UNIVERSITÉ DE NICE - SOPHIA ANTIPOLIS

ÉCOLE DOCTORALE STIC
SCIENCES ET TECHNOLOGIES DE L'INFORMATION ET DE LA COMMUNICATION

T H È S E

pour obtenir le titre de

Docteur en Sciences

de l'Université de Nice - Sophia Antipolis
Mention : Automatique Traitement du Signal et des Images

Présentée et soutenue par

Jean-Christophe SOUPLET

Évaluation de l'atrophie et de la charge lésionnelle sur des séquences IRM de patients atteints de sclérose en plaques

Thèse dirigée par Grégoire MALANDAIN
et co-dirigée par Nicholas AYACHE

préparée à l'INRIA Sophia Antipolis, Équipe-Projet ASCLEPIOS

et soutenue le 21-01-2009

Jury :

Rapporteurs :	Jean-Paul ARMSPACH	-	Univ. de Strasbourg (LINC)
	Christian BARILLOT	-	IRISA (VisAGeS)
Directeur :	Grégoire MALANDAIN	-	INRIA (Asclépios)
Examinateur :	Sébastien OURSELIN	-	UCL (CMIC)
Invitée :	Christine LEBRUN	-	CHU de Nice (Neurologie)

Je dédie cette thèse à mes parents.

Remerciements

Je souhaite tout d'abord remercier les différents membres de mon jury de thèse. Un grand merci à Jean-Paul Armspach et Christian Barillot d'avoir accepté de rapporter ce travail de thèse et pour leurs remarques constructives. Merci également à Sébastien Ourselin pour ses questions et ses apports notamment sur la prise en compte des problèmes liés à l'acquisition des images. Enfin merci à Christine Lebrun d'avoir apporté son point de vue applicatif sur ces travaux et pour toutes ses explications sur la sclérose en plaques et le domaine médical.

Pendant les trois années et les trois mois de cette thèse, j'ai particulièrement apprécié mon cadre de travail au sein de l'équipe-projet Asclépios de l'INRIA Sophia Antipolis. Je tiens donc à remercier tout particulièrement Grégoire Malandain et Nicholas Ayache de m'avoir accueilli au sein de cette équipe et d'avoir réciproquement dirigé et co-dirigé cette thèse. Merci pour leurs questions, leurs conseils, leurs critiques nombreuses et constructives ainsi que pour avoir su transmettre une partie de leur savoir. Je remercie ensuite les différents membres (anciens ou présents) de l'équipe. Leur accueil, leur joie de vivre et leur esprit d'équipe m'ont réellement touché et soutenu dans les moments heureux mais aussi ceux plus délicats de cette thèse. Je remercie plus particulièrement : Isabelle, notre assistante, pour son aide dans les difficultés administratives ; Olivier Commowick et Pierre Fillard, les anciens, pour leur aide dans la compréhension des différentes architectures logicielles ; Jean-Marc Peyrat et Ender Konukoglu, les compagnons de fortune, pour leur soutien ; Erik Pernod, mon stagiaire, pour son aide dans le développement logiciel ; Mauricio Reyes, Nicolas Toussaint et François Chung, mes co-bureaux, pour leur présence au quotidien.

Dans un contexte plus large, je remercie les différents collègues de l'INRIA et les camarades de l'ADSTIC. Je tiens également à remercier Sébastien Novellas, radiologue au CHU de Nice, et Mikael Cohen, interne en Neurologie, pour m'avoir expliqué leur travail respectif et pour leur aide à comprendre les problématiques auxquelles ils sont confrontés. Enfin, un énorme merci à Pierre Kornprost pour ses conseils et ses encouragements.

Cette thèse a aussi été l'occasion de découvrir l'enseignement, je renouvelle à ce sujet mes remerciements à Grégoire Malandain et Sabet Hachem du master de Génie Biomédical, à Mélanie Ciussi du CERAM et à Igor Litovsky de Polytech pour m'avoir donné ma chance. Un grand merci également, aux étudiants que j'ai eus, pour leurs retours constructifs.

Sur un plan plus personnel, merci aux amis (ceux d'avant la thèse comme les nouveaux). Pendant ces années, ils ont su m'apporter de la joie et des pauses reconstructives pour l'esprit. Les Mac-Pedro, les aventures folles du cycliste malchanceux, les cachets d'amitié d'une pharmacienne de choc, les jeux de cartes innovants du club karmax, les visites de la région organisées par Sansan et bien sûr la joie de vivre et l'hospitalité des *rami* resteront dans ma mémoire. Un grand merci également à JM, Vesti et Stéphane, le trio de coeur ainsi qu'à leur compagne respective pour avoir su garder un contact régulier malgré la distance.

Pour finir, merci à la famille. Tout d'abord mes parents, vraie source d'inspiration, de soutien et d'admiration. Leurs exemples dans la vie de tous les jours et leurs exemples sur l'attitude à prendre dans les moments difficiles sont pour beaucoup dans la réussite de ces travaux. Merci également à ma nouvelle famille (ma belle-famille) pour son accueil très chaleureux en son sein. Celui-ci m'a permis de me sentir chez moi même loin de ma terre d'origine. Merci notamment à mes beaux parents et ma belle soeur pour leurs conseils, pour la motivation sportive des dimanches matins et pour leur aide dans ces travaux. Enfin, merci à ma moitié, Stéphanie, qui a su m'ouvrir son coeur, m'écouter et comprendre mes difficultés. Ses conseils, son soutien et simplement sa présence m'ont énormément apporté pendant ces années de thèse.

Merci à tous et bonne lecture.

Table des matières

1 Introduction **1**

1.1 Cadre des travaux . . . 2

1.2 Organisation du manuscrit . . . 3

2 Contexte médical **7**

2.1 La sclérose en plaques (SEP) . . . 8

2.1.1 Symptômes et diagnostic . . . 10

2.1.2 Les différentes formes de SEP . . . 11

2.2 L'imagerie par résonance magnétique (IRM) . . . 11

2.2.1 Le principe de l'IRM . . . 12

2.2.2 Les différentes séquences IRM . . . 13

2.3 L'IRM et la sclérose en plaques . . . 16

2.3.1 Critères de diagnostic . . . 16

2.3.2 Observations IRM et handicap du patient . . . 17

2.3.3 Analyse à mener . . . 20

2.4 La base de données QUALICORE . . . 21

2.4.1 Protocole de récolte des données . . . 21

2.4.2 Protocole de suivi de la base de données . . . 24

3 Segmentation des compartiments cérébraux **25**

3.1 Notions de traitement d'images . . . 27

3.1.1 Volume partiel . . . 27

3.1.2 Segmentation . . . 28

3.1.3 Atlas . . . 28

3.2 Prétraitements . . . 29

3.2.1 Normalisation spatiale : recalage . . . 29

3.2.2 Normalisation en intensité . . . 30

3.2.3 Extraction du cerveau . . . 33

3.3 Segmentation du cerveau en différents compartiments . . . 34

3.3.1 État de l'art . . . 34

3.3.2 L'algorithme d'espérance maximisée (EM) . . . 36

3.3.3 Segmentation de tissus cérébraux et améliorations de l'EM dans ce cadre . . . 38

3.4 Chaîne de traitement, EMA . . . 45

3.4.1 Choix des données . . . 45

3.4.2 Enchaînement des données . . . 47

4 Extraction du cerveau et régions d'intérêt **49**
4.1 Méthodes d'extraction de l'encéphale . . . 51
4.1.1 AnaT1toBrainmask (ATB) . . . 51
4.1.2 Brain Extraction Tool (BET) . . . 52
4.1.3 Hybrid Watershed Algorithm (HWA) . . . 53
4.1.4 3dIntracranial (3DI) . . . 54
4.1.5 EM-BrainMask (EBM) . . . 55
4.2 Comparaison et nouvelle méthode . . . 56
4.2.1 Cadre de l'étude : l'algorithme STAPLE . . . 56
4.2.2 Comparaison des 5 méthodes sélectionnées et de 26 méthodes composées . . . 58
4.3 Segmentation en sous régions d'intérêt . . . 60

5 Segmentation des lésions **65**
5.1 État de l'art . . . 67
5.1.1 Différentes séquences IRM conventionnelles montrant différentes sortes de lésions . . . 67
5.1.2 Approches de segmentation des lésions . . . 69
5.1.3 Segmentation directe des lésions . . . 69
5.1.4 Classification avec segmentation des compartiments cérébraux 73
5.1.5 Validation des méthodes de segmentation . . . 77
5.1.6 Discussion . . . 80
5.2 Segmentation manuelle et/ou semi-automatique . . . 82
5.3 Segmentation automatique . . . 83
5.3.1 Méthode de segmentation de G. Dugas, PEMS-G . . . 83
5.3.2 Utilisation de la chaîne de traitement EMA . . . 85
5.3.3 Segmentation des lésions du cortex, PEMS-COR . . . 86
5.3.4 Segmentation des lésions du cervelet, PEMS-CER . . . 87
5.4 Workshop MICCAI : *MS lesion segmentation challenge 2008* . . . 88
5.4.1 Optimisation de l'algorithme de segmentation automatique . 88
5.4.2 Comparaison de méthodes sur les données d'entraînement . . 91
5.4.3 Résultats de PEMS-COR sur les données test . . . 91
5.4.4 Discussion . . . 95
5.4.5 Challenge sur site et résultats . . . 96
5.5 Choix de la méthode pour segmenter les lésions . . . 97
5.5.1 Comparaison de deux méthodes automatiques sur la base de données Qualicore . . . 97
5.5.2 Discussion . . . 102

6 Évaluation de l'atrophie **105**
6.1 État de l'art . . . 107
6.1.1 Mesures linéaires . . . 107
6.1.2 Mesures de surfaces ou de volumes . . . 108
6.1.3 Mesures de variations . . . 109

6.1.4 Discussion . . . 110
6.2 Amélioration de la reproductivité des mesures 2D . . . 110
6.2.1 Protocole de mesure . . . 110
6.2.2 Reproductibilité des méthodes . . . 112
6.3 Deux nouvelles méthodes d'évaluation de la FPC . . . 113
6.3.1 La méthode CLASS . . . 113
6.3.2 La méthode LCLASS . . . 114
6.3.3 Validation . . . 115
6.3.4 Apport des méthodes . . . 118
6.4 Comparaison et choix de méthodes d'évaluation de l'atrophie . . . 119
6.4.1 Analyse statistique . . . 119
6.4.2 Comparaison sur des données simulées . . . 120
6.4.3 Comparaison sur des données réelles . . . 121
6.4.4 Choix des méthodes pour analyser la base Qualicore . . . 125
6.4.5 Limites des méthodes choisies sur les images de notre base . . . 126

7 Cadre d'analyse de bases de données et application des algorithmes 127
7.1 Analyse de la base de données QUALICORE . . . 129
7.1.1 Analyse statique . . . 129
7.1.2 Analyse dynamique . . . 138
7.1.3 Analyse par sujet . . . 143
7.1.4 Conclusion . . . 153
7.2 Intégration et mise à disposition des outils d'analyse . . . 154
7.2.1 SepINRIA . . . 154
7.2.2 Web-services exécutés sur une grille de calcul . . . 157

8 Conclusion et perspectives 159
8.1 Synthèse des contributions . . . 160
8.1.1 La chaîne de traitement, EMA . . . 160
8.1.2 Segmentation des lésions . . . 161
8.1.3 Évaluation de l'atrophie . . . 162
8.1.4 Cadre d'analyse d'IRM de patients atteints de SEP : application à la base QUALICORE . . . 162
8.2 Perspectives . . . 163
8.2.1 Segmentation des lésions du tronc et sous classification . . . 163
8.2.2 Perfectionnement et évaluations locales d'atrophie . . . 163
8.2.3 Analyses complémentaires de la base QUALICORE et d'autres bases de données . . . 164
8.2.4 Intensifier l'intégration dans le domaine médical des outils développés . . . 164
8.3 Publications de l'auteur . . . 165
8.3.1 Articles de journaux . . . 165
8.3.2 Articles de congrès (avec actes et comité de sélection) . . . 165
8.3.3 Conférences sur résumé . . . 166

A Tableaux de suivi de la base QUALICORE **169**

B Exemple de segmentations des lésions **171**
B.1 Segmentation des lésions du cortex 171
B.2 Segmentation des lésions du cervelet 188

C Tableau récapitulatif des mesures effectuées sur les sujets de la base QUALICORE **193**

D Projet NeuroLOG et Web-services **197**
D.1 Le projet NeuroLOG 197
D.2 Aspects techniques 198
D.3 Web-services et chaîne de traitement 199

Bibliographie **203**

Liste des acronymes utilisés

3DI	3dIntracranial.
3e ven	Troisième ventricule.
ANOVA	Analysis of variance, soit en français : analyse de la variance.
ALES	Autres lésions (qui ne sont ni des lésions prenant le gadolinium ni des trous noirs en séquence T1).
ALS	Automatic Lesion Segmentation, soit en français : segmentation automatique des lésions. Fonctionnalité du logiciel SepINRIA.
ATB	AnaT1toBrainmask.
Auto	Segmentation automatique.
BAE	Brain Atrophy Evaluation, soit en français : évaluation automatique de l'atrophie. Fonctionnalité du logiciel SepINRIA.
BET	Brain Extraction Tool.
BE3	Méthode composée d'extraction du cerveau. BE3 est basée sur (BET, EM-BrainMask, 3dIntracranial) et a été développée lors des travaux de cette thèse.
BSI	Boundary Shift Integral, soit en français : intégrale sur les déplacements de la frontière.
CL	Charge Lésionnelle.
CLASS	Classification, méthode d'évaluation de l'atrophie basée sur un calcul de la fraction parenchymateuse cérébrale. Cette méthode a été développée lors des travaux de cette thèse.
CR	Carte de Répartition.
DP	Densité de Protons ou séquence de densité de protons.
EBM	EM-BrainMask.
EDSS	Expanded Disability Status Scale, soit en français : échelle étendue de l'état de déficience.
EM	Algorithme d'Espérance Maximisée.
EMA	Chaîne de traitement pour la segmentation des compartiments cérébraux développée lors des travaux de cette thèse.
EMAL	Chaîne de traitement EMA intégrant des modifications pour la prise en compte des informations longitudinales.
EMP	Chaîne de traitement pour la segmentation des compartiments cérébraux qui a été développée par G. Dugas dans [Dugas-Phocion 2006].
FN	Faux Négatif.
FP	Faux Positif.
FPC	Fraction Parenchymateuse Cérébrale.
HWA	Hybrid Watershed Algorithm.
IM	Instant de Mesure.

IR	Instant de Référence.
IRM	Imagerie par Résonance Magnétique ou Images de Résonance Magnétique.
IRMf	Imagerie par Résonance Magnétique fonctionnelle.
IS	Indice de Similarité.
ISC	Image or Segmentation Comparison, soit en français : comparaison d'images ou de segmentations. Fonctionnalité du logiciel SepINRIA.
Lat ven	Ventricules latéraux.
LCLASS	Longitudinale classification, méthode d'évaluation de l'atrophie basée sur un calcul de la fraction parenchymateuse cérébrale. Cette méthode a été développée lors des travaux de cette thèse. Contrairement à CLASS, LCLASS prend en compte des informations longitudinales.
LCR	Liquide Céphalo-Rachidien.
LESGADO	Lésions prenant le gadolinium.
LSE	Lésion Segmentation Edition, soit en français : édition de segmentations de lésions. Fonctionnalité du logiciel SepINRIA.
Manu X	Segmentation manuelle de la séquence X.
MICCAI	Medical Image Computing and Computer-Assisted Intervention.
MNI	Montreal Neurological Institute.
MS	Multiple Sclerosis soit en français : sclérose en plaques.
PEMS-G	Prétraitements puis EM puis application d'un Seuil de manière Globale sur la séquence T2-FLAIR. Cette version a été développée par G. Dugas dans [Dugas-Phocion 2006].
PEMS-CER	Prétraitements puis EM puis application d'un Seuil sur la séquence DP dans la région correspondant à la SB du Cervelet. Cette version a été développée lors des travaux de cette thèse.
PEMS-COMBI	Union des résultats de PEMS-CER et de PEMS-COR.
PEMS-COR	Prétraitements puis EM puis application d'un Seuil sur la séquence T2-FLAIR dans la région correspondant à la SB du Cortex. Cette version a été développée lors des travaux de cette thèse.
PPC	Pourcentage de Points Considérés.
PV	Pourcentage de Variation.
PVVC	Pourcentage de Variation du Volume Cérébral.
ROI	Region Of Interest, soit en français : région d'intérêt.
SA	Segmentation Automatique.
SB	Substance Blanche.
SEN	Sensibilité.
SEP	Sclérose En Plaques.
SG	Substance Grise.
SIENA	Structural Image Evaluation using Normalisation of Atrophy.
SIENAX	extension de SIENA aux études de cohortes.
SNC	Système Nerveux Central.

SPE	Spécificité.
SPG	Syndrome Pseudo Grippal.
SR	Segmentation de référence.
STAPLE	Simultaneous Truth And Performance Level Estimation.
T1	Séquence T1.
T1-GADO	Séquence T1 précédée d'une injection de gadolinium.
T2	Séquence T2.
T2-FLAIR	Séquence T2 avec annulation du signal des liquides.
TNT1	Trous Noirs en séquence T1.
VCN	Volume Cérébral Normalisé.
VPP	Valeur Prédictive Positive.

Chapitre 1

Introduction

Sommaire

1.1 Cadre des travaux 2
1.2 Organisation du manuscrit 3

ORGANISATION DU CHAPITRE : Ce chapitre présente le cadre des travaux de cette thèse. Puis, il dresse une description de l'organisation de ce manuscrit.

L'imagerie médicale a considérablement évolué ces dernières années. Elle permet une analyse non invasive de l'état d'un patient. Celle-ci peut se réaliser d'une part dans le cadre clinique, où elle se fait majoritairement de manière qualitative. Elle peut aussi d'autre part se réaliser de manière plus quantitative en laboratoire de recherche par des algorithmes de traitement d'images. Cette thèse de doctorat se veut un lien entre ces deux domaines. Elle vise à produire des outils de mesures fiables et reproductibles pouvant aider au diagnostic et au suivi d'une maladie, la sclérose en plaques, à partir d'images (séquences) IRM dites conventionnelles (cf. section 2.2.2).

1.1 Cadre des travaux

Le travail présenté dans ce document a été réalisé au sein de l'équipe-projet Asclepios de l'INRIA Sophia Antipolis. Il s'inscrit dans le cadre du projet de recherche clinique : QUALICORE.

Le projet QUALICORE est une évaluation des troubles cognitifs et de la qualité de vie chez les patients atteints de Sclérose en Plaques et traités par Interféron β. Ce travail consiste en une recherche de corrélation entre ces données de qualité de vie et les données d'imagerie. Le porteur du projet est le centre hospitalier universitaire de Clermont-Ferrand. Cette étude exploratoire ne modifie pas la prise en charge thérapeutique mais elle comprend un suivi régulier et standardisé par un neuropsychologue ainsi qu'un suivi en IRM. Cette étude prévoit également d'inclure des patients à partir de cinq centres hospitaliers : Clermont-Ferrand, Dijon, Marseille, Montpellier et Nice.

Ce projet a divers objectifs :

- Corréler l'importance des troubles cognitifs et les données de qualité de vie de patients présentant une sclérose en plaques évoluant par poussées et débutant un traitement par Interféron β.
- Déterminer les caractéristiques démographiques, épidémiologiques.
- Confronter les données du bilan cognitif et la mesure de qualité de vie obtenues auprès des patients avec l'évolution clinique (nombre de poussées, retentissement sur le handicap par échelle EDSS) sur une période de 36 mois.
- Corréler les mesures clinimétriques habituelles, les données du bilan cognitif et les mesures de qualité de vie aux examens IRM conventionnels (cf. section 2.2.2).
- Valider en multicentrique les outils de prétraitement d'images en IRM.

En conséquence, ce projet prévoit d'évaluer chaque patient avant l'inclusion puis une fois par an, soit 4 évaluations au total sur une durée de trois ans. L'objectif est de recruter soixante-quinze patients, soit quinze patients par centre.

Dans ce projet, l'équipe Asclepios est en charge de l'analyse des images IRM. Pour cela, elle bénéficie de son expertise dans l'analyse d'images médicales (les travaux présentés ici font suite à ceux de D. Rey [Rey 2002a] et à ceux de G. Dugas [Dugas-Phocion 2006]) ainsi que de l'expertise des neurologues Christine Lebrun et Mikael Cohen du CHU Pasteur de Nice avec qui ces travaux ont été réalisés.

1.2 Organisation du manuscrit

L'organisation de ce manuscrit est donnée par l'organigramme de la figure 1.1. Elle est relativement linéaire. Voici une brève description des différents chapitres :

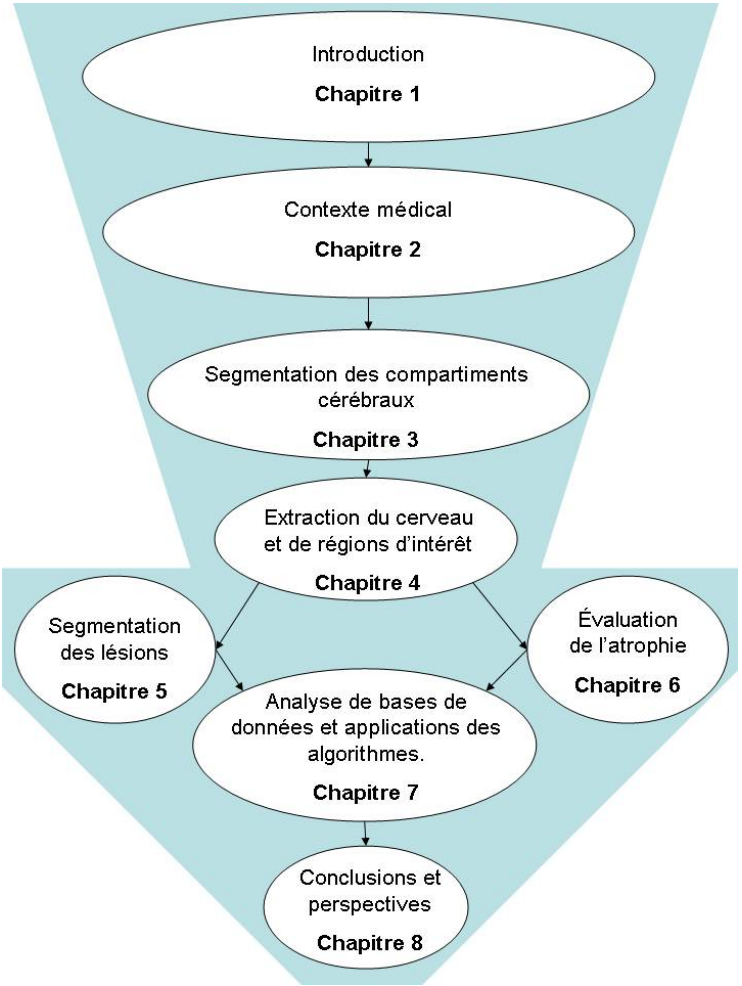


FIG. 1.1 – *Organigramme du manuscrit.*

Chapitre 2 : Contexte médical

Ce chapitre présente la sclérose en plaque, une maladie dégénérative du système nerveux. Il décrit ensuite l'IRM et donne les liens entre les observations qui peuvent y être faites et la maladie. Deux critères (la charge lésionnelle et l'atrophie) sont sélectionnés pour notre étude. Ce chapitre finit par la présentation de la base de données issues du projet QUALICORE (cf. section précédente). Les images de cette dernière seront utilisées pour tester et valider les différents algorithmes et seront analysées dans le chapitre 7.

Chapitre 3 : Segmentation des compartiments cérébraux

Le chapitre 3 commence par de brèves définitions de notions de traitement d'images. Puis, différentes étapes, correspondant à des prétraitements sont détaillées. Celles-ci permettent de normaliser les images avant de réaliser un traitement. Dans notre cas, ce dernier correspond à la segmentation des différents compartiments du cerveau. Un état de l'art des méthodes permettant d'effectuer ce traitement, ainsi qu'une description précise de la méthode retenue et de ses limites sont alors donnés. Cette dernière reprend les travaux de G. Dugas [Dugas-Phocion 2006] mais des améliorations y ont été apportées. Ce chapitre finit par la synthèse des différentes étapes et donne les liens qui existent entre elles, formant ainsi une chaîne de traitement.

Chapitre 4 : Extraction du cerveau et régions d'intérêt

L'extraction du cerveau fait partie des prétraitements qui ont été présentés dans le chapitre précédent. La méthode qui était utilisée jusque-là ne donnait pas pleinement satisfaction. Dans ce chapitre 4, une comparaison de différentes méthodes disponibles sous forme de logiciels est présentée. Celui-ci introduit aussi un cadre permettant de combiner ces méthodes (d'obtenir une segmentation "moyenne") grâce à l'algorithme STAPLE [Warfield 2004]. La comparaison de ces combinaisons et des méthodes sélectionnées a permis d'en sélectionner une. Ce chapitre termine sur la description d'une méthode originale permettant de diviser le masque du cerveau obtenu à l'étape précédente en trois régions d'intérêt : cortex, cervelet et tronc.

Chapitre 5 : Segmentation des lésions

Le chapitre 2 a fait ressortir la charge lésionnelle comme l'un des marqueurs IRM de la SEP. Après un état de l'art des méthodes de segmentation des lésions de SEP, ce chapitre 5 présente un nouvel algorithme semi-automatique de segmentation des lésions. Celui-ci permet un gain de temps pour la réalisation de segmentations de référence par un expert. Puis, le chapitre revient sur la méthode de segmentation des lésions proposée par G. Dugas. Il en donne les principaux défauts et présente deux nouvelles versions de cette méthode que nous avons mises en place pour y répondre. Celles-ci commencent toutes les deux par exécuter la chaîne de traitement présentée au chapitre 3. Ensuite, la première version permet de segmenter les lésions du cortex à partir de la séquence T2-FLAIR. La seconde segmente les lésions du cervelet à partir de la séquence DP. Les résultats de ces méthodes au workshop MICCAI *MS lesion segmentation challenge 2008* sont ensuite donnés. Une étude comparative

entre la méthode de G. Dugas et ces nouvelles versions est ensuite menée sur des données de notre base.

Chapitre 6 : Évaluation de l'atrophie

L'autre marqueur IRM mis en avant par l'étude menée dans le chapitre 2 est l'atrophie. Le chapitre 6 commence par un état de l'art des méthodes permettant de l'évaluer. Il présente ensuite une méthode semi-automatique originale permettant d'améliorer la reproductivité des mesures linéaires manuelles. Puis, deux nouvelles méthodes d'évaluation de la fraction parenchymateuse cérébrale sont présentées. Elles se servent toutes les deux des segmentations fournies par la chaîne de traitement détaillée au chapitre 3. Cependant, dans la deuxième méthode cette chaîne est modifiée pour prendre en compte les données longitudinales.

Pour finir, le chapitre expose : la difficulté de la validation de ces méthodes, l'absence de méthode de référence et le fait que leurs résultats peuvent être contradictoires. Une étude comparative des méthodes proposées et des méthodes disponibles sur internet a ensuite été menée sur les données de notre base.

Chapitre 7 : Cadre d'analyse et application des algorithmes

Ce chapitre est une mise en pratique des algorithmes développés dans les chapitres précédents. Tout d'abord, ces derniers ont été utilisés pour analyser la base de données QUALICORE. Cette analyse a permis de prouver l'homogénéité de la base. La méthode automatique de calcul de la charge lésionnelle est aussi corrélée avec les mesures manuelles. De même, la méthode d'évaluation de l'atrophie prenant en compte les données longitudinales se corrèle aussi avec les mesures manuelles. L'analyse par sujet introduit une discussion sur l'influence de certains paramètres cliniques ou radiologiques qui pourraient faire varier de façon erronée les différentes mesures.

Enfin, différents outils informatiques les utilisant ont été développés. Un logiciel SepINRIA vise à permettre d'analyser les IRM de patients atteints de SEP. Outre le fait de permettre la comparaison des algorithmes développés avec d'autres (non disponibles), ce logiciel peut déjà se révéler utile pour la prise de décisions thérapeutiques. Un dernier outil développé en collaboration avec un stagiaire, Erik Pernod, consiste en la réalisation de web-services permettant d'exécuter sur une grille de calcul la chaîne de segmentation présentée au chapitre 2. Cet outil permettra un partage des différentes briques du traitement entre plusieurs laboratoires et la puissance de calcul de la grille permettra d'analyser un grand nombre de données.

Chapitre 8 : Conclusion et perspectives

Ce manuscrit se terminera par une synthèse des contributions et des perspectives issues de ces travaux.

Chapitre 2

Contexte médical

Sommaire

2.1 La sclérose en plaques (SEP) 8
2.1.1 Symptômes et diagnostic 10
2.1.2 Les différentes formes de SEP 11
2.2 L'imagerie par résonance magnétique (IRM) 11
2.2.1 Le principe de l'IRM 12
2.2.2 Les différentes séquences IRM 13
2.3 L'IRM et la sclérose en plaques 16
2.3.1 Critères de diagnostic 16
2.3.2 Observations IRM et handicap du patient 17
2.3.3 Analyse à mener 20
2.4 La base de données QUALICORE 21
2.4.1 Protocole de récolte des données 21
2.4.2 Protocole de suivi de la base de données 24

Résumé : La sclérose en plaques (SEP) est une maladie auto-immune du système nerveux central (cerveau, mœlle épinière). Son diagnostic est facilité par l'utilisation de l'IRM. Cet examen sert également à prendre des décisions thérapeutiques dans le cadre du suivi du patient. Le marqueur IRM prédominant est la charge lésionnelle. Les neurologues cherchent sur l'imagerie conventionnelle des marqueurs d'activité (*i.e.* nouvelles lésions) qu'ils corrèlent à un évènement clinique péjoratif. Une segmentation précise des lésions permettra de les aider dans cette recherche. Toutefois, cette charge lésionnelle à un instant, T, est peu corrélée avec le handicap du patient (donné par le score EDSS). En conséquence, un autre marqueur est utilisé : l'atrophie. Elle peut être évaluée par différentes méthodes mais aucune ne s'impose comme une méthode de référence. En outre, leurs résultats peuvent être contradictoires. L'analyse de la base QUALICORE se fera donc par une évaluation de ces deux critères, au moyen de différentes méthodes. Le but est de trouver la méthode la mieux adaptée à l'analyse de nos données.

Mots clefs : Sclérose en plaques, Imagerie par Résonance Magnétique, Base QUALICORE.

Contributions :
- État de l'art sur les liens entre IRM et Sclérose en plaques.
- Protocoles de gestion (inclusion et suivi) de la base de données.

Organisation du chapitre : Tout d'abord, ce chapitre décrit la sclérose en plaques (section 2.1) et le principe de l'IRM (section 2.2). Puis, les liens entre les observations IRM et les effets de la maladie sont discutés dans le but de sélectionner les mesures à effectuer sur ces images (section 2.3). Pour finir, la base de données et les protocoles mis en place pour y inclure des données et en assurer le suivi sont détaillés (section 2.4).

2.1 La sclérose en plaques (SEP)

Avec 80 000 malades en France et 2 500 000 dans le monde, la sclérose en plaques (SEP) est la maladie neurologique de l'adulte jeune la plus fréquente après l'épilepsie. La SEP débute le plus souvent entre 20 et 40 ans et touche principalement les femmes (2/3 des malades). Même si les premières descriptions de la maladie remontent au XIV$^{\text{ème}}$ siècle, il faudra attendre le 14 mars 1868 pour que le docteur Jean Martin Charcot en réalise une description claire et précise des lésions [Charcot 1898]. Ces dernières, appelées aussi plaques, sont des zones où une perte de la gaine de myéline des axones d'un neurone s'est produite. Celle-ci entraîne une diminution de la vitesse de transmission d'un neurone à l'autre. Elle peut aussi conduire à la rupture de l'axone (cf. figure 2.1).

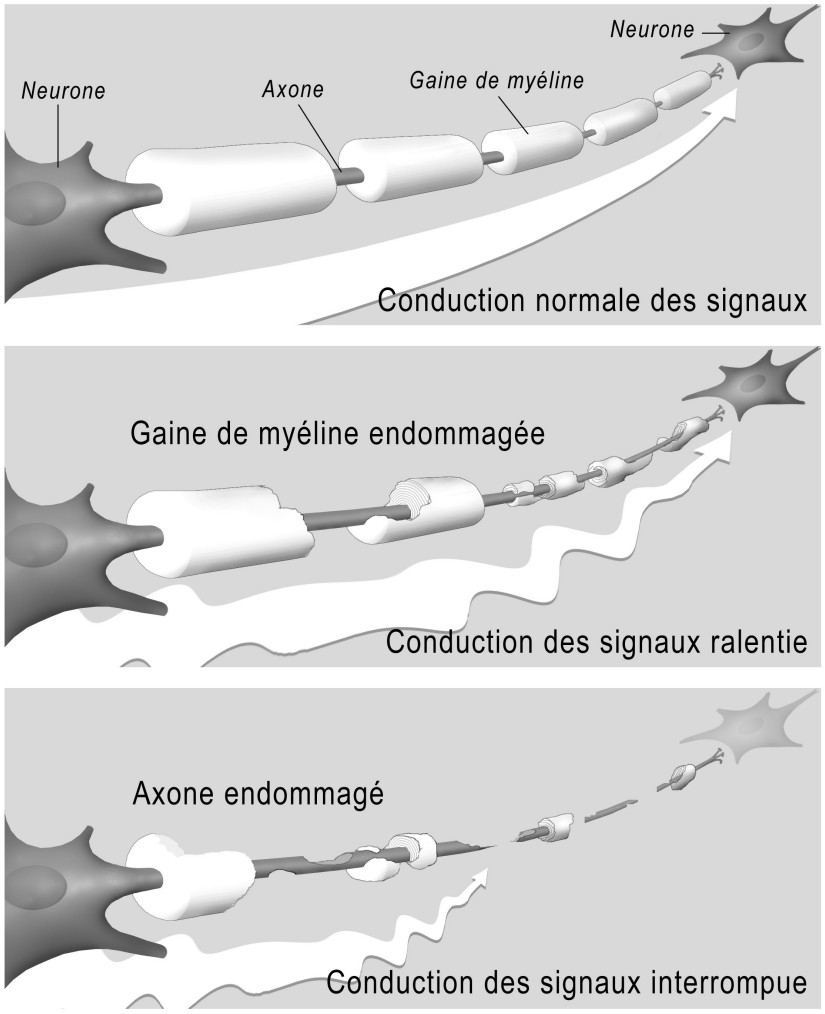


FIG. 2.1 – *Les plaques de sclérose en plaques se caractérisent par une perte de la gaine de myéline (source : site internet de la société suisse de la sclérose en plaques (http ://www.multiplesklerose.ch), remerciements à Frank Meyenberg pour son autorisation).*

La SEP correspond à une atteinte globale du système nerveux central (SNC) (encéphale, tronc cérébral, mœlle épinière). Ainsi, la maladie est caractérisée par de nombreux symptômes selon les régions du système nerveux atteintes. Elle débute le plus souvent par poussées, périodes de survenue ou d'aggravation de signes neurologiques durant plus de 24 heures et séparées de la précédente d'au moins un mois.

De nombreuses hypothèses existent sur l'origine de la maladie. A la fin du XIX^{ème} siècle, le neurologue Pierre Marie émettait la possibilité d'une origine infectieuse, comme si la maladie était le résultat tardif d'une maladie infectieuse banale de l'enfance. Depuis, plusieurs études ont été menées : l'influence, sur le déroulement de la maladie, de l'anxiété et des agressions psychologiques, la répartition géographique de la SEP (épidémiologie), l'incidence d'un rétrovirus (nouvelle classe de virus parasitant le système nerveux) propre à la SEP. Le plus probable est que toutes ces hypothèses s'associent entre elles pour le déclenchement de la maladie[1].

2.1.1 Symptômes et diagnostic

La SEP peut provoquer de multiples symptômes qui retentissent sur la vie quotidienne par les handicaps qu'ils entraînent : troubles de la motricité, troubles sensitifs (fourmillements), troubles de la vision, troubles neurologiques (maladresse des gestes) et troubles génito-urinaires. Une échelle pour quantifier la gravité de ces symptômes a été mise en place en 1983 par Kurtzke [Kurtzke 1983]. Cette échelle EDSS (Expanded Disability Status Scale, soit en français échelle étendue de l'état de déficience) s'avère toutefois être plus une échelle d'autonomie que de handicap. D'autres mesures de l'état d'un patient atteint de SEP ont donc été développées [Amato 2007].

Le diagnostic de la maladie n'est pas pour autant facile en raison des différentes formes qu'elle peut prendre (cf. section 2.1.2). Les médecins se basent alors sur divers critères : les symptômes décrits par le patient, le mode et l'âge de survenue et l'élimination de toute autre maladie de la liste des causes possibles des symptômes. Pour établir ces critères de diagnostic, les médecins disposent de plusieurs examens répartis en trois catégories : électrophysiologique, biologique et radiologique.

L'électrophysiologie se fait par l'étude des potentiels évoqués visuels et consiste à mesurer la vitesse de conduction de l'influx nerveux.

Les mesures biologiques se basent par exemple sur l'analyse du liquide céphalo-rachidien (prélevé chez le patient par ponction lombaire).

La radiologie concerne l'ensemble des examens produisant des images (dans notre cas, du cerveau ou de la mœlle épinière). L'une des procédures les plus utilisées est l'imagerie par résonance magnétique (IRM). Elle sert souvent de référence pour valider les études des autres domaines comme pour les travaux de G. Acar [Acar 2003] ou M. Filippi [Filippi 2004a] en immunologie. Elle permet d'obtenir des images précises du cerveau et d'identifier les plaques.

[1] Site internet de l'ARSEP : http ://www.arsep.org/

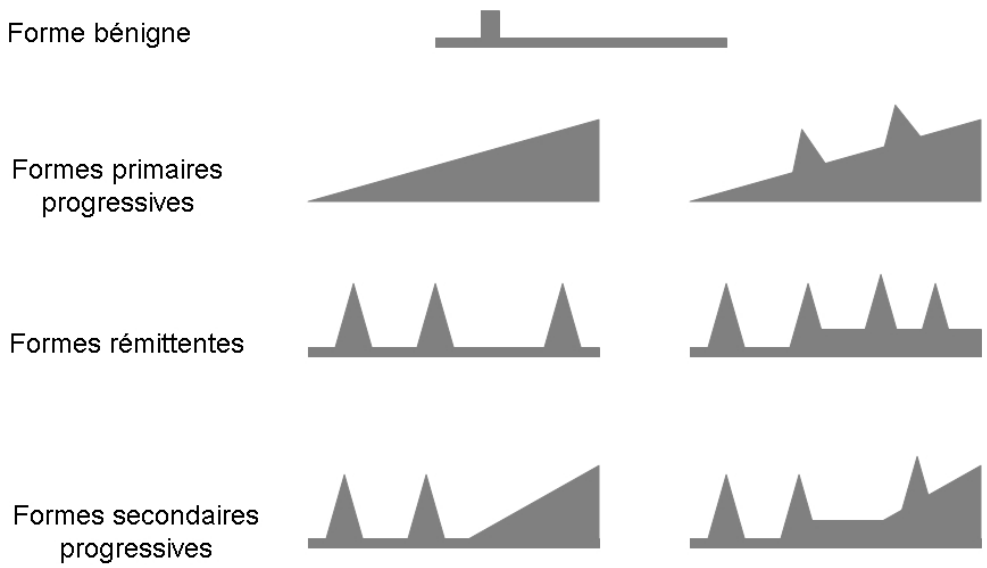


FIG. 2.2 – *Les différentes formes de SEP. Les courbes indiquent la progression des symptômes de la maladie. Les poussées apparaissent sous forme de pics.*

2.1.2 Les différentes formes de SEP

La sclérose en plaques s'exprime le plus souvent par poussées, augmentations rapides du nombre des symptômes pendant quelques jours. En fonction de ce phénomène, différentes formes de SEP sont définies (cf. figure 2.2) :

- les formes bénignes, $\sim$ 10% des cas, se caractérisent par un handicap non significatif après 10 ans d'évolution.
- les formes primaires progressives, $\sim$ 15% des cas, consistent en une aggravation progressive et continue des symptômes.
- les formes rémittentes, $\sim$ 75% des cas au début, présentent des poussées plus ou moins nombreuses avec ou non une invalidité résiduelle.
- les formes secondaires progressives, $\sim$ 50% des formes rémittentes après 10 ans, correspondent à un début de la maladie sous forme rémittente mais qui est suivi par une aggravation progressive et continue de la maladie.

2.2 L'imagerie par résonance magnétique (IRM)

L'IRM a considérablement amélioré la visualisation du SNC. Concernant la SEP, elle permet notamment de quantifier les lésions d'un patient de manière non invasive [Lublin 2005]. Ces lésions correspondent à des zones de l'image en hyposignal ou hypersignal en fonction de la séquence IRM choisie (cf. figure 2.6).

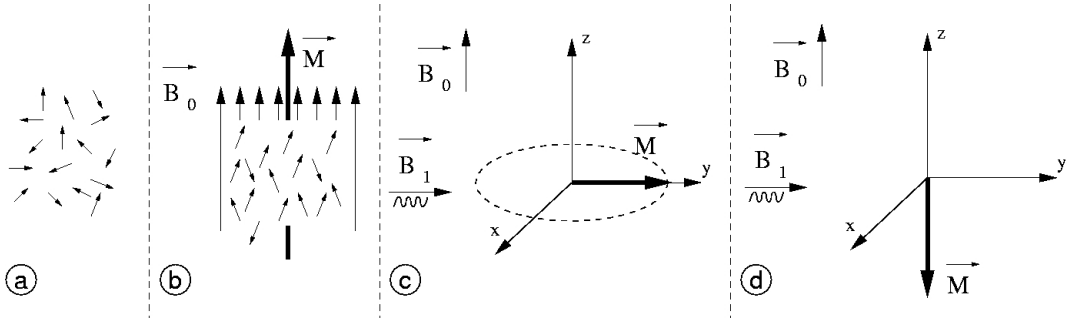


FIG. 2.3 – *Sans champ magnétique, la résultante des moments est nulle car l'orientation des spins est aléatoire (a). Sous l'effet d'un champ magnétique* $\overrightarrow{B_0}$*, il y a un excédent des spins en position parallèle (angle* α*) par rapport aux autres (angle* $\pi - \alpha$*), ce qui crée une résultante des moments* $\overrightarrow{M}$*, colinéaire à* $\overrightarrow{B_0}$ *(b). Il est possible d'obtenir, en fonction de la durée et de l'amplitude de l'excitation* $\overrightarrow{B_1}$*, un angle de basculement de 90 degrés (c) ou de 180 degrés (d) de la résultante* $\overrightarrow{M}$ *(source : [Dugas-Phocion 2006]).*

2.2.1 Le principe de l'IRM

L'IRM est basée sur le principe de la résonance magnétique nucléaire (RMN). La RMN est une technique en développement depuis une soixantaine d'années. Le phénomène physique a été conceptualisé en 1946 par Félix Bloch à Stanford et Edward Purcell à Harvard. Au début des années 70, les travaux de Paul Lauterbur, incluant une information spatiale grâce à des gradients dans le champ magnétique [Lauterbur 1973], ont permis de construire des images bidimensionnelles. Les premières images chez l'homme ont été réalisées en 1979. Aujourd'hui, après une trentaine d'années d'évolution, l'IRM est devenue une technique majeure de l'imagerie médicale moderne.

D'un point de vue technique, le principe de la RMN se découpe en quatre phases : le repos, la polarisation, la résonance, la relaxation. Dans le cas de l'IRM, il est appliqué sur les protons, notamment ceux présents dans l'eau. Ceux-ci sont alors assimilables à des aimants de moment magnétique $\overrightarrow{\mu}$.

- **Au repos**, la résultante (somme des moments magnétiques) $\overrightarrow{M} = \sum_v \overrightarrow{\mu}$ est nulle (figure 2.3, a).
- **En phase de polarisation**, les protons sont plongés dans un champ magnétique $\overrightarrow{B_0}$ et leurs moments magnétiques de spin $\overrightarrow{\mu}$ s'alignent localement sur la direction de $\overrightarrow{B_0}$. L'orientation parallèle au champ $\overrightarrow{B_0}$ est la plus probable, car le niveau d'énergie est plus bas qu'en position antiparallèle. Cet alignement fait apparaître une résultante $\overrightarrow{M}$ orientée dans la direction du champ $\overrightarrow{B_0}$ (figure 2.3, b). A cette étape, les spins sont animés d'un mouvement de précession autour de $\overrightarrow{B_0}$ à une fréquence précise dépendant directement de $\overrightarrow{B_0}$, la fréquence de Larmor.

- **En phase de résonance**, une onde électromagnétique $\overrightarrow{B_1}$ à la fréquence de résonance, c'est-à-dire à la fréquence de Larmor est appliquée. Cette onde est dite impulsion radio-fréquence et permet, en fonction de son amplitude et de sa durée, de faire basculer d'un angle choisi (par exemple 90 degrés (figure 2.3, c) ou 180 degrés (figure 2.3, d)) le moment magnétique macroscopique $\overrightarrow{M}$. Ce moment bascule en suivant une spirale [Bloch 1946].
- **La relaxation** commence lorsque l'impulsion radio-fréquence s'arrête. À ce moment, le retour des spins à l'état d'équilibre crée une onde électromagnétique qui peut être mesurée par une antenne. La mesure IRM correspond donc au temps de relaxation de ce signal. Ce dernier dépendant de l'intensité du champs $\overrightarrow{B_0}$ mais également de la nature des tissus. Cette relaxation correspond en fait à deux phénomènes : la relaxation longitudinale (de temps T1) correspondant aux échanges d'énergie lents entre le milieu et le système de spins, et la relaxation transverse (de temps T2) correspondant aux échanges d'énergie (plus rapides) à l'intérieur du système de spins. En fonction de leur composition, les compartiments (tissus, os, liquide céphalo-rachidien, ...) du cerveau ont des temps de relaxation T1 et/ou T2 différents.

