. (Cited on pages 34 and 91.)
- [138] Sluimer, J. D., Vrenken, H., Blankenstein, M. A., Fox, N. C., Scheltens, P., Barkhof, F., and van der Flier, W. M. (2008). Whole-brain atrophy rate in Alzheimer’s disease: Identifying fast progressors. *Neurology*, 70(19):1836–1841. (Cited on pages 1, 24 and 25.)
- [139] Smith, S. M. (2002). Fast robust automated brain extraction. *Human Brain Mapping*, 17(3):143–155. (Cited on pages 3, 42, 53, 55, 56, 171 and 172.)

- [140] Smith, S. M., De Stefano, N., Jenkinson, M., and Matthews, P. M. (2001). Normalized accurate measurement of longitudinal brain change. *Journal of Computer Assisted Tomography*, 25(3):466–475. (Cited on pages 2, 4, 40, 41 and 42.)
- [141] Smith, S. M., Jenkinson, M., Woolrich, M. W. and Beckman, C. F., Behrens, T. E. J., Johansen-Berg, H., Bannister, P. R., and et. al (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, 23(S1):208–219. (Cited on pages 3, 4, 43, 44, 49, 57, 58, 63 and 169.)
- [142] Smith, S. M., Rao, A., De Stefano, N., Jenkinson, M., Schott, J. M., Matthews, P. M., and Fox, N. C. (2007). Longitudinal and cross-sectional analysis of atrophy in Alzheimer’s disease: Cross validation of BSI, SIENA and SIENAX. *NeuroImage*, 36:1200–1206. (Cited on pages 4, 6, 49, 59, 61, 65, 92, 96, 108 and 160.)
- [143] Smith, S. M., Zhang, Y., Jenkinson, M., Chen, J., Matthews, P. M., Federico, A., and De Stefano, N. (2002). Accurate, robust and automated longitudinal and cross-sectional brain change analysis. *NeuroImage*, 17(1):479–489. (Cited on pages 2, 3, 4, 40, 41, 42, 49 and 58.)
- [144] Souplet, J.-C. (2009). *Evaluation de l’atrophie et de la charge lésionnelle sur des séquences IRM de patients atteints de sclérose en plaques*. PhD thesis, University of Nice-Sophia Antipolis. (Cited on page 56.)
- [145] Souplet, J.-C., Lebrun, C., Ayache, N., and Malandain, G. (2008). A new evaluation of the brain parenchymal fraction: Application in multiple sclerosis longitudinal studies. In *Proceedings of the IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI’08)*, pages 65–68, Paris, France. IEEE. (Cited on page 51.)
- [146] Studholme, C., Cardenas, V., Song, E., Ezekiel, F., Maudsley, A., and Weiner, M. (2004). Accurate template-based correction of brain MRI intensity distortion with application to dementia and aging. *Medical Imaging, IEEE Transactions on*, 23(1):99–110. (Cited on page 35.)
- [147] Studholme, C., Cardenas, V., and Weiner, M. (2001). Multiscale image and multiscale deformation of brain anatomy for building average brain atlases. *SPIE Medical Imaging*, 2(7):557–568. (Cited on pages 131 and 156.)
- [148] Styner, M., Brechbuhler, C., Szekely, G., and Gerig, G. (2000). Parametric estimate of intensity inhomogeneities applied to MRI. *IEEE Transactions on Medical Imaging*, 19:153–165. (Cited on pages 34, 46, 131, 138, 155 and 168.)
- [149] Taylor, J. R. (1996). *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*. University Science Books. (Cited on page 97.)
- [150] Thirion, J. P. (1998). Image matching as a diffusion process: an analogy with maxwell’s demons. *Medical Image Analysis*, 2(3):243–260. (Cited on page 99.)