La qualité des images produites dépend de nombreux facteurs (homogénéité du champ $\overrightarrow{B_0}$, mouvements du sujet, hypersignaux appelés artefacts de flux et causés par le mouvement de matière dans le volume observé, ...). Une description plus précise du fonctionnement de l'IRM est accessible sur le site internet du centre d'IRM fonctionnelle de Marseille[2], sur un site internet[3] décrivant "The Basics of MRI" et réalisé par Joseph P. Hornak, professeur de Chimie et de sciences de l'image à l'institut de technologies de Rochester, NY, USA et dans [Liang 1999].

L'IRM cérébrale permet d'observer de façon non invasive différentes zones et donc différents compartiments du cerveau (*i.e.* substance blanche (SB), substance grise (SG), liquide céphalo-rachidien (LCR)) (cf. figure 2.4).

2.2.2 Les différentes séquences IRM

La modification des paramètres d'acquisition IRM (*i.e.* le temps de répétition, Tr, entre deux excitations et le temps d'écho, Te, entre le signal d'excitation et la réception de l'écho) permet de faire apparaître les différences de temps T1 et T2 des différents compartiments observés. Par exemple, en pondération T1 (cf. figure 2.5, a), un temps de répétition court ne laisse le temps qu'aux atomes d'hydrogène de certains tissus de revenir en position d'équilibre, ce qui permet de distinguer ces derniers. En pondération T2 (cf. figure 2.5, b), un temps d'écho long permet d'observer les écarts de décroissance d'énergie et ainsi de différencier les tissus.

Une autre pondération fréquemment utilisée est celle dite de densité de protons (DP) (cf. figure 2.5, c). Elle s'obtient en utilisant un temps de répétition long et un temps d'écho court.

[2] http ://irmfmrs.free.fr/formation/cours_irm/cours_irm.htm
[3] http ://www.cis.rit.edu/htbooks/mri/

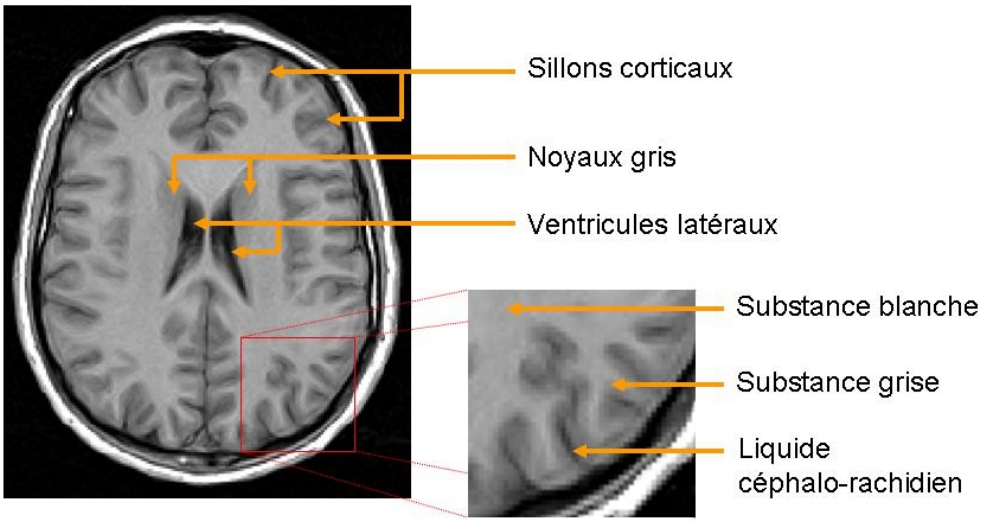


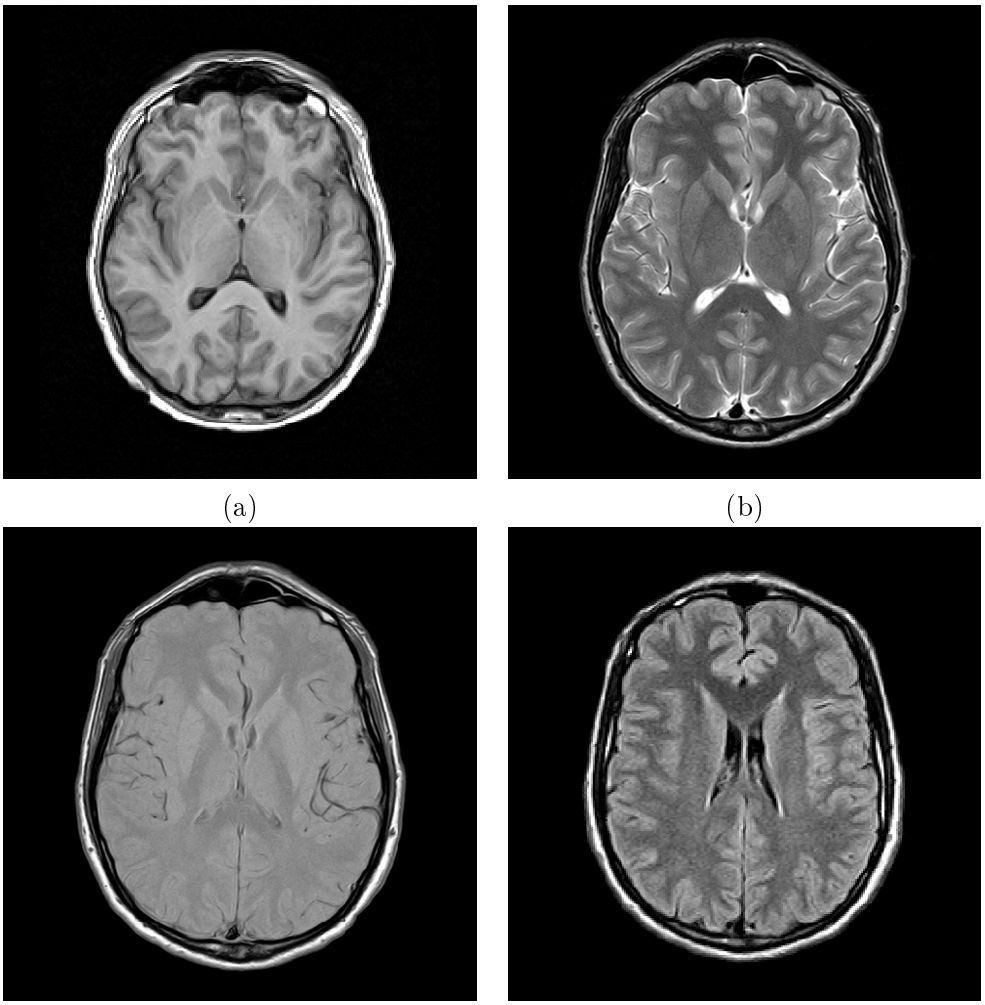
FIG. 2.4 – *Exemple de zones et de compartiments cérébraux pouvant être observés grâce à l'IRM.*

D'autres variations sont possibles : par exemple, l'imagerie T2-FLAIR (cf. figure 2.5, d) est un T2 avec annulation du signal des liquides par une première impulsion à 180 degrés.

Ces quatre séquences (T1, T2, DP, T2-FLAIR) ainsi que la séquence T1 précédée d'une injection de gadolinium (T1-GADO) sont dites des séquences conventionnelles par opposition à de nouvelles séquences IRM permettant d'observer d'autres phénomènes. Les paragraphes suivants présentent certaines de ces nouvelles séquences même si cette présentation ne se veut pas exhaustive.

L'IRM fonctionnelle, IRMf, mesure un changement de proportion entre les globules rouges oxygénés (contenant de l'oxyhémoglobine, molécule non visible en RMN) et les globules rouges désoxygénés (contenant de la désoxyhémoglobine paramagnétique visible en RMN). Une activation (par rapport à un repos) donne lieu à une consommation d'oxygène, donc une diminution du rapport oxygénés/désoxygénés (dans le compartiment veineux, en aval de l'activation). Cependant, cela entraîne un afflux de sang frais (oxygéné), et donc une augmentation de ce rapport (toujours dans les veines en aval de l'activation). La variation de ce rapport est appelée "réponse hémo-dynamique" et permet de mesurer (par corrélation) la réponse à l'activation. Ainsi, il est possible d'appréhender l'organisation fonctionnelle du cerveau. L'IRMf possède de nombreux domaines d'application comme la psychologie, la psychiatrie et la médecine. Toutefois, cette séquence exige de bons rapports signal-bruit et une très grande vitesse d'acquisition des images.

L'imagerie de diffusion mesure le sens de diffusion des molécules d'eau grâce à une variation du champ magnétique de l'IRM. On peut utiliser les valeurs obtenues pour dresser des cartes de diffusion moyenne ou de fraction d'anisotropie. En supposant



(a) (b)

(c) (d)

FIG. 2.5 – *Exemple de séquences IRM conventionnelles : (a) séquence pondérée T1, (b) séquence pondérée T2, (c) séquence pondérée en densité de protons (DP), (d) séquence pondérée T2-FLAIR.*

que la diffusion des molécules se fait principalement le long des fibres du cerveau, on peut visualiser la position, l'orientation et l'anisotropie de ces fibres. Ces dernières correspondent à des faisceaux regroupant plusieurs axones et facilitant la diffusion des molécules d'eau.

2.3 L'IRM et la sclérose en plaques

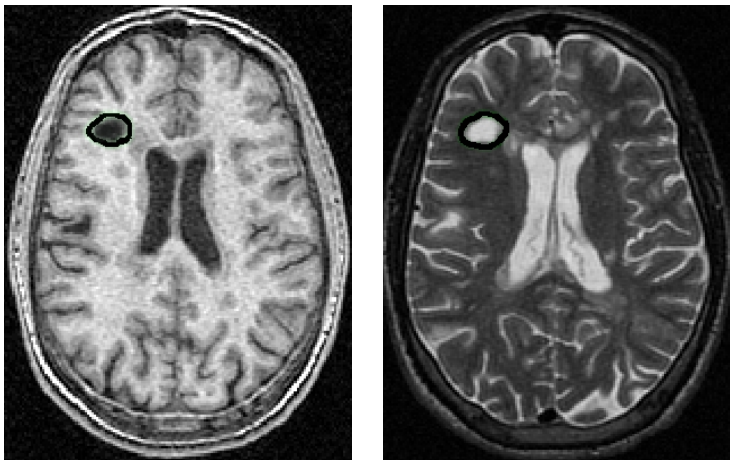
Un examen d'IRM est couramment réalisé en cas de suspicion de sclérose en plaques (SEP) chez un patient ou lors du suivi de cette maladie. En effet, le diagnostic de sclérose en plaques n'est pas évident. D'un point de vue clinique, celui-ci se base sur les critères de McDonald qui permettent de prendre en considération l'ensemble des examens pour poser le diagnostic. Concernant l'IRM, cet examen est suggestif (pouvant correspondre à un patient atteint de SEP) s'il répond aux critères de Barkhof. Le diagnostic de SEP peut ainsi être supposé à partir des images IRM. Cependant, les liens entre les observations possibles sur ces images et l'évolution de la maladie s'exprimant par des troubles physiques ne sont pas certains.

2.3.1 Critères de diagnostic

L'IRM possède un rôle primordial dans la détection de la sclérose en plaques. Des critères spécifiques ont été mis en place à partir de 1997 : les critères de Barkhof/Tintore [Barkhof 1997, Tintore 2000]. Ceux-ci sont basés sur la présence de lésions dans l'image et stipulent qu'un patient a une IRM dite suggestive si parmi les quatre critères suivants au moins trois sont avérés :

- une lésion réhaussée par le gadolinium ou 9 lésions hyperintenses T2 ;
- au moins une lésion juxta-corticale (zone à proximité du cortex) ;
- au moins une lésion sous-tentorielle (zone située en dessous de la tente du cervelet) ;
- au moins trois lésions périventriculaires (proches des ventricules qui sont des cavités contenant le liquide céphalo-rachidien).

Cet examen se retrouve dans les critères de McDonald [McDonald 2001] servant à poser le diagnostic de SEP. Ces derniers ont remplacé ceux de Poser [Poser 1983] utilisés jusque-là et suivent l'idée que la SEP présente à la fois une dissémination spatiale (les lésions peuvent apparaître à plusieurs endroits du centre nerveux (cerveau, mœlle épinière, ...) et une dissémination temporelle (les lésions sont évolutives, elles peuvent apparaître mais aussi disparaître). L'exemple des critères pour diagnostiquer une forme de poussée rémittente (cas de notre base QUALICORE) illustre ceci ; un tel diagnostic sera posé grâce à l'arbre de décision présenté dans la figure 2.7.



(a) (b)

FIG. 2.6 – *L'IRM permet d'observer les lésions de sclérose en plaques : zone en hyposignal à gauche sur une séquence T1 et en hypersignal à droite sur une séquence T2.*

En 2005, ces critères ont été complétés par Polman [Polman 2005]. Ceux-ci donnent deux nouvelles manières de montrer la dissémination en temps :

- détection d'une nouvelle lésion au gadolinium dans les 3 mois suivant l'évènement clinique si l'IRM de référence date de la poussée. Cette lésion ne doit pas être du côté correspondant à cet évènement ;
- détection d'une nouvelle lésion prenant le gadolinium ou T2 si l'IRM de référence a été réalisée plus d'un mois après la poussée.

Ces nouveaux critères prennent aussi en compte les lésions de la mœlle épinière : "une lésion de la mœlle épinière peut être considérée équivalente à une lésion sous-tentorielle du cerveau".

En conséquence, l'IRM semble être le meilleur outil permettant la prise en charge des patients atteints de SEP [Tourbah 2001].

2.3.2 Observations IRM et handicap du patient

Il existe une forte complexité entre les observations IRM et le niveau de handicap du patient [Barkhof 1999, Tourbah 2001]. Pour l'expliquer nous reprenons tout d'abord un article de 2001 de J. Grimaud [Grimaud 2001]. Celui-ci résume et explique ces liens ou plutôt cette absence de lien. Selon J. Grimaud, la corrélation entre cet handicap, mesuré à l'aide de l'échelle EDSS (Expanded Disability Status Scale [Kurtzke 1983]) et la charge lésionnelle observée en IRM T2 est trop faible pour avoir un intérêt clinique. De même, il n'existe aucune corrélation entre la charge des lésions actives obtenue sur une séquence T1 réhaussée par le gadolinium et le handicap [Simon 1998, O'Riordan 1998].

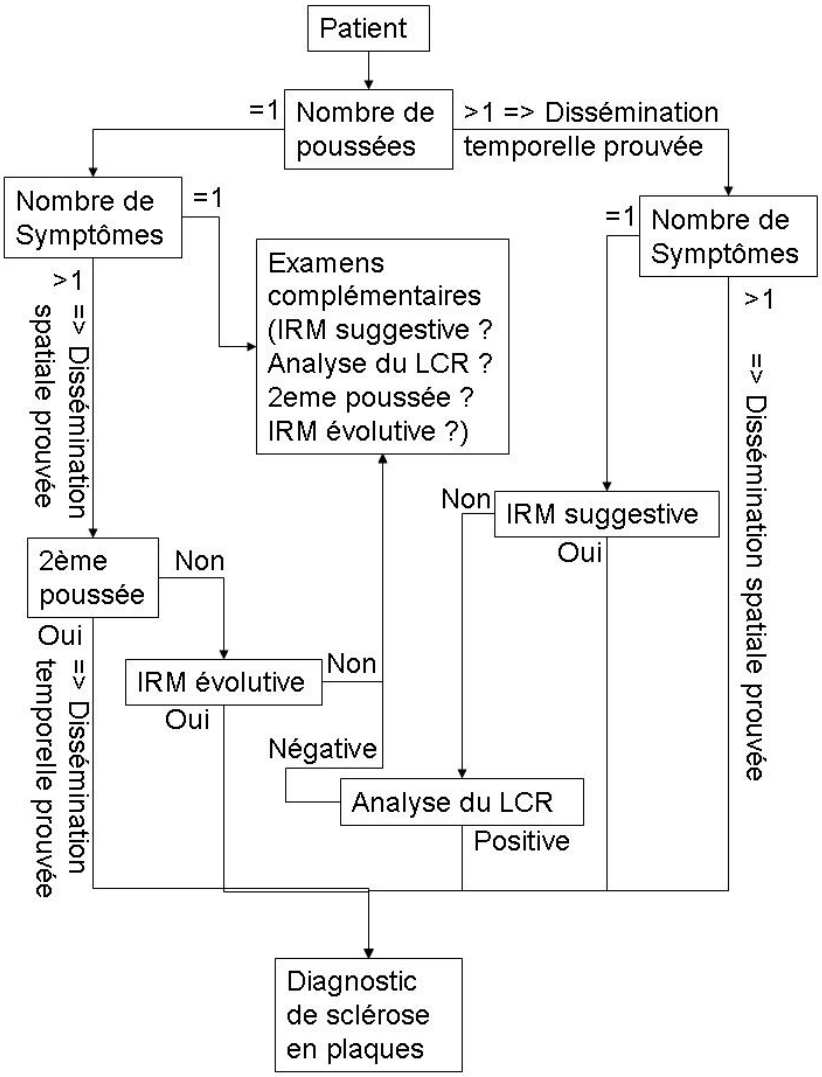


FIG. 2.7 – *Description des critères de McDonald pour diagnostiquer une forme de poussées rémittentes sous forme d'arbre de décision.*

La corrélation entre le handicap et la charge de lésions en hyposignal, appelées "trous noirs", sur les séquences IRM T1 est elle aussi faible. Pour deux études [Truyen 1996, van Walderveen 1999] donnant un coefficient de corrélation significatif, entre le volume de ces trous noirs et le score EDSS, cinq autres études [Giugni 1997, Nijeholt 1998, O'Riordan 1998, Dastidar 1999, van Walderveen 2001] affirment le contraire.

J. Grimaud précise qu'il existe diverses explications plausibles de ce paradoxe :

- l'imperfection de l'échelle EDSS (prise en compte des troubles moteurs et non des troubles cognitifs) ;
- l'absence de prise en compte de l'âge du patient (*i.e.* le nombre d'hypersignaux d'origine vasculaire augmente en T2 avec l'âge du patient, induisant un biais de mesure au niveau de la charge lésionnelle) ;
- l'absence de prise en considération de la vitesse d'apparition des lésions ;
- l'imprécision des segmentations (notamment du contour) des lésions (variabilité intra et inter-expert, fluctuation naturelle des lésions IRM, résolution et type de séquence d'acquisition des images, taille des lésions à segmenter, effet de volume partiel, ...) ;
- la présence de lésions remyélisées ou de nouveaux axones se formant pour permettre à l'information de contourner l'obstacle (dans ces deux cas une lésion reste visible à l'IRM même s'il n'y a pas de conséquence fonctionnelle) ;
- le fait que les lésions observées ne constituent qu'une partie des dégâts engendrés par la SEP.

Pour toutes ces raisons, il semble improbable de pouvoir augmenter la corrélation entre la charge lésionnelle et le handicap du patient avec les échelles de mesure actuelles. L'auteur préconise alors d'utiliser d'autres biomarqueurs comme l'atrophie pour obtenir une évaluation du handicap du patient par l'IRM.

Cette analyse se retrouve dans divers articles [Filippi 2003, Simon 2006]. Les lésions observées sur les IRM conventionnelles ne sont que la partie visible de l'iceberg et sont peu corrélées avec le handicap du patient [Giugni 1997, Simon 1998, O'Riordan 1998, Dastidar 1999, Nijeholt 1998, van Walderveen 2001, Rashid 2007]. Cependant, dans le cas du premier syndrome clinique isolé (première poussée), la charge lésionnelle a une valeur pronostique du handicap à plus long terme [Brex 2002, Korteweg 2006, Kappos 2006]. Des changements de charge lésionnelle observés à l'IRM pendant la première année d'un traitement à l'interféron-β semblent aussi pouvoir indiquer une évolution forte du handicap pendant l'année suivante [Rio 2008]. En outre, une étude suggère qu'il existe une limite supérieure de la charge lésionnelle (volume entre 13 000 et 17 000 mm^3) lorsque le score EDSS est élevé (> 4.5) [Li 2006].

Dans ce contexte, les nouvelles séquences IRM permettent de mettre en évidence des anomalies plus diffuses (ne touchant pas seulement la substance blanche mais aussi la substance grise) et microscopiques par l'analyse des temps de relaxa-

tion, par le transfert d'aimantation, par l'imagerie de diffusion [Zivadinov 2004]. De même, l'IRM fonctionnelle (IRMf) aide à la compréhension des phénomènes de remyélisation [Filippi 2004c]. Il faut noter que ces anomalies peuvent aussi se situer dans la substance grise. La spectroscopie par résonance magnétique fournit, par exemple, une mesure du taux de n-acetyl aspartate (NAA), marqueur des axones fonctionnants. En effet, une diminution de ce taux reflète une perte des axones [Barkhof 2004]. Ce taux est également un bon indicateur de l'évolution du volume cérébral après 18 mois [Coles 1999]. D'autres mesures sont aussi possibles avec la spectroscopie telle que l'évaluation du métabolisme énergétique selon la quantité de phosphocréatine et de créatine ou la sensibilité à la démyélinisation active en calculant la quantité de lipides libres [Bruneteau 2005].

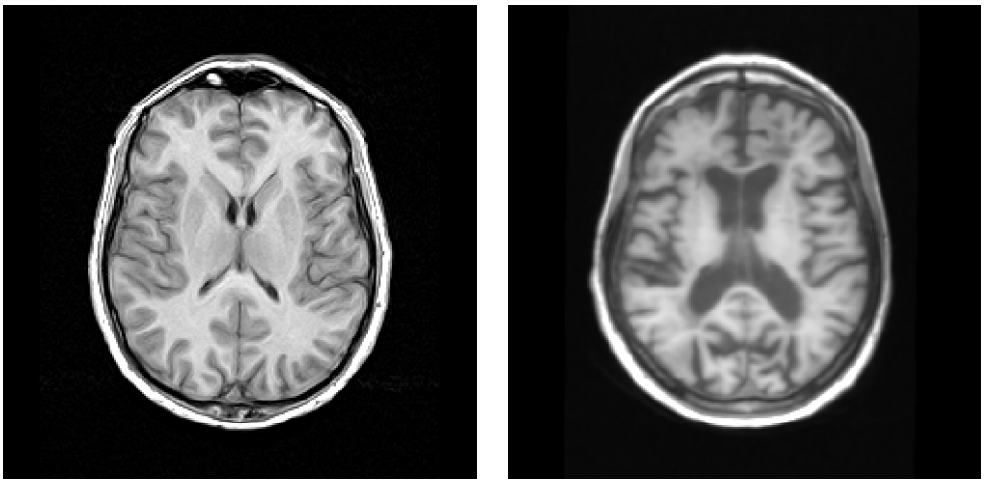
L'atrophie cérébrale est un autre biomarqueur observable à l'IRM qui représente les pertes du volume cérébral causées par divers processus de la maladie [Rudick 2000]. À l'IRM, une augmentation du volume des ventricules et un élargissement des sillons corticaux sont observés (cf. figure 2.8). L'atrophie apparaît dès le début de la phase rémittente de la maladie [Lin 2003] et contribuerait même à la dépression rencontrée chez certains patients [Barkhof 2004]. Certaines études trouvent aussi une corrélation (assez faible mais significative : $r \simeq 0.3, p < 0.001$) entre l'atrophie et le score EDSS (ainsi qu'entre leurs évolutions) [Fisher 2000]. Une corrélation plus forte ($r > 0.4, p < 0.0001$) est aussi démontrée entre l'atrophie et le *Multiple Sclerosis Functional Composite Score* [Cutter 1999, Fischer 1999, Rudick 2002] (niveau de handicap par une mesure composite basée sur la fonctionnalité des membres inférieurs, des membres supérieurs et des fonctions cognitives) [Fisher 2000, Furby 2008].

La pathologie de la SEP s'exprime donc par plusieurs aspects : lésions diffuses dans la matière blanche [Ge 2003], présence de lésions dans la matière grise, mesure d'atrophie, récupération fonctionnelle. De plus, à ces aspects s'ajoute la dégénérescence des axones arrivant dès la phase rémittente. Un manque de corrélation entre l'atrophie (neurodégénérescence) et l'observation des lésions (inflammation des tissus) est aussi confirmé [Filippi 2005].

2.3.3 Analyse à mener

La SEP n'est donc pas exclusivement une maladie de la substance blanche se révélant par des lésions focales. C'est une maladie diffuse produisant des inflammations et de la neurodégénérescence. Ces effets ne sont pas corrélés entre eux mais sont tous les deux présents dès le début de la maladie [Filippi 2005].

Dans le cas où uniquement des séquences IRM conventionnelles sont disponibles, il semble donc nécessaire pour caractériser l'évolution de la maladie de mesurer la charge lésionnelle ainsi que l'atrophie. Notre étude cherchera à mesurer ces deux phénomènes sur les images de notre base de données.



(a) (b)

FIG. 2.8 – *Exemple des effets de l'atrophie : (a) séquence T1 d'un sujet témoin, (b) séquence T1 d'un patient atteint de SEP (effet visible de l'atrophie : Augmentation du volume des ventricules et élargissement des sillons corticaux). La séquence (b) est issue de la base d'IRM du workshop MICCAI'08 MS lesion segmentation Challenge, remerciements à Martin Styner pour son autorisation.*

2.4 La base de données QUALICORE

Le projet QUALICORE (cf. section 1.1) prévoit la récolte d'un grand nombre de données (75 patients $\times$ 4 instants $\times$ 5 modalités = 1500 images). Celles-ci venant de cinq centres différents, des inhomogénéités de protocole d'acquisition et donc de stockage (toutes les images ne sont pas orientées de la même façon) pouvaient exister. De plus, les outils du projet Asclépios fonctionnent sur des images au format *.inr.gz*, il a donc fallu prévoir un passage du format DICOM à ce format. Dans un premier temps, un protocole visant à harmoniser et à convertir ces données a donc été mis en place. Dans un second temps, un protocole pour suivre le remplissage de la base de données a aussi été défini.

2.4.1 Protocole de récolte des données

Chaque centre acquiert pour chaque examen (instant) une série de cinq séquences : T1, couple T2-DP Fast Spin Echo (FSE), T2-FLAIR et T1-GADO. Ces acquisitions se font sur un appareil General Electric de 1.5 Tesla (modèle Signa Horizon LX, software version 9.0). En fonction du centre, elles sont réalisées selon un protocole proche de celui donné dans la table 2.1.

En outre, les différents centres nous envoient des CD-ROM contenant les images au format DICOM. Ce CD-ROM contient un fichier DICOM par coupe (plan) de chaque image 3D. Les images doivent être reconstruites en 3D. En fonction des centres, des variations de l'orientation des images peuvent être présentes. Une

TAB. 2.1 – *Protocole d'acquisition utilisé au centre de Nice.*

Séquence	T1 3D	T1-GADO	FSE T2/DP	T2-FLAIR
Temps d'écho (ms)	1,66	1,66	7,99/103,89	147
Temps de répétition (ms)	7,9	7,9	5000	10004
Angle de bascule (°)	20	20	90	90
Épaisseur (mm)	1,6	1,6	2,0	4,0
Distance intercoupe (mm)	0,8	0,8	2,0	4,0
Taille de la matrice	256×256	256×256	256×256	256×256
Nombre de coupes	152	64	64	28

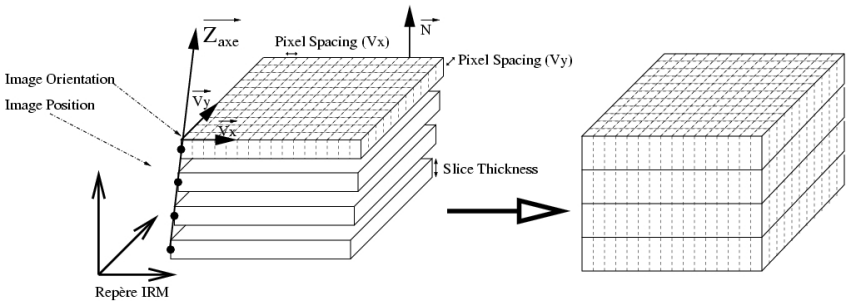


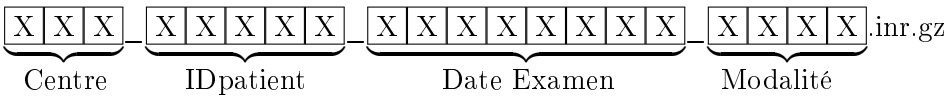
FIG. 2.9 – *Les images IRM peuvent ne pas être des pavés parfaits (légère inclinaison de l'axe $\overrightarrow{z}$) (source : [Dugas-Phocion 2006]).*

conversion du format DICOM au format inr.gz doit aussi être effectuée. Dans certains cas, les images ne sont pas des pavés parfaits [Dugas-Phocion 2006]. En effet, les coupes ne se superposent pas parfaitement et l'axe $\overrightarrow{z}$ est légèrement incliné par rapport au plan $(\overrightarrow{x}, \overrightarrow{y})$ (cf. figure 2.9). Enfin, une norme de nommage (avec anonymisation) et de stockage des données doit être définie.

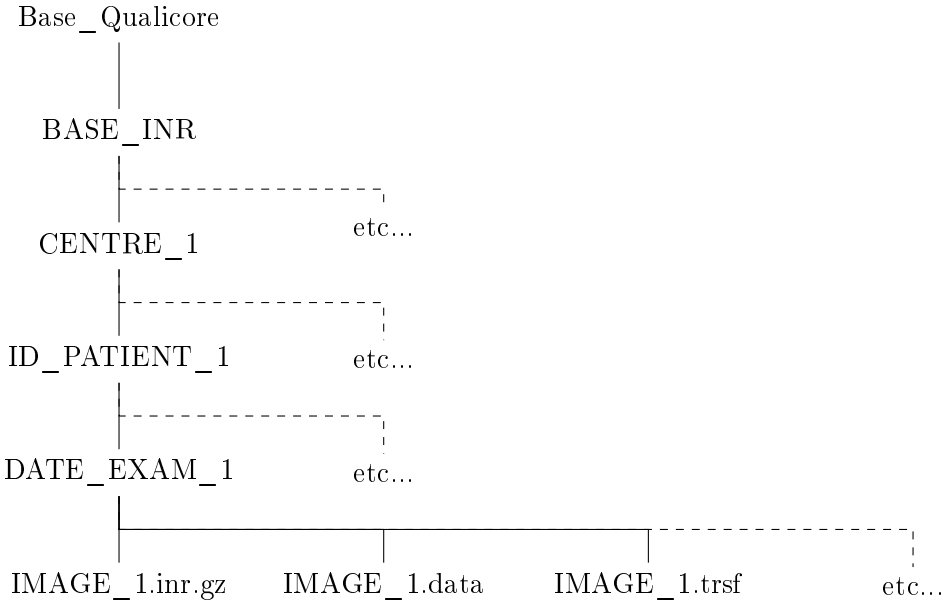
L'ouverture et la reconstruction d'images 3D à partir de fichiers DICOM 2D ne sont pas forcément évidentes. Divers champs DICOM doivent être lus et classés pour trier les coupes 2D. Au début de ma thèse, cette reconstruction se faisait à l'aide de différents scripts écrits dans divers langages et pouvant inclure des étapes réalisées manuellement. Pour automatiser ce pipeline, l'ensemble des traitements a été regroupé au sein du même script qui fournit en sortie l'image 3D (anonymisée automatiquement par lecture des champs DICOM), un fichier .trsf contenant une matrice à appliquer pour corriger l'inclinaison de l'axe $\overrightarrow{z}$ et un fichier .data contenant des informations sur l'image. En fonction du centre de provenance des images, le script les réoriente ou non automatiquement. Le nommage du fichier se fait donc automatiquement en fonction du centre de provenance, du prénom et du nom du patient et de la modalité IRM considérée. Au final, l'insertion de nouvelles données dans la base se réalise directement à partir du CD-ROM en appelant un script.

Voici la méthode qui a été retenue pour le nommage et le stockage des images.

Le nom des images sera de ce type :



Les données sont classées par répertoires suivant l'arborescence suivante :



- BASE_INR : Images brutes.
- CENTRE : Les trois premières lettres du centre (CLE,MAR,MON,NIC,DIJ).
- ID_PATIENT : 5 caractères (les trois premiers du nom et les deux premiers du prénom du sujet).
- DATE_EXAM : date de l'examen.
- IMAGE_1.inr.gz : L'image au format inr.gz.
- IMAGE_1.data : Les informations de l'image.
- IMAGE_1.trsf : La matrice pour corriger l'inclinaison de l'axe $\overrightarrow{z}$.

Dans les trois types de fichiers précédents, IMAGE_1 est remplacée par le nommage vu précédemment. Pour chaque répertoire DATE_EXAM, il y a cinq modalités (T1, T2, DP, T2-FLAIR, T1-GADO) x trois fichiers (.inr.gz, .data. trsf) soit quinze fichiers.

2.4.2 Protocole de suivi de la base de données

Pour suivre l'évolution de la base de données un tableau excel a été créé. Celui-ci contient : pour chaque centre les patients inclus, pour chaque patient les différentes dates d'examens déjà incluses dans la base ainsi que le nombre de jours restant avant la date du prochain examen. Ce tableau récapitulatif a été envoyé régulièrement aux différents partenaires du projet.

Ce bilan est toutefois mitigé. Sur cinq centres, uniquement trois ont inclus des patients (Montpellier : deux patients ; Clermont-Ferrand : huit patients et Nice : seize patients). Le nombre d'examens pour la majorité de ces patients se situe entre un et trois. Les tableaux de suivi pour ces trois centres sont en annexe A.

Nous avons aussi récupéré les images de la base MCI (base de données utilisée par G. Dugas dans [Dugas-Phocion 2006]). Les données disponibles pour les travaux de cette thèse sont répertoriées dans la table 2.2. Celle-ci donne la répartition des sujets (patients) de l'étude, en fonction de leur centre de provenance et du nombre d'instants (examens) disponibles. Pour chaque examen, les séquences T1, T2, DP, T2-FLAIR, T1-GADO sont disponibles sinon l'absence d'une de ces séquences est indiquée dans le tableau. Suite à des erreurs dans le protocole d'acquisition, des images (notées QUAL (erreur)) de certains patients ne respectent pas le protocole d'acquisition QUALICORE. Elles sont à des résolutions différentes. Ceci limite l'utilisation de ces dernières dans le cadre d'une étude portant sur l'atrophie.

TAB. 2.2 – *Nombres de sujets pour chaque série de données, classés selon le nombre d'instants uniquement disponibles.*

Serie	T0	T0+T1	T0+T1+T2	T0+T1+T2+T3
QUAL NIC	3	5	2	3
QUAL MON	2	0	1	0
QUAL CLE	3	4	1	0
QUAL (erreur)	0	3 (NIC)	0	0
MCI	16	11	0	0
QUAL TOTAL	8	9	4	3
Données TOTAL	24	20	4	3

Chapitre 3

Segmentation des compartiments cérébraux

Sommaire

3.1 Notions de traitement d'images 27
3.1.1 Volume partiel 27
3.1.2 Segmentation 28
3.1.3 Atlas 28
3.2 Prétraitements 29
3.2.1 Normalisation spatiale : recalage 29
3.2.2 Normalisation en intensité 30
3.2.3 Extraction du cerveau 33
3.3 Segmentation du cerveau en différents compartiments . . . 34
3.3.1 État de l'art 34
3.3.2 L'algorithme d'espérance maximisée (EM) 36
3.3.3 Segmentation de tissus cérébraux et améliorations de l'EM dans ce cadre 38
3.4 Chaîne de traitement, EMA 45
3.4.1 Choix des données 45
3.4.2 Enchaînement des données 47

Résumé : Les IRM contiennent des inhomogénéités (spatiales, d'intensité, ...). Celles-ci doivent être corrigées par des prétraitements avant de commencer l'analyse des images. Après ces corrections, un voxel d'une séquence IRM correspond physiquement à la même partie anatomique du cerveau que les voxels de mêmes coordonnées sur les autres séquences. Son intensité est proche de l'intensité des autres voxels représentant le même compartiment cérébral que lui. Le cerveau a été extrait de l'image (le crâne, les yeux et d'autres tissus ont été retirés).

Parmi différentes méthodes de segmentation des compartiments cérébraux (*i.e.* substance blanche, substance grise, liquide céphalo-rachidien), l'algorithme d'espérance maximisée, EM, a été selectionné pour diverses raisons (garantie de sa convergence, prise en compte possible des volumes partiels, modèle statistique de la répartition des données permettant d'exclure des points aberrants, ...). Cet algorithme a aussi certaines limites (dépendance du résultat à l'initialisation, fixation manuelle de paramètres, échec possible, ...) auxquelles nous avons proposé des réponses soit à l'aide de la littérature soit à l'aide de solutions originales (calcul d'une pseudo-inverse dans le cas d'une matrice de covariance singulière).

Des liens entre ces différentes boîtes à outils de traitement d'images sont ensuite définis. Ainsi, une chaîne de traitement complète, appelée EMA et permettant de segmenter à partir des images brutes les compartiments cérébraux, est obtenue.

Mots clefs : Prétraitements, Segmentation, Algorithme d'espérance maximisée (EM).

Contributions :

- État de l'art des prétraitements IRM.
- Améliorations de l'EM :
  - Appel à une matrice pseudo-inverse lorsque la matrice de covariance est singulière.
  - Étude de l'influence du pourcentage de points considérés (PPC) pour calculer les paramètres des classes dans l'EM sur la segmentation de la SB.

Organisation du chapitre : Après une définition des notions basiques de traitement d'images (section 3.1), ce chapitre décrit les différents prétraitements pouvant être appliqués sur les images (section 3.2). Puis, un bref état de l'art des algorithmes de segmentations d'images IRM cérébrales ainsi que les justifications du choix de l'un de ces algorithmes, l'algorithme d'espérance maximisée (EM), sont donnés (section 3.3.1). L'EM est ensuite décrit dans la section 3.3.2. Ses limites et les améliorations pouvant y être apportées sont données dans la section 3.3.3. Pour finir, les liens entre ces outils (définissant la chaîne de traitement) sont présentés dans la section 3.4.

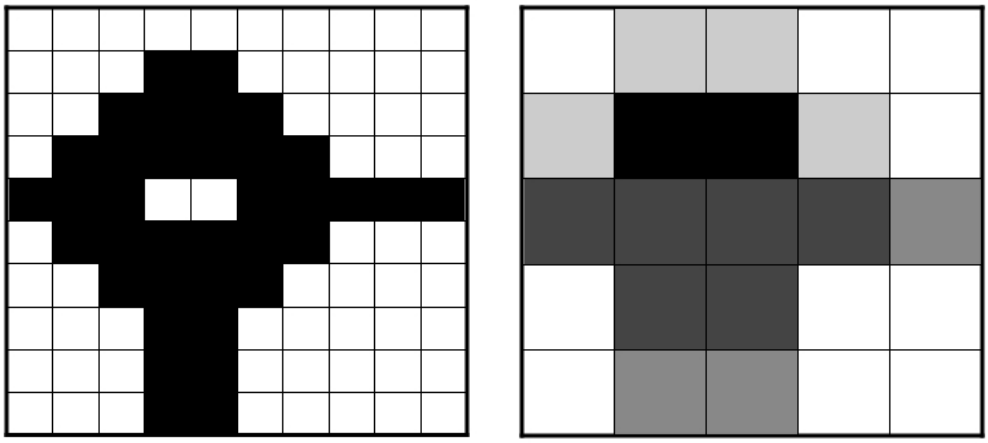


FIG. 3.1 – *Exemple de l'effet de volume partiel : à gauche, l'image (maillage de* 10×10 *pixels) est binaire (pixels blanc ou noir) et la forme est bien définie ; à droite, l'image (maillage de* 5×5 *pixels) possède 5 intensités différentes : blanc, gris clair, gris moyen, gris foncé et noir. L'intensité de chaque voxel de l'image de droite est la moyenne des intensités des 4 voxels de l'image de gauche auxquels il correspond. Les détails ne sont plus visibles sur l'image de droite.*

3.1 Notions de traitement d'images

Cette section vise à expliquer certains termes du traitement d'images qui sont couramment utilisés dans ce manuscrit.

3.1.1 Volume partiel

Les images correspondent à des maillages 2D ou 3D représentant un objet. Dans ces maillages, le plus petit élément est appelé un pixel s'il s'agit d'une image 2D ou un voxel dans le cas 3D. Il faut noter que l'on peut aussi parler d'image 4D dans le cas d'une succession temporelle d'images 3D. Dans la suite de ce manuscrit, nous utiliserons le terme de voxel.

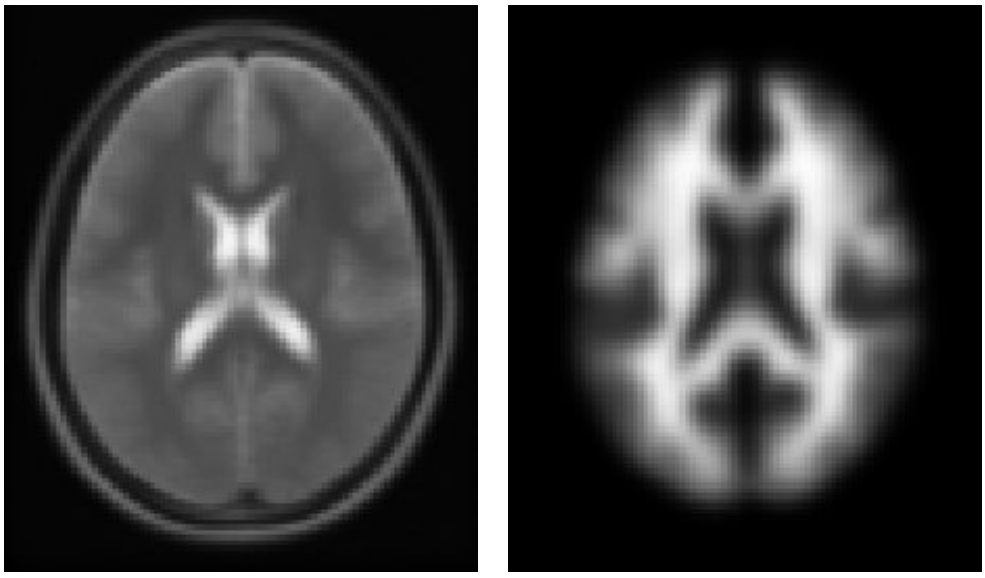
Un voxel possède une dimension dépendante du système ou du protocole d'acquisition. Son intensité correspond à l'intégration sur son volume des signaux envoyés par les tissus qui le composent. Dans le cas où seul un tissu est présent dans le voxel, l'intensité est donc obtenue par l'addition du signal correspondant au tissu et du signal correspondant au bruit. Dans le cas où plusieurs tissus sont présents, l'intensité obtenue correspond à un mélange des différents signaux correspondant aux divers tissus et au bruit. L'effet de volume partiel est illustré par la figure 3.1 : le contour d'une forme bien définie devient difficile à délimiter lors d'une diminution de la résolution de l'image.

3.1.2 Segmentation

Supposons qu'une image soit composée de trois tissus, A, B et C. La segmentation du tissu B correspond à une image donnant l'appartenance de chaque voxel à cette classe B. Cette segmentation peut être binaire : un 1 indique que le voxel appartient à la classe sinon le voxel prend la valeur nulle. Elle peut aussi être probabiliste, le voxel prend alors une valeur comprise entre 0 et 1 (cas où les valeurs sont normalisées).

3.1.3 Atlas

Un atlas correspond à un ensemble ordonné de "cartes" servant à représenter et à décrire des données. Dans la suite de ce manuscrit, le terme atlas se définit de la manière suivante : ensemble composé de deux images (cf. figure 3.2). La première correspond à une image IRM (pouvant être moyennée à partir de celles de plusieurs sujets). Ceci permet d'avoir une image le plus standard possible. La seconde image est une segmentation (binaire ou probabiliste). Dans notre cas, l'utilisation principale de cet atlas est de recaler (cf section 3.2.1) l'image moyenne sur l'image à traiter et d'appliquer la transformation obtenue sur l'image contenant la segmentation. Cette segmentation se retrouve alors dans le repère de l'image et peut apporter des informations pour un traitement ultérieur.



(a) (b)

FIG. 3.2 – *Exemple de l'atlas du Montreal Neurological Institute (MNI) : (a) séquence T1 moyenne, (b) segmentation probabiliste de la substance blanche.*

3.2 Prétraitements

3.2.1 Normalisation spatiale : recalage

L'exploitation des différentes séquences IRM nécessite d'avoir une correspondance point à point entre les différentes images. Par exemple, dans notre cas, nous avons pour chaque instant de l'étude plusieurs séquences IRM. Nous avons donc besoin que les voxels de coordonnées identiques sur ces différentes images correspondent à la même partie anatomique du cerveau. Cette relation entre les voxels des images est intrinsèque pour les séquences double-écho, qui donnent les images T2 et DP, mais pas pour les autres séquences : un déplacement (fictif) du patient peut apparaître entre les différentes images. Un tel déplacement existe aussi entre les séquences d'un même patient acquises à deux instants différents. Il faut donc calculer les transformations spatiales existant entre les images, c'est le but des méthodes dites de recalage. Pour reprendre la formulation proposée dans [Brown 1992] : une transformation $\hat{T}$ maximisant un critère S de correspondance entre une image de référence, R, et une image flottante, F doit être calculée :

$$\hat{T} = \arg\max_{T} S(R, F, T) \tag{3.1}$$

Le calcul de cette transformation peut se faire suivant différents critères. Les méthodes de la première catégorie sont dites géométriques car elles cherchent une mise en correspondance entre des primitives géométriques (points, courbes, surfaces, ...), appelées aussi amers géométriques. Ces méthodes peuvent être découpées en trois étapes : extractions des primitives, appariements de celles-ci (chaque primitive de l'image flottante, F, est associée à la primitive correspondante sur l'image de référence, R), calcul de la transformation, T. Le critère S correspond ici à la distance entre ces primitives. Les méthodes de la seconde catégorie correspondent aux méthodes iconiques. Elles sont basées sur l'intensité des voxels et cherchent à optimiser un critère de ressemblance appelé mesure de similarité. De nombreuses mesures de similarité sont proposées dans la littérature [Hill 2001].

Les difficultés majeures des méthodes géométriques consistent en l'extraction (segmentation) et l'appariement des primitives géométriques. Les méthodes iconiques permettent de s'affranchir de celles-ci mais le critère de similarité doit être adapté à la relation physique entre les images à recaler [Roche 2000]. Dans notre cas, on favorisera alors une troisième classe de méthodes, dites de recalage iconique local. Ces dernières appliquent le principe des méthodes iconiques globales indépendamment sur plusieurs régions, blocs, de l'image et cherchent ensuite à optimiser la somme de ces critères locaux. Une variante consiste à apparier les blocs entre eux, puis à calculer une transformation entre ces appariements [Ourselin 2000]. Dans cette méthode, le calcul de la transformation se fait par la méthode des moindres carrés tamisés pondérés (LTSW) qui permet de rejeter les appariements aberrants.

Suivant l'application, différents types de transformations peuvent être calculés. Dans le cadre de notre travail nous nous limiterons aux transformations linéaires.

Ces dernières se décomposent principalement en trois sous-groupes :

- les transformations rigides, 6 degrés de liberté (une rotation et une translation),
- les similitudes, 7 degrés de liberté (une rotation, une translation et un facteur d'échelle),
- les transformations affines, 12 degrés de liberté, (une rotation, une translation, des cisaillements et un facteur d'échelle différent selon les trois dimensions de l'espace).

Les autres transformations sont dites non-linéaires (transformations paramétriques ou transformations denses). Elles offrent plus de degrés de liberté et permettent de déformer davantage l'image flottante. Toutefois, ces transformations offrent trop de degrés de liberté dans notre cas. En effet, nos travaux recalent essentiellement des séquences IRM issues d'un même patient. Ceci ne nécessite donc pas une déformation importante de l'image.

Après le calcul de la transformation, l'image flottante est rééchantillonnée afin de correspondre point à point à l'image de référence.

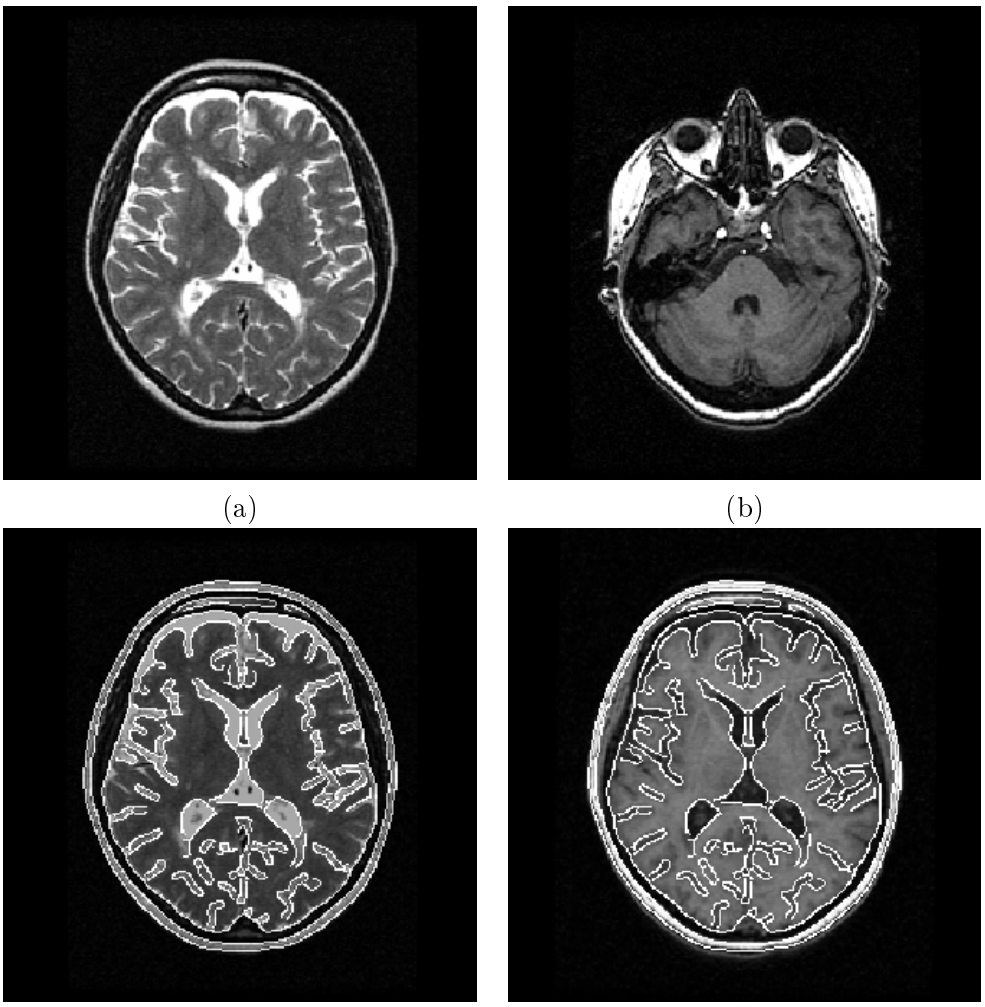
Le recalage permet donc de mettre en correspondance des images (cf. figure 3.3). La littérature propose un grand nombre de méthodes pour réaliser cette tâche. Les articles [Brown 1992, Van den Elsen 1993, Maintz 1998, Lester 1999, Hajnal 2001, Crum 2003, Zitova 2003] en dressent un état de l'art. L'utilité du recalage en routine clinique est de plus démontrée, notamment dans le cadre de la SEP [Lebrun 2004]. Aujourd'hui, le problème du recalage de séquences IRM conventionnelles cérébrales est globalement résolu. Les derniers travaux dans ce domaine s'appliquent aux nouvelles modalités IRM [Gholipour 2007, Yeo 2008].

Dans notre cas, la méthode de [Ourselin 2000] a été retenue pour recaler les images. Elle permet de réaliser des recalages rigides ou affines. Son calcul de transformation avec rejet des points aberrants est également un atout pour recaler des images de patients atteints de SEP. En effet, dans le cas de deux images acquises à des instants différents, la charge lésionnelle n'est pas la même. Une lésion a pu apparaître ou disparaître entre les deux images. Le rejet des appariements aberrants permettra de calculer une transformation correcte malgré ces variations de signal. Pour ces raisons et pour sa vitesse d'exécution, cette méthode a donc été choisie dans notre étude. Elle servira à effectuer les deux types de recalages cités en début de cette section et dans le cas d'une troisème application dont nous aurons besoin : recalage d'un atlas sur nos données (cf. section 3.3.2).

3.2.2 Normalisation en intensité

En IRM, deux voxels ayant la même composition biologique, *i.e.* tissulaire, dans la même séquence peuvent ne pas avoir la même intensité. Cette différence en intensité est appelée biais et peut être classée en deux catégories :

- Les inhomogénéités intra image dues :
  - aux inhomogénéités de l'acquisition radio fréquence [Narayana 1988],



(a) (b)

(c) (d)

FIG. 3.3 – *Recalage d'une séquence T1 sur une séquence T2 : (a) 32ème coupe d'une séquence T2 (image de référence), (b) 32ème coupe d'une séquence T1 (image flottante de résolution différente), (c) 32ème coupe de la séquence T2 et ses contours, (d) contours de la séquence T2 superposés sur la 32ème coupe de la séquence T1 après recalage.*

- au biais tissulaire (l'intensité d'un même tissu biologique varie autour d'une valeur moyenne [Prima 2001]).
- Les inhomogénéités inter images dues à l'absence d'échelle d'intensité absolue en IRM [Prima 2003]. Ces inhomogénéités doivent être corrigées dans les études longitudinales ou de cohortes.

Les inhomogénéités intra image, ou biais spatial, sont souvent modélisées en fonction de l'intensité "réelle" du tissu [Prima 2001]. Pour un voxel i, son intensité, y_i sur l'IRM est considérée comme étant son intensité réelle, x_i multipliée par un facteur, b_i (correspondant au biais, supposé multiplicatif [Sled 1998a]) plus un bruit de mesure (cf. équation 3.2).

$$y_i = b_i x_i + \varepsilon_i^{mes} \tag{3.2}$$

Pour tenir compte du biais tissulaire, on peut considérer que l'intensité "réelle", x_i est la somme de la moyenne de la classe, μ_k et d'un bruit biologique, ε_i^{bio} (cf. équation 3.3). La modélisation complète du biais de l'image est alors donnée par l'équation 3.4.

$$x_i = \mu_k + \varepsilon_i^{bio} \tag{3.3}$$

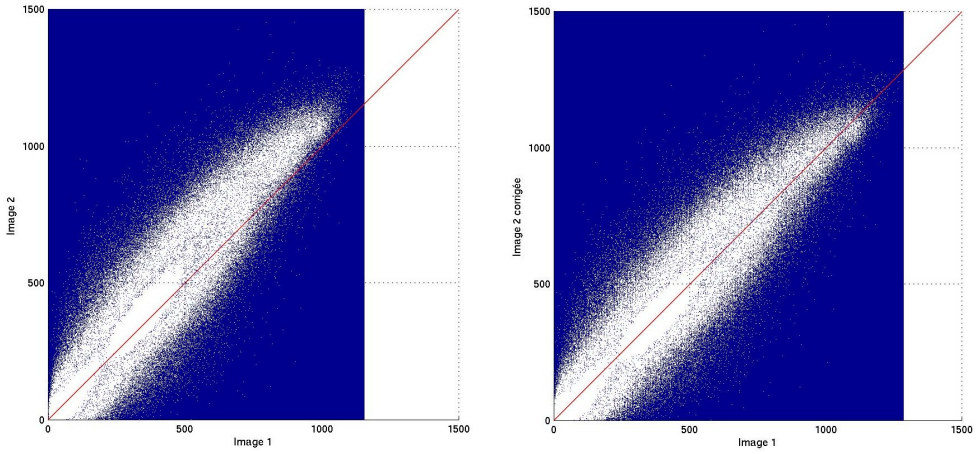
$$y_i = b_i(\mu_k + \varepsilon_i^{bio}) + \varepsilon_i^{mes}. \tag{3.4}$$

Les inhomogénéités inter images, ou biais temporel, sont dues à la sensibilité de l'IRM aux facteurs extérieurs (température, taux d'humidité, ...) et au manque d'échelle absolue d'intensité en IRM. Un même voxel représentant la même composition biologique n'aura pas la même intensité sur deux acquisitions différentes. Contrairement au biais spatial qui peut être considéré comme une fonction lente de l'espace, les variations induites par cet effet peuvent être plus rapides.

Ces deux catégories de biais ont des effets différents. En conséquence, nous avons choisi de les traiter séparément. La littérature propose de nombreuses méthodes pour corriger le biais spatial [Sled 1998b, Van Leemput 1999a, Mangin 2000, Prima 2001, Learned-Miller 2005, Milles 2007]. Dans la pratique, puisque les méthodes présentées dans les chapitres suivants sont basées sur l'algorithme d'espérance maximisée (EM), nous avons utilisé la méthode [Prima 2001]. Une première segmentation du cerveau est effectuée grâce à l'algorithme EM à partir de différentes séquences IRM préalablement recalées. Puis, pour chaque séquence IRM, un polynôme à appliquer pour corriger le biais est calculé à partir des segmentations obtenues.

Les méthodes de correction du biais temporel sont principalement basées sur l'égalisation des histogrammes entre deux images [Styner 2000]. Dans ce cas, l'histogramme de chaque image est modélisé. Puis, la transformation à appliquer pour égaliser les paramètres des deux modélisations obtenues est calculée. Ce principe peut aussi être appliqué aux corrections d'images de nature différente [Wang 1998a, Molyneux 1998, Beyer 2008]. Dans notre cas, nous utiliserons la méthode décrite dans [Rey 2002a]. Deux séquences IRM de différents instants sont

recalées entre elles. L'histogramme conjoint de ces deux images, I1 et I2, peut alors être tracé (cf. figure 3.4, a). En cas d'absence de biais temporel, les points de cet histogramme devraient être placés le long de la droite $y = x$. Ce n'est pas le cas ici. Pour redresser cette distribution, une fonction affine d'équation (y=ax+b) est appliquée sur les intensités de I2. Les valeurs a et b sont calculées par la méthode des moindres carrés utilisée avec une distance orthogonale (distance la plus courte entre un point et la droite de régression linéaire). L'histogramme conjoint de I1 et de I2 corrigée montre alors le redressement de la distribution le long de la droite $y = x$ (cf. figure 3.4, b).



(a) (b)

FIG. 3.4 – *Histogrammes conjoints : (a) de deux images : I1 et I2, (b) de I1 et I2 corrigée.*

3.2.3 Extraction du cerveau

La majorité des méthodes de classification (cf. section 3.3.1) ne considère que l'encéphale (cerveau). Elles requièrent donc de l'avoir précédemment isolé du reste de l'image. On exclut volontairement les autres tissus (peau, yeux, os, ...) qui peuvent perturber l'analyse. Plusieurs de ces méthodes, dénommées skull-stripping en anglais, sont disponibles [Sandor 1997, Ward 1999, Shattuck 2001, Smith 2002b, Segonne 2004, Zhuang 2006].

Certaines d'entre elles sont disponibles sous forme de logiciels. Nous avons pu les tester sur nos données. Cependant, aucune n'a donné satisfaction sur toutes les images. Certaines échouant là où les autres donnaient un résultat acceptable et inversement. Nous avons cherché à utiliser les avantages de chacune grâce au cadre de l'algorithme STAPLE [Warfield 2004], permettant d'obtenir à partir de plusieurs segmentations, une segmentation "vraie" et probabiliste. Cette étude ainsi que la description de la méthode d'extraction du cerveau que nous avons sélectionnée sont présentées dans le chapitre 4.

3.3 Segmentation du cerveau en différents compartiments

La segmentation du cerveau en différents compartiments (*i.e.* substance blanche (SB), substance grise (SG), liquide céphalo-rachidien (LCR)) est une étape clef de notre étude. Premièrement, la connaissance de ces segmentations permettra d'obtenir des informations directement exploitables dans la détection et la réalisation du contour des lésions (cf. chapitre 5). Deuxièmement, la segmentation de ces tissus permettra d'obtenir le volume du cerveau et d'en déduire une mesure d'atrophie (cf. chapitre 6).

3.3.1 État de l'art

En imagerie médicale cérébrale, le but de la segmentation est de regrouper les voxels représentant la même composition biologique, les mêmes régions anatomiques : SB, SG, LCR mais aussi vaisseaux, os, yeux, ... (si l'étape d'extraction du cerveau n'a pas été réalisée). De nombreux articles ou rapports de recherche dressent un état de l'art complet de ces méthodes [Pham 2000, Suri 2002a, Suri 2002b, Zaidi 2006, Lecoeur 2007]. L'objet de cette section n'est donc pas d'en faire la synthèse mais de présenter les principes des différentes approches sur lesquelles sont basés ces nombreux algorithmes.

Les méthodes du premier groupe cherchent à trouver et à découper un ou plusieurs éléments de l'image à traiter. Elles réalisent une segmentation.

Le seuillage est la méthode de segmentation la plus simple. Elle consiste à classer les voxels d'intensité inférieure à une valeur, un seuil, dans une classe et ceux d'intensité supérieure dans une autre. Le choix de ce seuil peut se faire manuellement mais aussi automatiquement en se basant par exemple sur une analyse de l'histogramme de l'image. L'application de cette méthode à la segmentation du cerveau ou aux lésions de SEP est possible [Grimaud 1996, Rovaris 1997]. Cependant, ce type de méthodes reste le plus souvent utilisé dans le cadre d'une segmentation semi-automatique réalisée sous le contrôle d'un expert.

Un algorithme de croissance de région est une autre méthode de segmentation. Celui-ci se décompose en trois étapes. La première consiste à choisir un voxel ou une région de départ. La seconde correspond au fait d'ajouter les voxels adjacents au(x) voxel(s) sélectionné(s) si ceux-ci respectent un ou des critères préalablement définis. La dernière étape consiste à itérer la seconde tant que de nouveaux voxels sont inclus dans la région. Le ou les critères d'inclusion peuvent être basés par exemple sur l'intensité du voxel ou sur la distance avec le point d'initialisation. Ces critères peuvent aussi être mis à jour automatiquement au cours des itérations. Comme pour le seuillage, ces méthodes sont le plus souvent semi-automatiques (avec une initialisation réalisée manuellement) notamment sur des images de patients atteints d'une pathologie [Sandor 1991, Filippi 2001].

Le recalage d'un atlas sur une image peut permettre de segmenter des régions d'intérêts (ROI). Par exemple, le recalage d'un atlas composé d'une image céré-

brale moyenne et de la segmentation associée du cervelet se décompose en deux étapes. Premièrement, la transformation permettant le passage du repère de l'image moyenne à celui de l'image à segmenter est calculée. Deuxièmement, cette transformation est appliquée à la segmentation du cervelet, donnant la correspondance de cette segmentation sur l'image à segmenter. Cette méthode souffre de la variabilité inter-sujets même si le recalage peut se faire suivant différents degrés de liberté (cf. section 3.2.1). Ces méthodes peuvent être totalement automatisées [Bondiau 2005, Commowick 2008]. Cependant, dans le cas de patients atteints de SEP, la présence de lésions peut nuire au recalage. De plus, le caractère diffus de la maladie empêche de construire un atlas précis des lésions.

Le principe des modèles déformables se décompose en deux étapes : choix et application sur l'image d'une forme (*i.e.* cercle, sphère), déplacement du modèle en fonction d'un ensemble de règles (*i.e.* minimisation d'une énergie) jusqu'à arrêt de cette évolution. Malgré la variabilité des replis du cortex, des méthodes de segmentation de tissus cérébraux utilisent ce principe [Pitiot 2003, Rousson 2004, Hu 2007]. Toutefois, leur mise en œuvre est souvent complexe et nécessite des temps de calculs conséquents. Les modèles déformables nécessitent aussi le plus souvent un *a priori* sur la forme de l'objet à segmenter.

Le second groupe de méthodes suppose de connaître à l'avance le nombre de parties de l'image, de catégories dans lesquelles classer les voxels de l'image. Ces méthodes réalisent donc une classification. Elles nécessitent une connaissance *a priori* du contenu de l'image.

Les réseaux de neurones sont des outils de classification mimant les réseaux de neurones biologiques. Ils sont composés d'unités, appelées neurones, reliées entre elles. La difficulté est de définir les règles fixant l'effet d'un neurone sur le neurone auquel il est relié. Ceci se fait principalement grâce à des données d'apprentissage, données pour lesquelles la réponse en sortie de la classification est connue. Dans le domaine médical, celles-ci sont le plus souvent obtenues par l'intervention d'un expert qui réalise manuellement la classification souhaitée sur ces données. Pour ces raisons, de telles méthodes de classification sont dites supervisées. Elles ont déjà été utilisées pour segmenter les tissus cérébraux [Wang 1998b, Song 2007] ou même directement les lésions de SEP [Alonge 2001, Admasu 2003]. Toutefois, leurs mises en œuvre ne restent possible que dans le cas d'une base de données d'importance suffisante pour pouvoir obtenir un maximum de données d'apprentissage.

Il existe ensuite de nombreux algorithmes de classification. Ceux-ci peuvent être paramétriques ou non paramétriques. Les algorithmes non paramétriques ne supposent aucune répartition statistique des données. Par exemple, l'algorithme du plus proche voisin classe chaque voxel dans la même classe que la donnée d'apprentissage la plus proche de lui, au sens d'un critère préalablement défini. Cet algorithme a ensuite été généralisé au k-plus proches voisins. Le voxel est alors classé dans la même classe que la majorité des k données d'apprentissage les plus proches de lui [Anbeek 2004a, Anbeek 2004b, Wu 2006].

Lorsque des données d'apprentissage ne sont pas disponibles, les voxels peuvent

être regroupés entre eux en itérant une étape de segmentation et une étape de redéfinition des classes en fonction des segmentations précédemment obtenues. Par exemple, l'algorithme des k-moyennes associe chaque voxel à la classe dont la moyenne des intensités est la plus proche de la sienne. Puis, en fonction de la classification obtenue les moyennes de chaque classe sont recalculées. Généralisant cette approche dans le cadre de la logique floue, l'algorithme des c-moyennes floues [Dunn 1973, Bezdek 1999] associe à chaque voxel une mesure floue, "de probabilité", d'appartenance aux différentes classes. Cette mesure est basée sur l'écart entre l'intensité du voxel et l'intensité moyenne de la classe. Dans le cadre de la logique floue, les résultats obtenus ne sont plus déterministes, ceux-ci sont donc souvent binarisés avant de poursuivre l'analyse des données. Il existe de très nombreuses techniques d'analyse d'images médicales basées sur l'algorithme des c-moyennes floues [Pham 1999, Boudraa 2000, Alonge 2001, Admiraal-Behloul 2005, Kawa 2007].

D'autres algorithmes supposent une distribution statistique des données. Ils sont dits paramétriques. Parmi ceux-ci, l'algorithme d'espérance maximisée (EM) [Dempster 1977] est un algorithme basé sur le même principe que celui des algorithmes des k-moyennes et des c-moyennes floues mais supposant une répartition des données (très souvent un mélange de gaussiennes). Des contraintes sur la position du voxel peuvent aussi être incluses dans l'EM avec par exemple des champs de Markov [Kindermann 1980]. Sa robustesse peut être améliorée par l'ajout de connaissances *a priori*. En imagerie médicale cérébrale, cet algorithme permet de segmenter les images en fonction de critères basés sur l'intensité et la localisation des voxels [Wells 1995, Van Leemput 1999b]. Une correction des inhomogénéités de l'image [Van Leemput 1999a, Zhang 2001] (cf. section 3.2.2) et la prise en compte des volumes partiels peuvent aussi y être incluses [Leemput 2001, Dugas-Phocion 2004a]. Des connaissances *a priori* peuvent être apportées par un atlas. Ce dernier garantissant la robustesse de l'algorithme à la condition qu'il soit représentatif de l'image traitée. Enfin, dans le cadre d'IRM de patients atteints de SEP, le modèle statistique utilisé nous permettra d'exclure des segmentations les lésions en tant que points aberrants du modèle [Van Leemput 2001, Dugas-Phocion 2004b].

Pour toutes ces raisons, l'algorithme EM est un excellent choix pour segmenter les tissus cérébraux. Nous l'avons donc choisi dans notre étude. Les segmentations produites seront utiles pour nos mesures. Le calcul de leur volume permettra de calculer le volume cérébral ainsi que d'évaluer l'atrophie. Les propriétés des classes segmentées (*i.e.* moyenne, écart type) fourniront les informations nécessaires à la segmentation des lésions.

3.3.2 L'algorithme d'espérance maximisée (EM)

L'EM a été introduit par Dempster dans [Dempster 1977] suite aux travaux de Hartley [Hartley 1958]. Il permet d'optimiser les paramètres d'un modèle en fonction d'une loi de vraisemblance de ces paramètres par rapport aux données traitées. Dans la version la plus simple de l'EM, cette optimisation se réalise en itérant deux étapes. La première, dite d'espérance, calcule l'attachement des don-

nées au modèle. La seconde, dite de maximisation, modifie les paramètres du modèle en fonction des données et de leur attachement au modèle. La convergence de cet algorithme est prouvée. Le lecteur cherchant la preuve de celle-ci peut se référer au papier de Dempster [Dempster 1977] ainsi qu'aux documents suivants qui la reprennent [Flandin 2004, Dugas-Phocion 2006, Aït-Ali 2006]. De nombreux tutoriaux présentent aussi l'EM ainsi que ses différentes variantes [Dellaert 2002, Roche 2003, Borman 2004].

En définissant $\mathbf{X}$ une variable aléatoire de réalisations $x_1, x_2, \ldots, x_n$ et $\boldsymbol{\theta}$ l'ensemble des paramètres du modèle servant à estimer la distribution statistique de $\mathbf{X}$, la log-vraisemblance des paramètres s'écrit :

$$L(\boldsymbol{\theta}) = \log P(\mathbf{X}|\boldsymbol{\theta}) \tag{3.5}$$

L'estimation de $\boldsymbol{\theta}$ est donnée par l'équation suivante :

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} L(\boldsymbol{\theta}) \tag{3.6}$$

Comme le vecteur des réalisations de $\mathbf{X}$ ne correspond qu'à une observation partielle (puisque de nombre fini) du phénomène considéré, l'EM introduit un ensemble de variables cachées, $\mathbf{Y}$ pour prendre en compte ces données manquantes (*i.e.* l'appartenance des données au modèle). À partir de ces définitions et des démonstrations présentes dans les documents précédemment cités, on peut aussi prouver que la maximisation de $L(\boldsymbol{\theta})$ revient à maximer une fonction dépendante du problème considéré que nous noterons $\mathcal{L}(\mathbf{Y^{t+1}}, \boldsymbol{\theta})$. Avec ces notations, le principe général de l'EM est donné dans l'algorithme 1.

Algorithme 1 Algorithme EM, principe général

1: Initialisation des paramètres $\boldsymbol{\theta}^0$ à l'itération 0.
2: **répéter**
3: Étape d'espérance : calcul de $P^{t+1}(\mathbf{Y}) = P(\mathbf{Y}|\mathbf{X}, \boldsymbol{\theta}^t)$
4: Étape de maximisation : calcul de $\boldsymbol{\theta^{t+1}} = \arg\max_{\boldsymbol{\theta}} \mathcal{L}(\mathbf{Y^{t+1}}, \boldsymbol{\theta})$
5: **jusqu'à** convergence : $|\mathcal{L}(\mathbf{Y}, \boldsymbol{\theta^{t+1}}) - \mathcal{L}(\mathbf{Y}, \boldsymbol{\theta^t})| < \varepsilon$ où ε est un paramètre de l'algorithme.

Dans cet algorithme, la maximisation de $\mathcal{L}(\mathbf{Y^{t+1}}, \boldsymbol{\theta})$ selon $\boldsymbol{\theta}$ peut aussi se faire en maximisant la fonctionnelle $Q(\boldsymbol{\theta})$ [Flandin 2004] :

$$Q(\boldsymbol{\theta}) = \sum_{\mathbf{Y}} P^{t+1}(\mathbf{Y}) \log P(\mathbf{X}, \mathbf{Y}|\boldsymbol{\theta}) \tag{3.7}$$

Dans le cadre des IRM, la distribution des intensités des images peut être modélisée par un mélange de gaussiennes [Gudbjartsson 1995, Sijbers 1998]. Nous reprenons donc ici la description de l'EM dans ce cadre.

Soit K, le nombre de classes (gaussiennes) du modèle. Pour chaque classe, k, la distribution des intensités des voxels qui la composent est modélisée par une

gaussienne, G_{μ_k,Σ_k} de moyenne μ_k et de matrice de covariance Σ_k. Avec ce modèle, l'algorithme EM permet à partir des intensités des voxels, x_i et des probabilités *a priori*, π_i^k (d'appartenir à la classe k pour chaque voxel, i) de calculer les paramètres des gaussiennes (μ_k, Σ_k). Il permet aussi de calculer la probabilité *a posteriori*, γ_i^k, d'obtenir la labélisation $\mathbf{Y_i} = k$ sachant x_i, l'échantillon et $\boldsymbol{\theta_k}$, les paramètres de la classe, k. Dans ce cas, les variables cachées, Y, représentent l'appartenance du voxel à chacune des classes.

Par définition on a :

$$\gamma_i^k = P(\mathbf{Y_i} = k|\mathbf{X} = x_i; \boldsymbol{\theta_k}) \tag{3.8}$$

D'où, selon une loi de Bayes (avec $\boldsymbol{\Theta}$ l'ensemble des paramètres du modèle) :

$$\gamma_i^k = \frac{P(\mathbf{Y_i} = k)P(\mathbf{X} = x_i|\mathbf{Y}_i = k; \boldsymbol{\theta_k})}{P(\mathbf{X} = x_i; \boldsymbol{\Theta})} \tag{3.9}$$

$$= \frac{P(\mathbf{Y_i} = k)P(\mathbf{X} = x_i|\mathbf{Y}_i = k; \boldsymbol{\theta_k})}{\sum_{j=1}^{K} P(\mathbf{Y_i} = j)P(\mathbf{X} = x_i|\mathbf{Y}_i = j; \boldsymbol{\theta_j})} \tag{3.10}$$

$$\tag{3.11}$$

Dans le cas d'un mélange de gaussiennes :

$$P\big(\mathbf{X} = x|\mathbf{Y} = k; \boldsymbol{\theta_k}\big) = \frac{1}{(2\pi)^{d/2}|\Sigma_k|^{1/2}} e^{-\frac{1}{2}(x-\mu_k)^T\Sigma_k^{-1}(x-\mu_k)} \tag{3.12}$$

$$= G_{\mu_k,\Sigma_k}(x) \tag{3.13}$$

D'où :

$$\gamma_i^k = \frac{\pi_i^k G_{\mu_k,\Sigma_k}(x_i)}{\sum_{j=1}^{K} \pi_i^j G_{\mu_j,\Sigma_j}(x_i)} \tag{3.14}$$

L'équation 3.14 permet de calculer l'étape d'espérance. L'étape de maximisation s'obtient en dérivant la fonctionnelle $Q(\boldsymbol{\theta})$ (équation 3.7). Ce calcul disponible dans [Flandin 2004] permet de calculer les paramètres des classes (équations présentes à l'étape de maximisation dans l'algorithme 2).

L'algorithme EM dans le cas d'un modèle de mélange de gaussiennes est présenté dans l'algorithme 2. Comme l'initialisation se fait sur les probabilités *a priori* et non sur les paramètres des classes, les étapes d'espérance et de maximisation sont inversées.

3.3.3 Segmentation de tissus cérébraux et améliorations de l'EM dans ce cadre

Comme le montre les sections précédentes, l'algorithme EM est particulièrement intéressant dans le cadre de la segmentation d'images IRM cérébrales. Il repose

Algorithme 2 Algorithme EM, cas d'un mélange de gaussiennes

1: Initialisation des γ_i^k avec les π_i^k à l'itération 0
2: **répéter**
3: Étape de maximisation :

$$\begin{aligned} \pi_k &= \frac{1}{N}\sum_{i=1}^{N}\gamma_i^k \\ \mu_k &= \frac{\sum_{i=1}^{N}\gamma_i^k x_i}{\sum_{i=1}^{N}\gamma_i^k} \\ \Sigma_k &= \frac{\sum_{i=1}^{N}\gamma_i^k (x_i-\mu_k)(x_i-\mu_k)^T}{\sum_{i=1}^{N}\gamma_i^k} \end{aligned}$$

4: Étape d'espérance :

$$\gamma_i^k = \frac{\pi_k G_{\mu_k,\Sigma_k}(x_i)}{\sum_{j=1}^{K}\pi_j G_{\mu_l,\Sigma_j}(x_i)}$$

où K, le nombre de classes, est un paramètre de l'algorithme.
5: **jusqu'à** convergence : $|\mathcal{L}(\mathbf{Y},\boldsymbol{\theta^{t+1}}) - \mathcal{L}(\mathbf{Y},\boldsymbol{\theta^{t}})| < \varepsilon$ où ε est un paramètre de l'algorithme.

sur des bases mathématiques solides. Toutefois, les principaux problèmes des algorithmes de classification se posent aussi dans son cas [Saint-Jean 2001] :

- dépendance du résultat à l'initialisation,
- fixation manuelle de paramètres (*i.e.* nombre K de classes),
- convergence lente de l'algorithme,
- échec possible (*i.e.* matrices singulières),
- sensibilité aux données aberrantes.

Dans le but d'obtenir un algorithme fiable et robuste, j'ai cherché à minimiser ces difficultés.

3.3.3.1 Initialisation par un atlas, apport de connaissances *a priori*

Les personnes atteintes de SEP ont en majorité entre 20 et 40 ans. Dans cette tranche d'âge, il existe des variations de la physionomie du cerveau d'un sujet à l'autre mais celles-ci restent faibles, si on les compare à l'écart entre le cerveau d'un nourrisson et d'un adulte. L'apport d'informations *a priori* par un atlas est donc possible et a été utilisé à différentes reprises [Van Leemput 1999b, Dugas-Phocion 2006].

Comme dans les cas précédents, nous utiliserons un atlas mis à disposition par le *Montreal Neurological Institute*[1] (MNI) et construit à partir de 305 IRM [Evans 1992, Evans 1993] pour donner les probabilités *a priori* de chaque voxel d'être de la substance blanche (SB), de la substance grise (SG) ou du liquide céphalo-rachidien (LCR). Les variabilités d'un patient à l'autre étant limitées, le recalage de l'atlas sur nos images se fera de manière affine pour limiter le nombre de degrés de liberté (cf section 3.2.1).

[1] http ://www.mni.mcgill.ca/

3.3.3.2 Nombre de classes et modèle de volume partiel

Habituellement, une segmentation du cerveau se réalise à l'aide de trois classes : SB, SG, LCR. Cependant, ce modèle, en adéquation avec le nombre de compartiments cérébraux, ne colle pas parfaitement à la réalité des images IRM. Celles-ci sont en effet acquises à une certaine résolution, ce qui entraîne la présence de volumes partiels (cf. section 3.1.1). Un voxel peut alors correspondre à plusieurs tissus et son intensité s'éloigne donc de l'intensité moyenne des classes "pures" (classes de SB, SG et LCR qui seraient obtenues en ne prenant en compte que les voxels "purs" correspondant entièrement au compartiment considéré). Dans notre cas l'algorithme de classification va cependant faire correspondre ce voxel de volume partiel à principalement l'une des classes. Les paramètres de celle-ci vont alors être biaisés pour l'itération suivante.

En imagerie cérébrale, ce phénomène est particulièrement important à la frontière entre la SG et le LCR à cause des replis du cortex et de la minceur des sillons corticaux. En conséquence, des points correspondants à ces volumes partiels vont être inclus dans les classes de SG et de LCR, biaisant ainsi les paramètres de ces classes (déplacement de la moyenne et augmentation de l'écart type). Dans le cas de la SG, ce phénomène reste limité car de nombreux voxels purs correspondent à cette classe. En revanche le nombre de voxels purs de LCR est lui beaucoup plus limité. Le biais induit par les volumes partiels augmente donc grandement l'écart type de cette classe, conduisant ainsi à une surestimation du LCR par rapport à la SG. Cette imprécision induit des erreurs sur le calcul des volumes des différents compartiments (de 20 à 60% selon [Niessen 1997, González Ballester 1999]). Dans ce cas, l'évaluation de l'atrophie pourrait donc être, elle aussi, biaisée. Il est à noter que l'effet de volume partiel est en revanche limité concernant l'interface SB/SG parce que ces deux classes contiennent de nombreux voxels purs.

Pour répondre au problème précédent, je propose d'utiliser le modèle de volumes partiels présenté par G. Dugas dans [Dugas-Phocion 2004a, Dugas-Phocion 2006]. Ce modèle de volumes partiels entre la SG et le LCR considère que l'intensité d'un voxel de volume partiel composé de $100 * \alpha\%$ du tissu 1 (d'intensité I_1) et de $100 * (1-\alpha)\%$ du tissu 2 (d'intensité I_2), avec $0 \leq \alpha \leq 1$, s'obtient par l'équation :

$$I_{PVE} = \alpha I_1 + (1-\alpha) I_2 \tag{3.15}$$

Il faut choisir un nombre de classes de volume partiel (PVE) et fixer les valeurs du paramètre α. Selon [Dugas-Phocion 2006], un nombre correct de classes est 6. Nous utiliserons donc cette valeur. Les paramètres des classes PVE sont alors obtenus par les équations suivantes, où (μ_1, Σ_1) sont les paramètres de la classe "pure" du tissu 1 et réciproquement (μ_2, Σ_2) ceux de la classe "pure" du tissu 2 :

$$\mu_\alpha = \alpha\mu_1 + (1-\alpha)\mu_2 \tag{3.16}$$

$$\Sigma_\alpha = \alpha^2\Sigma_1 + (1-\alpha)^2\Sigma_2 \tag{3.17}$$

Ce modèle de volumes partiels est particulièrement intéressant dans notre cas. Il permettra un calcul correct du volume du cerveau (SB + SG) et du volume du LCR en vue d'une évaluation de l'atrophie.

3.3.3.3 Rapidité de convergence de l'algorithme

Le principe de base pour augmenter la vitesse de l'algorithme est de diminuer le nombre de calculs. Ceci est possible en limitant le nombre de points devant être classifiés (*i.e.* seuls les voxels compris dans le masque du cerveau sont utilisés). Il est aussi possible d'effectuer ces calculs uniquement sur un échantillon tiré aléatoirement des points de l'image. Le principe est de diminuer le nombre de points utilisés, et ainsi de calculs, pour définir les paramètres des classes. Une fois le critère de vraisemblance atteint, ces paramètres sont calculés ainsi que la probabilité *a posteriori* de tous les voxels de l'image d'appartenir à chacune des classes. Un paramètre, PPC, correspondant au pourcentage de points considérés pour calculer les paramètres des classes a donc été ajouté à l'EM. Ce principe permet donc un gain de temps. Cependant l'effet de ce paramètre, PPC, sur la qualité des segmentations doit être étudié.

Cette étude a été menée en collaboration avec Erik Pernod. Grâce au déploiement de l'algorithme de segmentation sur une grille de calcul (cf. section 7.2.2 et le rapport de stage d'Erik Pernod [Pernod 2008]), nous avons pu tester l'influence du paramètre PPC sur la segmentation de la SB. Cette segmentation a été choisie car elle est binaire (contrairement aux autres qui dépendent des volumes partiels) mais aussi et avant tout car elle est utilisée pour définir la zone d'intérêt de recherche des lésions.

Exécutions : Pour quantifier l'impact de ce paramètre sur la qualité des segmentations obtenues, l'algorithme de segmentation a été appliqué successivement avec différentes valeurs de PPC. Les segmentations obtenues ont alors été comparées en terme de sensibilité et de spécificité (cf. section 5.1.5) avec la segmentation de référence (obtenue pour PPC=100%). Ces expériences ont été réalisées sur des images IRM d'un patient atteint de SEP rémittente et sur un sujet témoin.

Il doit être noté que les voxels sont choisis aléatoirement dans l'image 3D. En conséquence, des résultats différents peuvent être obtenus avec une même valeur de PPC. Pour minimiser l'influence de cet effet, plusieurs exécutions ont été réalisées pour calculer les valeurs moyennes de sensibilité et de spécificité pour différents PPCs. La figure 3.5 montre ces moyennes ainsi que les variations observées en fonction du pourcentage de voxels considérés.

Discussion : Grâce au calcul du masque du cerveau, la segmentation ne se fait sur ces images que sur approximativement 830.000 voxels. La figure 3.5 montre que dans les deux cas (sujet témoin et patient atteint de SEP), nous obtenons des résultats similaires. La sensibilité diminue avec le pourcentage de points considérés. La spécificité est plus stable mais ces deux mesures sont de plus en plus variables.

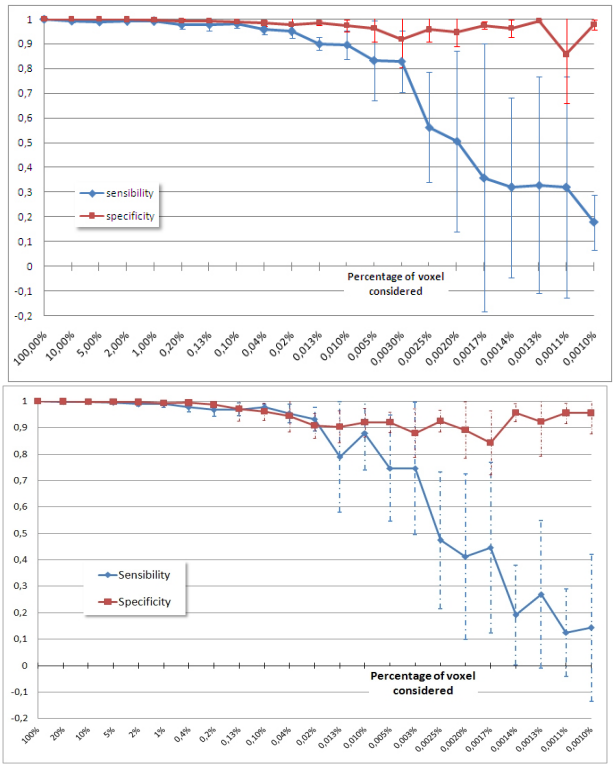
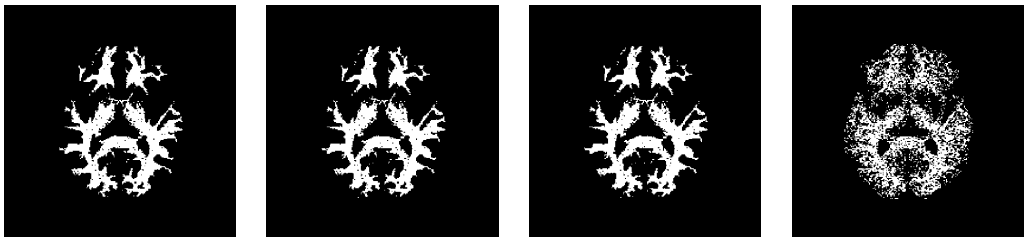


FIG. 3.5 – *Sensibilité moyenne et spécificité moyenne des segmentations de la SB en fonction du pourcentage de voxels considérés pour calculer les paramètres des classes. Les variations de ces moyennes sont aussi indiquées (chaque point correspond à la moyenne obtenue pour neuf exécutions de l'algorithme). En haut : sur un sujet témoin ; en bas : sur un patient atteint de SEP.*

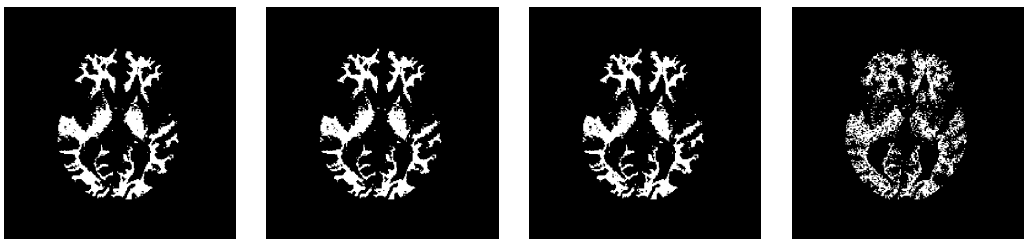
En effet, si l'algorithme considère moins de 1% des voxels du cerveau les résultats deviennent très variables.

La décroissance de ces deux mesures indique vraisemblablement qu'après un certain seuil, il n'y a plus assez de voxels correspondants à de la SB pour définir correctement les paramètres de la gaussienne modélisant cette classe. Les résultats de cette étude montrent qu'en prenant moins d'1% des voxels de l'image, les segmentations obtenues (au moins de la SB) ne sont plus fiables. Pour cette valeur critique, nous avons pu observer une division par 3 ou 4 du temps de calcul de l'EM. La figure 3.6 montre la segmentation de SB obtenue avec différentes proportions de points à partir des IRM du sujet témoin (en haut) et d'un patient atteint de SEP (en bas). Il reste donc à confirmer les valeurs trouvées par cette étude dans le cas des segmentations de SG et de LCR.



(a) (b) (c) (d)

Cas du sujet témoin



(a) (b) (c) (d)

Cas d'un patient atteint de SEP

FIG. 3.6 – *Segmentations de SB obtenues avec différentes proportions de points : (a) 100%, (b) 1%, (c) 0.1%, (d) 0.005%.*

3.3.3.4 Échecs, non inversibilité de la matrice de covariance

L'EM peut dans certains cas ne pas aboutir. C'est l'une des difficultés majeures que nous avons rencontrée. En effet, le calcul des probabilités *a posteriori* des voxels nécessite d'inverser la matrice de covariance de chacune des classes (équation 3.12). Cette matrice est symétrique et ses valeurs propres sont positives ou nulles mais son inversibilité n'est pas garantie. Dans le cas, où elle est non-inversible, singulière, il

faut trouver une solution pour obtenir les segmentations recherchées.

Dans le cas d'un algorithme EM, ayant un paramètre PPC servant à définir le nombre de points utilisés pour calculer les paramètres des classes du modèle (cf section précédente), le fait de relancer l'algorithme donne une nouvelle série de voxels pour ce calcul. La non-inversibilité de la matrice observée avec la première série de données peut ne pas se reproduire avec la seconde. L'algorithme peut donc finalement aboutir dans ce deuxième cas.

Dans le cas où tous les points sont pris en considération, cette possibilité n'existe plus. Une autre solution, inspirée de [Raudys 1998], que nous proposons dans ce cas, est de calculer la pseudo-inverse de Moore-Penrose.

Une matrice carrée, M, de taille $n \times n$ présente dans un espace ayant une base orthonormée, A, peut être décomposée en valeurs singulières, c'est-à-dire diagonalisée, et peut s'écrire $M = UEU^T$, où U est une matrice regroupant les vecteurs orthonormés servant à décrire A. E est la matrice diagonale des valeurs propres de M [Godement 2005].

$$M = U \begin{pmatrix} \lambda_1 & 0 & ... & 0 \\ 0 & \lambda_2 & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & \lambda_n \end{pmatrix} U^T \tag{3.18}$$

Selon la théorie du calcul de la pseudo-inverse de Moore-Penrose [Hogben 2007], simplifiée dans le cas d'une matrice carrée, la matrice inverse de M, M^{-1} s'écrit :

$$M^{-1} = U \begin{pmatrix} \alpha_1 & 0 & ... & 0 \\ 0 & \alpha_2 & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & \alpha_n \end{pmatrix} U^T \tag{3.19}$$

où

$$\begin{aligned} \alpha_i &= 0 \quad \text{si } \lambda_i = 0 \\ &= 1/\lambda_i \quad \text{sinon} \end{aligned} \tag{3.20}$$

La matrice M^{-1} est donc utilisée à la place de Σ_k^{-1} dans l'équation 3.12. Enfin la valeur du déterminant de Σ_k^{-1} est remplacée par le produit des valeurs propres, λ_i, non nulles.

Dans la pratique, la non-inversibilité de la matrice de covariance signifie que même si les données sont apportées par N images, elles sont en fait de dimension $N-1$ au plus. Le calcul de la pseudo-inverse revient alors à calculer l'inverse dans la variété de dimension inférieure sous-jacente.

Pour tester cet algorithme modifié, celui-ci a été appliqué sur une série de trois images composées de la séquence T2 et deux séquences DP. Si l'algorithme n'avait pas été modifié, le doublon de la séquence DP garantit de produire l'erreur de non-inversibilité de la matrice. Dans notre cas, l'algorithme modifié n'échoue pas. Nous

obtenons les mêmes segmentations que si nous n'avions utilisé que la séquence T2 et une seule séquence DP.