- [151] Thompson, P. M., Hayashi, K. M., De Zubicaray, G., Janke, A. L., Rose, S. E., Semple, J., Herman, D., Hong, M. S., Dittmer, S. S., Doddrell, D. M., and Toga, A. W. (2003). Dynamics of gray matter loss in Alzheimer’s disease. *The Journal of Neuroscience*, 23:994–1005. (Cited on page 75.)
- [152] Tierney, L. (1994). Markov chains for exploring posterior distributions. *Annals of Statistics*, 22:1701–1762. (Cited on page 136.)
- [153] Toews, M. and Wells, Iii, W. M. (2009). Bayesian registration via local image regions: Information, selection and marginalization. In *Proceedings of the 21st International Conference on Information Processing in Medical Imaging, IPMI ’09*, pages 435–446, Berlin, Heidelberg. Springer-Verlag. (Cited on pages 103 and 108.)
- [154] Tomanek, B., Ryner, L., Hoult, D. I., Kozlowski, P., and Saunders, J. K. (1997). Dual surface coil with high-B1 homogeneity for deep organ MR imaging. *Magnetic Resonance Imaging*, 15(10):1199 – 1204. (Cited on page 33.)
- [155] Tustison, N. J., Avants, B. B., Cook, P. A., Yuanjie, Z., Egan, A., Yushkevich, P. A., and Gee, J. C. (2010). N4ITK: Improved N3 bias correction. *Medical Imaging, IEEE Transactions on*, 29(6):1310 –1320. (Cited on pages 16, 34, 127, 138, 154, 155 and 161.)
- [156] van Heteren, J. G., Henkelman, R. M., and Bronskill, M. J. (1987). Equivalent circuit for coil-patient interactions in magnetic resonance imaging. *Magnetic Resonance Imaging*, 5(2):93 – 99. (Cited on page 33.)
- [157] Vemuri, B. C., Ye, J., Chen, Y., and Leonard, C. M. (2003). Image registration via level-set motion: Applications to atlas-based segmentation. *Medical Image Analysis*, 7(1):1 – 20. (Cited on pages 7, 9, 16, 97, 99, 111, 116, 117, 124, 127, 143, 145, 160, 170 and 171.)
- [158] Vovk, U., Pernus, F., and Likar, B. (2007). A review of methods for correction of intensity inhomogeneity in MRI. *Medical Imaging, IEEE Transactions on*, 26(3):405 –421. (Cited on pages 34 and 36.)
- [159] Wahba, G. (1990). *Spline models for observational data*, volume 59 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA. (Cited on page 130.)
- [160] Wang, D., Strugnell, W., Cowin, G., Doddrell, D. M., and Slaughter, R. (2004a). Geometric distortion in clinical MRI systems: Part I: Evaluation using a 3D phantom. *Magnetic Resonance Imaging*, 22(9):1211 – 1221. (Cited on page 37.)
- [161] Wang, D., Strugnell, W., Cowin, G., Doddrell, D. M., and Slaughter, R. (2004b). Geometric distortion in clinical MRI systems: Part II: Correction using a 3D phantom. *Magnetic Resonance Imaging*, 22(9):1223 – 1232. (Cited on page 37.)

- [162] Wang, Y., Song, Y., Rajagopalan, P., An, T., Liu, K., Chou, Y.-Y., Guttman, B., Toga, A. W., and Thompson, P. M. (2011). Surface-based TBM boosts power to detect disease effects on the brain: An N-804 ADNI study. *NeuroImage*, 56(4):1993 – 2010. (Cited on pages 49 and 50.)
- [163] Ward, B. D. (1999). Intracranial segmentation. Technical report, Medical College of Wisconsin. (Cited on pages 55 and 56.)
- [164] Warfield, S., Zou, K., and Wells, W. (2004). Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *Medical Imaging, IEEE Transactions on*, 23(7):903 –921. (Cited on page 56.)
- [165] Whitwell, J. L. . (2008). Longitudinal imaging: change and causality. *Current Opinion in Neurology*, 21(4):1350–7540. (Cited on page 26.)
- [166] Wicks, D. A., Barker, G. J., and Tofts, P. S. (1993). Correction of intensity nonuniformity in MR images of any orientation. *Magnetic Resonance Imaging*, 11(2):183 – 96. (Cited on page 33.)
- [167] Wiest-Daesslé, N., Prima, S., Coupé, P., Morrissey, S. P., and Barillot, C. (2008). Rician noise removal by non-local means filtering for low signal-to-noise ratio MRI: Applications to DT-MRI. In *MICCAI, Part II*, pages 171–179, Berlin, Heidelberg. Springer-Verlag. (Cited on pages 32 and 33.)
- [168] Woods, R. P., Grafton, S. T., Holmes, C. J., Cherry, S. R., and Mazziotta, J. C. (1998). Automated image registration: I. General methods and intrasubject, intramodality validation. *Journal of Computer Assisted Tomography*, 22:139–152. (Cited on page 71.)
- [169] Xue, Z., Shen, D., and Davatzikos, C. (2006). Statistical representation of high-dimensional deformation fields with application to statistically constrained 3D warping. *Medical Image Analysis*, 10(5):740 – 751. (Cited on page 161.)
- [170] Yanovsky, I., Leow, A. D., Lee, S., Osher, S. J., and Thompson, P. M. (2009). Comparing registration methods for mapping brain change using tensor-based morphometry. *Medical Image Analysis*, 13(5):679 – 700. (Cited on pages 2, 40 and 109.)
- [171] Yanovsky, I., Thompson, P. M., Osher, S., Hua, X., Shattuck, D. W., Toga, A. W., and Leow, A. D. (2008). Validating unbiased registration on longitudinal MRI scans from the Alzheimer’s disease neuroimaging initiative (ADNI). In *ISBI*, pages 1091–1094. (Cited on pages 2, 40, 41, 49 and 57.)
- [172] Yetik, I. and Nehorai, A. (2006). Performance bounds on image registration. *Signal Processing, IEEE Transactions on*, 54(5):1737 – 1749. (Cited on page 99.)