3.3.3.5 Prise en compte des points aberrants

Après l'étape de segmentation du cerveau (cf. section 3.2.3), les images IRM cérébrales sont composées principalement de trois compartiments : SB, SG, LCR mais pas seulement. Il y a en effet d'autres compartiments comme les vaisseaux sanguins. Ceux-ci, même si de volume négligeable par rapport à ceux des autres compartiments, peuvent gêner la classification. Une classe de points aberrants visant à les segmenter peut être ajoutée dans le modèle utilisé [Dugas-Phocion 2006].

La prise en compte de la présence de lésions de SEP peut se faire par différents moyens. Un des plus utilisés [Van Leemput 2001, Dugas-Phocion 2006, Aït-Ali 2006] est le calcul de la distance de Mahalanobis entre le vecteur d'intensité du point et la moyenne de chaque classe. Ceci permet d'extraire les données atypiques du modèle de manière simple et robuste. Pour cette raison, cette dernière méthode a été sélectionnée dans notre cas.

3.3.3.6 L'algorithme EM appliqué à la segmentation d'IRM cérébrale de patients atteints de SEP

L'algorithme modifié pour prendre en compte les différents critères précédemment expliqués est décrit dans l'algorithme 3.

3.4 Chaîne de traitement, EMA

Cette section présente les liens entre les divers outils (prétraitements, algorithme de segmentation) introduits précédemment. Ces liens décrivent une chaîne de traitement partant des images brutes et produisant les segmentations souhaitées.

3.4.1 Choix des données

Comme l'a présenté la section 2.2.2, différentes séquences IRM sont disponibles. Leur existence se justifie par le fait qu'elles permettent de voir des choses différentes (*i.e.* type de lésions) sur les images. Les contrastes entre les compartiments du cerveau y sont aussi différents. Toutefois, la résolution varie aussi entre les différentes séquences. La classification doit se baser sur les séquences de plus forte résolution pour limiter les influences des volumes partiels dus nativement à l'acquisition ou causés par le recalage inter-séquences.

Dans notre base de données, le couple de séquences T2/DP présente plusieurs avantages. Sa résolution est parmi les plus élevées. De plus, ces deux séquences sont intrinsèquement recalées. Il n'y a donc pas besoin de rééchantillonner ces images et aucun volume partiel n'est ajouté. Dans l'étude QUALICORE, la résolution des IRM T1 est supérieure à celle du couple T2-DP. Le recalage de cette séquence sur

Algorithme 3 Algorithme EM, appliqué à la segmentation d'IRM cérébrale de patients atteints de SEP

1: Tirage aléatoire de N voxels servant au calcul. N est fixé en fonction du paramètre PPC de l'algorithme.
2: Initialisation des γ_i^k par les informations données par l'atlas recalé à l'itération 0
3: **répéter**
4: Étape de Maximisation :
   - Calcul des paramètres des classes pures :

$$\begin{aligned} \pi_k &= \frac{1}{N}\sum_{i=1}^{N}\gamma_i^k \\ \mu_k &= \frac{\sum_{i=1}^{N}\gamma_i^k x_i}{\sum_{i=1}^{N}\gamma_i^k} \\ \Sigma_k &= \frac{\sum_{i=1}^{N}\gamma_i^k (x_i-\mu_k)(x_i-\mu_k)^T}{\sum_{i=1}^{N}\gamma_i^k} \end{aligned}$$

   - Calcul des paramètres des classes partielles de SG-LCR de proportion α de SG :

$$\begin{aligned} \mu_\alpha &= \alpha\mu_1 + (1-\alpha)\mu_2 \\ \Sigma_\alpha &= \alpha^2\Sigma_1 + (1-\alpha)^2\Sigma_2 \end{aligned}$$

5: Étape d'espérance : Calcul des γ_i^k pour les classes pures et les classes de volumes partiels
   $\gamma_i^k = \frac{\pi_k G_{\mu_k,\Sigma_k}(x_i)}{\sum_{j=1}^{K}\pi_j G_{\mu_l,\Sigma_j}(x_i)}$ où K=10 (4 classes pures et 6 classes de volumes partiels) et où Σ_k^{-1} est remplacée par la pseudo-inverse de Moore-Penrose si elle est singulière.
6: **jusqu'à** convergence : $|\mathcal{L}(\mathbf{Y}, \boldsymbol{\theta^{t+1}}) - \mathcal{L}(\mathbf{Y}, \boldsymbol{\theta^t})| < \varepsilon$ où ε est un paramètre de l'algorithme.
7: Calcul des points aberrants par la distance de Mahalanobis.

le T2 se fait en incluant peu de volumes partiels. En revanche, les séquences T2-FLAIR ou T1-GADO sont à une résolution inférieure ou similaire à celle du T2, le recalage de ces séquences sur le T2 augmentera donc l'effet de volume partiel qui y est observé.

Pour ces raisons, j'ai choisi de réaliser l'étape de classification à partir des séquences T1, T2 et DP.

3.4.2 Enchaînement des données

Notre segmentation se base sur trois séquences IRM : T1, T2 et DP (cf. section précédente). Dans notre cas, les séquences T2 et DP sont intrinsèquement recalées mais ce n'est pas le cas de la séquence T1 qui est donc recalée sur le T2. Comme ces séquences sont issues du même patient, ce recalage est effectué par une transformation rigide à l'aide de [Ourselin 2000] (voir section 3.2.1).

Le biais des images est corrigé à l'aide de [Prima 2001] (voir section 3.2.2). Le cerveau est extrait des images à l'aide des travaux présentés dans le chapitre 4. Cette dernière étape nécessitant un recalage affine de l'atlas du MNI sur les données, cet atlas recalé est gardé pour initialiser l'algorithme de classification.

L'algorithme EM amélioré (cf. section 3.3.3) est ensuite appliqué sur ces données prétraitées. Après convergence, dix segmentations probabilistes sont obtenues : SB, SG, LCR, 6 classes de volumes partiels SG/LCR (avec différentes proportions), et une classe de points aberrants (vaisseaux + lésions). Excepté pour les points aberrants déjà binarisés par le calcul de la distance de Mahalanobis, les segmentations finales sont obtenues en classant chaque voxel dans la classe la plus probable. Cette chaîne de traitement, que nous appellerons EMA dans la suite de ce manuscrit, est résumée par la figure 3.7.

Même si chaque étape de EMA a été étudiée lors de cette thèse, EMA reprend et améliore les étapes de la chaîne de traitement qui avait été proposée par G. Dugas dans [Dugas-Phocion 2006]. Cette précédente chaîne de traitement sera appelée EMP dans la suite de ce manuscrit. Par rapport à EMP, EMA se base sur trois séquences IRM (T1, T2, DP) et non uniquement le couple (T2, DP). Elle utilise une nouvelle méthode d'extraction du cerveau (cf. chapitre 4). Les limites de l'EM qu'elle utilise ont été étudiées et des réponses ont été apportées en cas de non-inversibilité de la matrice de covariance.

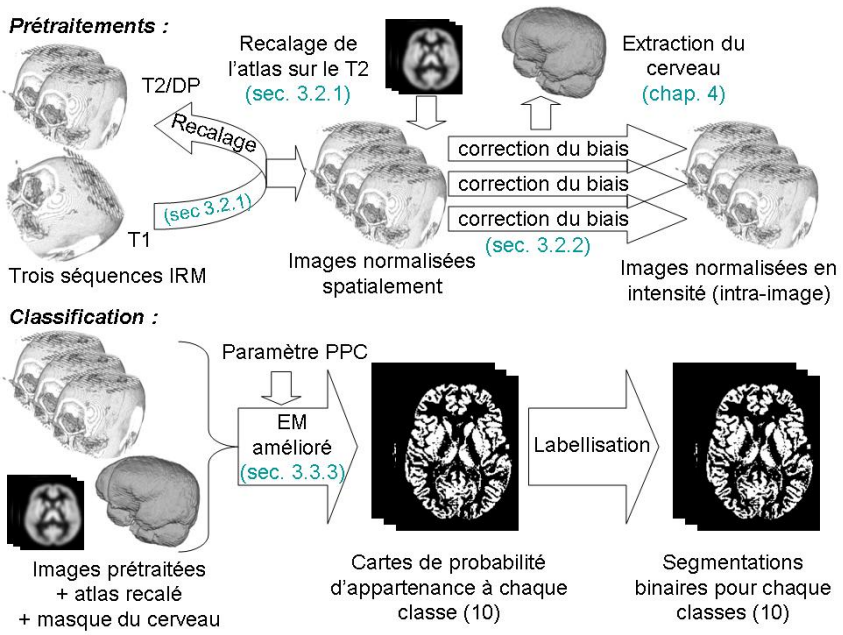


FIG. 3.7 – *Chaîne de traitement, EMA, pour la segmentation des compartiments cérébraux.*

Chapitre 4

Extraction du cerveau et régions d'intérêt

Sommaire

4.1 Méthodes d'extraction de l'encéphale 51
4.1.1 AnaT1toBrainmask (ATB) 51
4.1.2 Brain Extraction Tool (BET) 52
4.1.3 Hybrid Watershed Algorithm (HWA) 53
4.1.4 3dIntracranial (3DI) 54
4.1.5 EM-BrainMask (EBM) 55
4.2 Comparaison et nouvelle méthode 56
4.2.1 Cadre de l'étude : l'algorithme STAPLE 56
4.2.2 Comparaison des 5 méthodes sélectionnées et de 26 méthodes composées 58
4.3 Segmentation en sous régions d'intérêt 60

Résumé : Pour commencer, cinq méthodes d'extraction du cerveau sont présentées. Elles ont été choisies pour leur notoriété et leur disponibilité sous forme logicielle. L'algorithme STAPLE suivi d'un seuillage automatique permet de combiner les différentes segmentations binaires fournies par une sélection de ces méthodes pour produire une nouvelle segmentation binaire. Ce même algorithme permet aussi, à partir de l'utilisation des cinq méthodes, de produire une segmentation probabiliste de référence. La comparaison des 5 méthodes et des 26 combinaisons possibles par rapport à cette segmentation de référence permet de classer ces 31 algorithmes par ordre de performance. Ensuite, la prise en compte de différents critères (robustesse de la méthode, temps de calcul, ...) permet de sélectionner la méthode optimale. Dans notre cas, nous avons choisi une combinaison de trois méthodes (BET, EBM, 3dIntracranial).

L'étape précédente fournit un masque global de l'encéphale (union cervelet, cortex et tronc). Je propose alors une nouvelle méthode pour découper ce dernier dans ces trois régions d'intérêt. Premièrement, un atlas contenant un *a priori* sur ces régions est recalé de manière multi-affine sur l'image. Chaque région est alors représentée par une segmentation binaire mais certains points du masque global ne sont pas classés. Ceux-ci sont inclus par une croissance par zone d'influence des trois régions dans le masque global. L'algorithme permet alors d'obtenir les trois régions d'intérêt recherchées. Cet algorithme est amélioré pour pouvoir segmenter le tronc sur des images IRM allant jusqu'au bas du cou.

Mots clefs : Segmentation du cerveau, STAPLE, Segmentation du cortex, Segmentation du cervelet, Segmentation du tronc.

Contributions :

- Mise en place d'un cadre pour combiner des segmentations binaires (utilisation de STAPLE + seuillage automatique).
- Test de cinq méthodes de segmentation du cerveau + 26 combinaisons pour des images de patients atteints de SEP.
- Sélection d'une méthode combinée comme méthode de référence.
- Une nouvelle méthode de segmentation du cerveau en trois régions d'intérêt.

Organisation du chapitre : Ce chapitre présente 5 différentes méthodes d'extraction du cerveau dans la section 4.1. Puis, la section 4.2 présente la comparaison de ces méthodes ainsi qu'une manière originale de les combiner en utilisant le cadre de l'algorithme STAPLE. Finalement, un algorithme novateur est proposé dans la section 4.3 pour segmenter le cerveau en trois régions d'intérêt : cortex, tronc et cervelet.

4.1 Méthodes d'extraction de l'encéphale

Comme présenté dans la section 3.2.3, l'extraction de l'encéphale consiste à retirer les autres tissus (peau, yeux, os, ...) de l'image. Cette étape enlève les voxels ne correspondant pas au modèle utilisé pour la segmentation et limite en conséquence le nombre de points aberrants. Les classes sont ainsi mieux définies et les segmentations supposées de meilleure qualité. En conséquence, cette étape semble primordiale notamment pour l'évaluation de l'atrophie.

Elle peut se faire manuellement mais avec un coût en temps important (15 minutes à 2 heures par volume 3D). Le technicien doit aussi être expérimenté et maintenir son attention pendant l'exercice (pour ne pas, par exemple, faire d'erreur entre le bas du cervelet et les veines). L'utilisation d'une méthode complètement automatique et robuste semble donc incontournable.

La littérature offre de nombreuses méthodes automatiques d'extraction du cerveau [Sandor 1997, Ward 1999, Shattuck 2001, Segonne 2004, Zhuang 2006]. Certaines existent sous forme de logiciels disponibles sur internet. Cependant, ces algorithmes ont été testés dans un cadre particulier et sur des données dont le protocole d'acquisition peut différer de celui utilisé dans notre étude. Une étude comparative de ces méthodes sur nos données est donc nécessaire.

Cette section présente les méthodes d'extraction de l'encéphale que nous avons sélectionnées pour notre étude comparative. Ce choix s'est fait selon des critères de notoriété de l'algorithme, son utilisation dans des cas similaires au nôtre mais aussi sur la disponibilité de ce dernier sous forme logicielle.

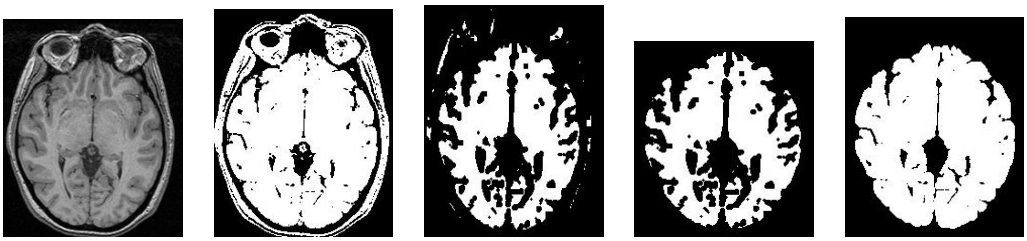
4.1.1 AnaT1toBrainmask (ATB)

AnaT1toBrainmask[1] (ATB) est une chaîne de traitement proposée par JF. Mangin [Mangin 1998] du Laboratoire de Neuroimagerie Assistée par Ordinateur[2] et présente dans la plateforme logiciel Brainvisa[3]. Le principe de Brainvisa est de fournir un cadre simple pour l'utilisation de différentes chaînes de traitements sur une série d'images. La majorité des algorithmes ainsi disponibles vise à la segmentation de séquences IRM T1.

AnaT1toBrainmask (ATB) génère un masque du cerveau à partir de cette séquence. Son principe se découpe en différentes étapes (cf. figure 4.1) :

- binarisation de l'image en appliquant un seuil,
- érosion de l'image binaire pour obtenir plusieurs morceaux déconnectés,
- sélection du plus gros d'entre eux en tant que graine du cerveau,
- croissance de cette graine pour récupérer la forme du cerveau.

[1] http ://brainvisa.info/doc/brainvisa-3.1/en/processes/AnaT1toBrainMask.html
[2] http ://lnao.fr/
[3] http ://brainvisa.info



(a) (b) (c) (d) (e)

FIG. 4.1 – *Étapes de la méthode ATB : (a) séquence T1, (b) binarisation, (c) érosion, (d) sélection de la plus grande zone connexe, (e) croissance de cette zone (source : http ://brainvisa.info/doc/brainvisa-3.1/en/processes/AnaT1toBrainMask.htm, remerciements à JF. Mangin pour son autorisation).*

4.1.2 Brain Extraction Tool (BET)

La méthode BET (Brain Extraction Tool) a été développée par Stephen M. Smith en 2002 au centre d'IRM fonctionnelle du cerveau d'Oxford [Smith 2002b]. Elle fait partie du package FSL[4] de logiciels de ce centre.

Les étapes de la méthode BET sont les suivantes (cf. figure 4.2) :

- segmentation approximative du cerveau grâce à l'application d'un seuil obtenu par l'histogramme des intensités de l'image,
- une sphère maillée en triangles et de rayon approximatif de la moitié de celui du cerveau est positionnée au centre de gravité de ce dernier,
- les contours de la sphère sont alors déplacés en direction de la limite du cerveau en respectant des critères de lissage et de seuillages locaux en intensité.

BET a été comparée à d'autres méthodes dans différents articles. BET s'est différenciée par une très bonne corrélation entre les volumes de ses segmentations et des segmentations manuelles obtenues sur les mêmes données [Hartley 2006]. BET permet de classer différents cerveaux par taille. Selon cette même étude, BET a aussi très peu de faux négatifs, ce qui garantit l'inclusion du cerveau dans le masque obtenu. Toutefois, BET a aussi tendance à inclure dans le masque des tissus non cérébraux notamment dans le cas d'IRM de sujets âgés [Fennema-Notestine 2006].

Note : Une nouvelle version de BET a été développée depuis cette étude. Celle-ci est une fusion du logiciel BET2 (reprogrammation du logiciel BET en C++) et du logiciel BETSURF calculant la surface interne et externe du crâne depuis une séquence T1 et une séquence T2.

[4] http ://www.fmrib.ox.ac.uk/fsl/

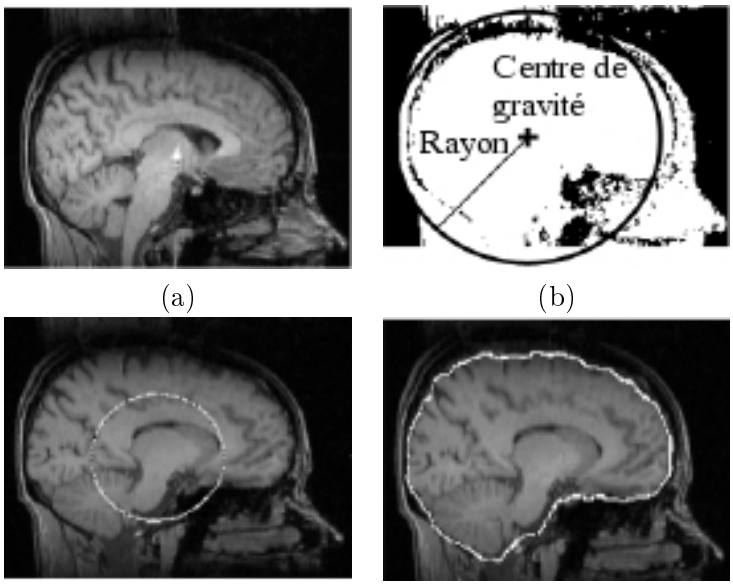


FIG. 4.2 – *Étapes de la méthode BET : (a) séquence T1, (b) binarisation (seuil basé sur l'histogramme) et calcul du centre de gravité et du rayon d'une sphère équivalente au volume, (c) initialisation de la surface, (d) surface finale (source ; [Smith 2002b], remerciements à S. Smith pour son autorisation).*

4.1.3 Hybrid Watershed Algorithm (HWA)

L'Hybrid Watershed Algorithm est une méthode de 2004 [Segonne 2004] intégrée au logiciel FreeSurfer[5] de l'université d'Harvard. Cette méthode combine un algorithme de ligne de partage des eaux à un modèle déformable de surface. L'algorithme de partage des eaux est basé sur les intensités de l'image et fonctionne en supposant la connexité de la matière blanche. Il donne alors une première segmentation du cerveau mais qui contient des imperfections. Pour les corriger, un modèle déformable de surface est appliqué.

HWA doit être appliquée sur une séquence T1. Les étapes qui la composent sont les suivantes.

- **Localisation d'un voxel de la matière blanche** : Les étapes préliminaires de BET sont réutilisées ici pour définir un point de la SB. Le centre de gravité du cerveau est calculé. Le centre d'un cube est alors placé sur ce point. A partir d'informations données par l'histogramme de la région d'intérêt définie par ce cube, des paramètres (moyenne, écart type) de la SB sont calculés et un voxel correspondant à cette classe est choisi.
- **Application de la ligne de partage des eaux** : Premièrement, la séquence T1 est inversée. La SB qui était en hypersignal dans l'image est donc mainte-

[5] http ://surfer.nmr.mgh.harvard.edu/

nant en hyposignal. De même, La SG et le LCR qui entourent cette dernière et qui étaient en hyposignal sur l'image sont maintenant en hypersignal. Si l'on considère l'intensité des points comme étant la hauteur du voxel dans le plan de l'image, la SB devient donc une vallée entourée de montagnes. Le principe de partage des eaux permet alors de regrouper les voxels en bassins. Ces bassins sont initiés par la détection des minima locaux. Le bassin comprenant le point de SB précédemment défini est une première segmentation du cerveau.

– **Application du modèle déformable** : Le volume obtenu à l'étape précédente est utilisé pour initialiser un modèle déformable. Celui-ci est déformé selon trois forces : une force interne basée sur la courbure, une force externe dépendant de l'intensité du voxel et une force contraignant la forme du modèle par rapport à celle donnée par un atlas.

HWA est, selon [Fennema-Notestine 2006], l'une des méthodes les plus prometteuses. Certains voxels ne faisant pas partie du cerveau sont tout de même inclus dans le masque.

4.1.4 3dIntracranial (3DI)

3dIntracranial a été développée par Ward en 1999 [Ward 1999] et est incluse dans le logiciel Analysis of Functional NeuroImage[6] (AFNI) des instituts nationaux américains de la santé (*National Institutes of Health*).

Les étapes de la méthode 3DI sont les suivantes :

– **Estimation par densité de probabilité** : L'histogramme de l'image est identifié comme étant la composition de trois gaussiennes : SB, SG et fond (cf. équation 4.1).

$$f(x) = k_{SB} * G_{\mu_{SB},\sigma_{SB}}(x) + k_{SG} * G_{\mu_{SG},\sigma_{SG}}(x) + k_{fond} * G_{\mu_{fond},\sigma_{fond}}(x) \quad (4.1)$$

avec :
– k_X = coefficient de la classe X,
– μ_X = moyenne de la classe X,
– σ_X = écart type de la classe X,
– $G_{\mu_X,\sigma_X}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$.

Ces 9 paramètres sont estimés par une méthode d'optimisation non linéaire : l'algorithme du simplexe. Celui-ci approxime par moindres carrés l'histogramme de l'image à la densité de probabilité du modèle des trois gaussiennes. Ces paramètres doivent être réestimés pour chaque image.
On seuille ensuite l'image pour garder les voxels du cerveau ($min = \mu_{SG} - 2\sigma_{SG}$, $max = \mu_{SB} + 2\sigma_{SB}$).

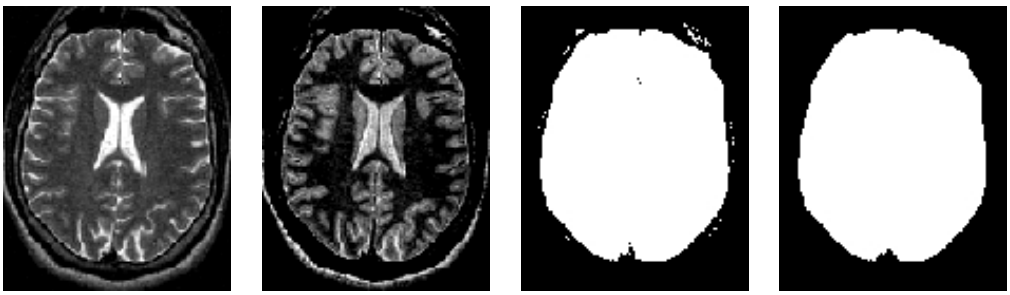
[6] http ://afni.nimh.nih.gov/afni/

- **Segmentation coupe par coupe** : En commençant par la coupe axiale centrale, l'ensemble E des voxels du cerveau est considéré. E est en général non connexe, on extrait donc le plus grand sous ensemble connexe C de E. Si C a des trous, ils sont remplis pour donner l'ensemble R. Ce traitement est renouvelé sur les coupes adjacentes préalablement masquées par R' qui correspond à une dilatation de R. Comme nous recherchons une connectivité 3D ce traitement est renouvelé en sagittal et en coronal.
- **Création de l'enveloppe 3D** : À partir des résultats précédents, une enveloppe 3D du cerveau est créée. Elle est enregistrée dans un tableau T tel que $T(\theta, \phi)$ est la distance maximale du centre du cerveau aux voxels cérébraux selon la direction (θ, ϕ), à partir du centre du cerveau. T est ensuite lissé par moyennage local en T'. T' est une nouvelle enveloppe du cerveau. Les voxels dont le milieu y est inclus sont considérés comme appartenant au cerveau.
- **Connexité** : La dernière étape élimine les voxels du cerveau ayant peu de voisins (au sens de la 6 connexité) et rajoute ceux du fond ayant beaucoup de voisins labelisés comme étant des voxels du cerveau.

De même que BET, la méthode 3DI a été décrite comme moins sensible que HWA et ayant tendance à inclure dans le masque des tissus non cérébraux, notamment dans le cas d'IRM de sujets âgés [Fennema-Notestine 2006]

4.1.5 EM-BrainMask (EBM)

Cette méthode a été mise en place et utilisée par G Dugas [Dugas-Phocion 2004b, Dugas-Phocion 2006] et est similaire à [Kapur 1996]. Un algorithme EM à quatre classes (SB, SG, LCR, autre) est appliqué sur le couple de séquences T2/DP. Ceci fournit une première segmentation du cerveau. L'application d'outils de morphologie mathématique permet alors d'obtenir la région d'intérêt recherchée. Cette dernière correspond à la plus grande composante connexe de l'image dont les cavités ("trous") ont été comblées (cf. figure 4.3).



(a) (b) (c) (d)

FIG. 4.3 – *Nécessité des outils de morphologie mathématique dans la méthode EBM : (a) séquence T2, (b) séquence DP, (c) masque obtenu sans morphologie mathématique, (d) masque après morphologie mathématique (source [Dugas-Phocion 2006]).*

La méthode EBM a été utilisée par G. Dugas pour analyser une série d'images IRM de patients atteints de SEP avec un protocole d'acquisition proche du nôtre. Selon [Dugas-Phocion 2006], de bons résultats ont été obtenus. Cependant, cet auteur indique que "des sur-segmentations dans les zones grasses entre le parenchyme et le crâne sont possibles" ainsi que des sous-segmentations dans les autres zones.

4.2 Comparaison et nouvelle méthode

Nous recherchons une méthode de skull-stripping fiable et produisant des résultats acceptables (qualifiés de corrects par un neurologue) sur l'ensemble de notre base d'IRM de patients atteints de SEP. La section précédente en a sélectionné cinq. Certaines propriétés (sur- ou sous- segmentations) de ces méthodes ont pu être déduites de la littérature [Hartley 2006, Fennema-Notestine 2006]. Cependant, aucune méthode ne semble s'imposer comme un "gold-standard". De plus, excepté pour EBM, ces méthodes n'ont que peu été étudiées dans le cadre de la SEP, à notre connaissance. Une étude comparative sur nos données doit donc être menée.

Dans cette étude, les méthodes ont été utilisées à partir de la suite logicielle dont elles dépendent et avec leurs paramètres par défaut. Concernant HWA, son utilisation s'est faite partiellement. En effet, le logiciel nous renvoyait une erreur sur nos données à partir de l'étape de modèle déformable. La cause de cette erreur semble être due à la résolution des images, même si d'autres possibilités existent. Toutefois la méthode basée uniquement sur la méthode de lignes de partage des eaux donne des résultats préliminaires que nous avons souhaité inclure dans cette étude.

4.2.1 Cadre de l'étude : l'algorithme STAPLE

L'algorithme STAPLE a été proposé par S. Warfield dans [Warfield 2004]. Grâce à l'utilisation d'un algorithme EM (cf. section 3.3.2), il fournit, à partir de segmentations binaires d'entrée, une segmentation probabiliste (cf. figure 4.4) ainsi que des paramètres de qualité de chaque segmentation d'entrée.

Ces paramètres correspondent à des mesures de qualité très employées pour la comparaison de segmentations ou pour la validation de tests : la sensibilité et la spécificité. Ces valeurs sont comprises entre 0 et 1. Une sensibilité proche de 1 signifie que la grande majorité des voxels appartenant à la segmentation de référence sont aussi segmentés sur la segmentation à comparer. Une spécificité proche de 1 indique que le masque de la segmentation à comparer contient peu de voxels qui n'appartiennent pas à celui de la segmentation de référence. La segmentation est donc jugée de meilleure qualité lorsque ces valeurs sont conjointement proches de 1.

Le paragraphe précédent a introduit la signification des critères de mesure de la qualité des segmentations obtenues. Le lecteur intéressé par le calcul de ces valeurs peut se référer au chapitre 5 (page 79) pour leurs calculs dans le cas de deux segmentations binaires et au chapitre 6 (page 116) ou à [Commowick 2007] dans le cas où l'une des segmentations est probabiliste.

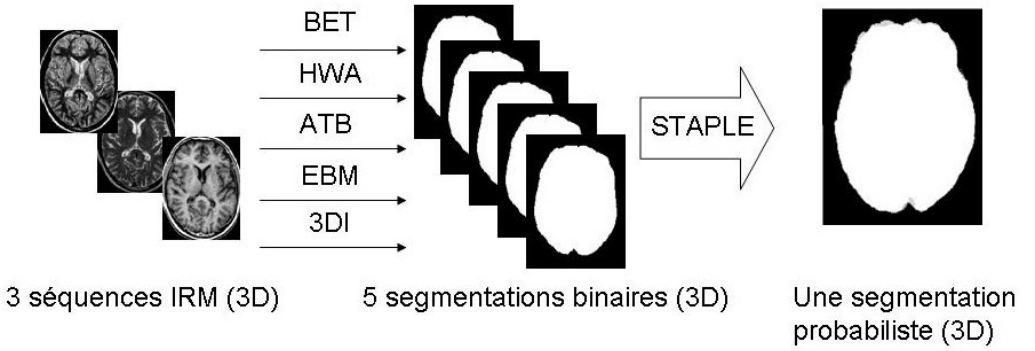


FIG. 4.4 – *Réalisation d'une segmentation probabiliste de référence avec STAPLE.*

Dans la suite de cette section, les valeurs de sensibilité et de spécificité seront regroupées au sein d'un même critère appelé distance moyenne : moyenne des distances, D, entre un couple (sensibilité, spécificité) et le point de valeur (1,1) (cf. équation 4.2).

$$D = \sqrt{(1 - Sen)^2 + (1 - Spe)^2} \tag{4.2}$$

La segmentation probabiliste obtenue par STAPLE peut soit être seuillée pour fournir une nouvelle segmentation binaire (une segmentation composée des segmentations d'entrée) soit servir de segmentation de référence par rapport à laquelle on peut comparer les segmentations d'entrée ou d'autres segmentations.

Dans le cas où STAPLE est utilisé pour composer des segmentations (cf. figure 4.5), j'ai proposé un calcul automatique du seuil à appliquer sur la segmentation probabiliste. Différents seuils sont appliqués. Celui pour lequel la distance moyenne entre la segmentation binaire et la segmentation probabiliste est la plus élevée, est gardé.

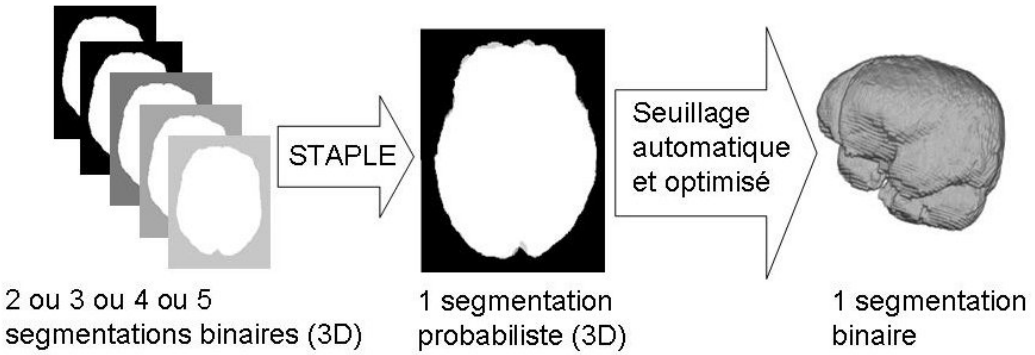


FIG. 4.5 – *Combinaison de segmentations binaires avec STAPLE.*

STAPLE permet aussi de comparer des segmentations. Premièrement, comme indiqué dans les paragraphes précédents, l'algorithme fournit pour chaque segmentation binaire d'entrée sa sensibilité et sa spécificité par rapport à la segmentation probabiliste obtenue. Deuxièmement, d'autres segmentations binaires peuvent aussi être comparées à cette segmentation probabiliste. La segmentation probabiliste joue alors le rôle de segmentation de référence.

4.2.2 Comparaison des 5 méthodes sélectionnées et de 26 méthodes composées

En utilisant le cadre de STAPLE, nous avons comparé les cinq méthodes précédemment présentées : BET, HWA, AnaT1toBrainMask (ATB), EM-BrainMask (EBM) et 3dIntracranial (3DI) sur 30 séries de séquences IRM (T1, T2, DP).

Les couples (sensibilité, spécificité), (Sen, Spe), de chacune des combinaisons étudiées sont donnés dans la table 4.1. Ceux-ci vont de (0.948-0.368) à (0.985-0.993) et sont classés en fonction d'un critère, appelé distance moyenne : moyenne des distances D entre un couple (sensibilité, spécificité) et le point de valeur (1,1) (cf. équation 4.2).

L'utilisation de HWA sans l'étape de modèle déformable conduit à une sur-segmentation, expliquant la faible spécificité de la méthode dans ce cas. Les propriétés de BET trouvées par [Hartley 2006] et [Fennema-Notestine 2006] sont ici confirmées. En effet une forte sensibilité révèle qu'il y a peu de faux négatifs et que le cerveau est bien inclus dans le masque. La spécificité plus faible de BET confirme que le masque peut encore contenir des voxels ne correspondant pas au cerveau. EBM n'avait jamais été comparée à d'autres méthodes. Cette étude révèle que cette méthode possède peu de faux négatifs mais qu'elle inclut des tissus extérieurs au cerveau (spécificité plus faible). À l'inverse ATB est plus spécifique que sensible. Elle garantit donc que les voxels du masque qu'elle produit correspondent bien au cerveau mais pas l'inclusion de tous les voxels ayant cette même propriété. 3DI se caractérise par une sensibilité et une spécificité correctes. Toutefois, certaines segmentations binaires obtenues par la combinaison de plusieurs méthodes permettent de s'approcher encore mieux de la segmentation probabiliste de référence. La méthode la plus proche de cette dernière est donc la segmentation binaire issue de la combinaison des cinq méthodes. Ce résultat peut sembler logique mais il était difficile à affirmer avant cette étude. Par exemple, il pourrait sembler logique que la combinaison de deux méthodes soit meilleure que de prendre une seule méthode or la table 4.1 montre que ce n'est pas toujours le cas.

Considérant des informations supplémentaires (temps moyens d'exécution, facilité d'installation du logiciel, robustesse, ...), la combinaison des méthodes (BET, EM-BrainMask, 3dIntracranial) avec (0.980-0.951) a été sélectionnée dans notre cas. Celle-ci sera appelée BE3 dans la suite de ce manuscrit. Les résultats de BE3 sur notre base de données ont été validés par un expert : Christine Lebrun, neurologue au CHU de Nice. Nous avons ainsi obtenu la méthode recherchée, étape préliminaire et essentielle pour une mesure de la charge lésionnelle et de l'atrophie. En se servant

TAB. 4.1 – *Comparaison entre 31 segmentations binaires (5 méthodes + 26 combinaisons) et une segmentation probabiliste de référence.*

Méthodes	Sensibilité Moyenne	Spécificité Moyenne	Distance Moyenne	Rang
EBM-BET-3DI-HWA-ATB	0.985	0.993	0.018	1
EBM-BET-3DI-ATB	0.971	0.997	0.030	2
EBM-3DI-HWA-ATB	0.970	0.986	0.035	3
BET-3DI-HWA-ATB	0.982	0.971	0.039	4
BET-3DI-ATB	0.982	0.969	0.042	5
EBM-3DI-ATB	0.963	0.986	0.043	6
EBM-BET-3DI	0.980	0.951	0.060	7
EBM-3DI	0.954	0.974	0.064	8
BET-3DI	0.989	0.938	0.067	9
EBM-BET-HWA-ATB	0.987	0.920	0.082	10
EBM-BET-3DI-HWA	0.991	0.914	0.087	11
EBM-BET-ATB	0.982	0.919	0.088	12
EBM-BET	0.958	0.926	0.090	13
3DI-HWA-ATB	0.952	0.940	0.100	14
3DI	0.943	0.930	0.122	15
3DI-HWA	0.904	0.948	0.129	16
3DI-ATB	0.938	0.926	0.131	17
EBM-3DI-HWA	0.966	0.873	0.138	18
EBM-HWA-ATB	0.966	0.864	0.143	19
ATB	0.867	0.961	0.149	20
EBM-HWA	0.920	0.873	0.154	21
EBM-BET-HWA	0.976	0.849	0.156	22
EBM	0.961	0.848	0.161	23
EBM-ATB	0.854	0.982	0.161	23
BET-HWA	0.949	0.841	0.172	25
BET-3DI-HWA	0.975	0.842	0.176	26
BET-HWA-ATB	0.995	0.823	0.178	27
BET	0.995	0.814	0.186	28
BET-ATB	0.979	0.824	0.194	29
HWA-ATB	0.838	0.763	0.389	30
HWA	0.948	0.368	0.636	31

des atouts de chacune des méthodes, BE3 semble plus robuste. Son utilisation donne de meilleurs résultats que EBM pour certains patients (cf. figure 4.6).

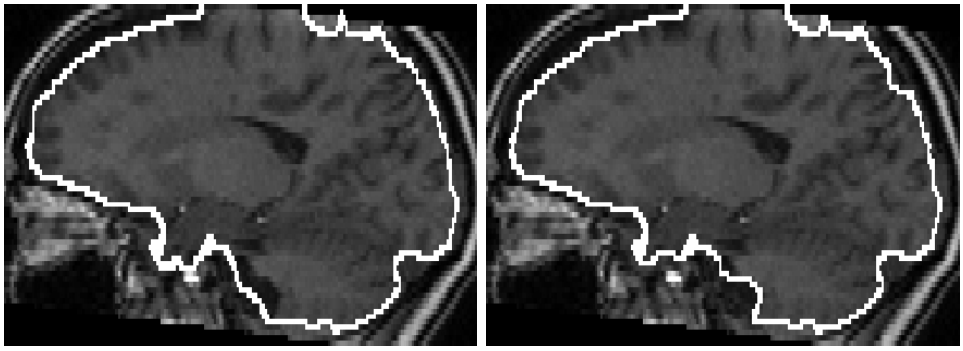


FIG. 4.6 – *À gauche : masque obtenu par EBM (Problème sur la segmentation en avant du cervelet) ; à droite : masque obtenu avec la nouvelle méthode (la segmentation est plus proche du cervelet).*

La méthode combinée, BE3, se comporte mieux que les trois autres méthodes prises individuellement. Par exemple, la figure 4.7 montre des différences importantes entre les segmentations obtenues sur les images de BrainWeb[7] (Base de données simulées). Le masque obtenu par la méthode combinée n'inclut pas le LCR à l'avant du tronc contrairement à BET et EBM. Les contours de la nouvelle méthode sont moins irréguliers que ceux de 3DI.

En conséquence, BE3 a été utilisée à la place de EBM dans la chaîne de traitement EMA.

4.3 Segmentation en sous régions d'intérêt

Le travail présenté dans la section précédente permet d'obtenir un masque global du cerveau. Une division de ce masque en trois régions d'intérêt (cortex, tronc, cervelet) est intéressante pour potentiellement donner des valeurs régionales d'atrophie ou de charge lésionnelle (cf. figure 4.8).

Cette étape est réalisée de la manière suivante :

- recalage multi-affine d'un atlas contenant ces trois structures sur l'image correspondante,
- croissance par zone d'influence de ces trois régions dans le masque global précédemment décrit.

Le recalage multi-affine utilisé ici a été développé par O. Commowick dans [Commowick 2006, Commowick 2007]. Il se découpe en différentes étapes :

- définition de N régions de l'image,
- appariement de manière séparée des N régions définies (blocs de l'image) avec l'image de référence et calcul des N transformations affines en découlant,

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

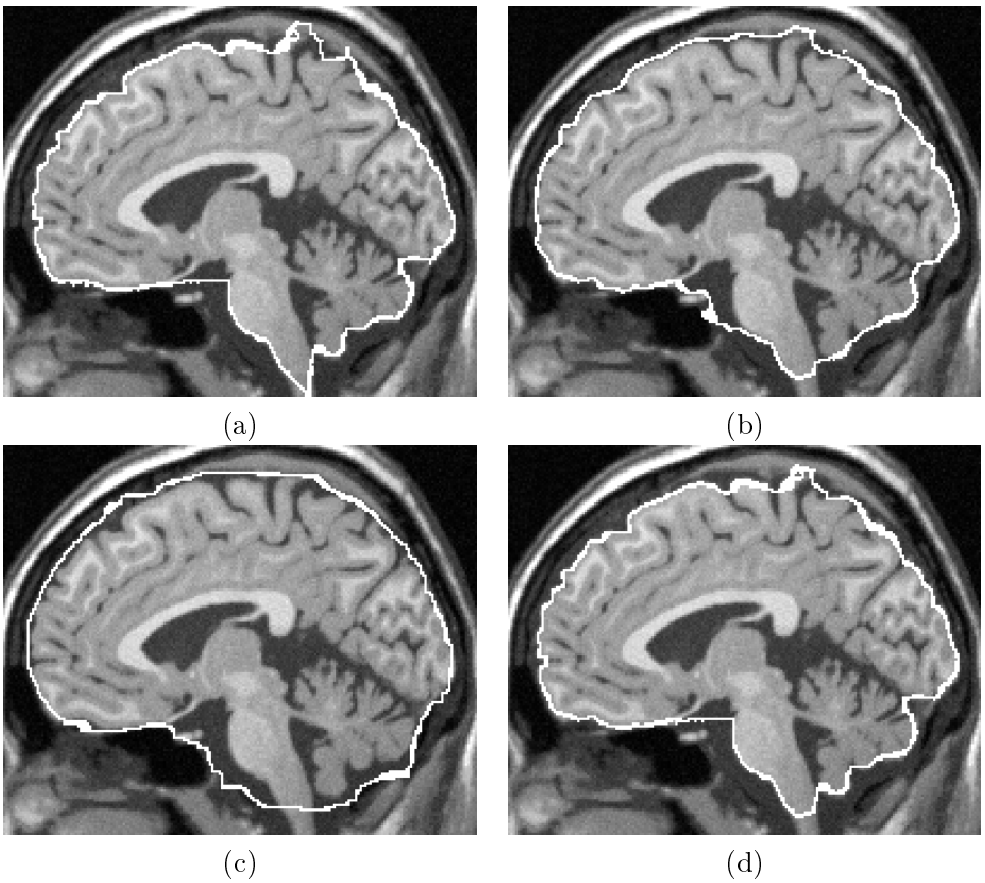


FIG. 4.7 – *Contour du masque du cerveau obtenu par diverses méthodes : (a) 3DI, (b) BET, (c) EBM, (d) BE3 (combinaison de BET, EBM et 3DI).*

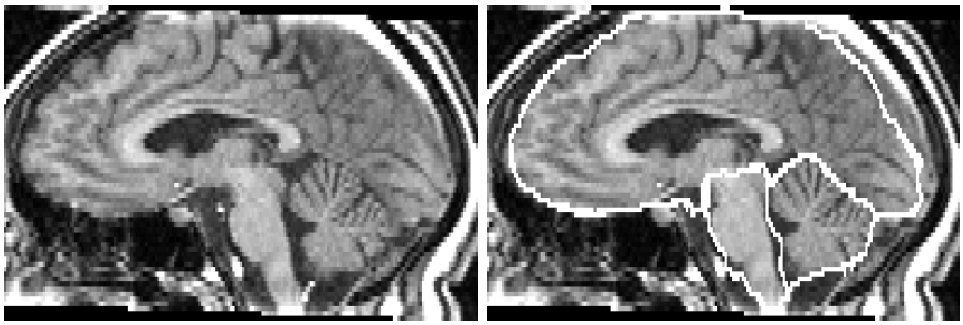


FIG. 4.8 – *À gauche : IRM T1 d'un patient atteint de SEP ; à droite : les trois régions d'intérêt dans le masque du cerveau sur cette même image.*

– régularisation de ces transformations pour limiter de brutales déformations entre les régions de l'image.

Dans notre cas, cet algorithme est utilisé pour recaler l'atlas présenté dans [Bondiau 2005]. On obtient alors une première segmentation de l'image en trois zones : cervelet, cortex et tronc cérébral. Cette segmentation est le plus souvent une sous-segmentation de ces différentes régions. De plus, cette segmentation n'inclut pas le CSF entourant ces régions. Les trois segmentations obtenues sont donc non adjacentes. Pour obtenir uniquement trois zones dans notre masque global du cerveau nous avons donc proposé de réaliser une croissance par zone d'influence (cf. algorithme 4) de ces trois régions à l'intérieur du masque global.

Algorithme 4 Croissance par zone d'influence, contrainte par un masque

Initialisation de N régions, R, à l'intérieur d'un masque
répéter
  pour tout R **faire**
    pour tout voxel, v de R **faire**
      pour tout voxel adjacent, va, de v **faire**
        si va non inclus dans R ou une autre région **alors**
          va appartient à R
jusqu'à Tous les points sont classés.

Les IRM T1 de Clermont-Ferrand et de Montpellier vont jusqu'au bas du cou. Nous avons proposé une méthode particulière pour segmenter le tronc dans ce cas. Une fois le recalage multi-affine des trois structures réalisé, la segmentation obtenue sert d'initialisation à un algorithme de croissance de région dont les paramètres, T et N ont été fixés empiriquement (respectivement à 5 mm et 80%) (cf. algorithme 5). La figure 4.9 présente la segmentation obtenue avant et après croissance de région.

Algorithme 5 Croissance de région du tronc

Dans le système de coordonnées $(\overrightarrow{x}, \overrightarrow{y}, \overrightarrow{z})$. L'image est orientée de manière à ce que le bas du tronc soit en bas de l'axe z.
Initialisation de la zone tronc par recalage multi-affine d'un atlas. Une première segmentation, S, est obtenue.
Calcul du maximum, max et du minimum, min, des intensités des voxels de S. Détection du voxel, vi, de S ayant la coordonnée en z la moins élevée.
Initialisation d'une région, R, par le voxel vi.
pour tout voxel, v de la liste des voxels appartenants à R **faire**
  pour tout voxel adjacent, va, de v de coordonnée z inférieure ou égale à la coordonnée en z de v **faire**
    si l'intensité de va est comprise dans l'intervalle $[min, max]$ **alors**
      si Plus de N % des points au voisinage (de taille T) de va ont une intensité comprise dans l'intervalle $[min, max]$ **alors**
        va est ajouté à la liste des voxels de R

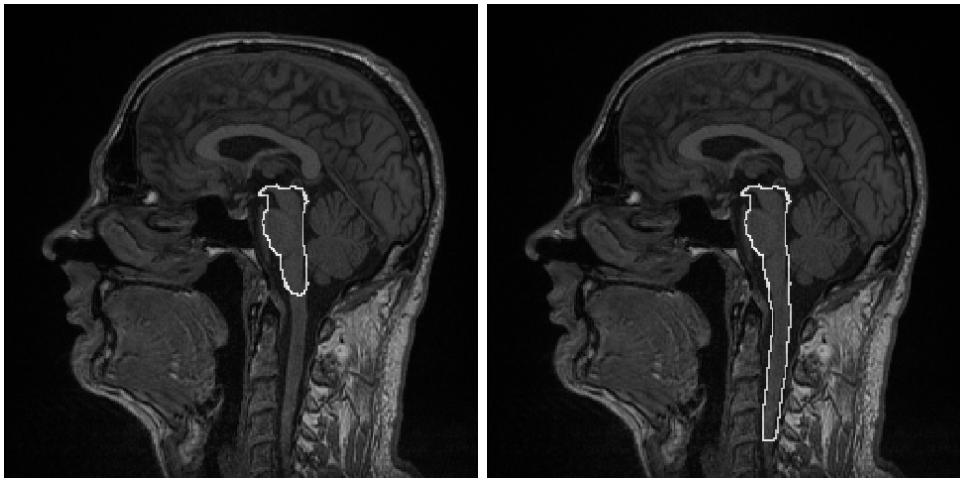


FIG. 4.9 – *À gauche : IRM T1 d'un patient atteint de SEP avec une première segmentation du tronc cérébral ; à droite : la segmentation du tronc est complétée par croissance de région.*

Chapitre 5

Segmentation des lésions

Sommaire

5.1 État de l'art 67
5.1.1 Différentes séquences IRM conventionnelles montrant différentes sortes de lésions 67
5.1.2 Approches de segmentation des lésions 69
5.1.3 Segmentation directe des lésions 69
5.1.4 Classification avec segmentation des compartiments cérébraux 73
5.1.5 Validation des méthodes de segmentation 77
5.1.6 Discussion 80
5.2 Segmentation manuelle et/ou semi-automatique 82
5.3 Segmentation automatique 83
5.3.1 Méthode de segmentation de G. Dugas, PEMS-G 83
5.3.2 Utilisation de la chaîne de traitement EMA 85
5.3.3 Segmentation des lésions du cortex, PEMS-COR 86
5.3.4 Segmentation des lésions du cervelet, PEMS-CER 87
5.4 Workshop MICCAI : *MS lesion segmentation challenge 2008* 88
5.4.1 Optimisation de l'algorithme de segmentation automatique . 88
5.4.2 Comparaison de méthodes sur les données d'entraînement . . 91
5.4.3 Résultats de PEMS-COR sur les données test 91
5.4.4 Discussion 95
5.4.5 Challenge sur site et résultats 96
5.5 Choix de la méthode pour segmenter les lésions 97
5.5.1 Comparaison de deux méthodes automatiques sur la base de données Qualicore 97
5.5.2 Discussion 102

Résumé : De nombreuses méthodes de segmentation des lésions de SEP sont présentes dans la littérature. Ce chapitre en dresse un état de l'art et en propose une classification. Puis, un nouvel algorithme semi-automatique est présenté. Celui-ci permet un gain de temps pour la réalisation de segmentations de référence par un expert. Deux nouvelles versions de l'algorithme décrit dans [Dugas-Phocion 2004b] sont alors introduites. Leurs premières étapes sont similaires à [Dugas-Phocion 2004b], mise à part l'algorithme d'extraction du cerveau (cf. chapitre 4), la prise en compte d'une troisième image pour segmenter les différents compartiments cérébraux et les améliorations de l'EM présentée au chapitre 3. En revanche, les post-traitements (comme par exemple l'extraction des lésions à partir des segmentations du cortex et/ou du cervelet) sont différents. Les résultats au workshop MICCAI "challenge de segmentation des lésions de SEP" de la version segmentant les lésions du cortex sont ensuite donnés. Puis, les différentes versions de l'algorithme sont comparées entre elles sur les données de la base QUALICORE.

Mots clefs : Segmentation des lésions de SEP, État de l'art, Algorithme semi-automatique, Algorithme automatique.

Contributions :

- Algorithme semi-automatique (croissance de région) pour la segmentation des lésions.
- Algorithmes automatiques pour la segmentation des lésions (du cortex et du cervelet).
- Comparaison de méthodes automatiques (sur des données de la base QUALICORE et participation au challenge de segmentation des lésions).

Organisation du chapitre : Ce chapitre commence par un état de l'art complet des méthodes de segmentation des lésions de SEP sur des IRM ainsi que des critères de validation de ces méthodes (section 5.1). Il présente ensuite un outil de segmentation manuel que j'ai développé ainsi qu'un nouvel algorithme de segmentation semi-automatique (section 5.2). Ces outils permettent d'obtenir des segmentations de référence. En troisième partie, deux nouvelles versions d'un algorithme précédemment développé sont décrites (section 5.3). Les différentes versions de l'algorithme ont ensuite été comparées sur les données d'entraînement du Workshop MICCAI *Multiple sclerosis lesion segmentation Challenge 08*. La version donnant les meilleurs résultats a été choisie pour participer au challenge (section 5.4). Le chapitre finit sur une étude comparative des diverses méthodes sur les IRM de notre base de données (section 5.5).

5.1 État de l'art

Un examen IRM est fréquemment requis pour poser un diagnostic de SEP selon les critères de McDonald [McDonald 2001]. De plus, cet examen est souvent pratiqué dans le suivi d'un patient (aide à la décision thérapeutique) ou dans des études cliniques [Filippi 2004b]. Comme indiqué dans le chapitre 2, cette analyse IRM est basée sur les critères de Barkhof/Tintore [Barkhof 1997, Tintore 2000] qui incluent le nombre de signaux hyperintenses en T2, leur emplacement, les lésions prenant le gadolinium sur les séquences T1-GADO et les lésions de la mœlle épinière [Polman 2005].

Le nombre de lésions T2, leur volume et l'atrophie cérébrale sont considérés comme des biomarqueurs précoces de la sévérité de la maladie et de l'efficacité d'un traitement en recherche clinique [Ge 2006]. Des études ont montré que la charge lésionnelle est très faiblement corrélée avec le handicap physique du patient [Nijeholt 1998, Dastidar 1999, Grimaud 2001, van Walderveen 2001] excepté pour le syndrôme cliniquement isolé (la toute première manifestation des symptômes) [Brex 2002, Korteweg 2006, Kappos 2006]. Il y a aussi de fortes variations en nombre et en emplacement des lésions T2 en fonction du moment où l'examen IRM est réalisé. Un autre marqueur de la SEP est l'atrophie cérébrale. Celle-ci peut être évaluée par différentes techniques (cf. chapitre 6) mais il n'y a pour l'instant aucune méthode de référence.

La connaissance d'une segmentation précise des lésions peut être très utile pour évaluer ces biomarqueurs (même pour l'atrophie où une telle segmentation permet de diminuer l'impact des points aberrants causés par les lésions sur la qualité des segmentations). Le masque des lésions est obtenu par une classification des voxels du cerveau dans différentes classes (par exemple "lésion" ou "non-lésion"). La réalisation d'une segmentation manuelle est une tâche fastidueuse et est sujette aux variabilités intra- et inter-experts [Grimaud 1996, Zijdenbos 1998, Ashton 2003]. Pour répondre à ce problème, de nombreuses méthodes de segmentations (assistées par ordinateurs ou semi-automatiques ou entièrement automatiques) ont été développées ces vingt dernières années.

Cette section va donc dresser un état de l'art de ces méthodes en se focalisant principalement sur celles entièrement automatiques et qui s'appliquent sur des séquences IRM conventionnelles (*i.e.* T1, T1-GADO, T2, DP et T2-FLAIR). Après une description des aspects cliniques des lésions, cet état de l'art propose une classification de ces méthodes. Il résume les forces et les faiblesses de chaque approche et présente un processus de validation des méthodes.

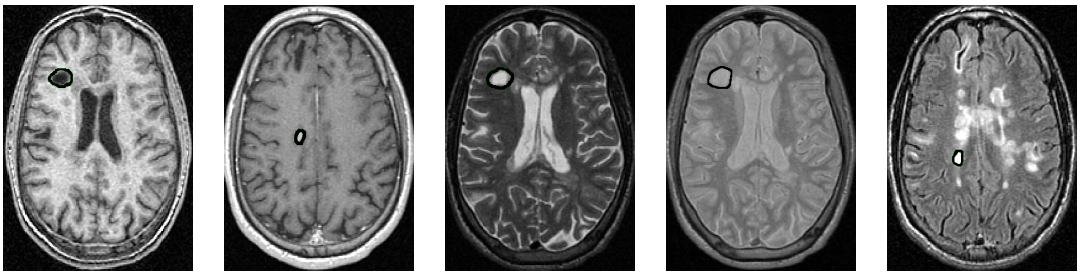
5.1.1 Différentes séquences IRM conventionnelles montrant différentes sortes de lésions

Les lésions de SEP peuvent avoir différentes formes et être localisées à différents endroits du cerveau [Bakshi 2005]. Une même lésion peut apparaître comme un signal hyperintense ou un signal hypointense en fonction de la séquence IRM utilisée.

Les lésions de SEP sont des signaux hyperintenses en T2, DP et en T2-FLAIR (cf. figure 5.1 (c,d,e)). La séquence T2-FLAIR permet une meilleure différenciation entre le LCR et les lésions mais des artéfacts osseux ou de flux sont présents dans l'image. D'autres séquences IRM permettent de labéliser les lésions en trois sous-types [Barkhof 2001, Paty 2001].

- **Les lésions prenant le gadolinium (LESGADO)** sont la marque d'une cassure de la barrière hémato-encéphalique et reflètent les lésions actives. Ces lésions sont des signaux hyperintenses en T2, DP et en T1-GADO (cf. figure 5.1 (b)).
- **Les trous noirs en séquence T1 (TNT1)** sont le reflet d'une atteinte sévère et chronique des tissus. Ces lésions sont des signaux hyperintenses en T2 et en DP mais hypointenses en T1 (voir figure 5.1 (a)). Même si les hypointensités en T1 sont majoritairement considérées comme des TNT1, certaines correspondent à des lésions œdémateuses transitoires.
- **Les autres lésions (ALES)** correspondent à des signaux hyperintenses en T2 et en DP mais ne correspondent ni à des signaux hyperintenses en T1-GADO ni à des signaux hypointenses en T1.

Dans la suite de ce manuscrit, je me réfèrerai à ces sous-types de lésions par leur acronyme (LESGADO, TNT1, ALES). La table 5.1 résume les différences d'intensité pouvant être observées sur les diverses séquences IRM entre les sous-types de lésions et les compartiments cérébraux (SB, SG, LCR). D'autres modalités (imagerie de diffusion du tenseur [Goldberg-Zimring 2005], spectroscopie par résonance magnétique [Narayana 2005, Gonzalez-Toledo 2006], IRM fonctionnelle [Buckle 2005]) peuvent aussi aider à la segmentation des lésions mais ne sont pas le propos de cet état de l'art.



(a) (b) (c) (d) (e)

FIG. 5.1 – *Exemples d'apparences des lésions de SEP sur les séquences (a) T1, (b) T1-GADO, (c) T2, (d) DP, (e) T2-FLAIR.*

TAB. 5.1 – *Différences d'intensité pouvant être observées sur les diverses séquences IRM entre les sous-types de lésions et les compartiments cérébraux (SB, SG, LCR). Hypo, Hyper et Iso signifient respectivement que le signal du sous-type de lésion est plus sombre, plus clair ou similaire au signal considéré.*

	T1	T1-GADO	T2	T2-FLAIR	DP
LESGADO/LCR	Iso	Hyper	Iso	Hyper	Iso
LESGADO/SG	Iso	Hyper	Iso	Hyper	Iso
LESGADO/SB	Hypo	Hyper	Hyper	Hyper	Hyper
TNT1/LCR	Iso	Iso	Iso	Hyper	Iso
TNT1/SG	Iso	Iso	Iso	Hyper	Iso
TNT1/SB	Hypo	Hypo	Hyper	Hyper	Hyper
ALES/LCR	Iso	Iso	Iso	Hyper	Iso
ALES/SG	Iso	Iso	Iso	Hyper	Iso
ALES/SB	Iso	Iso	Hyper	Hyper	Hyper

5.1.2 Approches de segmentation des lésions

Les méthodes de segmentation des lésions de SEP peuvent être classées en deux catégories.

1. Le premier type de méthodes segmente directement les lésions en se basant sur les signaux caractéristiques de celles-ci. Les voxels sont labélisés comme lésion ou comme non-lésion en fonction d'une classification locale (*i.e.* l'application d'un seuil) (les lésions sont vues comme des valeurs caractéristiques d'intensité) ou en fonction d'une analyse de l'évolution de leur intensité (les lésions sont vues comme des évolutions caractéristiques d'intensité).
2. Le second classe les voxels du cerveau dans les classes de compartiments cérébraux. Dans cette approche, les lésions sont détectées en tant que points ne pouvant être classés (les lésions sont vues comme des points aberrants) ou sont définies comme une classe supplémentaire dans l'algorithme de segmentation (les lésions sont vues comme une classe de tissu).

La figure 5.2 montre ces différentes approches de segmentation. Dans la suite de cette section, la description des différentes méthodes ne prend pas en compte les prétraitements. Ceux-ci ont été décrits dans le chapitre 3.

5.1.3 Segmentation directe des lésions

Dans cette partie, les méthodes qui segmentent les lésions de SEP comme des signaux caractéristiques vont être présentées. Un voxel est labélisé comme lésion ou non en fonction de son intensité ou de l'évolution de cette dernière.

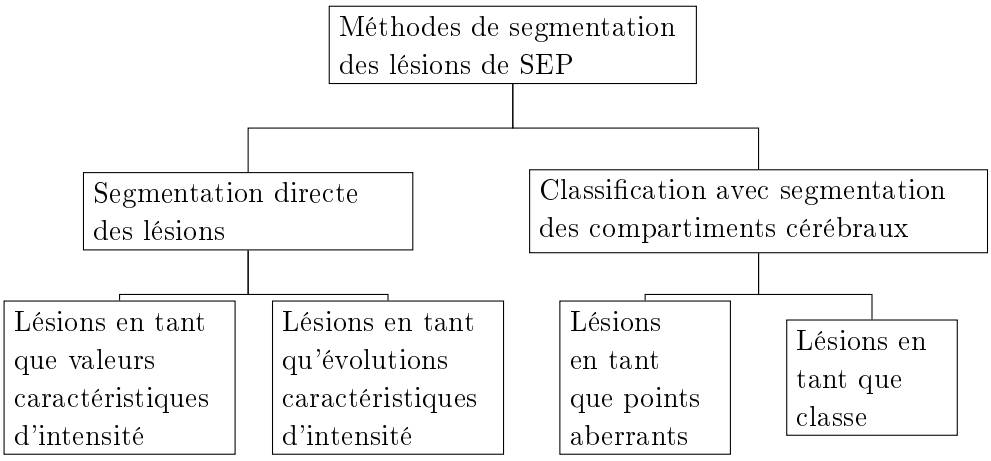


FIG. 5.2 – *Arbre de segmentation des lésions de SEP.*

5.1.3.1 Lésions en tant que valeurs caractéristiques d'intensité

Selon la séquence IRM utilisée et le sous-type de lésions observé, les lésions correspondent à des signaux hyperintenses ou hypointenses. Basées sur ces caractéristiques, différentes approches permettent de les segmenter. Tout d'abord, une classification locale peut être effectuée. Par exemple, un opérateur manuel clique au centre d'une lésion puis un seuil est appliqué localement autour de ce point [Grimaud 1996, Achiron 2002, Horsfield 2007]. Toutefois, ces algorithmes de segmentation sont semi-automatiques et dépendent d'un expert.

Dans le cas de la segmentation de sous-types de lésions, cette approche peut être entièrement automatisée. Par exemple, les lésions prenant le gadolinium (LESGADO) sont des signaux hyperintenses sur la séquence T1-GADO. Une segmentation de ces signaux peut être obtenue par l'application d'un seuil (calculé automatiquement depuis l'histogramme de l'image [Samarasekera 1997, He 2000, He 2002]) ou par la détection des maxima régionaux grâce à une dilatation géodésique [Datta 2007]. De la même manière, les trous noirs en T1 (TNT1) qui sont des signaux hypointenses sur cette séquence peuvent être segmentés comme des minima régionaux grâce à une érosion géodésique [Datta 2006].

Une autre approche ne classe pas les voxels directement mais les regroupe en fonction de diverses propriétés (par exemple d'intensité, de volume, ou d'emplacement) [Akselrod-Ballin 2006]. Ce regroupement de régions permet de fusionner les informations et d'effectuer la classification sur plus de données qu'uniquement l'intensité du voxel. Dans le cas d'études longitudinales, cette fusion d'informations peut être réalisée à partir des données de chaque instant. Par exemple, des régions d'intérêt sont modélisées dans un espace 4D (intensité, position (x,y) et temps) avant d'effectuer la classification dans [Shahar 2004]. Dans cette approche, l'auteur utilise des connaissances *a priori* (*i.e.* intensité moyenne des régions et évolution temporelle de leur taille) pour détecter les lésions.

Post-traitements

Pour détecter les lésions, les méthodes précédentes sont basées sur l'étude des signaux caractéristiques. Cependant, deux effets opposés doivent être pris en considération. D'un côté, les lésions n'ont pas un contour homogène en terme d'intensité. En conséquence, la valeur du seuil doit être diminuée pour obtenir un maximum de voxels appartenant à la lésion et pour améliorer la sensibilité de la méthode. D'un autre côté, les intensités de la SG et/ou des volumes partiels SG/LCR sont proches de l'intensité des lésions (cf. table 5.1). Ceci implique d'augmenter la valeur du seuil pour éliminer les voxels n'appartenant pas à une lésion et pour améliorer la spécificité de la méthode. Pour ces raisons, les diverses méthodes sont des compromis entre sensibilité et spécificité.

Pour améliorer la délimitation d'une lésion informe, des post-traitements comme ceux basés sur de la connectivité floue sont communément utilisés [Samarasekera 1997, Datta 2006, Datta 2007]. Ce cadre de la connectivité floue peut également inclure des connaissances *a priori* données par un atlas des lésions [Horsfield 2007].

Pour diminuer le nombre de faux positifs (voxels qui ne devraient pas être classés comme lésion), différents post-traitements peuvent être utilisés. Les faux positifs peuvent être rejetés parce qu'ils correspondent à un trop grand objet ou grâce à un opérateur humain [Samarasekera 1997]. Les lésions erronées qui sont situées dans les régions plausibles de volume partiel (*i.e.* connectées avec la surface extérieure du cerveau) peuvent également être éliminées par des opérations morphologiques [He 2000, He 2002] ou par connectivité [Shahar 2004]. La segmentation obtenue peut aussi être comparée avec les résultats d'une autre méthode. Par exemple, les trous noirs en T1 et les lésions prenant le gadolinium doivent correspondre aux signaux hyperintenses en T2 obtenus par [Sajja 2006] dans [Datta 2006] et dans [Datta 2007].

Discussion

La segmentation des lésions basée sur les aspects caractéristiques de leurs signaux implique un compromis entre la sensibilité et la spécificité. En conséquence, la segmentation obtenue doit être post-traitée pour réduire le nombre de lésions en faux positifs et pour améliorer le tracé des lésions restantes. Toutefois, ces méthodes restent relativement simples et faciles à réaliser.

5.1.3.2 Lésions en tant qu'évolutions caractéristiques d'intensité

Dans le cas d'études longitudinales (au moins 2 points de temps), il pourrait être plus facile de segmenter les lésions évolutives. La détection des zones évolutives de l'image peut se faire grâce à deux approches.

La première étudie l'évolution en intensité de chaque voxel. Si seulement deux instants sont disponibles, cette évolution est obtenue par le calcul de l'image de

différence [Duan 2008]. Sur cette image, les changements peuvent être obtenus par l'application d'un seuil ou en utilisant le test du rapport de vraisemblance généralisé, détectant les changements significatifs d'intensité dans une fenêtre 3D de l'image [Bosc 2003]. Si davantage d'instants sont disponibles, l'évolution en intensité d'un voxel est représentée grâce à une courbe. L'analyse porte alors sur la recherche d'évolutions caractéristiques de ces courbes d'intensité. Par exemple, les lésions se caractérisent par une fluctuation de la luminosité de l'image dans [Gerig 2000] ou par une détection de points correspondant à un changement d'évolution sur ces courbes [Prima 2002]. Pour obtenir des évolutions plus caractéristiques, le processus peut également prendre en compte l'information spatiale comme les courbes d'intensité des voxels voisins [Welti 2001, Rey 2001b] ou les données issues des différentes séquences [Prima 2003].

La deuxième approche recherche des évolutions de formes. De telles méthodes sont basées sur l'analyse du champ de déformation obtenu par un recalage non rigide entre deux images. Des seuils sont alors appliqués sur la valeur du jacobien du champ de déformation local pour détecter les changements significatifs de volume [Thirion 1999, Rey 2002b].

Modèles d'évolution de l'intensité des lésions

Certaines des méthodes précédentes comparent la courbe d'évolution de l'intensité d'un voxel appartenant à une lésion avec un modèle *a priori*. Ceci peut être utilisé pour diminuer le nombre de faux positifs [Rey 2001a, Rey 2001b], mais aussi pour comprendre la maladie. En effet, les courbes d'évolution en intensité semblent refléter deux processus : le processus inflammatoire/dégénératif qui augmente l'intensité et le processus de réparation/de résorption qui diminue l'intensité. En fonction de ces modèles, trois types de courbes d'évolution en intensité peuvent être définis : avec un processus de recouvrement total, partiel ou inexistant [Meier 2003, Meier 2006].

Discussion

La classification effectuée par les méthodes ci-dessus est basée sur la détection d'un processus évolutif. Ceci peut aider à rendre plus robuste la segmentation des lésions. Toutefois, les lésions ne sont pas le seul processus dynamique dans le cerveau (il y a par exemple aussi l'atrophie). En conséquence, les critères correspondant à une évolution significative doivent être définis. Ces derniers impliquent que seules les lésions évolutives seront détectées. La méthode ne sera pas en mesure d'évaluer la charge lésionnelle. En outre, ces méthodes nécessitent un recalage précis pour garantir l'étude de l'évolution d'un voxel et non des transitions en intensité entre différents voxels. Comme il a été expliqué au chapitre 3, les intensités des images des différents instants doivent être normalisées. Enfin, une série temporelle de séquences IRM doit également être disponible et cette analyse ne peut être effectuée qu'à la fin de l'étude.

5.1.4 Classification avec segmentation des compartiments cérébraux

Une autre façon de segmenter les lésions est de commencer par segmenter les compartiments cérébraux sains (SB, SG, LCR). Dans ce cas, les lésions peuvent être obtenues en tant que points aberrants ou être segmentées directement en tant que classe supplémentaire.

5.1.4.1 Lésions en tant que points aberrants

Les méthodes de cette catégorie segmentent les compartiments cérébraux (SB, SG, LCR) sur une ou plusieurs séquences IRM. À partir de ces segmentations, quatre approches différentes permettent d'identifier les lésions.

La première approche assimile les lésions aux voxels restants de la classification précédente. Dans ce cas, un algorithme supervisé de connectivité floue est souvent utilisé pour segmenter les compartiments cérébraux. Puis, les voxels non classés sont regroupés. Les agrégats obtenus peuvent alors être étiquetés comme des lésions ou rejetés, soit par un opérateur humain [Udupa 1997] soit par un réseau de neurones [Admasu 2003]. Un réseau de neurones est également utilisé dans [Ardizzone 1999] et dans [Alonge 2001] pour classer des agrégats non étiquetés par l'application d'un algorithme des c-moyennes floues.

La deuxième approche segmente les compartiments indépendamment sur différentes séquences IRM. Dans ce cas, les lésions correspondent aux voxels qui sont classés différemment en fonction des séquences. Par exemple, un algorithme des c-moyennes floues peut être appliqué pour classer tous les voxels en tant que LCR ou SG ou SB indépendamment sur une séquence T1 et une séquence T2. Les voxels qui sont semblables à de la SG sur la séquence T1 et similaires à du LCR sur la séquence T2 sont alors classés comme lésions [Yang 2004]. De la même manière, les segmentations obtenues par un algorithme des k-moyennes indépendamment sur chaque séquence IRM (T1, T2, PD) sont fusionnées pour obtenir une unique segmentation pour chaque classe (LCR, SB et SG). Enfin, les cavités de la SB (trous dans la segmentation de la SB) sont classées dans la catégorie lésions en fonction de leur intensité moyenne sur les différentes séquences [Zhu 2003].

La troisième approche utilise les segmentations pour chercher les lésions dans des parties spécifiques du cerveau. La plupart des méthodes de cette approche proposent de segmenter les lésions uniquement dans le masque de la "SB supposée". Cette "SB supposée" correspond à la segmentation de la SB qui devrait être obtenue si aucune lésion n'était présente dans l'image. Par exemple dans [Warfield 1995], un modèle est déformé pour extraire un masque de la cavité intra-crânienne, un masque du cortex et une classification des structures sous-corticales. Ensuite, une classification est réalisée dans le masque de la SB entre la SB et les lésions grâce à un classificateur de distance minimale. Dans [Kamber 1995], quatre arbres de décision probabilistes sont comparés et appliqués sur des données multi-canaux (T2, DP). L'arbre fournissant les résultats de meilleure qualité cherche les lésions dans les zones plausibles de SB.

La dernière approche identifie les lésions à partir de la distribution des paramètres des différentes classes. Un algorithme couramment utilisé pour segmenter les compartiments sains est l'algorithme d'espérance maximisée (EM) (cf. chapitre 3). Dans [Van Leemput 2001], les paramètres de distribution des différentes classes (LCR, SB, SG, points aberrants) sont calculés grâce à un algorithme EM multicanaux (T1, T2, DP). Ensuite, la probabilité de chaque voxel d'être une lésion est calculée en fonction de son classement précédent et de critères *a priori*. Cet algorithme est une amélioration de celui proposé dans [Van Leemput 1999b]. Il peut également être utilisé pour corriger le biais des images IRM [Van Leemput 1999a]. Pour améliorer la classification, un algorithme EM est modifié pour prendre en considération les effets de volumes partiels entre les classes de SG et de LCR dans [Dugas-Phocion 2004b]. Dans ce dernier article, les lésions sont obtenues en appliquant un seuil dont la valeur est calculée automatiquement à partir des paramètres des classes sur la séquence T2-FLAIR. Dans [Aït-Ali 2005], l'algorithme EM est appliqué en une seule fois sur tous les instants disponibles. Les paramètres des classes dépendent des informations fournies par chaque instant et sont calculés à partir d'un histogramme multidimensionnel des images. Aucun atlas n'est utilisé pour introduire une probabilité spatiale *a priori*. Cet algorithme EM modifié est combiné avec un algorithme de Mean Shift dans [Garcia-Lorenzo 2008]. L'algorithme de Mean Shift permet de définir des régions d'intérêt. Ces agrégats sont ensuite classés et regroupés à partir des paramètres des classes calculés par l'algorithme EM. Cette approche (segmentation des lésions comme points aberrants d'un modèle) peut aussi être utilisée avec d'autres algorithmes de segmentation que l'EM comme par exemple une chaîne de Markov cachée [Bricq 2008].

Post-traitements

Les méthodes de cette partie segmentent les lésions grâce à un processus hiérarchique. Tout d'abord, elles segmentent les voxels ne correspondant pas à des lésions. Puis, elles utilisent les informations fournies par les segmentations obtenues pour caractériser les lésions. Néanmoins, ces segmentations sont toujours le résultat d'un compromis entre la sensibilité et la spécificité. Pour les améliorer, des post-traitements fondés sur les signaux caractéristiques (comme dans la section 5.1.3.1) peuvent être appliqués. La délimitation des lésions peut être améliorée par connectivité floue [Udupa 1997, Admasu 2003] ou grâce à des modèles déformables [Yang 2004]. Le nombre de lésions en faux positifs peut également être réduit par un opérateur humain [Udupa 1997] ou des opérations de morphologie mathématique [Dugas-Phocion 2004a].

Ce type de méthodes offre également une nouvelle correction. Chercher les lésions uniquement dans la SB permet de diminuer le nombre de faux négatifs (lésions qui sont classées comme SG et qui ne sont donc pas détectées) [Warfield 1995].

Discussion

Le type de méthodes décrites dans cette partie classe les compartiments du cerveau dans au moins trois catégories : LCR, SB, SG. À partir de ces segmentations, les méthodes s'intéressent aux voxels non étiquetés ou utilisent ces segmentations pour définir des critères de détection des lésions. Comme le cerveau contient en volume davantage de compartiments sains que de lésions, cette approche pourrait être plus fiable que les approches précédentes. Cependant, les algorithmes utilisés sont souvent plus complexes et difficiles à mettre en œuvre. En outre, les algorithmes paramétriques supposent que les distributions des intensités des différentes classes soient connues. Ces algorithmes doivent également être robustes aux effets de volumes partiels et de SB "sale" (appelée "dirty white matter" en anglais).

5.1.4.2 Lésions en tant que classe

Au lieu de segmenter les lésions à la suite des compartiments sains, une classe, lésions, peut être incluse directement dans l'algorithme. L'objectif des méthodes de cette catégorie est de définir les propriétés de cette classe lésions.

Tout d'abord, la classe lésions peut être définie par des intensités caractéristiques sur les différentes séquences IRM. Par exemple, des seuils sont utilisés dans [Krishnan 1998]. Dans cette méthode, ces seuils sont déduits de l'histogramme de l'image de ratio T2-DP. Leur application, sur la séquence T2 et indépendamment sur la séquence DP, donne un masque des lésions, un masque du parenchyme et un masque de la SB pour chaque séquence. Le masque final des lésions est obtenu par une combinaison des deux segmentations précédentes des lésions. Une autre approche se base sur l'information multi-séquences pour isoler les lésions des autres compartiments grâce au calcul d'un gradient multi-spectral et l'utilisation de graph cuts [Lecoeur 2008].

Selon les séquences IRM, les valeurs d'intensité du LCR et/ou des volumes partiels LCR/SG peuvent aussi être proches des valeurs d'intensité des lésions. Cette propriété est utilisée dans [Boudraa 2000]. Premièrement, la séquence DP est prétraitée pour augmenter le contraste entre le LCR et les lésions. Deuxièmement, un algorithme des c-moyennes floues est appliqué sur cette séquence pour donner le masque de la classe composée LCR/lésions. Ce masque est alors superposé sur le DP. Dans cette région masquée, une nouvelle application de l'algorithme des c-moyennes floues classe les voxels en lésion ou en LCR.

Comme pour les compartiments sains (cf. section 3.3.1), la classe lésions peut être définie par un apprentissage. Dans [Zijdenbos 1994], un réseau de neurones supervisé classe les voxels dans les catégories suivantes : LCR, SB, SG, lésion de la SB ou *fond* à partir de l'intensité des voxels sur les séquences T1, DP et T2. Cette méthode est comparée avec des segmentations d'experts dans [Zijdenbos 1998]. Elle est validée sur une importante base de données dans [Zijdenbos 2002]. En utilisant cette méthode, [Charil 2003] conclut que la charge lésionnelle est peu corrélée avec le score EDSS, mais qu'il existe une meilleure corrélation entre l'emplacement des lésions et le handicap des malades.

Une autre manière d'estimer la distribution des intensités correspondant à une classe est de réaliser un fenêtrage de Parzen (initialisé manuellement). On peut alors classer les voxels à partir de ces distributions. Dans [Sajja 2006], l'application de ce traitement donne une première segmentation du LCR, des lésions et du parenchyme à partir des informations du T2 et du T2-FLAIR. Ensuite, un algorithme EM avec des champs Markov [Zhang 2001] est appliqué pour diviser la classe du parenchyme en deux classes (SB et SG) à partir des images T2 et DP. Enfin, les groupements non connectés de SG et situés dans le masque de la SB sont réexaminés en fonction des résultats de la première classification en vue d'une éventuelle reclassification en tant que lésions.

Les lésions peuvent aussi correspondre à une classe dans un algorithme EM. Par exemple, la cavité intra-craniale est segmentée par un algorithme EM multi-canaux (T2, DP) en 4 classes : LCR, SG, SB et lésions de SB dans [Kikinis 1999, Guttmann 1999]. Dans cette dernière méthode, des opérations morphologiques sont utilisées pour corriger les faux positifs qui sont causés par des effets de volumes partiels. Dans [Li 2003] et dans [Johnston 1996], la même approche est utilisée mais la classification est effectuée respectivement par un algorithme EM avec des champs de Markov et un algorithme du mode conditionnel itéré. En outre, chaque voxel est considéré comme un mélange de compartiments et est appelé mixel dans [Li 2003]. Dans [Wei 2002], trois chaînes de traitements sont comparées. Pour chaque chaîne, un algorithme EM multi-canaux (T2, DP) est appliqué pour classer les voxels dans l'une des catégories suivantes : fond, peau, LCR, SB, SG, lésions de la SB. Dans la première chaîne de traitement, cette étape est suivie par la minimisation des faux négatifs qui a été présentée dans [Warfield 2000] et dans la section 5.1.4.1. Dans la seconde, l'étape de l'EM est suivie par la réduction du nombre de faux positifs qui a été proposée dans [Kikinis 1999]. La dernière chaîne de traitement applique l'algorithme EM et les deux algorithmes de minimisation d'erreur. Cette dernière correction est également utilisée dans [Wu 2006] où un algorithme de classification aux k plus proches voisins segmente les classes suivantes : fond, LCR, SB, SG, LESGADO, TNT1 et les signaux hyperintenses en T2 à partir des séquences T1, T2 et DP.

Post-traitements

Les étapes de post-traitements utilisées précédemment peuvent encore être utilisées avec les méthodes de cette catégorie. Par exemple, le nombre de faux positifs est diminué par des opérations morphologiques dans [Krishnan 1998], par des connaissances *a priori* dans [Zijdenbos 1994] et par un seuil sur la taille des composantes connexes obtenues dans [Boudraa 2000]. De la même manière, une correction des faux négatifs est également effectuée dans [Kikinis 1999]. L'étude de Wei *et al.* compare l'influence de chacune de ces corrections. Elle conclut sur le fait que l'utilisation conjointe des deux corrections (des faux positifs et des faux négatifs) améliore la précision et la reproductibilité des mesures du volume des lésions de SB.

Discussion

Les méthodes de cette section segmentent les lésions comme une classe. La principale difficulté de cette approche est de définir les caractéristiques de cette classe (seuils [Krishnan 1998], paramètres de distribution des intensités correspondant à la classe [Kikinis 1999, Wei 2002], ...). Dans ce cas, les lésions sont segmentées comme "un compartiment sain". Cependant, il ne faut pas oublier que les lésions sont des objets non connectés, et que leur volume est très souvent bien plus faible que celui du LCR, de la SB ou de la SG. En outre, la phase de réduction des erreurs est encore nécessaire. Son importance est montrée dans [Wei 2002]. De plus, ces méthodes sont complexes à mettre en œuvre.

5.1.5 Validation des méthodes de segmentation

Les sections précédentes montrent que la littérature offre un grand nombre de méthodes de segmentation automatiques des lésions. Toutefois, ces méthodes sont très difficiles à comparer. Elles ont été validées avec des images de résolutions différentes et grâce à des critères différents. Cette section passe en revue ces critères et présente un processus pour valider de telles méthodes.

Lors du développement d'une chaîne de traitement visant à segmenter les lésions, la validation est une des étapes les plus difficiles. Deux problèmes doivent être résolus :

- obtenir une segmentation "vérité terrain" des lésions ;
- définir les critères permettant de comparer la segmentation automatique à la vérité terrain.

5.1.5.1 La vérité terrain

Il s'agit de la vraie segmentation 3D des lésions. Toute segmentation (manuelle, semi-automatique ou entièrement automatique) en est une approximation.

Segmentation manuelle réalisée par un expert

Une segmentation obtenue grâce à un algorithme est communément comparée à une segmentation manuelle réalisée par un expert. Toutefois, les variabilités intra et inter experts peuvent être importantes [Grimaud 1996, Zijdenbos 1998, Ashton 2003]. En conséquence, cette référence reste une approximation de la vérité terrain même si une segmentation manuelle et précise des lésions peut prendre plusieurs heures. Les méthodes semi-automatiques peuvent également aider l'expert à réaliser cette segmentation de référence. Cependant, les résultats restent encore améliorables [Parodi 2002].

Moyenne de segmentations manuelles réalisées par un/des expert(s)

Dans [Warfield 2004], Warfield *et al.* présentent l'algorithme STAPLE. À partir de différentes segmentations (obtenues par des experts ou par plusieurs méthodes), cet algorithme donne une image représentant la probabilité de chaque voxel d'être

réellement une lésion. Cette approche donne une segmentation de la vérité terrain plus probable puisqu'elle est basée sur plusieurs segmentations auxquelles elle associe des poids de vraisemblance (cf. chapitre 4). Toutefois, elle nécessite plusieurs segmentations manuelles qui sont fastidieuses et difficiles à obtenir.

Utilisation de données simulées

Une autre solution pour obtenir la vérité terrain est d'utiliser les données simulées à partir d'un modèle. BrainWeb est une base de données de volumes simulés, proposée par le Montréal Neurological Institute[1] et disponible en ligne[2]. À partir de modèles anatomiques où des lésions de SEP sont présentes, un simulateur produit des volumes IRM en utilisant différents paramètres [Collins 1998].

La segmentation de ces données simulées par une méthode automatique peut être comparée avec le modèle anatomique ayant servi à produire ces données. Dans ce cas, ce modèle correspond à la vérité terrain. Trois types de modèles contenant différentes charges lésionnelles sont disponibles. La simulation peut être faite dans les séquences IRM T1, T2 ou DP.

Ce type de validation est plus facile à mettre en œuvre, mais la vraisemblance du simulateur et le choix des paramètres restent un problème.

5.1.5.2 Critères de comparaison entre une segmentation et la vérité terrain

Différents critères peuvent être utilisés pour comparer des segmentations d'images médicales [Fenster 2005, Udupa 2006]. Dans le cas de segmentation de lésions, diverses valeurs sont traditionnellement calculées pour comparer deux segmentations (Seg et Ref). Seg est l'ensemble des voxels qui ont été classés comme "lésions" par un algorithme de segmentation. Ref est l'ensemble des voxels représentant une lésion dans la vérité terrain (par exemple, dans la segmentation réalisée par un expert). La figure 5.3 illustre les différentes catégories de voxels (vrais positifs, faux positifs, vrais négatifs, faux négatifs).

La table 5.2 regroupe les différents critères qui peuvent être calculés à partir de ces catégories de voxels.

5.1.5.3 Processus de validation

Les mesures présentées dans la section 5.1.5.2 sont traditionnellement utilisées pour comparer deux segmentations. Dans le cas de segmentations de lésions cette approche peut être améliorée par une validation en deux étapes [Dugas-Phocion 2006] : la détection des lésions et la délimitation des lésions. En effet, ces deux processus sont différents et peuvent avoir aussi des applications différentes. La détection des lésions est nécessaire dans les critères de McDonald. Le calcul du volume des lésions peut être utilisé pour l'aide à la décision thérapeutique et dans la recherche pharmaceutique.

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

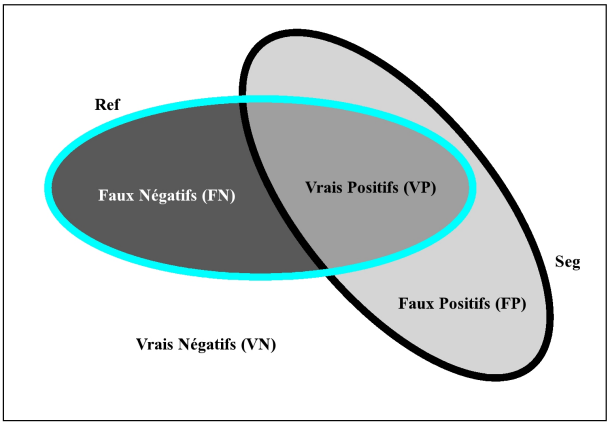


FIG. 5.3 – *Catégories de voxels (vrais positifs, faux positifs, vrais négatifs, faux négatifs) pour la comparaison d'une segmentation (Seg) avec une segmentation de référence (Ref).*

TAB. 5.2 – *Critères de comparaison de segmentation.*

Critères	Représentations topologiques	Représentations par catégories de voxels
Coefficient de similarité de Jaccard (CSJ) [Jaccard 1912, Zou 2004a, Zou 2004b]	$\frac{Card(Ref \cap Seg)}{Card(Ref \cup Seg)}$	$\frac{TP}{FN+TP+FP}$
Coefficient de similarité de Dice (CSD) ou Indice de similarité (IS) [Dice 1945, Zijdenbos 1994, Anbeek 2004a, Aït-Ali 2005]	$\frac{2Card(Ref \cap Seg)}{Card(Ref)+Card(Seg)}$	$\frac{2TP}{FN+2TP+FP}$
Fraction de sur-estimation [Datta 2006, Anbeek 2004b]	$\frac{Card(\overline{Ref} \cap Seg)}{Card(Ref)}$	$\frac{FP}{FN+TP}$
Fraction de sous-estimation [Datta 2006]	$\frac{Card(Ref \cap \overline{Seg})}{Card(Ref)}$	$\frac{FN}{FN+TP}$
Fraction d'estimation correcte (FEC) ou fraction de chevauchement (FC) ou Sensibilité (Sen) [Datta 2006, Anbeek 2004b, Warfield 2004]	$\frac{Card(Ref \cap Seg)}{Card(Ref)}$	$\frac{TP}{FN+TP}$
Spécificité (Spe) [Warfield 2004]	$\frac{Card(\overline{Ref \cup Seg})}{Card(\overline{Ref})}$	$\frac{TN}{TN+FP}$
Valeur prédictive positive (VPP) [Altman 1994]	$\frac{Card(Ref \cap Seg)}{Card(Seg)}$	$\frac{TP}{TP+FP}$
Valeur prédictive négative (VPN) [Altman 1994]	$\frac{Card(\overline{Ref \cup Seg})}{Card(\overline{Seg})}$	$\frac{TN}{TN+FN}$

Mesure de qualité de la détection des lésions

Dans ce cas, l'unité de mesure est une lésion. Une lésion est définie comme un ensemble de voxels connectés dans l'une des segmentations Seg ou Ref. Avec cette définition, et en considérant qu'une lésion de Ref est détectée si elle s'intersecte avec au moins une lésion de Seg, nous avons :

- nombre de vrais positifs (VP), nombre de lésions de Seg qui s'intersectent avec au moins une lésion de Ref,
- nombre de faux négatifs (FN), nombre de lésions de Ref qui ne s'intersectent avec aucune lésion de Seg,
- nombre de faux positifs (FP), nombre de lésions de Seg qui ne s'intersectent avec aucune lésion de Ref,
- nombre de vrais négatifs (VN), n'est pas défini, parce que les voxels qui n'ont pas été labellisés comme lésions sont connectés entre eux et ne forment qu'un seul groupe.

Comme VN n'est pas défini, la sensibilité et la valeur prédictive positive nous donnent des informations sur la qualité de détection des lésions.

Mesure de qualité de la délimitation des lésions

Dans ce cas, l'unité de mesure est le voxel. Premièrement, toutes les lésions de Ref ou de Seg qui ne s'intersectent pas avec une lésion de l'autre segmentation sont retirées des segmentations (tous les voxels de ces lésions sont mis à la valeur d'intensité du fond). Par conséquent, deux nouvelles segmentations ont été créées Ref' et Seg'. Avec ces notations, nous avons :

- nombre de vrais positifs (VP), nombre de voxels de Seg' étiquetés comme lésion à la fois dans Seg et Ref.
- nombre de faux négatifs (FN), nombre de voxels étiquetés comme lésion dans Ref mais non dans Seg,
- nombre de faux positifs (FP), nombre de voxels étiquetés comme lésion dans Seg mais non dans Ref,
- nombre de vrais négatifs (VN), nombre de voxels non étiquetés comme lésion ni dans Ref ni dans Seg.

La qualité de délimitation des lésions est donnée par la sensibilité et la valeur prédictive positive. Dans ce cas, la spécificité est toujours proche de 1 parce que TN est très élevé par rapport à FN.

5.1.6 Discussion

Différents algorithmes de segmentation des lésions de SEP ont été présentés. L'utilisation de séries d'images multi-séquences et/ou longitudinales peut améliorer la segmentation et permet de classer les lésions en sous-types. Cette classification peut se faire directement dans le processus de l'algorithme [Wu 2006] ou en reclassant les signaux hyperintenses en T2 en sous-types grâce à des connaissances *a priori* [Samarasekera 1997, He 2000, Datta 2006] ou grâce à une analyse de texture [Yu 1999]. La table 5.3 indique les séquences IRM utilisées et les types de lésions segmentées pour une sélection de méthodes représentatives.

TAB. 5.3 – Séquences IRM utilisées, types de lésions segmentées et valeur d'indice de similarité (IS) (entre les segmentations obtenues par la méthodes et une segmentation de référence) pour une sélection représentative des diverses méthodes de segmentation des lésions de SEP.