- [173] Yin, Y., Hoffman, E. A., Ding, K., Reinhardt, J. M., and Lin, C.-L. (2011). A cubic B-spline based hybrid registration of lung CT images for a dynamic airway geometric model with large deformation. *Physics in Medicine and Biology*, 56(1):203–218. (Cited on page 128.)
- [174] Yushkevich, P. A., Avants, B. B., Das, S. R., Pluta, J., Altinay, M., and Craige, C. (2010). Bias in estimation of hippocampal atrophy using deformation-based morphometry arises from asymmetric global normalization: An illustration in ADNI 3 T MRI data. *NeuroImage*, 50(2):434 – 445. (Cited on pages 7, 109, 110, 111 and 124.)
- [175] Zhang, Y., Brady, M., and Smith, S. M. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation maximization algorithm. *IEEE Transactions on Medical Imaging*, 20(1):45–57. (Cited on pages 3, 4, 42, 49, 81 and 172.)
- [176] Zitova, B. (2003). Image registration methods: A survey. *Image and Vision Computing*, 21(11):977–1000. (Cited on page 40.)
- [177] Zivadinov, R. (2005). Steroids and brain atrophy in multiple sclerosis. *Journal of the Neurological Sciences*, 233(1-2):73–81. (Cited on page 25.)
- [178] Zivadinov, R. and Bakshi, R. (2004). Role of MRI in multiple sclerosis II: Brain and spinal cord atrophy. *Frontiers in Bioscience*, (9):647–664. (Cited on page 24.)
- [179] Zivadinov, R., Locatelli, L., Cookfair, D., Srinivasaraghavan, B., Bertolotto, A., Ukmar, M., Bratina, A., Maggiore, C., Bosco, A., Grop, A., Catalan, M., and Zorzon, M. (2007). Interferon beta-1a slows progression of brain atrophy in relapsing-remitting multiple sclerosis predominantly by reducing gray matter atrophy. *Multiple Sclerosis*, 13(4):490–501+. (Cited on page 26.)
- [180] Zivadinov, R., Sepcic, J., Nasuelli, D., De Masi, R., Bragadin, L. M., Tommasi, M. A., Zambito-Marsala, S., Moretti, R., Bratina, A., Ukmar, M., Pozzi-Mucelli, R. S., Grop, A., Cazzato, G., and Zorzon, M. (2001). A longitudinal study of brain atrophy and cognitive disturbances in the early phase of relapsing-remitting multiple sclerosis. *Journal of Neurology, Neurosurgery and Psychiatry*, 70(6):773–780. (Cited on page 25.)

Résumé

Cette thèse est consacrée à l'estimation de l'atrophie cérébrale longitudinale pour les patients atteints de sclérose en plaques. Tout d'abord, sur la base de leur robustesse à diverses sources d'erreur (inhomogénéité en intensité, bruit, distorsions géométriques, artefacts d'interpolation et présence de lésions), une évaluation des principales approches d'estimation de l'atrophie cérébrale est réalisée à l'aide de simulations d'une vérité terrain.

Ensuite, une analyse statistique est effectuée afin d'estimer les incertitudes associées à l'atrophie mesurée. Un cadre générique est proposé pour construire des intervalles de confiance de l'atrophie estimée dans une région d'intérêt du cerveau. Une base d'apprentissage est utilisée pour obtenir ces intervalles de confiance pour un algorithme quelconque d'estimation de l'atrophie.

Enfin, un cadre bayésien est proposé pour l'estimation conjointe d'une transformation non-rigide (à partir de laquelle l'atrophie est calculée) et du champ d'inhomogénéité en intensité présent dans les images. Cette approche bayésienne couplée aux techniques MCMC fournit un cadre rigoureux pour la construction d'intervalles de confiance de l'atrophie cérébrale.

Mots-clés: imagerie médicale, atrophie du cerveau, validation, incertitude, intervalle de confiance, recalage non-rigide.

Abstract

This dissertation is dedicated to estimation of longitudinal brain atrophy in patients with multiple sclerosis. First, an evaluation of widely used brain atrophy estimation approaches is carried out using simulated gold standards in order to test their robustness to various sources of error (bias-field inhomogeneity, noise, geometrical distortions, interpolation artefacts and presence of lesions).

Next, statistical analysis of maps of volumetric change in an individual brain is performed in order to estimate the uncertainties associated with the measured atrophy. A generic framework for constructing confidence intervals for atrophy estimated in any region of interest in the brain is developed. With the help of a learning database, confidence intervals can be obtained for any atrophy estimation algorithm.

Finally, a Bayesian framework is developed that allows for the estimation of a non-rigid transformation (from which atrophy is obtained through its Jacobian) and bias field inhomogeneity. MCMC sampling methods coupled with this framework make the construction of confidence intervals possible for the estimated atrophy.

Keywords: medical imaging, brain atrophy, evaluation, uncertainty, confidence intervals, non-rigid registration.