	T1	T1 GADO	T2	T2 FLAIR	DP	LESGADO	TNT1	ALES	Signaux hyperintenses en T2	lésions évolutives	Valeur moyenne d'IS (coupes,examens)
[Bricq 2008]	+		+						+		
[Garcia-Lorenzo 2008]	+		+		+				+		
[Lecoeur 2008]	+		+		+				+		
[Samarasekera 1997]		+				+					
[He 2002]	+	+	+		+	+					
[Datta 2006]	+	+	+	+	+		+				0.73(X,14)
[Saija 2006]			+	+	+				+		0.78(X,23)
[Akselrod-Ballin 2006]	+		+	+	+				+		0.6(X,10)
[Shahar 2004]			+						+		
[Zhu 2003]	+		+		+				+		0.8(17,X)
[Yang 2004]	+		+						+		
[Van Leemput 2001]	+		+		+				+		0.45(X,20)
[Dugas-Phocion 2004a]	+		+	+	+				+		
[Udupa 1997]			+		+				+		
[Admasu 2003]	+		+		+				+		
[Ardizzone 1999]			+		+				+		
[Alonge 2001]			+		+				+		
[Aït-Ali 2005]	+		+		+				+		0.76 (X,BrainWeb)
[Kamber 1995]			+		+				+		
[Warfield 1995]			+		+				+		
[Krishnan 1998]			+		+				+		0.55(9,X)
[Zijdenbos 2002]	+		+		+				+		0.6(10,X)
[Charil 2003]	+		+		+				+		
[Boudraa 2000]			+		+				+		
[Kikinis 1999]			+		+				+		
[Li 2003]	+		+	+					+		
[Johnston 1996]			+		+				+		0.5(10,X)
[Wei 2002]			+		+				+		
[Wu 2006]		+	+		+	+	+	+	+		
[Thirion 1999]			+		+					+	
[Rey 2002b]			+		+					+	
[Gerig 2000]	+		+		+					+	
[Welti 2001]	+	+	+	+	+					+	
[Prima 2003]	+		+		+					+	

La plupart des méthodes segmentent les signaux hyperintenses en T2. Dans ce cas, deux types d'erreurs peuvent se produire dans les segmentations et doivent être corrigés. Ils sont tous les deux dus au fait que l'intensité des lésions est proche de l'intensité de la SG ou des volumes partiels de SG/LCR.

- Le premier type d'erreurs correspond aux faux positifs (*i.e.* des voxels de SG ou de volumes partiels de SG/LCR peuvent être classés comme des lésions). Les approches proposées pour corriger ce type d'erreurs sont les suivantes : demander son avis à un opérateur humain [Samarasekera 1997] ou ne pas considérer les voxels qui sont situés dans des zones plausibles de volumes partiels (par des opérations morphologiques [He 2002], par connectivité [Kikinis 1999, Shahar 2004]).
- Le deuxième type d'erreurs correspond aux faux négatifs (*i.e.* des voxels correspondant à des lésions peuvent être classés comme de la SG ou du LCR). Une approche proposée pour corriger ce type d'erreurs est de chercher les lésions uniquement dans un masque de la SB [Warfield 1995] même si les lésions peuvent également être présentes dans la SG [Inglese 2004, Filippi 2005].

La minimisation de ces erreurs dans des segmentations de sous-types de lésions peut aussi se réaliser en comparant cette segmentation avec une segmentation des signaux hyperintenses en T2 obtenue par une autre méthode [Datta 2006].

La validation des méthodes de segmentation des lésions de SEP doit se faire qualitativement et quantitativement. Si la segmentation est obtenue à partir d'IRM simulées, les résultats peuvent être comparés avec le modèle anatomique utilisé pour la simulation. Les segmentations obtenues à partir d'IRM réelles peuvent être comparées à une segmentation réalisée par un expert ou à une moyenne de segmentations réalisées par un/des expert(s). Différents critères sont utilisés pour comparer les segmentations entre elles et la validation peut être réalisée en deux étapes : la détection et la délimitation. La table 5.3 donne la valeur moyenne de l'indice de similarité pour diverses méthodes validées en utilisant cet indice. Les validations des diverses méthodes restent cependant difficiles à comparer parce que celles-ci n'ont pas été réalisées sur le même nombre de coupes ou d'examens et parce que les étapes et les méthodes de prétraitements utilisées par ces chaînes de traitements sont aussi différentes (cf. section 3.2). Néanmoins, les neurologues et les médecins sont en attente d'un algorithme automatique qui sera utile dans leur pratique quotidienne.

5.2 Segmentation manuelle et/ou semi-automatique

La section précédente a introduit différentes méthodes de segmentation des lésions entièrement automatiques. La validation de ces méthodes requiert une segmentation de référence. Le plus souvent celle-ci est obtenue par une segmentation manuelle réalisée par un expert. Cependant, très peu de logiciels sont disponibles pour effectuer cette segmentation et ces derniers sont le plus souvent payants et non conçus spécifiquement pour la SEP. Ils ne permettent pas, par exemple, de visualiser plusieurs images en même temps et dans le même repère géométrique.

Pour permettre aux neurologues de nous fournir une segmentation de référence pour chacune des images de notre base, un logiciel a été créé. Il est accessible sous forme d'une fonctionnalité nommée Lesion Segmentation Edition (LSE) du logiciel SepINRIA (cf. section 7.2.1). Cette fonctionnalité permet de recaler les séquences IRM entre elles, de les afficher simultanément et de segmenter manuellement les lésions sur les différentes séquences IRM. Cette segmentation peut être différente d'une séquence à l'autre.

Pour aider la segmentation, j'ai aussi développé un algorithme semi-automatique (cf. algorithme 6). Celui-ci prend un minimum de paramètres. Deux versions en sont disponibles. Dans la première, les paramètres sont fixés manuellement par l'utilisateur. Dans la seconde, ils ont été fixés empiriquement.

Cette approche permet donc de recaler les images entre elles, de les visualiser simultanément et de segmenter plus rapidement les lésions à l'aide d'un algorithme semi-automatique. En conséquence, la réalisation de segmentations de référence est facilitée et améliorée par cette solution que j'ai proposée et mise en œuvre.

De plus, l'algorithme semi-automatique proposé ici est générique. Il a par exemple été utilisé avec mon aide par le radiologue Sébastien Novellas pour segmenter les lymphomes médiastinaux en tomodensitométrie [Novellas 2007].

5.3 Segmentation automatique

De nombreuses méthodes entièrement automatiques de segmentation des lésions sont présentes dans la littérature (cf. section 5.1). À notre connaissance, aucune d'elles n'est présente sous la forme d'un logiciel disponible gratuitement sur internet. Les tester sur nos données semble donc difficile. Néanmoins, nous disposons d'une méthode qui a été testée sur des images dont le protocole d'acquisition était similaire au nôtre, celle de G. Dugas [Dugas-Phocion 2006]. Cette méthode sera appelée PEMS-G (Prétraitements, EM, Seuil - Global) dans la suite de ce document.

PEMS-G possède cependant différentes limites (cf. section 5.3.1). En conséquence, tout en gardant le principe de cette méthode (segmentation du cerveau en différents compartiments, puis calcul automatique en se servant des segmentations précédentes d'un seuil à appliquer sur une séquence), j'en ai proposé deux nouvelles versions. Celles-ci utilisent la chaîne de traitement EMA et non EMP. La première, PEMS-COR, permet d'extraire les lésions du cortex, la seconde, PEMS-CER, celles du cervelet. Il suffit alors de combiner leurs résultats pour obtenir une segmentation globale, PEMS-COMBI.

5.3.1 Méthode de segmentation de G. Dugas, PEMS-G

PEMS-G est basée sur la chaîne de traitement EMP (version préliminaire de EMA, cf. section 3.4). PEMS-G se sert des segmentations obtenues par EMP pour calculer automatiquement un seuil à appliquer sur la séquence T2-FLAIR. Ceci produit une image binaire sur laquelle les points appartenant au masque du LCR pur ou au masque de la SB (donnés par EMP) sont retirés. La segmentation finale

Algorithme 6 *Segmentation semi-automatique des lésions.*

Précondition : L'image à segmenter, ima ; un seuil sur le volume, $Vmax$; un paramètre d'amplitude des intensités, Tx ; un seuil sur la distance des points d'un voisinage, $Dmax$.
Le contraste de l'image à traiter ima, est rehaussé avec l'algorithme 7 pour donner l'image $imaC$.
L'image du gradient morphologique haut, $imaU$ de $imaC$ est calculée (cf. algorithme 8).
pour tout L, appartenant à l'ensemble des lésions **faire**

Précondition : Un clic de l'utilisateur sur la lésion.
Initialisation de la zone lésion en fonction de la position du clic de l'utilisateur.

pour tout vzl, appartenant à l'ensemble de la liste des voxels de la zone lésion **faire**
Calcul des propriétés de la zone lésion :
moyenne de l'intensité sur $imaC$, $MoyZ(imaC)$,
écart type de l'intensité sur $imaC$, $EctZ(imaC)$,
valeur médiane du gradient ($imaU$), $MedG$,
écart type du gradient ($imaU$), $EctG$.
Calcul de seuils :
$LTh = MoyZ(imaC) - Tx * EctZ(imaC)$,
$HTh = MoyZ(imaC) + Tx * EctZ(imaC)$,
$LThG = MedG - EctG$,
$HThG = MedG + EctG$.
pour tout $vvzl$, appartenant à l'ensemble des voxels voisins (voxels dont la distance à vzl est inférieure à $Dmax$) de vzl **faire**
si $LTh <$ Intensité$_{(vvzl)} < HTh$ **alors**
Calcul de la valeur médiane, $VMED$, des valeurs de gradient des voisins (connexité 26) de $vvzl$.
si $LThG < VMED < HThG$ **alors**
Ajout de $vvzl$ dans la liste des voxels appartenant à la zone lésion
Calcul du volume, Vol, de la zone lésion
si $Vol \geq Vmax$ **alors**
Arrêt de la boucle For

des lésions est alors obtenue en retirant les composantes connexes de faible taille de cette image.

Toutefois, PEMS-G a des limites :

- non-détection des lésions de la fosse postérieure (ces chiffres sont donnés sans prendre en compte les lésions segmentées manuellement de la fosse postérieure),
- sensibilité de détection des lésions de 74.9 % sur ces données (les lésions détectées représentent 95.4% du volume lésionnel mais ceci ne signifie pas que

Algorithme 7 *Rehaussement du contraste d'une image.*

Précondition : L'image à traiter, *ima*.

Soit *imaC*, une image vide de même taille que *ima*.
Soit *imaD*, l'image obtenue par une dilatation de *ima* en niveau de gris.
Soit *imaE*, l'image obtenue par une érosion de *ima* en niveau de gris.
pour tout i, appartenant à l'ensemble des voxels de *ima*. **faire**
  si $imaD_{(i)} - ima_{(i)} \leq ima_{(i)} - imaE_{(i)}$ **alors**
    $imaC_{(i)} = imaD_{(i)}$
  sinon si $ima_{(i)} - imaE_{(i)} \leq imaD_{(i)} - ima_{(i)}$ **alors**
    $imaC_{(i)} = imaE_{(i)}$
  sinon
    $imaC_{(i)} = ima_{(i)}$
Retourner *imaC*, l'image avec un contraste réhaussé.

Algorithme 8 *Gradient morphologique haut.*

Précondition : L'image à traiter, *ima*.

Soit *imaH*, une image vide de même taille que *ima*.
Soit *imaD*, l'image obtenue par une dilatation de *ima* en niveau de gris.
pour tout i, appartenant à l'ensemble des voxels de *ima*. **faire**
  $imaU_{(i)} = Dima_{(i)} - ima_{(i)}$
Retourner *imaU*, l'image correspondant à un gradient morphologique haut de *ima*.

la méthode segmente correctement plus de 95% de la charge lésionnelle),
- 30 % des lésions détectées sont des faux positifs (des points qui ne correspondent pas à une lésion),
- en ne considérant plus que les lésions bien détectées, seulement 40 % des voxels des lésions sont segmentés (ce pourcentage chute à 20 % dans le cas des lésions nécrotiques),
- sur ces mêmes lésions, 30 % des voxels segmentés sont des faux positifs (ils ne correspondent pas à une lésion).

Même si l'auteur de cette méthode dit qu'elle ne détecte pas les lésions de la fosse postérieure, il ne propose aucun moyen de limiter sa méthode au cortex (excepté manuellement). Pour garder le caractère automatique de la méthode (critère primordial, comme il a été expliqué dans la section 5.1) et pour répondre à notre but d'obtenir une segmentation complète des différentes lésions, PEMS-G sera donc considérée dans la suite de ce manuscrit comme une méthode de segmentation globale des lésions.

5.3.2 Utilisation de la chaîne de traitement EMA

La chaîne de traitement EMA (cf. chapitre 3) est composée de diverses étapes. Tout d'abord, les différentes séquences IRM sont prétraitées (*i.e.* normalisation spa-

tiale et en intensité, extraction du masque du cerveau). Puis, l'algorithme EM amélioré est utilisé pour obtenir les différentes segmentations.

Les segmentations obtenues serviront à définir une région d'intérêt et au calcul d'un seuil. Une très légère fluctuation de ces segmentations n'est donc pas gênante pour notre application. Pour cette raison et pour gagner du temps de calcul, j'ai choisi de fixer le paramètre de pourcentage des points considérés (PPC) pour calculer les paramètres des classes à 20% dans ce cas.

5.3.3 Segmentation des lésions du cortex, PEMS-COR

Mise à part les lésions nécrotiques, les lésions sont des signaux hyperintenses sur le T2-FLAIR. Les étapes suivantes utilisent cette propriété pour segmenter les lésions.

5.3.3.1 Extraction des lésions à partir de la séquence T2-FLAIR

L'application des segmentations binaires du cerveau (section 3.3) sur la séquence T2-FLAIR permet de calculer les propriétés (moyenne, μ, écart type, σ) des compartiments sains sur cette séquence. Comme les lésions sont des signaux hyperintenses sur le T2-FLAIR, un seuil qui permettra d'obtenir une segmentation de ces signaux peut être calculé de manière automatique à partir des propriétés de la SG (équation 5.1).

$$T = \mu_{SG} + 2\sigma_{SG} \tag{5.1}$$

L'application de ce seuil permet de détecter la majorité des lésions (la plupart des lésions ont au moins un voxel avec une intensité supérieure à ce seuil.). Toutefois, les voxels d'une même lésion ont des intensités inhomogènes. En conséquence, la délimitation des lésions n'est pas simple même si un voxel de cette lésion est connu. Pour ces raisons, j'ai proposé de réhausser le contraste de la séquence T2-FLAIR avant d'appliquer le seuil, T. Comme pour la méthode semi-automatique, cette étape est réalisée à l'aide de l'algorithme 7. L'application de T sur la séquence T2-FLAIR avec le contraste réhaussé nous donne alors une segmentation préliminaire des lésions, S_1.

5.3.3.2 Affinement de la segmentation en utilisant les résultats de la classification

S_1 contient les lésions mais aussi d'autres hypersignaux comme les artéfacts osseux ou de flux. Pour retirer les voxels correspondant à ces faux positifs, une région d'intérêt est définie. Comme dans [Kamber 1995, Warfield 1995], j'ai proposé de chercher les lésions dans la "SB supposée". Ce "compartiment supposé" correspond à la SB que nous devrions observer si aucune lésion n'était présente. Une segmentation de cette classe peut être approximée par le masque de la SB (donné par l'algorithme EM) dans lequel les "trous" (cavités) sont remplis par des opérations de morphologie mathématique (dilatation + remplissage des trous en 3D (extraction

des composantes connexes en connexité 26) + érosion). L'application de ce masque sur S_1 donne une seconde segmentation des lésions, S_2.

Selon [Dugas-Phocion 2006], les lésions ne sont classées ni comme SB ni dans la classe "pure" de LCR par l'algorithme EM. Les voxels appartenant à l'une de ces segmentations sont donc retirés de S_2.

Pour finir, les trous présents dans S_2 sont remplis par morphologie mathématique (application dans les 3 directions de l'algorithme de remplissage des trous en 2D (extraction des composantes connexes en connexité 8)) pour améliorer la délimitation des lésions. Le fait de remplir les trous en 2D et non en 3D permet de segmenter les lésions pour lesquelles l'algorithme n'a détecté qu'une partie du contour (par exemple les lésions (signal hyperintense en T2-FLAIR) avec un centre nécrotique qui apparaît comme un signal hypointense en T2-FLAIR).

5.3.3.3 Limitation de la région d'intérêt

La segmentation précédente repose sur le fait que les lésions sont des hypersignaux en T2-FLAIR. Ceci est vrai pour les lésions de la région sus-tentorielle (au-dessus de la tente du cervelet) mais pas pour les lésions de la fosse postérieure (tronc+cervelet). En conséquence, ces lésions ne sont pas incluses dans S_1 ni dans S_2 dans la majorité des cas. Pour éviter la présence de faux positifs dans cette région, on ne garde dans S_2 que les voxels présents dans la région sus-tentorielle. Le masque de cette région est obtenu grâce aux travaux présentés précédemment dans le chapitre 4.

5.3.4 Segmentation des lésions du cervelet, PEMS-CER

Les lésions du cervelet sont peu visibles sur les séquences IRM de la base QUALICORE dont nous disposons. Selon mes observations, la séquence IRM où elles se voient le mieux est la séquence DP. En conséquence, je propose une nouvelle version de l'algorithme PEMS pour segmenter les lésions du cervelet à partir de cette séquence.

5.3.4.1 Extraction des lésions à partir de la séquence DP

Comme précisé précédemment PEMS-CER repose sur les résultats de la chaîne de traitement EMA. De la même manière que sur la séquence T2-FLAIR, l'application des segmentations binaires du cerveau sur la séquence DP permet de calculer les propriétés (moyenne, μ, écart type, σ) des compartiments sains sur cette séquence. Comme les lésions du cervelet sont des signaux hyperintenses sur cette séquence, une première segmentation de celles-ci peut être obtenue en appliquant un seuil. Ce seuil est calculé de la même manière que dans la section 5.3.3.1 mais cette fois à partir des propriétés des classes sur la séquence DP. Ce seuil est aussi appliqué non sur la séquence DP brute mais sur la séquence DP réhaussée précédemment par l'algorithme 7.

5.3.4.2 Affinement de la segmentation et limitation de la région d'intérêt

Les lésions de la fosse postérieure sont visibles sur la séquence DP. Toutefois les contrastes entre ces lésions et les autres compartiments sont plus faibles sur cette séquence que ceux existant entre les lésions et les autres compartiments sur le T2-FLAIR. L'affinement de la segmentation se justifie donc pleinement, même si les artefacts osseux ou de flux ne sont pas ou peu visibles sur la séquence DP.

Les étapes décrites dans les sections 5.3.3.2 et 5.3.3.3 sont donc reproduites dans ce cas, excepté que le masque utilisé pour définir la région d'intérêt est celui du cervelet.

5.4 Workshop MICCAI : *MS lesion segmentation challenge 2008*

Ce challenge[3] de segmentation des lésions de SEP vise à comparer les méthodes de segmentation des lésions de SEP. Pour ce défi deux types de données étaient disponibles. Le premier correspondait à une série d'IRM pour lesquelles une segmentation de référence réalisée par un expert était disponible. Ces données permettaient donc d'optimiser la méthode à tester en fonction du protocole d'acquisition des données. Le second type de données correspondait à des données test pour lesquelles il fallait envoyer les segmentations binaires obtenues par la méthode testée. Les résultats des différentes méthodes étaient alors comparés avec les segmentations manuelles réalisées par deux experts par les organisateurs du "workshop".

5.4.1 Optimisation de l'algorithme de segmentation automatique

Les séries de données du challenge sont des images recalées entre elles et isotropiquement rééchantillonnées d'une séquence T1, d'une séquence T2, d'une séquence T2-FLAIR, de la carte des fractions d'anisotropie (FA) et de la carte de diffusion moyenne (DM) issues de la modalité DTI. La séquence DP n'est donc pas disponible. La chaîne de traitement EMA a été modifiée pour ne pas prendre en compte cette séquence. Cependant, la séquence DP est nécessaire dans PEMS-CER. Son absence rend donc la segmentation des lésions du cervelet irréalisable dans ce cas. L'algorithme modifié est donc très proche de PEMS-COR et se divise en différentes étapes :

- prétraitements, ils sont identiques excepté que le recalage n'est pas nécessaire (cf. figure 5.4),
- classification, elle ne se fait que sur les séquences T1 et T2 (cf. figure 5.5),
- extraction des lésions à partir du T2-FLAIR (cf. figure 5.6, a et b),
- affinement de la segmentation (cf. figure 5.6, c),
- limitation de la région d'intérêt (cf. figure 5.7),

[3] http ://www.ia.unc.edu/MSseg/

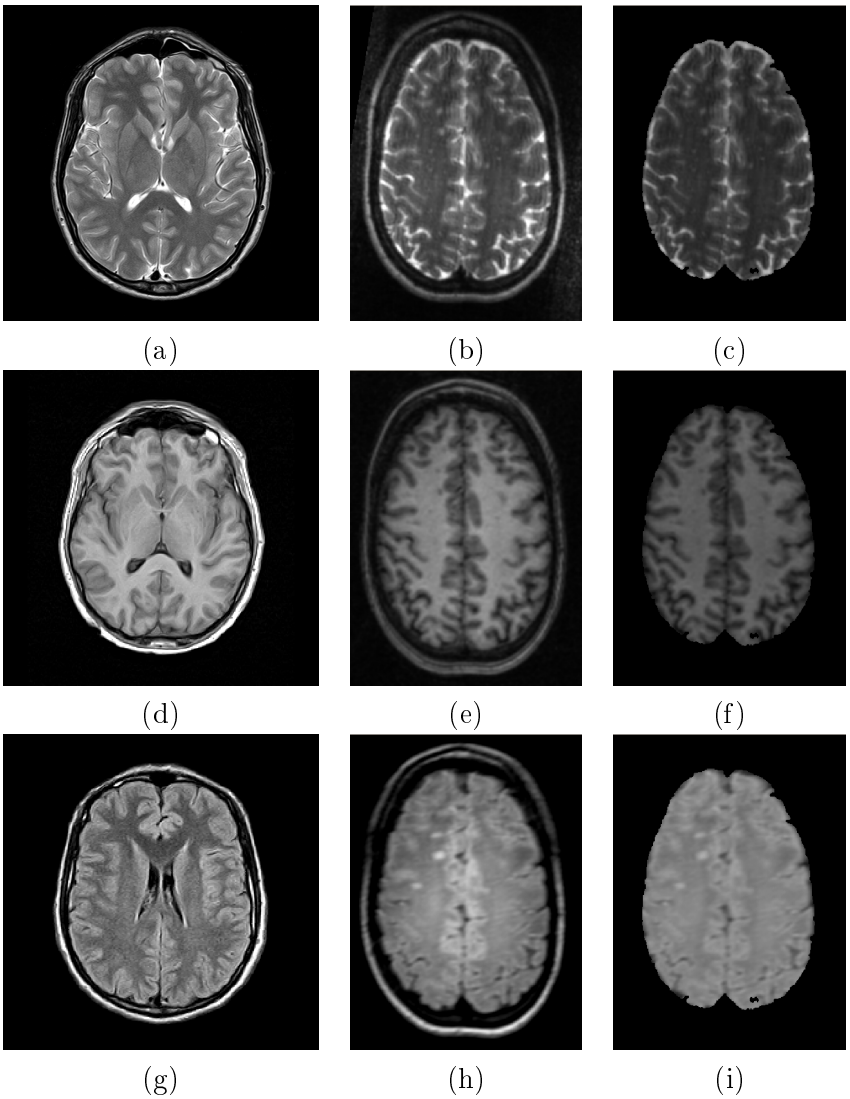
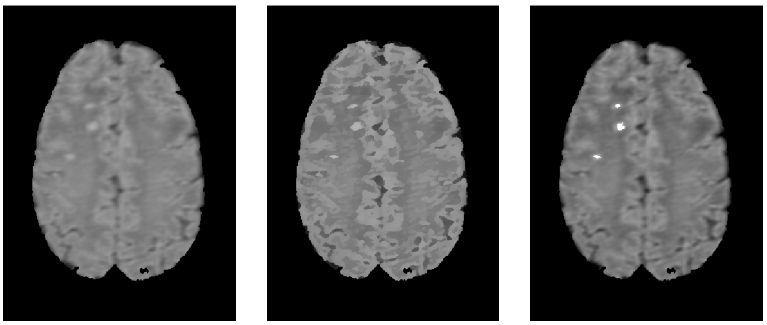


FIG. 5.4 – *Exemple de prétraitements sur CHB_ test1_ Case01 : (a) séquence T2 disponible, (b) séquence T2 avec région d'intérêt restreinte, (c) séquence T2 débiaisée et avec le cerveau extrait, (d) séquence T1 disponible, (e) séquence T1 avec région d'intérêt restreinte, (f) séquence T1 débiaisée et avec le cerveau extrait, (g) séquence T2-FLAIR disponible, (h) séquence T2 avec région d'intérêt restreinte, (i) séquence T2 débiaisée et avec le cerveau extrait.*



(a) (b) (c) (d) (e)

FIG. 5.5 – *Classification obtenue sur CHB_test1_Case01 test data : (a) LCR, (b) SG, (c) SB, (d) union des 6 classes de volumes partiels, (e) union des points aberrants.*



(a) (b) (c)

FIG. 5.6 – *Extraction des lésions à partir du T2-FLAIR : (a) séquence T2-FLAIR, (b) séquence T2-FLAIR avec le contraste réhaussé, (c) segmentation finale des lésions.*

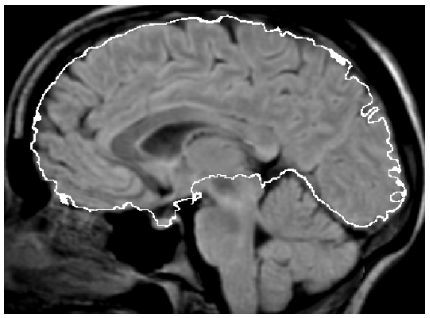
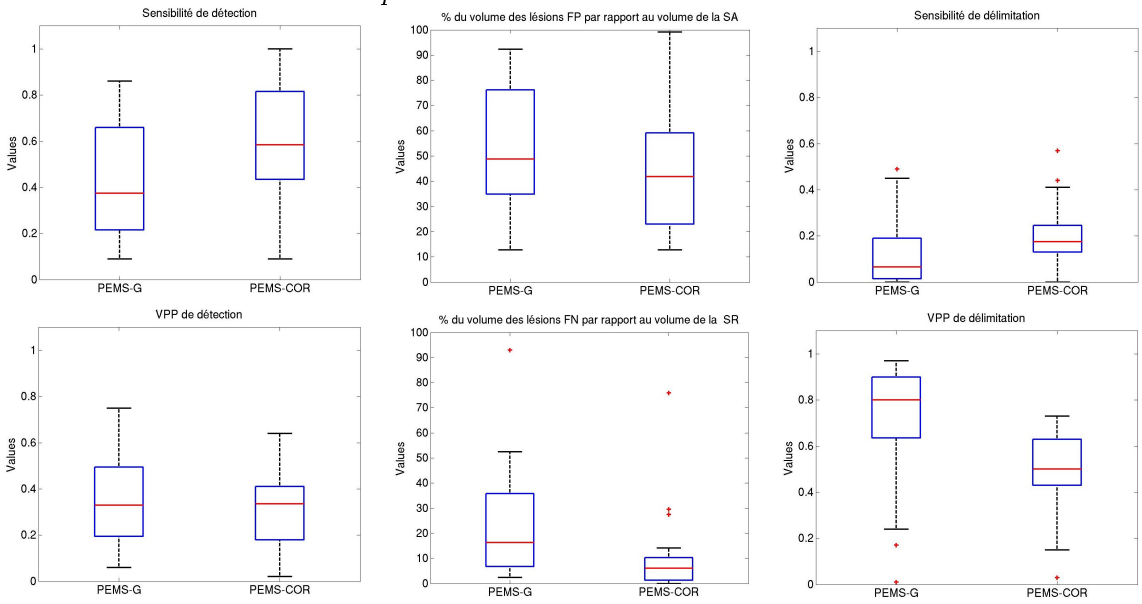


FIG. 5.7 – *Contours du masque sus-tentoriel sur la séquence T2-FLAIR.*

FIG. 5.8 – *Boîtes à moustaches reprenant les résultats des tables 5.4, 5.5 et 5.6. Les notations utilisées sont reprises de ces tables.*



5.4.2 Comparaison de méthodes sur les données d'entraînement

J'ai appliqué les prétraitements et la méthode de classification précédemment définie sur les données d'entraînement. J'ai ensuite exécuté d'une part les post-traitements proposés par G. Dugas (dans la version PEMS-G de l'algorithme) et d'autre part ceux que j'ai proposés précédemment (dans la version PEMS-COR).

La table 5.4 montre que PEMS-COR est bien plus sensible que PEMS-G en terme de détection. Ce gain de près de 20% en sensibilité se fait avec une perte de seulement 3% en valeur prédictive positive. De plus, le volume des lésions segmentées à tort ne représente plus que 44% du volume de la segmentation automatique contre 52% avec PEMS-G (table 5.5). Dans le même temps, le pourcentage de volume des lésions non segmentées par rapport à la segmentation de référence chute de 24,2% avec PEMS-G à 10,67 % avec PEMS-COR (table 5.5). La sensibilité de délimitation des lésions est aussi améliorée (gain de proche de 10%) malgré une perte plus importante de la valeur prédictive positive (proche de 20%).

Les résultats précédents ainsi que leur reprise sous forme de boîtes à moustaches (cf. figure 5.8) montrent une supériorité de PEMS-COR par rapport à PEMS-G sur les données d'entraînement du challenge. Les résultats de PEMS-COR sont en effet meilleurs (moyennes plus élevées) et plus homogènes (écarts types égaux ou plus faibles). En conséquence, j'ai utilisé PEMS-COR pour effectuer le challenge.

5.4.3 Résultats de PEMS-COR sur les données test

Les résultats de PEMS-COR ont été envoyés aux organisateurs du challenge. Une comparaison entre la segmentation automatique et les segmentations manuelles

TAB. 5.4 – *Sensibilité et valeur prédictive positive (VPP) de détection pour PEMS-G et PEMS-COR sur les 20 données d'entraînement du "MS lesion Segmentation Challenge 08".*

Centre	Sujet	Sensibilité de détection PEMS-G	Sensibilité de détection PEMS-COR	VPP de détection PEMS-G	VPP de détection PEMS-COR
CHB	train_Case01	0,67	0,86	0,75	0,40
CHB	train_Case02	0,50	0,65	0,49	0,33
CHB	train_Case03	0,24	0,59	0,36	0,23
CHB	train_Case04	0,43	0,57	0,29	0,61
CHB	train_Case05	0,46	0,86	0,67	0,39
CHB	train_Case06	0,13	0,38	0,31	0,42
CHB	train_Case07	0,32	0,43	0,54	0,37
CHB	train_Case08	0,24	0,44	0,56	0,64
CHB	train_Case09	0,16	0,44	0,14	0,36
CHB	train_Case10	0,50	0,67	0,07	0,17
UNC	train_Case01	0,09	0,09	0,06	0,02
UNC	train_Case02	0,67	0,78	0,49	0,45
UNC	train_Case03	0,23	0,58	0,21	0,30
UNC	train_Case04	0,25	0,45	0,35	0,60
UNC	train_Case05	0,76	0,76	0,19	0,12
UNC	train_Case06	0,20	0,24	0,20	0,21
UNC	train_Case07	0,86	1,00	0,19	0,13
UNC	train_Case08	0,65	0,85	0,20	0,34
UNC	train_Case09	0,14	0,29	0,50	0,16
UNC	train_Case10	0,77	0,92	0,36	0,19
Moyenne		0,41	0,59	0,35	0,32
Ecart type		0,25	0,25	0,20	0,17

TAB. 5.5 – *Pourcentage du volume des lésions fausses positives (FP) par rapport au volume de la segmentation automatique (SA) et pourcentage du volume des lésions fausses négatives (FN) par rapport au volume de la segmentation de référence (SR) pour PEMS-G et PEMS-COR sur les 20 données d'entraînement du "MS lesion Segmentation Challenge 08".*

Centre	Sujet	% du volume des lésions FP par rapport au volume de la SA, PEMS-G	% du volume des lésions FP par rapport au volume de la SA, PEMS-COR	% du volume des lésions FN par rapport au volume de la SR, PEMS-G	% du volume des lésions FN par rapport au volume de la SR, PEMS-COR
CHB	train_Case01	18,87	18,24	2,47	0,26
CHB	train_Case02	40,35	44,81	2,35	1,64
CHB	train_Case03	46,52	56,57	39,94	6,26
CHB	train_Case04	48,56	15,28	14,96	12,18
CHB	train_Case05	12,72	25,81	28,95	1,79
CHB	train_Case06	63,55	35,45	17,57	8,29
CHB	train_Case07	49,01	42,94	7,26	4,34
CHB	train_Case08	24,50	20,27	12,15	6,47
CHB	train_Case09	86,28	61,72	39,42	4,06
CHB	train_Case10	83,37	66,07	7,26	0,48
UNC	train_Case01	92,31	99,08	93,00	75,92
UNC	train_Case02	15,09	18,07	6,26	5,98
UNC	train_Case03	74,22	73,26	52,52	27,51
UNC	train_Case04	58,65	12,71	32,24	14,13
UNC	train_Case05	79,17	83,22	5,33	7,06
UNC	train_Case06	38,96	42,85	30,53	29,54
UNC	train_Case07	58,65	55,68	4,36	0,00
UNC	train_Case08	78,30	40,29	25,51	0,87
UNC	train_Case09	47,79	40,69	50,48	6,10
UNC	train_Case10	30,73	33,06	11,47	0,55
Moyenne		52,38	44,30	24,20	10,67
Ecart type		24,61	23,82	22,79	17,44

TAB. 5.6 – *Sensibilité et valeur prédictive positive (VPP) de délimitation pour PEMS-G et PEMS-COR sur les 20 données d'entraînement du "MS lesion Segmentation Challenge 08".*

Centre	Sujet	Sensibilité de délimitation PEMS-G	Sensibilité de délimitation PEMS-COR	VPP de délimitation PEMS-G	VPP de délimitation PEMS-COR
CHB	train_Case01	0,18	0,21	0,74	0,50
CHB	train_Case02	0,06	0,18	0,76	0,52
CHB	train_Case03	0,05	0,18	0,93	0,48
CHB	train_Case04	0,19	0,13	0,94	0,65
CHB	train_Case05	0,03	0,22	0,71	0,57
CHB	train_Case06	0,02	0,13	0,82	0,71
CHB	train_Case07	0,01	0,13	0,90	0,67
CHB	train_Case08	0,02	0,13	0,84	0,68
CHB	train_Case09	0,00	0,02	0,24	0,46
CHB	train_Case10	0,01	0,05	0,62	0,50
UNC	train_Case01	0,01	0,00	0,01	0,03
UNC	train_Case02	0,49	0,39	0,65	0,47
UNC	train_Case03	0,09	0,15	0,88	0,59
UNC	train_Case04	0,18	0,44	0,90	0,61
UNC	train_Case05	0,19	0,41	0,91	0,45
UNC	train_Case06	0,07	0,11	0,17	0,15
UNC	train_Case07	0,45	0,57	0,78	0,39
UNC	train_Case08	0,21	0,27	0,59	0,33
UNC	train_Case09	0,01	0,17	0,97	0,73
UNC	train_Case10	0,22	0,22	0,88	0,41
Moyenne		0,12	0,21	0,71	0,49
Ecart type		0,14	0,14	0,27	0,18

fournies par deux experts a été réalisée suivant les critères suivants :

- la différence de volume, différence en pourcentage et en valeur absolue par rapport à la segmentation manuelle considérée,
- la distance moyenne, distance moyenne entre la surface de la segmentation par rapport à la segmentation manuelle considérée,
- le taux de vrais positifs, nombre de lésions de la segmentation automatique qui s'intersectent avec au moins une lésion de la segmentation de référence, divisé par le nombre de lésions de cette segmentation de référence,
- le taux de faux positifs, nombre de lésions de la segmentation automatique qui ne s'intersectent avec aucune lésion de la segmentation de référence, divisé par le nombre de lésions de la segmentation automatique,
- la sensibilité, rapport du nombre de voxels vrais positifs sur le nombre de voxels de la segmentation de référence,
- la spécificité, rapport du nombre de voxels vrais négatifs sur la somme du nombre de voxels vrais négatifs et du nombre de voxels faux positifs,
- la valeur prédictive positive, rapport du nombre de voxels vrais positifs sur le nombre de voxels de la segmentation automatique.

Ces critères sont donc similaires à ceux que nous avons proposés dans la section 5.1.5. Le taux de vrais positifs correspond à ce que nous avons appelé sensibilité de détection. Le taux de faux positifs correspond à la valeur prédictive positive.

TAB. 5.7 – *Résultats de la méthode automatique proposée sur les données test du challenge "MS lesion Segmentation Challenge 08".*

Ground Truth	UNC Rater								CHB Rater								STAPLE		
All Dataset	Volume Diff.		Avg. Dist.		True Pos.		False Pos.		Volume Diff.		Avg. Dist.		True Pos.		False Pos.		Total	Specificity	Sensitivity	PPV
	[%]	Score	[mm]	Score	[%]	Score	[%]	Score	[%]	Score	[mm]	Score	[%]	Score	[%]	Score				
UNC test1 Case01	56.6	92	4.8	90	51.2	81	75.7	64	36.5	95	5.3	89	53.1	82	69.2	68	82	0.9936	0.3248	0.7049
UNC test1 Case02	7.7	99	4.1	92	54.4	82	72.5	65	87.8	87	3.5	93	40.9	75	31.9	90	85	0.9949	0.1782	0.8373
UNC test1 Case03	46.2	93	1.5	97	53.5	82	49.4	80	30.5	96	1.3	97	55.9	83	40.6	85	89	0.9947	0.6529	0.8708
UNC test1 Case04	90.6	87	8.6	82	34.2	71	86.4	57	85.5	87	7.1	85	48.1	79	86.4	57	76	0.9955	0.0587	0.5161
UNC test1 Case05	53.4	92	6.0	88	52.4	81	75.2	64	5.3	99	6.4	87	78.3	96	77.9	62	84	0.9906	0.3695	0.6164
UNC test1 Case06	81.1	88	7.0	86	41.4	75	71.4	66	15.5	98	19.4	60	62.5	87	86.9	57	77	0.9932	0.1438	0.5770
UNC test1 Case07	45.6	93	2.6	95	55.7	83	58.3	74	26.2	96	6.4	87	80.0	97	71.5	66	86	0.9883	0.5412	0.6463
UNC test1 Case08	21.1	97	3.9	92	63.8	88	65.2	70	28.9	96	10.6	78	94.4	100	76.1	63	85	0.9865	0.5983	0.5587
UNC test1 Case09	157.9	77	44.0	9	33.3	70	96.6	51	263.9	61	53.0	0	0.0	51	100.0	49	46	0.9733	0.1110	0.0873
UNC test1 Case10	73.0	89	13.0	73	45.0	77	85.2	58	525.2	23	22.5	54	83.3	99	93.8	53	66	0.9847	0.6041	0.5892
CHB test1 Case01	70.3	90	3.9	92	52.0	81	27.6	93	57.6	92	2.6	95	83.9	99	35.3	88	91	0.9990	0.2075	0.8462
CHB test1 Case02	32.7	95	6.2	87	63.6	88	84.1	58	71.3	90	2.3	95	63.2	87	41.3	84	86	0.9965	0.2960	0.8138
CHB test1 Case03	48.9	93	11.6	76	50.0	80	87.7	56	75.3	89	12.2	75	33.3	70	90.4	55	74	0.9958	0.2282	0.4562
CHB test1 Case04	50.2	93	5.2	89	72.7	93	81.0	60	76.1	89	3.8	92	72.2	93	50.0	79	86	0.9992	0.1735	0.9193
CHB test1 Case05	36.8	95	8.8	82	48.1	79	92.5	53	74.0	89	3.2	93	69.6	91	71.0	66	81	0.9940	0.1492	0.5690
CHB test1 Case06	69.9	90	6.7	86	16.7	61	93.0	53	68.6	90	6.6	86	27.3	67	90.6	54	73	0.9819	0.0284	0.1001
CHB test1 Case07	62.7	91	5.5	89	48.3	79	85.6	58	77.3	89	4.8	90	55.3	83	74.0	65	80	0.9913	0.1085	0.4856
CHB test1 Case08	49.6	93	2.1	96	81.5	98	74.8	64	66.3	90	2.3	95	61.8	87	62.9	71	87	0.9964	0.3129	0.8219
CHB test1 Case09	56.3	92	3.6	92	34.9	71	69.6	67	63.1	91	2.6	95	35.8	72	55.4	76	82	0.9949	0.2781	0.7949
CHB test1 Case10	3.4	100	7.4	85	68.4	90	93.4	53	49.4	93	3.9	92	79.3	97	76.4	63	84	0.9908	0.1984	0.5123
CHB test1 Case11	16.3	98	6.0	88	47.7	79	91.4	54	72.9	89	3.6	93	51.7	81	81.7	60	80	0.9920	0.1657	0.5317
CHB test1 Case12	86.0	87	7.1	85	12.0	58	81.7	60	86.1	87	6.8	86	15.4	60	77.7	62	73	0.9872	0.0200	0.1141
CHB test1 Case13	50.0	93	5.4	89	50.0	80	85.4	58	69.3	90	4.6	91	76.2	95	64.6	70	83	0.9938	0.3095	0.7506
CHB test1 Case15	71.9	89	2.6	95	53.4	82	44.6	82	62.9	91	1.9	96	57.4	84	49.7	79	87	0.9965	0.2639	0.8947
All Average	55.8	92	7.4	85	49.4	80	76.2	63	86.5	87	8.2	84	57.5	84	69.0	68	80	0.9919	0.2634	0.6089
All UNC	63.3	91	9.5	80	48.5	79	73.6	65	110.5	84	13.5	73	59.7	85	73.4	65	78	0.9895	0.3583	0.6004
All CHB	50.4	93	5.9	88	50.0	80	78.0	62	69.3	90	4.4	91	55.9	83	65.8	70	82	0.9935	0.1957	0.6150

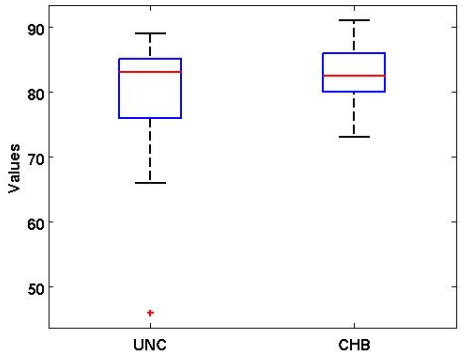


FIG. 5.9 – *Boîtes à moustaches des scores totaux pour les 24 patients en fonction du centre d'acquisition : un point aberrant (avec la valeur 46) est observé.*

TAB. 5.8 – *Exemple de temps d'exécution sur CHB_ test1_ Case01.*

Étape	temps (min)
Définition de la région d'intérêt global	24
Extraction du cerveau	2
Normalisation en intensité	30
Segmentation du cerveau	16
Extraction des lésions à partir du T2-FLAIR	3
Affinement en utilisant les résultats de la classification	3
Limitation de la région d'intérêt	12
Total	90 (1h30)

La table 5.7 donne les résultats de notre méthode sur les données test du challenge. La figure 5.9 présente la répartition des scores totaux pour chaque patient sous forme de boîtes à moustaches en fonction du centre d'acquisition. La table 5.8 donne un exemple de temps d'exécution pour le patient CHB_test1_Case01.

5.4.4 Discussion

L'analyse des données sur les critères pris individuellement est compliquée car ceux-ci sont dépendants entre eux. Par exemple, le taux de vrais positifs (sensibilité de détection) donne le nombre de lésions correctement détectées mais sa valeur ne reflète pas le volume des lésions non segmentées. Pour cette raison, la discussion suivante est basée uniquement sur les scores globaux.

La comparaison entre les deux experts donne un score global de 90/100. Notre méthode avec un résultat global de 80/100 est proche de cette variabilité inter-expert. La boîte à moustaches (figure 5.9) montre que les scores de notre mé-

thode sont entre 66 et 91 et similaires d'un centre à l'autre, excepté pour un point aberrant. En effet, le score de notre méthode est seulement de 46 pour le patient UNC_test1_Case09. Ceci peut être expliqué par le fait que ce patient semble avoir des lésions nécrotiques, un sous-type de lésions qui n'est pas détecté par notre méthode sur les données du challenge. En effet, les lésions nécrotiques sont des signaux hypointenses en T1 mais ne sont pas ou peu visibles en T2-FLAIR. Dans le cas de la base QUALICORE, ces lésions peuvent être détectées car elles sont, le plus souvent, entourées d'une couronne d'hypersignaux mais je n'ai pas observé ce phénomène sur les données du challenge. Le protocole d'acquisition de cette séquence semble différent dans le cas du challenge.

Concernant le patient UNC_test1_Case09, aucune autre lésion n'est de plus observée en T2-FLAIR. En conséquence, la méthode proposée détecte de nombreux faux positifs. Cet effet est aussi observé pour UNC_test1_Case04, pour CHB_test1_Case06 et pour CHB_test1_Case12. Dans ces trois cas, le score de notre méthode est en dessous de 80.

Notre méthode a aussi un score en dessous de 80 pour trois autres séries de données : UNC_test1_Case06, UNC_test1_Case10 et CHB_test1_Case03. Sans la segmentation de référence, la meilleure explication semble être que ces données correspondent à des patients avec une charge lésionnelle faible. Dans ce cas, la moindre erreur de segmentation diminue dramatiquement le score. De plus, même si la résolution d'origine (avant qu'elle soit recalée et rééchantillonnée) de la séquence T2-FLAIR n'a pas été donnée, cette séquence est acquise en général à une résolution plus faible que les autres séquences. Ceci peut diminuer la précision des segmentations obtenues.

Les résultats de STAPLE montrent que notre méthode est plus spécifique que sensible. Ceci révèle une sous-segmentation des lésions. Cet effet peut aussi être observé dans la figure 5.6.

La méthode a besoin d'environ 1 heure 30 minutes pour segmenter les lésions (cf. table 5.8). C'est souvent moins long qu'une segmentation manuelle. Le temps d'exécution de la méthode n'a, de plus, pas été optimisé.

La méthode proposée et optimisée sur les données du challenge permet de segmenter les lésions visibles sur la séquence T2-FLAIR. Cette méthode est donc incapable de détecter les lésions nécrotiques ou sous-tentorielles. Elle est aussi basée sur une bonne estimation des différents paramètres de classes. Dans le cas d'images IRM de patients très atteints par la maladie, on peut supposer que cette estimation peut être biaisée (*i.e.* s'il y a plus de lésions que de SG ces deux classes peuvent s'inverser dans la classification). Toutefois, je n'ai jamais observé cet effet sur les IRM que j'ai segmentées.

5.4.5 Challenge sur site et résultats

Le challenge de segmentation a continué pendant la tenue du workshop. En effet, sept données restaient à segmenter pendant une durée limitée de trois heures. Ces données avaient pu être téléchargées en amont mais restaient inaccessibles avant le

challenge grâce à un mot de passe. Une fois ce dernier donné lors du workshop, j'ai pu exécuter l'algorithme PEMS-COR sur ces données. Les scores globaux des méthodes participantes au challenge ont été fournis (cf. table 5.9). Sur les différentes séries de données PEMS-COR a été classée dans le groupe de tête.

TAB. 5.9 – *Table des résultats au challenge des différentes méthodes en fonction de la série de données traitées.*

1st Author	Site	Rank Pre	Rank Onsite	Rank Final
S Bricq	MIV	1 / 82	9 / 54	5 / 68
JC Souplet PEMS-COR	INRIA	2 / 80	2 / 73	1 / 77
N Sheen	JHU	3 / 80	3 / 69	3 / 75
D Garcia	Visage	4 / 79	5 / 62	4 / 71
P Anbeek	Utrecht	5 / 78	1 / 75	2 / 76
J Morra	UCLA	6 / 77	6 / 61	5 / 68
M Scully	MRN	7 / 67 / 73	7 / 59	7 / 66
M Prastawa	Utah	8 / 66	4 / 65	7 / 66
DJ Kroon	Twente	9 / 66	8 / 56	9 / 61

5.5 Choix de la méthode pour segmenter les lésions

La section précédente montre que PEMS-COR donne de meilleurs résultats que PEMS-G sur les IRM du challenge de segmentation. Cependant, ces IRM n'ont pas été acquises avec le même protocole que les IRM de la base QUALICORE. De plus la séquence DP est maintenant disponible. J'ai donc retesté PEMS-G et PEMS-COMBI (union de PEMS-COR et PEEMS-CER) sur cette base. Leurs résultats ont cette fois été comparés avec les segmentations manuelles fournies par notre expert Mikael Cohen.

5.5.1 Comparaison de deux méthodes automatiques sur la base de données Qualicore

Les tables 5.10 et 5.11 donnent respectivement la sensibilité et la valeur prédictive positive de détection sur 19 sujets de la base QUALICORE et pour les différentes méthodes automatiques de segmentation.

Sur les mêmes données et pour les mêmes méthodes, la table 5.12 donne le pourcentage du volume des lésions fausses positives (FP) par rapport au volume de la segmentation automatique (SA) et la table 5.13 donne le pourcentage du volume des lésions fausses négatives (FN) par rapport au volume de la segmentation de référence (SR).

TAB. 5.10 – *Sensibilité de détection sur 19 sujets de la base QUALICORE pour les différentes méthodes automatiques de segmentation.*

Sujet	Sensibilité détection PEMS-G	Sensibilité détection PEMS-COR	Sensibilité détection PEMS-COMBI
CLE_CELFR_20051028	0,556	0,556	0,556
CLE_CELFR_20061013	0,462	0,538	0,538
CLE_CELFR_20071012	0,111	0,667	0,667
CLE_DELEM_20051028	0,200	0,325	0,325
CLE_DELEM_20061027	0,303	0,424	0,424
CLE_LAMAN_20061124	0,323	0,387	0,387
CLE_PAUCH_20051230	0,333	0,500	0,500
CLE_PAUCH_20070117	0,381	0,762	0,762
CLE_ROCCA_20051202	0,412	0,471	0,500
CLE_ROCCA_20061201	0,318	0,318	0,318
MON_CAKMA_20071022	0,389	0,500	0,556
MON_SAIVE_20060831	0,094	0,219	0,219
MON_SAIVE_20070618	0,069	0,172	0,172
NIC_ALLJE_20051209	0,417	0,583	0,583
NIC_ALLJE_20061006	0,182	0,545	0,545
NIC_CALCE_20060512	0,714	0,571	0,571
NIC_CALCE_20070525	0,640	0,680	0,680
NIC_DAZIN_20040910	0,484	0,548	0,581
NIC_DRAJA_20060120	0,359	0,478	0,478
Moyenne	0,355	0,487	0,493
Ecart type	0,176	0,152	0,154

TAB. 5.11 – *Valeur prédictive positive (VPP) de détection sur 19 sujets de la base QUALICORE pour les différentes méthodes automatiques de segmentation.*

Sujet	VPP détection PEMS-G	VPP détection PEMS-COR	VPP détection PEMS-COMBI
CLE_CELFR_20051028	0,200	0,147	0,143
CLE_CELFR_20061013	0,273	0,135	0,127
CLE_CELFR_20071012	1,000	0,159	0,149
CLE_DELEM_20051028	0,929	0,848	0,848
CLE_DELEM_20061027	0,976	0,724	0,724
CLE_LAMAN_20061124	0,323	0,250	0,250
CLE_PAUCH_20051230	0,258	0,250	0,242
CLE_PAUCH_20070117	0,409	0,325	0,325
CLE_ROCCA_20051202	0,514	0,424	0,441
CLE_ROCCA_20061201	0,556	0,480	0,462
MON_CAKMA_20071022	0,533	0,600	0,588
MON_SAIVE_20060831	0,333	0,444	0,400
MON_SAIVE_20070618	0,286	0,250	0,227
NIC_ALLJE_20051209	0,345	0,324	0,300
NIC_ALLJE_20061006	0,100	0,115	0,105
NIC_CALCE_20060512	0,407	0,205	0,195
NIC_CALCE_20070525	0,565	0,219	0,206
NIC_DAZIN_20040910	0,717	0,450	0,455
NIC_DRAJA_20060120	0,889	0,750	0,750
Moyenne	0,506	0,374	0,365
Ecart type	0,276	0,222	0,225

TAB. 5.12 – *Pourcentage du volume des lésions fausses positives (FP) par rapport au volume de la segmentation automatique (SA) sur 19 sujets de la base QUALICORE pour les différentes méthodes automatiques de segmentation.*

Sujet	% du volume des lésions FP par rapport au volume de la SA, PEMS-G	% du volume des lésions FP par rapport au volume de la SA, PEMS-COR	% du volume des lésions FP par rapport au volume de la SA, PEMS-COMBI
CLE_CELFR_20051028	40,694	47,561	47,913
CLE_CELFR_20061013	30,945	64,430	65,438
CLE_CELFR_20071012	0,000	49,386	50,527
CLE_DELEM_20051028	0,364	2,133	2,133
CLE_DELEM_20061027	0,459	4,264	4,264
CLE_LAMAN_20061124	24,769	33,363	33,363
CLE_PAUCH_20051230	40,619	26,265	26,591
CLE_PAUCH_20070117	30,640	36,765	36,765
CLE_ROCCA_20051202	18,127	36,644	36,234
CLE_ROCCA_20061201	17,857	29,725	31,101
MON_CAKMA_20071022	23,448	19,279	19,573
MON_SAIVE_20060831	44,690	25,520	28,429
MON_SAIVE_20070618	33,493	47,749	53,073
NIC_ALLJE_20051209	72,623	34,426	35,863
NIC_ALLJE_20061006	91,837	56,597	59,331
NIC_CALCE_20060512	12,060	27,907	29,181
NIC_CALCE_20070525	7,897	20,904	26,643
NIC_DAZIN_20040910	16,546	16,053	16,465
NIC_DRAJA_20060120	3,932	9,007	9,007
Moyenne	26,895	30,946	32,205
Ecart type	24,260	17,169	17,681

TAB. 5.13 – *Pourcentage du volume des lésions fausses négatives (FN) par rapport au volume de la segmentation de référence (SR) sur 19 sujets de la base QUALICORE pour les différentes méthodes automatiques de segmentation.*

Sujet	% du volume des lésions FN par rapport au volume de la SR, PEMS-G	% du volume des lésions FN par rapport au volume de la SR, PEMS-COR	% du volume des lésions FN par rapport au volume de la SR, PEMS-COMBI
CLE_CELFR_20051028	14,268	37,485	37,485
CLE_CELFR_20061013	16,916	32,120	32,120
CLE_CELFR_20071012	46,337	17,123	17,123
CLE_DELEM_20051028	4,861	3,954	3,954
CLE_DELEM_20061027	5,688	5,353	5,353
CLE_LAMAN_20061124	22,431	29,572	29,572
CLE_PAUCH_20051230	25,819	19,655	19,655
CLE_PAUCH_20070117	24,428	6,533	6,533
CLE_ROCCA_20051202	11,500	13,724	12,534
CLE_ROCCA_20061201	20,441	24,956	24,956
MON_CAKMA_20071022	21,617	31,836	28,579
MON_SAIVE_20060831	44,439	24,248	24,248
MON_SAIVE_20070618	88,505	37,172	37,172
NIC_ALLJE_20051209	21,885	10,965	10,965
NIC_ALLJE_20061006	42,128	18,339	18,339
NIC_CALCE_20060512	1,172	11,484	11,484
NIC_CALCE_20070525	6,282	6,619	6,619
NIC_DAZIN_20040910	11,342	11,468	8,962
NIC_DRAJA_20060120	17,902	20,536	20,536
Moyenne	23,577	19,113	18,747
Ecart type	20,367	10,845	10,804

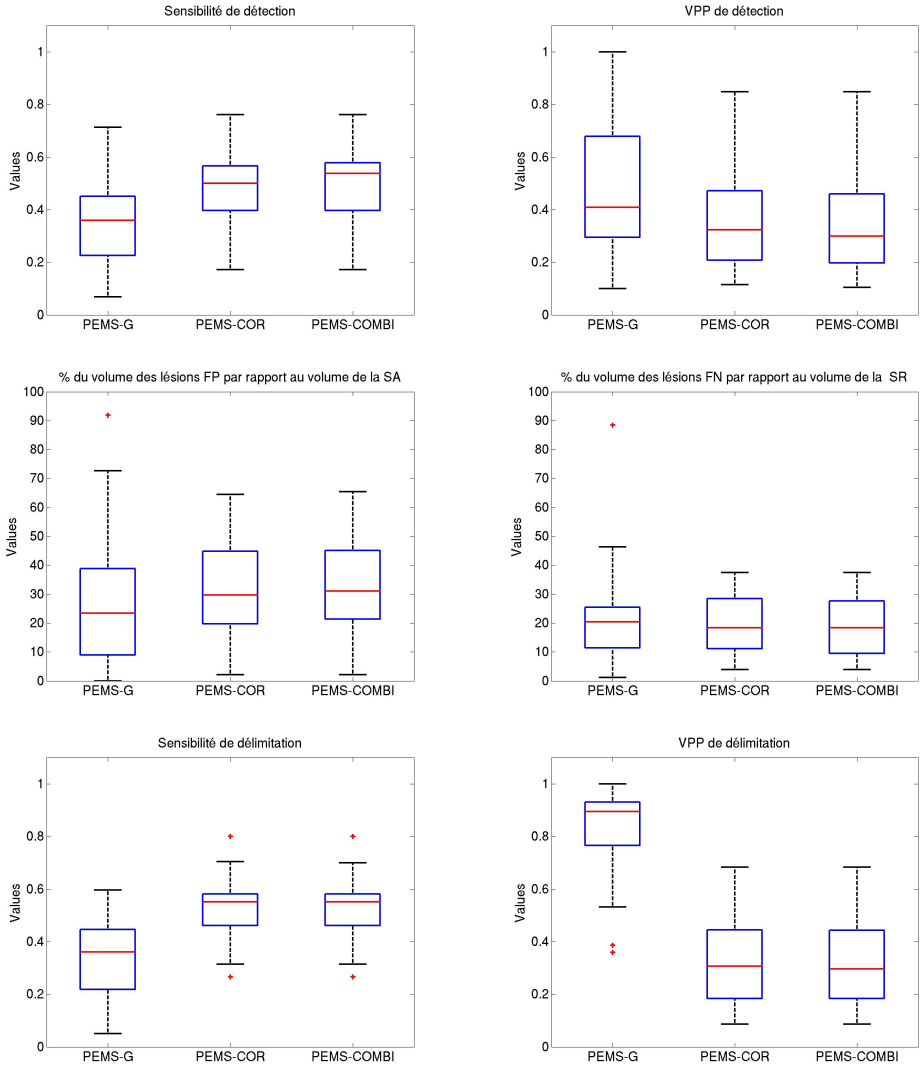
TAB. 5.14 – *Sensibilité de délimitation sur 19 sujets de la base QUALICORE pour les différentes méthodes automatiques de segmentation.*

Sujet	Sensibilité délimitation PEMS-G	Sensibilité délimitation PEMS-COR	Sensibilité délimitation PEMS-COMBI
CLE_CELFR_20051028	0,597	0,801	0,801
CLE_CELFR_20061013	0,460	0,565	0,565
CLE_CELFR_20071012	0,077	0,573	0,573
CLE_DELEM_20051028	0,407	0,460	0,460
CLE_DELEM_20061027	0,371	0,644	0,644
CLE_LAMAN_20061124	0,486	0,466	0,466
CLE_PAUCH_20051230	0,356	0,552	0,552
CLE_PAUCH_20070117	0,291	0,506	0,506
CLE_ROCCA_20051202	0,361	0,439	0,436
CLE_ROCCA_20061201	0,364	0,582	0,582
MON_CAKMA_20071022	0,277	0,509	0,518
MON_SAIVE_20060831	0,050	0,339	0,339
MON_SAIVE_20070618	0,216	0,267	0,267
NIC_ALLJE_20051209	0,224	0,632	0,632
NIC_ALLJE_20061006	0,111	0,515	0,515
NIC_CALCE_20060512	0,546	0,314	0,314
NIC_CALCE_20070525	0,477	0,579	0,579
NIC_DAZIN_20040910	0,396	0,704	0,700
NIC_DRAJA_20060120	0,178	0,576	0,576
Moyenne	0,329	0,527	0,528
Ecart type	0,156	0,131	0,131

TAB. 5.15 – *Valeur prédictive positive (VPP) de délimitation sur 19 sujets de la base QUALICORE pour les différentes méthodes automatiques de segmentation.*

Sujet	VPP délimitation PEMS-G	VPP délimitation PEMS-COR	VPP délimitation PEMS-COMBI
CLE_CELFR_20051028	0,359	0,121	0,121
CLE_CELFR_20061013	0,386	0,211	0,211
CLE_CELFR_20071012	1,000	0,145	0,145
CLE_DELEM_20051028	0,972	0,684	0,684
CLE_DELEM_20061027	0,934	0,601	0,601
CLE_LAMAN_20061124	0,895	0,199	0,199
CLE_PAUCH_20051230	0,532	0,129	0,129
CLE_PAUCH_20070117	0,750	0,255	0,255
CLE_ROCCA_20051202	0,953	0,390	0,386
CLE_ROCCA_20061201	0,923	0,514	0,514
MON_CAKMA_20071022	0,872	0,307	0,296
MON_SAIVE_20060831	0,984	0,463	0,463
MON_SAIVE_20070618	0,856	0,464	0,464
NIC_ALLJE_20051209	0,902	0,179	0,179
NIC_ALLJE_20061006	0,925	0,086	0,086
NIC_CALCE_20060512	0,768	0,349	0,349
NIC_CALCE_20070525	0,766	0,220	0,220
NIC_DAZIN_20040910	0,826	0,375	0,377
NIC_DRAJA_20060120	0,903	0,379	0,379
Moyenne	0,816	0,319	0,319
Ecart type	0,190	0,171	0,171

FIG. 5.10 – *Boîtes à moustaches reprenant les résultats des tables 5.10, 5.11, 5.12,5.13, 5.14 et 5.15. Les notations utilisées sont reprises de ces tables.*



Pour finir, les tables 5.14 et 5.15 donnent respectivement la sensibilité et la valeur prédictive positive de délimitation.

Sur la base QUALICORE, PEMS-COR ou PEMS-COMBI sont plus sensibles que PEMS-G (gain moyen de 13,8%). De même, le volume des lésions non détectées ne représente plus que 18,7% du volume de la segmentation de référence avec PEMS-COMBI contre 24,6% avec PEMS-G. En revanche, ce gain se fait au détriment de la valeur prédictive positive et du pourcentage du volume des lésions fausses positives par rapport au volume de la segmentation automatique qui passe de 26,9% à 32.2% (perte de 5,3%).

Cet effet se retrouve aussi dans la délimitation. Un gain de 20% est constaté entre PEMS-G et PEMS-COMBI pour la sensibilité mais dans le même temps la valeur prédictive positive chute de 50%.

Globalement (cf. figure 5.10), la segmentation des lésions du cervelet améliore la sensibilité de détection des lésions et diminue la valeur prédictive positive. Elle diminue aussi le volume des lésions non détectées même si le volume des faux positifs augmente. Cependant, toutes ces évolutions restent faibles (proche de 1%). L'ajout de la segmentation du cervelet n'a quasiment aucune influence sur la qualité de la délimitation des lésions. Ces faibles variations s'expliquent par le fait que le nombre de lésions du cervelet est en général bien plus petit que le nombre de lésions de la zone cortex et que les lésions du cervelet sont aussi de volume moyen plus faible.

5.5.2 Discussion

Sur les données d'entraînement du challenge, la méthode PEMS-COR s'est révélée meilleure que PEMS-G. PEMS-COR a eu de bons résultats sur les diverses données de ce challenge puisqu'elle y a été classée à chaque fois dans le groupe de tête.

Les images de la base QUALICORE ont toutefois été acquises avec un protocole différent de celui du challenge. La comparaison entre les résultats des méthodes automatiques et une segmentation manuelle faite par notre expert Mikael Cohen démontre que PEMS-COMBI est la méthode la plus sensible aussi bien en détection qu'en délimitation. Elle permet de détecter une lésion sur deux contre une lésion sur trois avec PEMS-G. Les lésions détectées représentent maintenant 81,6 % de la charge lésionnelle contre seulement 76,4% avec PEMS-G.

Le gain en sensibilité de PEMS-COMBI s'est fait au détriment de la valeur prédictive positive. La diminution de cette valeur peut être forte. Par exemple, elle est de 50% dans le cas de la délimitation. Cependant, les résultats de délimitation de PEMS-COMBI se font sur plus de lésions que ceux de PEMS-G. En effet, la première méthode est plus sensible que la seconde (cf. paragraphe précédent). Cette perte en VPP de délimitation est donc relative car ces valeurs ne sont pas obtenues sur les mêmes lésions. À l'opposé, le gain en sensibilité de délimitation est d'autant plus remarquable parce qu'il est fait sur plus de lésions. Ces résultats font aussi, peut être, apparaître un compromis entre les valeurs de détection et celles de délimitation.

Pour effectuer l'analyse complète de la base QUALICORE (qui reste l'un des

buts de cette thèse), PEMS-COMBI a donc été préférée pour sa sensibilité plus forte en terme de détection et de délimitation et pour le fait qu'elle détecte les lésions du cervelet. Globalement, ses résultats semblent aussi plus homogènes car les écarts types des diverses valeurs sont égaux ou inférieurs à ceux de PEMS-G. Des exemples de segmentations obtenues avec PEMS-COMBI sont donnés en annexe B.

Chapitre 6

Évaluation de l'atrophie

Sommaire

6.1 État de l'art 107
6.1.1 Mesures linéaires 107
6.1.2 Mesures de surfaces ou de volumes 108
6.1.3 Mesures de variations 109
6.1.4 Discussion 110
6.2 Amélioration de la reproductivité des mesures 2D 110
6.2.1 Protocole de mesure 110
6.2.2 Reproductibilité des méthodes 112
6.3 Deux nouvelles méthodes d'évaluation de la FPC 113
6.3.1 La méthode CLASS 113
6.3.2 La méthode LCLASS 114
6.3.3 Validation 115
6.3.4 Apport des méthodes 118
6.4 Comparaison et choix de méthodes d'évaluation de l'atrophie 119
6.4.1 Analyse statistique 119
6.4.2 Comparaison sur des données simulées 120
6.4.3 Comparaison sur des données réelles 121
6.4.4 Choix des méthodes pour analyser la base Qualicore 125
6.4.5 Limites des méthodes choisies sur les images de notre base . . 126

Résumé : Dans la recherche sur la SEP, le niveau de la maladie et l'efficacité d'un traitement sont principalement évalués avec la charge lésionnelle et l'atrophie. La première étant pauvrement corrélée avec le handicap du patient, une évaluation précise de la seconde est d'autant plus importante. De très nombreuses méthodes proposées pour réaliser cette évaluation sont présentes dans la littérature.

Les premières études comparatives de ces différentes méthodes exposent la difficulté de leur validation, l'absence de méthode de référence et aussi le fait que leurs résultats peuvent être contradictoires. Parmi ces mesures, l'évaluation de la fraction parenchymateuse cérébrale (FPC) est possible à partir de la segmentation des compartiments cérébraux. Cette technique a déjà été utilisée à partir de différents algorithmes de segmentation mais en général en se limitant à un certain nombre de coupes des images. Je propose de la calculer, grâce à la chaîne de traitement EMA, sur l'ensemble de l'image excepté certains points considérés comme aberrants. Cette méthode est appelée CLASS dans ce manuscrit. Cependant, l'évaluation de la fraction parenchymateuse est supposée être moins précise que certaines méthodes sur les études longitudinales. Je propose une deuxième chaîne de traitement, LCLASS, qui évalue la FPC en prenant en compte les informations longitudinales pour réaliser la classification. Comme la validation de ces divers algorithmes est difficile, ceux-ci ont été comparés entre eux sur des données simulées mais aussi sur les données de la base QUALICORE.

Mots clefs : Atrophie, État de l'art, Fraction parenchymateuse, Algorithme automatique.

Contributions :

- État de l'art des mesures d'atrophie.
- Développement d'une méthode automatique de sélection de coupe 2D pour réaliser des mesures linéaires.
- Développement de deux nouvelles méthodes d'évaluation de la fraction parenchymateuse.
- Comparaison des différentes méthodes sur des données simulées et des données réelles.

Organisation du chapitre : Une description de différentes méthodes d'évaluation de l'atrophie cérébrale est donnée dans la section 6.1. La section 6.2 présente une nouvelle méthode pour choisir automatiquement les coupes 2D sur lesquelles effectuer des mesures linéaires. Ensuite, deux nouvelles chaînes de traitements entièrement automatiques (CLASS et LCLASS) pour évaluer la fraction parenchymateuse cérébrale (FPC) sont proposées dans la section 6.3. Enfin, une comparaison des diverses méthodes est menée sur des données simulées et sur les données de notre base (section 6.4).

6.1 État de l'art

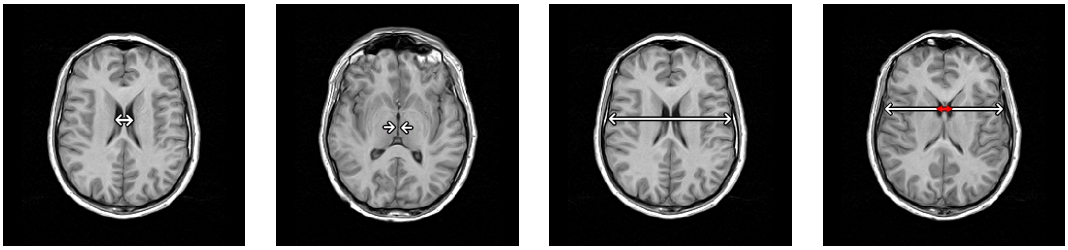
Selon les critères de McDonald modifiés [Polman 2005], le nombre de lésions visibles en IRM et leur localisation sont nécessaires pour poser le diagnostic de SEP. Cette mesure de charge lésionnelle est aussi utilisée dans des études de suivi des patients et de recherches pharmaceutiques pour évaluer l'effet d'un médicament. Cependant, excepté dans le cas du syndrome clinique isolé [Korteweg 2006], des études cliniques ont montré que cette mesure est peu corrélée avec le handicap du patient (cf. section 2.3.2 et chapitre 5). En conséquence, d'autres approches comme la mesure de l'atrophie globale du cerveau sont étudiées.

La littérature offre de très nombreuses méthodes pour évaluer l'atrophie. Un état de l'art de ces dernières est proposé dans [Miller 2002, Bermel 2006, Anderson 2006, Stefano 2007]. Elles peuvent être manuelles, semi-automatiques ou entièrement automatiques et être classées en trois catégories : mesures linéaires (1D), mesures réalisées à partir d'une segmentation et mesures d'une évolution (d'une déformation) entre deux images.

6.1.1 Mesures linéaires

Les mesures linéaires sont les mesures les plus simples à mettre en œuvre. Il s'agit de mesurer la distance séparant deux points spécifiques sur une coupe 2D d'une séquence IRM. Suivant les auteurs, différentes mesures ont été définies (liste non-exhaustive) :

- La largeur des ventricules latéraux (figure 6.1, a), largeur à mi-hauteur des ventricules sur la coupe où l'espace inter-ventriculaire est le plus fin.
- La largeur du troisième ventricule (figure 6.1, b), largeur à mi-hauteur du troisième ventricule sur la coupe où il se voit le mieux.
- La largeur du cerveau (figure 6.1, c), largeur du cerveau à mi-hauteur des ventricules sur la coupe où l'espace inter-ventriculaire est le plus fin.
- Le rapport bicaudal (figure 6.1, d), rapport de la distance entre les noyaux caudaux sur la largeur du cerveau à la même hauteur. Ces distances sont mesurées sur la coupe où les noyaux caudaux se voient le mieux.



(a) (b) (c) (d)

FIG. 6.1 – *Exemple de mesures linéaires : (a) largeur des ventricules latéraux, (b) largeur du troisième ventricule, (c) largeur du cerveau, (d) rapport bicaudal.*

Ces mesures ont été utilisées dans de nombreuses études [Simon 1999, Turner 2001, Bermel 2002, Sharma 2004, Butzkueven 2008]. Leur difficulté majeure est leur reproductibilité réduite. Par exemple, le choix de la coupe 2D utilisé répond souvent à la définition : "coupe où le compartiment mesuré est le plus visible", ce qui le rend très subjectif.

6.1.2 Mesures de surfaces ou de volumes

La plupart des méthodes d'évaluation de l'atrophie se basent sur une segmentation du cerveau en différents compartiments. Suivant le cas, divers algorithmes de segmentation peuvent être utilisés. Cette segmentation peut aussi se faire seulement sur une coupe 2D, sur une sélection de coupes 2D ou sur un volume 3D complet.

À partir des segmentations, diverses mesures peuvent être réalisées (*i.e.* la surface du corpus callosum [Simon 1987, Barkhof 1998, Lin 2008], le volume cérébral central (volume du cerveau sur une sélection de coupes 2D centrales) [Losseff 1996, Furby 2008], le volume des différents compartiments cérébraux (LCR, SG, SB) [Dastidar 1999, Quarantelli 2003, Horakova 2008], ...). Dans le but de normaliser ces mesures, les séquences peuvent être d'abord recalées dans un espace standard avant de calculer la segmentation. Par exemple, SIENAX est une méthode automatique pour évaluer les volumes des différents compartiments cérébraux [Smith 2002a]. Elle commence par extraire le cerveau de l'image. Puis, une classification des différents compartiments est réalisée sur l'image du cerveau extraite [Zhang 2001]. Le volume correspondant aux différentes segmentations est calculé. Dans un même temps, la méthode recale l'image dans l'espace de l'atlas "MNI152" [Mazziotta 2001]. Ce recalage permet de calculer un facteur de normalisation à appliquer sur les volumes obtenus pour obtenir un volume cérébral normalisé.

La normalisation des volumes mesurés peut aussi être obtenue en les divisant par un autre volume. Par exemple, la fraction parenchymateuse cérébrale (FPC) (Eq. 6.1) est le rapport du volume des tissus cérébraux (SB+SG) sur le volume inclus dans la surface du cerveau (SB+SG+LCR des ventricules+LCR des sillons corticaux) [Rudick 1999, Guttmann 2000, Ge 2002, Kassubek 2003, Juengling 2003]. Une approche similaire à cette FPC est le BICCR [Collins 2001]. Ce dernier repose sur le même principe sauf qu'il prend également en compte le LCR présent entre la surface du cerveau et la dure mère.

$$FPC = 100 \times \frac{\text{Volume du cerveau}}{\text{Volume du cerveau} + \text{Volume de LCR}} \tag{6.1}$$

Dans le cas d'études longitudinales, ces mesures permettent de suivre l'évolution de volume de différents compartiments. De plus, si elles sont normalisées, des comparaisons de séries de patients peuvent être effectuées. Par exemple, on peut différencier des patients traités aux interférons $\beta - 1a$ de patients sous placébo [Rudick 1999]. Néanmoins, ces méthodes se basant sur un algorithme de segmentation, leurs résultats dépendent de la précision de cet algorithme. Elles semblent donc être moins robustes dans le cas d'un suivi longitudinal [Bermel 2002, Stefano 2007].

6.1.3 Mesures de variations

Plutôt que de mesurer un volume à différents instants et de calculer ses variations en fonction de ces résultats, certaines méthodes quantifient directement les variations observées entre deux images. Après un recalage des images acquises à deux instants différents, ces méthodes cherchent les régions où l'intensité de l'image a changé. Par exemple, la méthode de "Boundary Shift Integral" (BSI) permet de calculer les variations d'un volume à partir des déplacements de la frontière de ce volume sur deux instants. En partant du tracé manuel du volume recherché sur les deux instants, la méthode calcule une région d'intérêt, appelée frontière, entre le volume et l'extérieur. Le suivi de l'intensité des voxels présents le long d'une ligne passant au travers de cette frontière montre une transition de l'intensité des voxels entre deux valeurs. Dans le cas du deuxième instant, cette transition est décalée car la frontière s'est déplacée (cf. figure 6.2). La variation globale de volume est alors obtenue en intégrant les décalages de cette transition sur l'ensemble du volume. Cette méthode a été appliquée aux mesures de variations de volume du cerveau et des ventricules. Elle est présentée dans [Freeborough 1997].

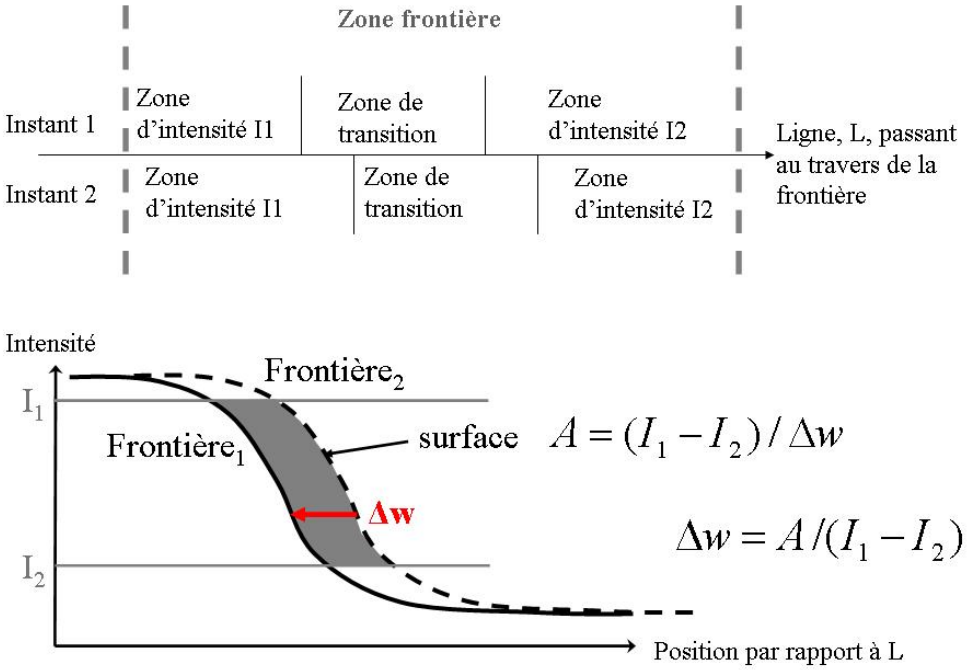


FIG. 6.2 – *Calcul des variations de surface à partir de la variation d'intensité des voxels situés le long d'une ligne passant au travers de la frontière, méthode "Boundary Shift Integral" [Freeborough 1997].*

Une méthode proche de celle-ci est SIENA [Smith 2002a]. Cette méthode calcule le pourcentage de variation du volume cérébral (PVVC) entre deux instants. SIENA commence par extraire le cerveau sur chacune des images et par recaler celles-ci dans un espace situé à mi-distance de chacune d'elles. Puis, une segmentation en tissus est réalisée [Zhang 2001] pour trouver les points situés à la limite cerveau/non-cerveau. SIENA estime ensuite le déplacement perpendiculaire de ces points entre les deux instants. Enfin, le déplacement moyen est converti en un pourcentage global de variation du volume cérébral.

6.1.4 Discussion

Toutes ces mesures sont disponibles pour évaluer l'atrophie sur la base QUALICORE. Cependant, leurs résultats peuvent être contradictoires [Zivadinov 2005] et aucune méthode ne s'est imposée comme méthode de référence. En outre, la validation de ces différentes méthodes est rendue complexe par l'absence de vérité terrain. Dans ces conditions, il semble nécessaire d'améliorer la reproductibilité des méthodes linéaires, pour qu'elles puissent servir de mesures de référence. De même, la mesure de FPC est présentée comme moins fiable que les mesures de variation pour des études longitudinales. Peut-on améliorer cette mesure ?

6.2 Amélioration de la reproductivité des mesures 2D

Parmi les mesures présentées dans la section 6.1.1, trois mesures ont été sélectionnées : largeur du cerveau, largeur des ventricules latéraux et largeur du troisième ventricule. Pour améliorer leur reproductibilité, un protocole de mesure a été mis en place puis testé sur notre base de données.

6.2.1 Protocole de mesure

L'amélioration majeure de ce protocole réside dans un choix automatique de la coupe sur laquelle est effectuée la mesure. Le protocole se base ensuite sur une approche présente dans [Benedict 2004] pour le positionnement des points de mesures.

6.2.1.1 Sélection de la coupe 2D

Les mesures linéaires dépendent du choix de la coupe 2D utilisée [Bermel 2002, Benedict 2004, Anderson 2006]. Cette sélection peut donc varier d'un expert à l'autre ou même pour un même expert d'un instant à l'autre. Pour répondre à ce problème, ce choix a été automatisé. Premièrement, une image de référence est sélectionnée parmi les images de la base de données. Sur cette image, les deux coupes 2D (où le troisième ventricule et les ventricules latéraux sont les plus visibles) sont manuellement sélectionnées. La première coupe est utilisée pour mesurer la largeur du troisième ventricule ; la seconde pour mesurer la largeur du cerveau et la largeur des ventricules latéraux.

Dans un second temps, les séquences IRM d'un patient (de même modalité que l'image de référence et acquises à des instants différents) sont recalées sur cette image de référence. A chaque instant, les deux coupes de ces séquences correspondant aux deux coupes de références sont utilisées pour réaliser la mesure (cf. figure 6.3).

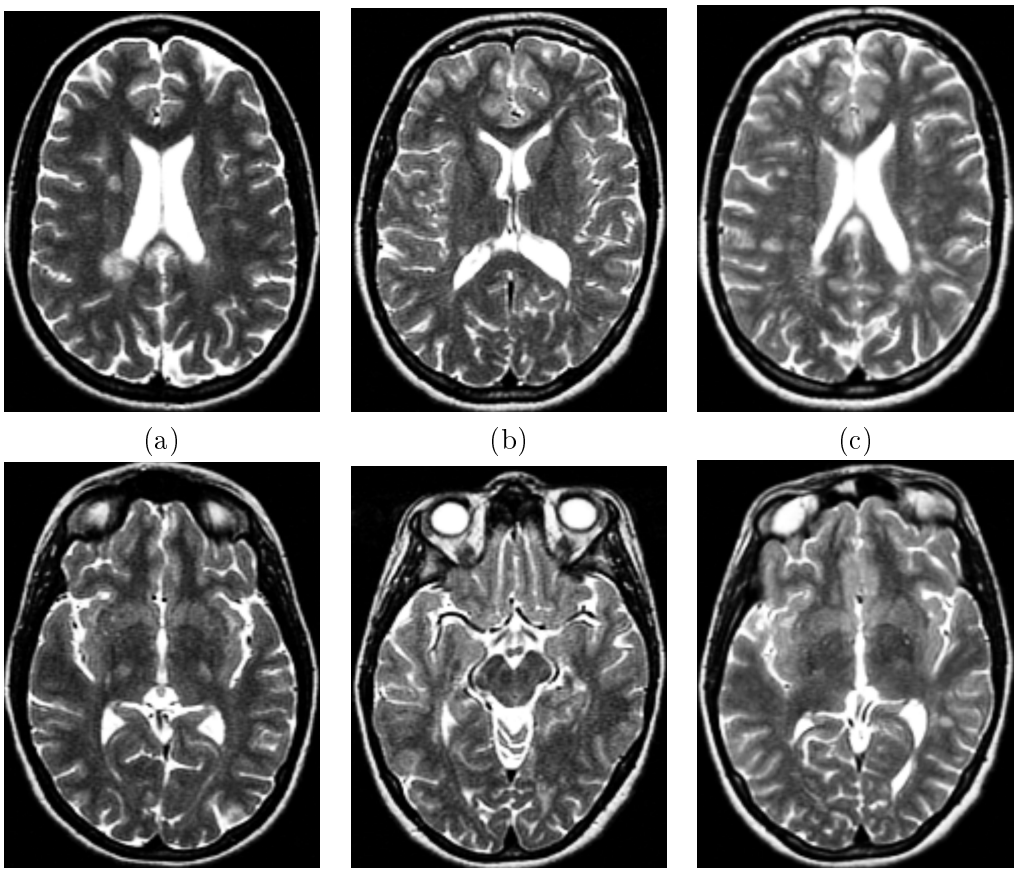
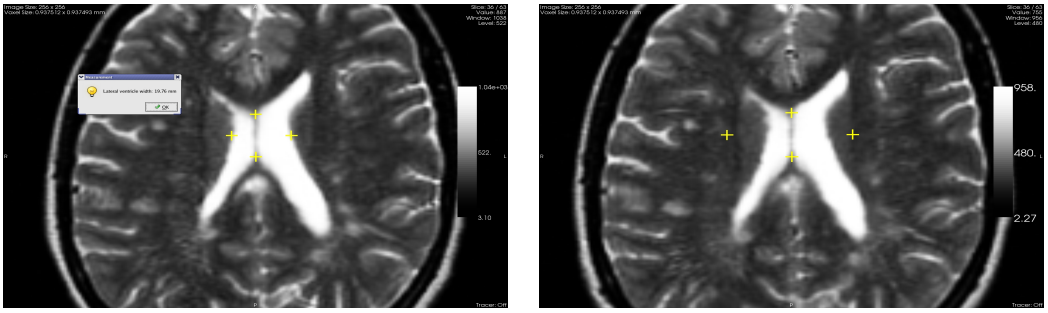


FIG. 6.3 – *Exemple de coupes 2D choisies automatiquement pour mesurer la largeur des ventricules latéraux ((a),(b),(c)) et du troisième ventricule ((d),(e),(f)) : (a) et (d) coupes sélectionnées manuellement à partir de l'image de référence, (b) et (e) coupes correspondantes dans l'image à traiter avant le recalage, (c) et (f) coupes sélectionnées automatiquement à partir de l'image à traiter.*

Dans cette approche, l'expert n'intervient que pour choisir l'image et les coupes de référence. Ce choix est conservé pour traiter tous les patients de la base de données. Si un autre expert veut analyser cette même base de données, il peut s'il le souhaite garder ces paramètres. Dans le cas d'un nouveau protocole d'acquisition des images avec une résolution différente, l'image de référence peut être changée pour avoir une résolution plus proche de celle des images à analyser et ainsi diminuer l'effet de volume partiel.

6.2.1.2 Positionnement de points et calcul de la distance

Sur les coupes sélectionnées, l'expert doit identifier des points spécifiques. Comme dans [Benedict 2004], il doit définir deux points de référence situés l'un en haut et l'autre en bas du/des ventricule(s) le long de la fissure interhémisphérique. Puis, deux autres points de mesures peuvent être déplacés, le long d'une ligne perpendiculaire en son milieu au segment défini par les points de référence. Ces points de mesure doivent être placés l'un à la limite gauche et l'autre à la limite droite du/des ventricule(s) ou du cerveau. La distance entre ces points de mesure est alors calculée et donne la mesure linéaire recherchée (cf. figure 6.4, a). Dans le cas où différents instants sont disponibles, les points de référence utilisés pour le premier instant sont automatiquement repositionnés sur les images des instants suivants (figure 6.4, b). L'expert reste libre de les changer.



(a) (b)

FIG. 6.4 – *Principe de mesure de la largeur des ventricules latéraux : (a) l'expert définit les deux points de référence ainsi que les deux points de mesure, (b) les deux points de référence sont automatiquement repositionnés sur l'image du deuxième instant et peuvent être réutilisés.*

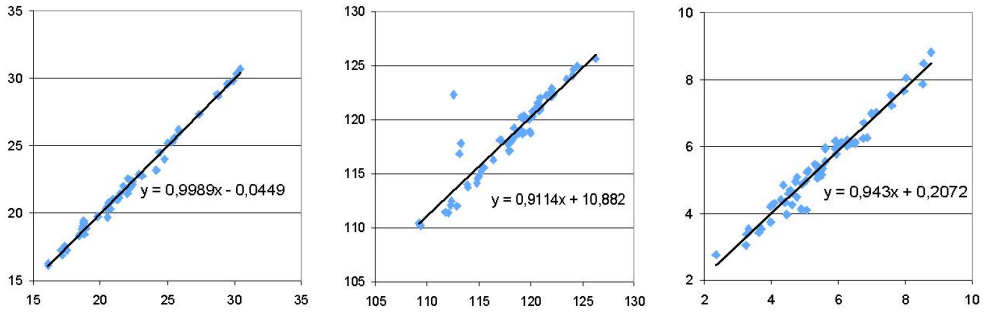
Ces trois mesures (largeur du cerveau, largeur des ventricules latéraux et largeur du troisième ventricule) ont été rendues disponibles dans le logiciel SepINRIA[1] (cf. section 7.2.1).

6.2.2 Reproductibilité des méthodes

Malgré le protocole précédent, les mesures utilisées sont manuelles. En conséquence, elles sont sujettes à une variabilité intra- et inter-experts. Grâce à deux séries de mesures effectuées à plus d'une semaine d'intervalle par Mikael Cohen, un interne en neurologie, nous avons pu tester la variabilité intra-expert de ces mesures. Ces dernières ont été effectuées sur 25 patients de notre base. La corrélation entre la première et la seconde série de mesures est donnée dans la figure 6.5. On peut voir que ces coefficients sont élevés. Ils sont aussi similaires à ceux trouvés dans d'autres études [Butzkueven 2008]. Néanmoins, il reste difficile de quantifier l'amélioration

[1] http ://www-sop.inria.fr/asclepios/software/SepINRIA/

apportée à la mesure (par le choix automatique de la coupe) car les images de ces études sont de résolutions différentes des nôtres.



(a) (b) (c)

FIG. 6.5 – *Coefficient de corrélation entre les deux séries de mesures linéaires : (a) largeur des ventricules latéraux (R=0.99), (b) largeur du cerveau (R=0.92), (c) largeur du troisième ventricule (R=0.98).*

6.3 Deux nouvelles méthodes d'évaluation de la FPC

Le calcul de la FPC (cf. section 6.1.2) nécessite une segmentation précise du cerveau et du LCR. Cependant, des artéfacts comme l'effet de volume partiel (VP) peuvent gêner cette classification. Selon divers articles [Bermel 2002, Stefano 2007], l'évaluation de l'atrophie par la FPC peut aussi être moins précise dans les études longitudinales. Pour répondre à ces problèmes, j'ai proposé deux nouvelles méthodes d'évaluation de la FPC. La première, CLASS, se base sur la chaîne de traitement présentée dans le chapitre 3. La seconde, LCLASS, inclut des modifications de cette chaîne pour améliorer sa fiabilité dans les études longitudinales.

6.3.1 La méthode CLASS

La méthode proposée évalue l'atrophie à partir des variations de la fraction parenchymateuse cérébrale. Cette dernière est calculée à partir des volumes des segmentations obtenues par la chaîne de traitement EMA (cf. chapitre 3).

6.3.1.1 Utilisation de la chaîne de traitement EMA

Le chapitre 3 a introduit la chaîne de traitement EMA. Celle-ci segmente les différents compartiments cérébraux. EMA possède de nombreux avantages pour permettre une évaluation optimale de la fraction parenchymateuse. Après une phase de prétraitements des images, un algorithme de segmentation fiable et robuste à la présence de points aberrants (*i.e.* des lésions) a été défini. De plus, ce dernier présente l'avantage de posséder un modèle de volumes partiels SG-LCR. Grâce à ce modèle, la classification des voxels situés à l'interface entre la SG et le LCR (malgré

les volumes partiels induits par la minceur des sillons corticaux) est espérée plus précise puisque ceux-ci sont classés majoritairement comme volumes partiels.

Pour donner une segmentation la plus fiable possible et reproductible, le paramètre de pourcentage des points considérés (PPC) pour calculer les paramètres des classes a été fixé à 100% dans ce cas.

6.3.1.2 Calcul du volume et de la fraction parenchymateuse cérébrale (FPC)

Pour prendre en compte les volumes partiels dans le calcul des volumes, nous avons généré les cartes de répartitions du LCR et du cerveau (SG+SB). Ces cartes ne sont pas des segmentations probabilistes mais donnent la proportion de chaque compartiment dans chaque voxel. Les cartes de répartition du LCR et du cerveau ($CR_{(LCR)}$, $CR_{(C)}$) sont obtenues respectivement par les équations 6.2 et 6.3 où $SEG_{(VP\alpha)}$ représente la segmentation des volumes partiels avec la proportion α de SG.

$$CR_{(LCR)} = \sum_{i=1}^{6} \frac{7-i}{7} \times SEG_{VPi} + SEG_{LCR} \tag{6.2}$$

$$CR_{(C)} = \sum_{i=1}^{6} \frac{i}{7} \times SEG_{VPi} + SEG_{SG} + SEG_{SB} \tag{6.3}$$

Les volumes du LCR et du cerveau sont alors obtenus par l'addition de toutes les valeurs des voxels de la carte de répartition correspondante, multipliée par le volume d'un voxel. Ces volumes donnent alors la FPC (cf. équation 6.1).

6.3.1.3 Calcul de l'atrophie

La FPC permet de comparer différents groupes de populations (par exemple : les patients atteints de SEP avec des sujets sains). Dans le cas où pour un même patient différents examens (réalisés à différents instants) sont disponibles, l'atrophie est donnée par la différence des fractions parenchymateuses successives. Dans ce cas, CLASS est appliquée indépendamment sur les images de chaque instant.

6.3.2 La méthode LCLASS

La méthode LCLASS est une évolution de CLASS prenant en compte les acquisitions successives pour améliorer la robustesse de la classification.

Pour permettre une prise en compte des informations longitudinales, certaines parties de la chaîne de traitement EMA doivent être changées :

- Toutes les images (de tous les instants) doivent être recalées sur la séquence T2 du premier instant.
- Pour chaque séquence IRM (T1, T2 ou DP), les intensités des images doivent être égalisées entre les différents instants.

- Le masque du cerveau obtenu à partir du premier instant est aussi utilisé pour les instants suivants.
- Dans la classification par l'algorithme EM, les paramètres des classes sont calculés à partir des images de chaque instant disponible. Ceci permet de calculer les paramètres des classes en prenant en compte les informations apportées par les images de tous les instants en même temps et non pas indépendamment comme dans CLASS. Cette approche issue de [Aït-Ali 2005] repose sur le principe que plus de points sont disponibles mieux les classes sont définies. Les paramètres obtenus sont alors utilisés pour classer les voxels de chaque instant.
- Comme les points aberrants ne sont pas considérés dans le calcul de la FPC, l'union des points aberrants des différents instants est prise en compte pour chaque instant.

Cette chaîne de traitement modifiée pour la prise en compte des informations longitudinales sera appelée EMAL dans la suite de ce manuscrit. Les prétraitements de EMAL sont illustrés par la figure 6.6.

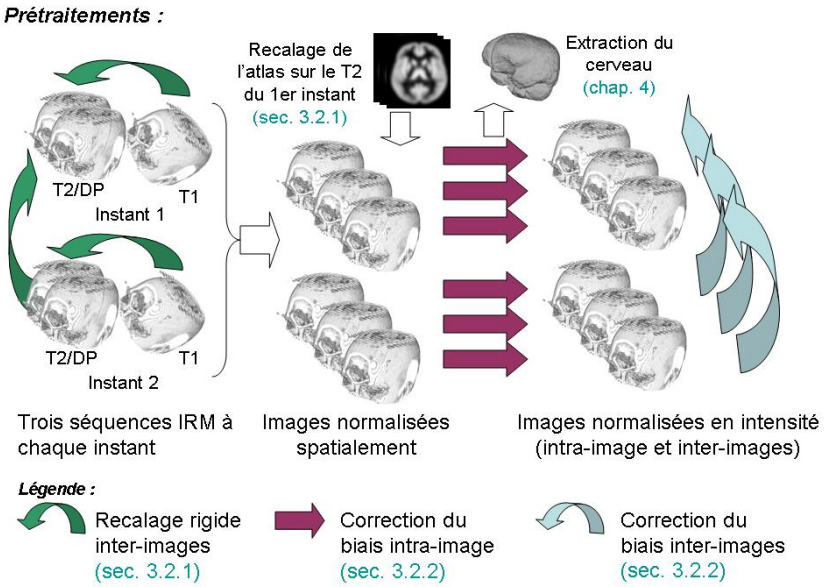


FIG. 6.6 – *Prétraitements de la chaîne de traitement EMAL (EMA avec modification pour la prise en compte des informations longitudinales).*

6.3.3 Validation

La validation des segmentations de volumes partiels semble peu réalisable puisque leurs courbes de densité de probabilité se superposent. En conséquence, un expert a réalisé une première validation qualitative des segmentations obtenues et aucune erreur significative n'a été notée. Dans un second temps, nous avons réalisé une évaluation quantitative sur des données simulées.

TAB. 6.1 – *Critères de comparaison de segmentations.*

Nom du critère	Expression conventionnelle	Expression généralisée
Indice de Similarité (SI)	$\frac{2Card(Ref \times Seg)}{Card(Ref)+Card(Seg)}$	$\frac{2\sum_i \min(Ref_{(i)}, Seg_{(i)})}{\sum_i Ref_{(i)}+\sum_i Seg_{(i)}}$
Sensibilité (SEN)	$\frac{Card(Ref \times Seg)}{Card(Ref)}$	$\frac{\sum_i \min(Ref_{(i)}, Seg_{(i)})}{\sum_i Ref_{(i)}}$
Spécificité (SPE)	$\frac{Card(\overline{Ref \times Seg})}{Card(\overline{Ref})}$	$\frac{\sum_i \min(1-Ref_{(i)}, 1-Seg_{(i)})}{\sum_i (1-Ref_{(i)})}$

6.3.3.1 Critères de comparaison

La première étape pour valider notre méthode est d'identifier des critères de comparaison entre nos résultats et la vérité terrain (connue puisque nous travaillons avec des données simulées). La méthode de classification ne donne pas des segmentations binaires du LCR et du cerveau mais des cartes de répartition. Pour comparer ces cartes, nous utilisons une version généralisée de l'indice de similarité (SI), appelé aussi coefficient de similarité de Dice [Dice 1945], de la sensibilité (SEN) et de la spécificité (SPE) [Crum 2006]. Ces critères sont donnés dans la table 6.1 pour une segmentation (Seg) et une segmentation de référence (Ref). $Ref_{(i)}$ et $Seg_{(i)}$ représentent l'intensité du voxel i dans l'image correspondante.

6.3.3.2 Validation des segmentations

Les segmentations données par notre méthode ont été validées sur des images simulées fournies par BrainWeb[1]. Les séquences IRM utilisées correspondent au modèle anatomique d'un patient ayant une charge lésionnelle modérée. Elles ont été générées avec les paramètres suivants : bruit 3%, non-uniformité en intensité 20%, épaisseur de coupes 1 mm et 3 mm. Les résultats de cette validation sont donnés dans la table 6.2.

TAB. 6.2 – *Résultats de la comparaison des segmentations pour le modèle anatomique d'un patient atteint de SEP (Brainweb).*

épaisseur de coupe	Classes	IS	SEN	SPE
1mm	LCR	0.81	0.79	0.99
1mm	Cerveau	0.98	0.98	0.99
3mm	LCR	0.82	0.81	0.99
3mm	Cerveau	0.95	0.98	0.97

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

Les résultats obtenus indiquent des segmentations correctes ($IS > 0.8$). Les valeurs de sensibilité sont correctes ($SEN > 0.79$) même si elles semblent indiquer une légère sous segmentation. Les valeurs de spécificité ($SPE > 0.97$) indiquent qu'il y a relativement peu de faux positifs. Les résultats concernant le LCR sont aussi légèrement plus faibles que ceux du cerveau. Une des raisons de cette différence est que le LCR représente le plus petit volume comparé à ceux de la SB et de la SG. Un voxel mal classé introduira une plus grande erreur relative sur la segmentation du LCR que sur la segmentation du cerveau (SB+SG).

6.3.3.3 Validation de la mesure de volume

La table 6.3 donne les erreurs relatives sur les mesures du volume de LCR et du cerveau. Les erreurs sur l'estimation du volume de SG et de SB sont assez importantes mais l'erreur sur le volume du cerveau est inférieure à 5%. De plus nous avons une bonne estimation du volume de LCR (erreur moyenne 4.5%) alors que le volume de LCR est le plus petit.

TAB. 6.3 – *Erreurs relatives sur les volumes obtenus.*

	LCR	SG	SB	Cerveau
1 mm	-4 %	+9 %	-11 %	+1 %
3 mm	-5 %	+16 %	-9 %	+5 %

6.3.3.4 Validation de la mesure d'atrophie

Des simulations d'atrophie sur des IRM ont été proposées dans [Studholme 2006, Karacali 2006, Sharma 2008]. Pour les travaux décrits dans cette section, les images d'atrophie simulée à partir d'IRM, elles aussi simulées, d'un patient sain ont été utilisées [Camara 2006]. Cette méthode de simulation d'atrophie se décompose en plusieurs étapes. Premièrement, un maillage du cerveau est calculé à partir de l'image IRM de l'atlas du MNI. Deuxièmement, chaque nœud de ce maillage est labellisé grâce à l'image de labels de l'atlas. En fonction de la classe à laquelle un nœud correspond, des coefficients d'expansion contrôlant le changement de volume de cet élément par rapport aux autres lui sont attribués. Le maillage est alors déformé produisant ainsi un champ de déformation. Puis, ce champ est appliqué sur l'image IRM de l'atlas qui est ensuite rééchantillonnée et bruitée.

Nous avons vérifié le comportement de nos méthodes sur ces données. La figure 6.7 montre la mesure d'atrophie en fonction de l'atrophie simulée pour les deux méthodes proposées ainsi que pour SIENA, méthode basée sur le recalage des images et décrite dans la section 6.1.3. Pour cette étude SIENA a été utilisée à partir de la suite logiciel FSL[2] [Smith 2004] et avec ses paramètres par défaut.

[2] http ://www.fmrib.ox.ac.uk/fsl/

Nous pouvons voir que CLASS ne permet pas de suivre l'atrophie simulée. Son utilisation dans des études longitudinales doit donc se faire avec précaution. En revanche, LCLASS est fortement corrélée à l'atrophie simulée. Cette étude montre aussi que LCLASS sous estime l'atrophie simulée contrairement à SIENA qui la surestime (une mesure parfaite étant obtenue pour la droite $y = x$). Les deux méthodes ont une forte corrélation avec l'atrophie simulée, la nôtre étant légèrement supérieure ($R^2 = 0,99, p < 0,05$). Il doit aussi être noté que l'approximation linéaire des résultats de notre méthode passe logiquement par l'origine alors que ce n'est pas le cas de l'approximation linéaire des résultats de SIENA.

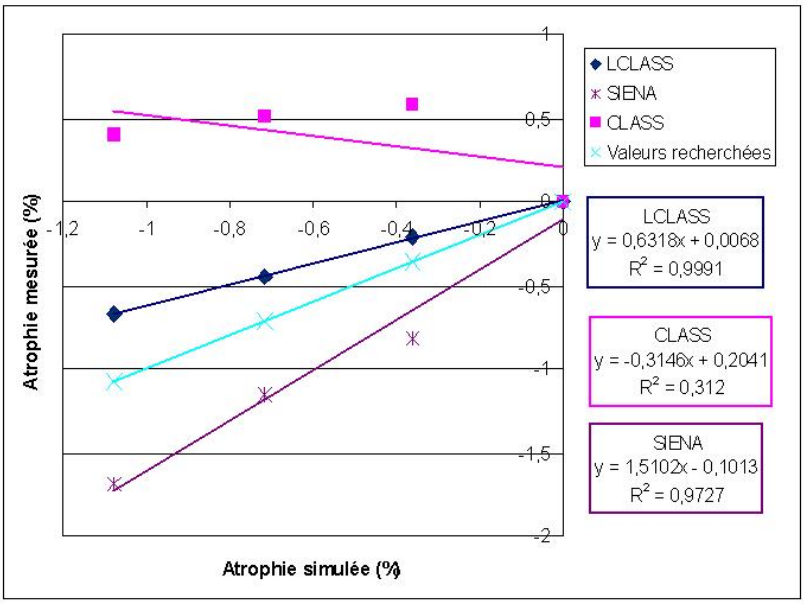


FIG. 6.7 – *Atrophie mesurée en fonction de l'atrophie simulée par rapport à l'instant d'origine, T0, pour les deux méthodes proposées ainsi que pour SIENA.*

6.3.4 Apport des méthodes

Deux nouvelles méthodes d'évaluation de la FPC ont été proposées : CLASS et LCLASS. Elles se distinguent des autres méthodes déjà présentes dans la littérature sur différents points. L'algorithme de segmentation utilisé est entièrement automatique et il inclut un modèle de volume partiel (LCR-SG) [Dugas-Phocion 2004a]. Son application se fait sur la globalité du cerveau (contrairement à d'autres algorithmes [Collins 2001]). En effet, tous les voxels du cerveau sont utilisés pour calculer la FPC excepté des voxels que l'algorithme ne "sait" pas classer et qui sont regroupés sous l'étiquette "points aberrants".

La méthode LCLASS inclut aussi des modifications pour permettre un suivi longitudinal de l'atrophie. En effet, la comparaison des résultats de LCLASS avec ceux de SIENA sur des données simulées montre que dans ce cas LCLASS est au

moins aussi précise que SIENA. Ceci semble donc indiquer qu'à la condition de prendre en compte l'aspect longitudinal (grâce aux modifications proposées), une méthode d'évaluation de l'atrophie basée sur le calcul de volume peut être aussi performante qu'une méthode basée sur le calcul des variations de ce volume.

6.4 Comparaison et choix de méthodes d'évaluation de l'atrophie

Des améliorations des méthodes de mesures linéaires et des modifications à inclure pour prendre en compte l'aspect longitudinal dans une méthode basée sur un calcul de volume ont été proposées. Cependant, la question de trouver une méthode fiable permettant de mesurer l'atrophie sur notre base de données reste entière. J'ai donc mené une étude comparative de huit mesures d'atrophie précédemment présentées : la FPC donnée par CLASS, celle donnée par LCLASS, celle donnée par SIENAX, le volume cérébral normalisé (VCN) donné par SIENAX, le pourcentage de variation du volume cérébral (PVVC) donné par SIENA et les trois mesures linéaires définies dans la section 6.2. Ces méthodes ont été sélectionnées sur des critères de disponibilité des logiciels et/ou d'utilisation préalable des méthodes dans des études similaires à la nôtre.

Après un choix des critères de comparaison de ces méthodes, celle-ci a été réalisée sur les données simulées de [Camara 2006] et sur 25 patients de notre base de données.

6.4.1 Analyse statistique

Les mesures obtenues seront comparées entre elles en utilisant le coefficient de corrélation de Pearson. Cette approche est valable lorsque l'on suppose une relation linéaire entre les mesures, ce qui est le cas des mesures 3D entre elles et des mesures linéaires entre elles. En revanche, cette supposition ne peut pas être faite sur la relation entre une mesure linéaire et une mesure 3D. La corrélation entre ces mesures sera calculée à partir du coefficient de corrélation de Spearman [Spearman 1904]. Celui-ci est similaire à celui de Pearson, excepté qu'il n'utilise pas directement les valeurs mais leur rang dans son calcul. Ce coefficient est aussi dit significatif si sa valeur dépasse une valeur de référence (dépendante du nombre de données et de la probabilité α de précision recherchée) donnée par les tables de Spearman. Dans notre cas, j'ai pris la valeur $\alpha = 0.05$. L'atrophie moyenne donnée par ces méthodes sera aussi comparée avec l'atrophie simulée ou avec celles trouvées dans des études similaires à la nôtre.

Premièrement, cette analyse sera faite sur les données d'atrophie simulée présentées précédemment [Camara 2006]. Dans ces images, la taille des voxels est de 1mm x 1mm x 1mm. Deuxièmement, le taux d'atrophie entre deux instants sera évalué sur 25 patients atteints de SEP rémittente et traités aux interférons $\beta - 1a$ de notre base de données. Ces deux instants sont séparés par une période d'un an. Pour ces

données, la taille des voxels est de 0,94mm x 0,94mm x 2mm.

Les mesures de FPC et de VCN donnent une valeur à chaque instant. Leur variation permet d'obtenir l'atrophie. Cette évolution, Δ, entre l'instant 1, I1, et l'instant 2, I2, est calculée avec l'équation 6.4.

$$\Delta = 100 \times \frac{\text{Mesure(I2)} - \text{Mesure(I1)}}{\text{Mesure(I1)}} \tag{6.4}$$

L'élargissement des ventricules et l'atrophie sont des effets opposés. En conséquence, la valeur $-\Delta$ sera prise dans la suite de ce manuscrit pour la largeur des ventricules latéraux et celle du troisième ventricule.

6.4.2 Comparaison sur des données simulées

Nous reprenons ici les données d'atrophie simulée [Camara 2006]. Nous utilisons cette simulation pour valider et comparer nos différentes mesures globales d'atrophie. La figure 6.8 montre l'atrophie mesurée en fonction de l'atrophie simulée pour les mesures suivantes : FPC de CLASS, FCP de LCLASS, FCP de SIENAX, VCN de SIENAX et le PVVC de SIENA.

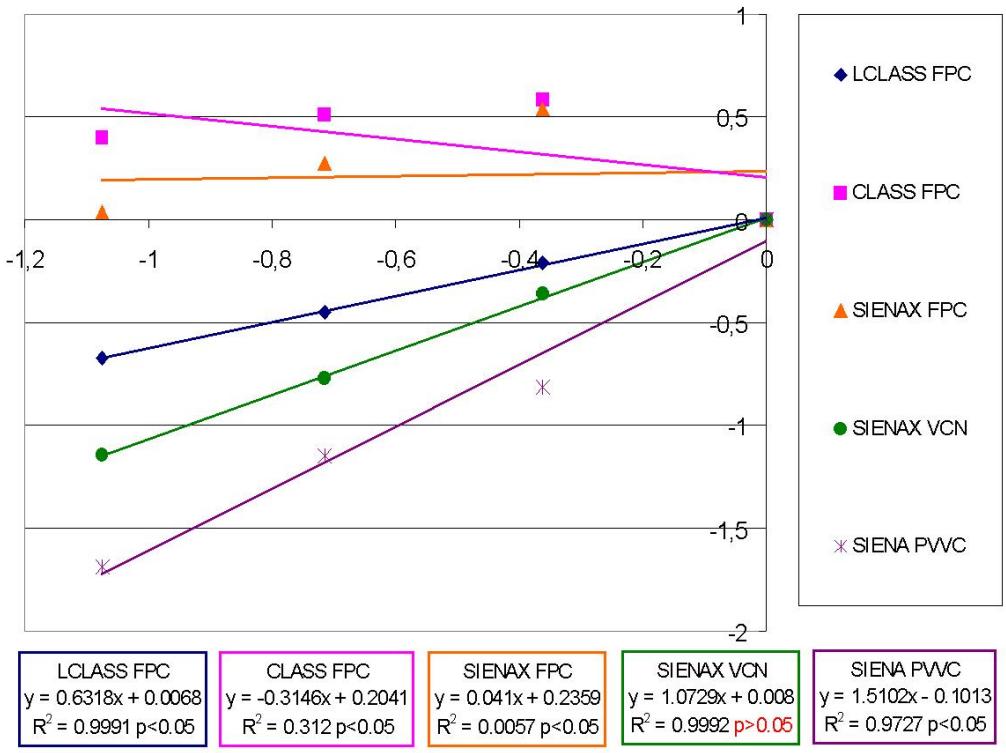


FIG. 6.8 – *Atrophie mesurée en fonction de l'atrophie simulée, par rapport à l'instant initial, T0, pour l'ensemble des méthodes testées.*

Nous pouvons observer que les résultats donnés par la FPC de CLASS et la FPC de SIENAX sont incorrects. Comme vu précédemment, la FPC de LCLASS et le PVVC de SIENA ont une bonne corrélation avec l'atrophie simulée. Ceci est aussi vrai pour le VCN de SIENAX. La FPC de LCLASS et le VCN de SIENAX ont une corrélation légèrement plus forte que le PVVC de SIENA. Leur approximation linéaire passe aussi logiquement par le point de coordonnées $(0, 0)$ contrairement à celle du PVVC de SIENA.

Le coefficient de corrélation de Spearman (noté *rho* pour ne pas confondre avec la valeur du test de student p) a été calculé entre les mesures linéaires et l'atrophie simulée :

- $-\Delta$(Largeur des ventricules latéraux) : $rho = 1,0$; $\alpha < 0,2$
- Δ(Largeur du cerveau) : $rho = 0,8$; Non significatif.
- $-\Delta$(Largeur du troisième ventricule) : $rho = 0,8$; Non significatif.

Ces résultats montrent une bonne corrélation entre la largeur du cerveau et l'atrophie simulée. Les mesures des largeurs des ventricules sont elles en revanche logiquement anticorrélées. Toutefois, pour ces mesures, les résultats ne sont pas ou très faiblement significatifs.

Dans le but de quantifier si une méthode sur ou sous estime l'atrophie, le taux d'atrophie moyen obtenu par les différentes mesures 3D a été calculé. Ces résultats sont donnés dans la table 6.4. Il s'agit d'une moyenne longitudinale, taux d'atrophie moyen par an. La valeur donnée par le VCN de SIENAX est la plus proche de celle de l'atrophie simulée. Nous retrouvons aussi que le PVVC donné par SIENA surestime l'atrophie (ce qui est aussi montré dans [Camara 2007, Sharma 2008]) alors que la FPC de LCLASS la sous-estime. Les résultats des FPC de CLASS et de SIENAX ne sont pas réalistes car positifs.

TAB. 6.4 – *Taux d'atrophie moyen sur les données simulées.*

	Moyenne (%) ± écart type
Δ(Atrophie simulée)	-0,72 ± 0,36
Δ(FPC de LCLASS)	-0,44 ± 0,22
Δ(FPC de CLASS)	0,50 ± 0,1
Δ(FPC de SIENAX)	0,29 ± 0,24
Δ(VCN de SIENAX)	-0,76 ± 0,39
Δ(PVVC de SIENA)	-1,22 ± 0,44

6.4.3 Comparaison sur des données réelles

J'ai aussi comparé les méthodes sélectionnées sur 25 sujets de notre base de données. Dans ce cas, à la différence des données d'atrophie simulée, nous n'avons pas accès à la vérité terrain. Nous ne pouvons donc que comparer les mesures entre elles.

Comme le montre la table 6.5, les seules corrélations significatives ($p < 0,05$) sont entre la FPC de SIENAX, le PVVC de SIENA et la FPC de LCLASS. Les mesures 2D ne se corrèlent pas ensemble (cf. table 6.6). La comparaison entre ces mesures 2D et les mesures 3D montre une anti-corrélation significative entre la largeur des ventricules latéraux, celle du troisième ventricule, la FPC de LCLASS et la FPC de SIENAX (cf. table 6.7).

TAB. 6.5 – *Corrélation entre les différentes mesures 3D.* (*) *signifie que* $p < 0,05$ *;* (*$*$) *signifie que la corrélation est égale à 1 par définition.*

	Δ(FPC de LCLASS)	Δ(FPC de CLASS)	Δ(FPC de SIENAX)	Δ(VCN de SIENAX)	Δ(PVVC de SIENA)
Δ(FPC de LCLASS)	1*$*$				
Δ(FPC de CLASS)	0,1271*	1*$*$			
Δ(FPC de SIENAX)	0,4832*	-0,3919	1*$*$		
Δ(VCN de SIENAX)	0,4566	-0,2015	0,3589	1*$*$	
Δ(PVVC de SIENA)	0,4733*	-0,1667	0,4302	0,5145	1*$*$

Les boîtes à moustaches de la figure 6.9 montrent la répartition des taux d'atrophie sur les 25 patients de notre base pour les diverses mesures. Au total, 16 points aberrants sont observés. Ceux-ci ne représentent en fait que 8 patients différents. La table 6.8 reprend les informations des boîtes à moustaches et donne les moyennes et les écarts types de ces valeurs en prenant en compte ou non les sujets correspondants au moins à un point aberrant. Ces moyennes sont ici faites sur plusieurs sujets. CLASS BPF, SIENAX VCN, la largeur des ventricules latéraux et la largeur du troisième ventricule ont un écart type élevé. La moyenne de la FPC de LCLASS est positive. Les moyennes de la FPC de SIENAX et du PVVC de SIENA sont proches de la valeur de 0.61 donnée dans [Filippi 2004a].

TAB. 6.6 – *Corrélation entre les différentes mesures 2D.* (*) *signifie que* $p < 0,05$ *;* (**) *signifie que la corrélation est égale à 1 par définition*

	$-\Delta$(largeur des ventricules latéraux)	Δ(largeur du cerveau)	$-\Delta$(largeur du troisième ventricule)
$-\Delta$(largeur des ventricules latéraux)	1**		
Δ(largeur du cerveau)	-0,0067*	1**	
$-\Delta$(largeur du troisième ventricule)	0,1963*	-0,1706*	1**

TAB. 6.7 – *Coefficient de corrélation de Spearman entre les mesures 2D et les mesures 3D.* (*) *signifie que* $\alpha < 0,05$

	Δ(FPC de LCLASS)	Δ(FPC de CLASS)	Δ(FPC de SIENAX)	Δ(VCN de SIENAX)	Δ(PVVC de SIENA)
$-\Delta$(largeur des ventricules latéraux)	0,5662*	-0,0323	0,5362*	0,5162*	0,2000
Δ(largeur du cerveau)	-0,0238	0,0577	-0,0738	0,1069	0,0369
$-\Delta$(largeur du troisième ventricule)	0,3938*	0,2723	0,4515*	0,0746	0,2877

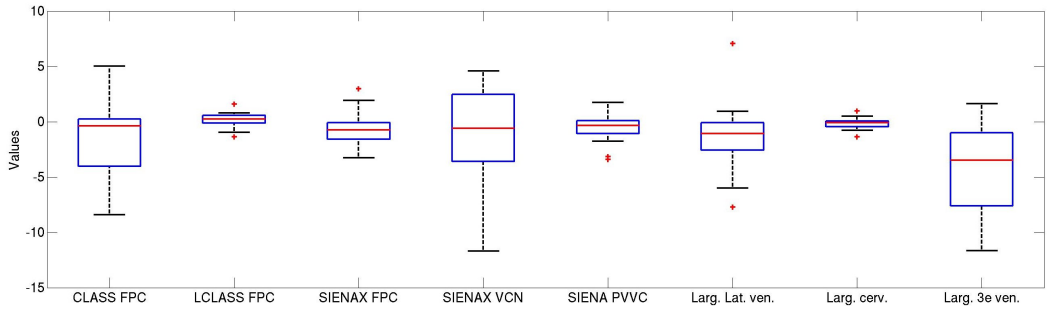


FIG. 6.9 – *Boîtes à moustaches des taux d'atrophie moyen ($-\Delta$ pour la largeur du troisième ventricule et des ventricules latéraux et $+\Delta$ pour les autres mesures) sur 1 an et sur 25 patients atteints de SEP rémittente pour les différentes méthodes. Pour une meilleure représentation, 4 points aberrants de la largeur du troisième ventricule ne sont pas représentés (valeurs : $-84,74$; $-22,81$; $-20,14$; $-17,68$).*

TAB. 6.8 – *Taux d'atrophie moyen sur 1 an et sur 25 patients atteints de SEP rémittante (en prenant ou sans prendre en considération les sujets de la base correspondant à au moins un point aberrant).*

	Moyenne (%) ± écart type	Moyenne (%) ± écart type (sans les points aberrants)
Δ(FPC de LCLASS)	0,12 ± 0,62	0,26 ± 0,39
Δ(FPC de CLASS)	-1,55 ± 3,2	-1,07 ± 3,19
Δ(FPC de SIENAX)	-0,77 ± 1,49	-0,80 ± 1,09
Δ(VCN de SIENAX)	-1,40 ± 4,73	0,12 ± 0,62
Δ(PVVC de SIENA)	-0,56 ± 1,11	-0,42 ± 0,71
$-\Delta$(largeur des ventricules latéraux)	-1,47 ± 2,81	-1,21 ± 1,57
Δ(largeur du cerveau)	-0,17 ± 0,45	-0,21 ± 0,27
$-\Delta$(largeur du troisième ventricule)	-8,54 ± 17,14	-3,04 ± 3,45

6.4.4 Choix des méthodes pour analyser la base Qualicore

Les expériences sur les données simulées montrent des corrélations (ou des anti-corrélations) entre les mesures 2D et la vérité terrain. Cependant, celles-ci ne sont pas significatives ($\alpha > 0,05$), excepté pour la mesure de la largeur des ventricules latéraux. Sur les données réelles, il a été montré que ces mesures sont reproductibles (cf. section 6.2.2). La largeur des ventricules latéraux est aussi significativement anti-corrélée avec la FPC de LCLASS, la FPC de SIENAX et le VCN de SIENAX. La largeur du troisième ventricule est seulement anti-corrélée avec la FPC de LCLASS et la FPC de SIENAX. Aucune corrélation n'est observée entre la largeur du cerveau et les mesures 3D. Ceci suggère une relation entre l'atrophie cérébrale centrale et l'atrophie cérébrale globale. Cette dernière avait du reste, déjà été observée dans [Sharma 2004]. Les largeurs des ventricules peuvent donc être utilisées pour évaluer l'atrophie globale [Turner 2001, Butzkueven 2008].

Concernant l'application des mesures 3D sur les données simulées, la vérité terrain est corrélée avec le VCN de SIENAX (non significativement, $p > 0,05$) ainsi qu'avec la FPC de LCLASS et le PVVC de SIENA (significativement, $p < 0,05$) mais pas avec les FPC de SIENAX ou de CLASS. Ceci suggère que l'étape d'extraction du cerveau est primordiale dans l'évaluation de l'atrophie.

Sur les données réelles, CLASS n'est corrélée avec aucune autre méthode. Cela confirme le fait que cette méthode n'est pas adéquate pour évaluer l'atrophie. Les trois mesures (FPC de LCLASS, FPC de SIENAX et le VCN de SIENAX) sont significativement corrélées avec la largeur des ventricules latéraux. Les deux premières mesures le sont aussi avec la largeur du troisième ventricule.

Le VCN de SIENAX, la FPC de SIENAX et le PVVC de SIENA donnent, comme il est attendu, un taux d'atrophie moyen négatif alors que cette valeur est étonnamment positive pour la FPC de LCLASS. L'écart type élevé du VCN de SIENAX suggère que la robustesse de cette mesure sur nos données est limitée. Des écarts types plus faibles sont obtenus pour la FPC de SIENAX, le PVVC de SIENA et la FPC de LCLASS. Dans le cas des deux premières mesures, le taux d'atrophie moyen est respectivement de -0.77% et de -0.56%. Ces résultats sont proches de ceux reportés dans la littérature (-0.61% dans [Filippi 2004a]).

J'ai pu comparer différentes méthodes d'évaluation de l'atrophie sur des données simulées et des données réelles. La méthode CLASS semble être inappropriée pour évaluer l'atrophie tout au moins longitudinalement. Les FPC données par LCLASS et SIENAX ainsi que le VCN se corrèlent bien avec les mesures 2D de largeurs des ventricules sur les données réelles. Toutefois, la FPC donnée par SIENAX donne des résultats aberrants sur les données simulées. La corrélation du VCN de SIENAX avec la vérité terrain de ces mêmes données est aussi non significative. De plus, l'écart type de cette mesure sur les données de notre base est élevé. Ceci semble indiquer que la méthode LCLASS est la méthode se corrélant le mieux avec le phénomène d'atrophie sur nos données.

Cependant, le fait que le taux moyen d'atrophie trouvé par cette méthode est positif signifie que son utilisation doit se faire prudemment. Je propose donc d'as-

socier cette mesure à celle de SIENA. En effet, même si le PVVC de SIENA est peu corrélé avec les mesures 2D sur les données réelles, cette méthode se corrèle significativement avec l'atrophie simulée. De plus, le taux d'atrophie moyen qu'elle donne sur les données réelles est vraisemblable selon la littérature. Je mesurerai donc l'atrophie sur notre base par ces deux méthodes.

6.4.5 Limites des méthodes choisies sur les images de notre base

Deux méthodes ont été sélectionnées pour analyser notre base : LCLASS et SIENA. LCLASS est une nouvelle méthode. Sa précision n'est donc pas connue. SIENA a été utilisée précédemment. Toutefois, sa précision, tout comme celle des autres méthodes de mesure d'atrophie, peut dépendre de la résolution des images [Smith 2002a]. En conséquence, cette mesure doit être testée sur nos données.

Dans la base de données, 4 patients avec 3 instants sont disponibles. SIENA donne le PVVC entre deux instants. En conséquence, l'addition du PVVC entre l'instant 1 et l'instant 2 et du PVVC entre l'instant 2 et l'instant 3 devrait donner le même résultat que le calcul du PVVC entre l'instant 1 et l'instant 3. Il faut toutefois noter que SIENA donne un pourcentage de variation par rapport à un instant. L'approche ci-dessus a donc une limite : le calcul du PVVC entre l'instant 2 et l'instant 3 se fait avec une référence différente que le calcul des deux autres mesures. Cependant, ce test est réalisé par S. Smith, le créateur de la méthode et peut tout au moins donner une borne supérieure de l'erreur de mesure. En réalisant ce test sur les données, j'ai obtenu une erreur sur le PVVC inférieure à 1.2%. Ceci est plus que les 0.2% donnés dans [Smith 2002a] mais moins que les 3.6% d'erreur donnés dans [Tobie 2007].

Dans l'étape de classification de LCLASS, les paramètres de classes sont calculés à partir des séquences IRM de chaque instant disponible. Les paramètres de classification obtenus sur les images de deux instants peuvent donc être modifiés par l'ajout d'un troisième instant. Les segmentations et par conséquent la valeur de la FPC mesurée changent aussi. En réalisant ce test sur les données de la base, j'ai obtenu une différence de FPC sur les deux premiers instants, comprise dans un intervalle de 0.46%. De plus, le taux d'atrophie moyen entre l'instant 1 et l'instant 2, sur ces quatre patients est de -0.25%, si seulement ces deux instants sont pris en considération et -0.35%, si le troisième instant est pris en considération. Cette deuxième valeur d'atrophie entre l'instant 1 et l'instant 2 (séparés de 1 an) semble plus réaliste, comparé au -0.33% trouvé dans [Hardmeier 2005] sur une base de patients similaires à la nôtre. Selon ces résultats, il semble donc que l'ajout d'un instant permet d'améliorer les résultats. Ces conclusions restent à confirmer sur une base de données de plus grande ampleur.

Chapitre 7

Cadre d'analyse de bases de données et application des algorithmes

Sommaire

7.1 Analyse de la base de données QUALICORE 129
7.1.1 Analyse statique 129
7.1.2 Analyse dynamique 138
7.1.3 Analyse par sujet 143
7.1.4 Conclusion 153
7.2 Intégration et mise à disposition des outils d'analyse 154
7.2.1 SepINRIA 154
7.2.2 Web-services exécutés sur une grille de calcul 157

Résumé : Ce chapitre a permis de mettre en pratique les algorithmes développés dans les chapitres précédents. Premièrement, je me suis servi de ces derniers pour analyser les sujets de la base QUALICORE. Parmi ces mesures, le volume des lésions est évalué par un expert (grâce aux outils manuels et semi-automatiques) ainsi que par notre méthode (automatiquement). Ensuite, l'atrophie est obtenue par LCLASS et SIENA. LCLASS fournit une fraction parenchymateuse cérébrale (FPC) à chaque instant. Ainsi, l'atrophie est obtenue par le calcul des variations de FPC. SIENA évalue une variation du volume du cerveau. Les mesures linéaires sont aussi données. Cette analyse a permis de prouver l'homogénéité de la base. La méthode automatique de calcul de la charge lésionnelle est aussi corrélée avec les mesures manuelles. De même, LCLASS se corrèle aussi avec les mesures de référence contrairement à SIENA. L'analyse par sujet introduit une discussion sur l'influence de certains paramètres cliniques ou radiologiques qui pourraient faire varier de façon erronée les différentes mesures.

Deuxièmement, un logiciel SepINRIA a été créé et développé pour diffuser ces algorithmes. Les buts de celui-ci sont d'aider les neurologues au diagnostic et à la prise de décisions thérapeutiques concernant des patients atteints de SEP. Ce logiciel permet aussi la comparaison des algorithmes que j'ai développés, avec d'autres algorithmes qui ne sont pas disponibles. Dans l'étude précédente, SepINRIA a permis de réaliser les différentes mesures manuelles et semi-automatiques. Les algorithmes automatiques y sont aussi présents. Même si SepINRIA ne possède pas encore d'outils précis pour l'aide au diagnostic (excepté le recalage d'images), il peut déjà se révéler utile pour la prise de décision thérapeutique. Par ailleurs, il commence à être utilisé dans différents centres cliniques ou de recherche. Il a été téléchargé plus de 3200 fois et son code a été déposé à l'agence pour la protection des programmes.

Enfin, un dernier outil développé consiste en la réalisation de web-services permettant d'exécuter sur une grille de calcul la chaîne de segmentation des différents compartiments cérébraux. Ce travail a déjà permis d'estimer l'influence d'un paramètre de cette chaîne. Il permettra aussi de rendre disponible les différentes briques du traitement. Ces briques pourront alors être remplacées par des traitements proposés par d'autres laboratoires et de nouvelles chaînes de traitements pourront être créées. La puissance de calcul de la grille permet aussi d'analyser un grand nombre de données.

Mots clefs : Qualicore, Corrélation, EDSS, Charge lésionnelle, Atrophie, SepINRIA, NeuroLOG.

Contributions :

- Mise en place d'un cadre pour l'analyse d'une base d'IRM de patients atteints de SEP.
- Application du cadre d'analyse sur la base QUALICORE.
- Développement d'un logiciel d'analyse d'images IRM de patients atteints de SEP : SepINRIA.

– Développement de web-services pour partager des ressources (données + algorithmes) et pour analyser d'importantes bases de données.

Organisation du chapitre : La section 7.1 présente les résultats des différentes méthodes (automatiques ou non) sur notre base de données. Elle donne aussi la valeur du coefficient de corrélation entre ces résultats et le score EDSS. Puis, les outils logiciels mis en place pour réaliser les mesures manuelles et semi-automatiques ainsi que pour diffuser les algorithmes de segmentation automatique sont décrits. La section 7.2.1 présente SepINRIA, un logiciel pour l'analyse d'images IRM de patients atteints de SEP. La section 7.2.2 explique la mise en place de services pour une exécution sur grille de calcul, de l'algorithme de segmentation du cerveau.

7.1 Analyse de la base de données QUALICORE

Le tableau récapitulatif de toutes les mesures effectuées sur tous les sujets de la base de données est présent en annexe C. L'analyse de ces mesures est, ici, présentée. Celle-ci est articulée autour de trois axes : analyse statique (par instants), analyse dynamique (des variations par rapport à T_0) et une analyse par sujet (évolution des mesures pour un même sujet). Ces axes permettent de traiter l'intégralité de la base alors que le nombre de sujets par instants varie.

7.1.1 Analyse statique

Les résultats des différentes mesures sont regardés dans cette section de manière statique, *i.e.* instant par instant. Ceci permet de vérifier et de prouver l'homogénéité de la base Qualicore au premier instant, T_0. Les coefficients de corrélation (Pearson ou Spearman selon le cas (cf. chapitre 6)) entre les diverses mesures seront aussi calculés. Nous pourrons ainsi vérifier par exemple que le score EDSS n'est corrélé ni avec la charge lésionnelle ni avec l'atrophie dans notre cas.

7.1.1.1 Instant T_0

La base de données contient 26 sujets avec des IRM au premier instant, T_0, de l'étude. Les boîtes à moustaches de la figure 7.1 donnent la distribution des différentes mesures à cet instant.

Les mesures de charge lésionnelle (figure 7.1, a) ont des moyennes différentes. En effet, la mesure automatique surestime cette valeur par rapport à celle mesurée manuellement en T2 et qui est communément utilisée comme référence. Cette même figure permet aussi de vérifier que la charge lésionnelle est de plus en plus élevée lorsqu'on passe de la séquence T1 à la séquence T2 puis à la séquence T2-FLAIR. Ceci se justifie pleinement par les propriétés des différents sous-types de lésions (cf. table 5.1). Concernant ces mesures manuelles, deux à trois points aberrants (croix rouges sur le graphique) sont présents par mesure. Il s'agit de trois mêmes sujets

MARTH, SCHIS et DELEM qui ont une charge lésionnelle élevée par rapport aux autres.

Une analyse de variance (ANOVA) à un facteur montre qu'il n'y a pas de différence significative entre les distributions du nombre de lésions pour les différentes méthodes (contrairement à la mesure de charge lésionnelle). Les valeurs médianes du nombre de lésions pour les différentes méthodes sont assez proches. La valeur maximale de la distribution du nombre de lésions sur la séquence T2-FLAIR est plus faible que celle de la distribution sur la séquence T1. Ceci indique qu'il existe au moins un cas où le nombre de lésions diminue entre la séquence T1 et la séquence T2-FLAIR. Cette diminution peut s'expliquer par l'augmentation apparente du volume des lésions entre les différentes séquences. En effet, lors de cette augmentation (visuelle), des lésions qui étaient proches sur une séquence peuvent se retrouver collées (fusionnées) sur une autre et le nombre de lésions peut ainsi diminuer.

Pour cet instant, les valeurs de FPC se situent dans l'intervalle $[87; 94]$, de la largeur du troisième ventricule dans $[2; 8]$, de la largeur des ventricules latéraux dans $[16; 28]$ (un point aberrant à 28,7 : sujet SCHIS) et les valeurs d'EDSS dans $[0; 3]$ (un point aberrant à 5 : sujet GUENA).

La table 7.1 donne les coefficients de corrélation (Pearson ou Spearman selon le cas) entre les différentes mesures et les mesures de charges lésionnelles (automatique ou manuelle en T2) pour l'instant T_0. On peut y voir que même si la charge lésionnelle mesurée par la méthode est plus élevée que celles mesurées manuellement (cf. paragraphes précédents), elle est corrélée assez fortement ($r>0,7$) et significativement aux charges lésionnelles mesurées manuellement. Ces dernières sont du reste corrélées à la charge lésionnelle en T2. Excepté entre la charge lésionnelle T2 et la largeur des ventricules latéraux, il n'y a pas de corrélation significative (p ou $\alpha < 0,05$) entre la charge lésionnelle et les mesures d'atrophie.

La table 7.2 montre pour les différentes segmentations des lésions une corrélation significative entre le volume des lésions et leur nombre. Cette corrélation est plus faible dans le cas de la méthode automatique.

La table 7.3 montre que l'EDSS n'est corrélé avec aucune mesure. J'ai utilisé dans ce cas le coefficient de corrélation de Spearman car l'échelle EDSS est non paramétrique (par exemple, une variation de 1 point entre 3 et 4 n'a pas la même signification clinique qu'une variation entre 7 et 8). C'est aussi pour cette raison que le score EDSS ne peut pas se moyenner.

7.1.1.2 Instant T_1

La base de données contient 16 sujets avec des IRM au second instant, T_1, de l'étude. Les boîtes à moustaches de la figure 7.2 donnent la distribution des différentes mesures à cet instant. Les résultats montrés par les différentes boîtes à moustaches pour cet instant sont très similaires à ceux du premier instant, T_0 et les points aberrants correspondent à deux mêmes sujets (SCHIS et DELEM). Le sujet MARTH n'a pas de deuxième instant.

De même que pour T_0, une analyse ANOVA à un facteur ne donne pas de diffé-

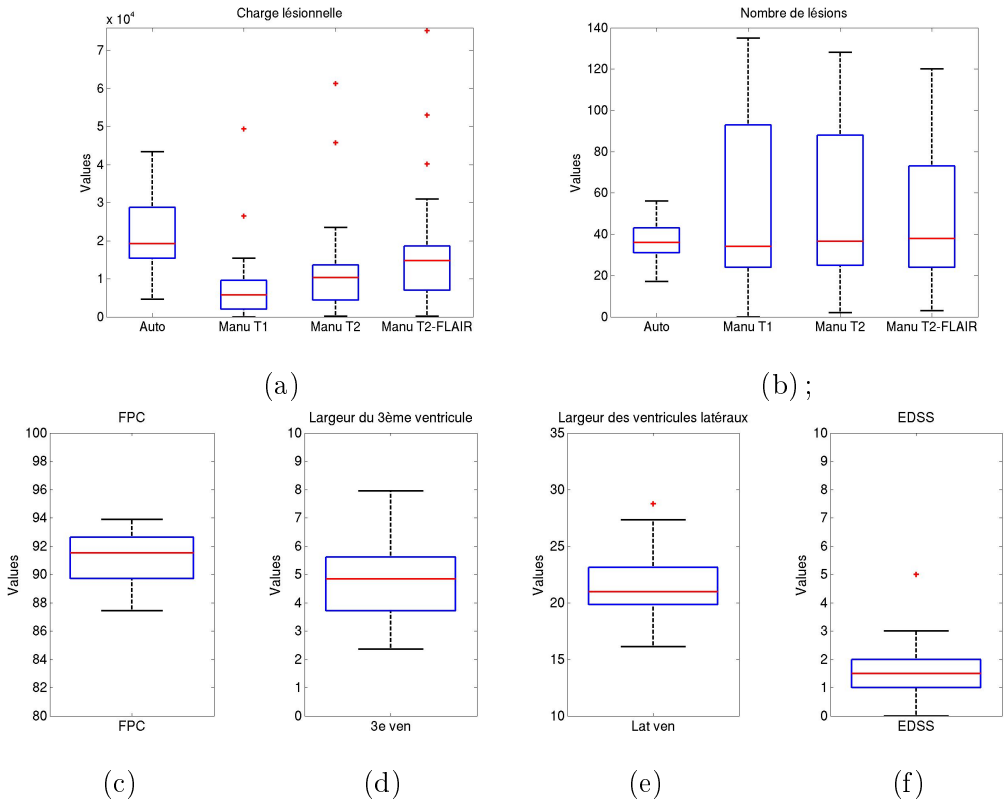


FIG. 7.1 – *Boîtes à moustaches des valeurs : (a) de charge lésionnelle, (b) du nombre des lésions, (c) de la fraction parenchymateuse cérébrale (FPC), (d) de la largeur du 3ème ventricule, (e) de la largeur des ventricules latéraux et (f) du score EDSS pour l'instant* T_0.

TAB. 7.1 – *Valeurs du coefficient de corrélation (de Pearson ou de Spearman selon le cas) entre les différentes mesures et les mesures de charges lésionnelles (automatique ou manuelle en T2) pour l'instant T_0.*

	Charge Lésionnelle Auto	Charge Lésionnelle T2
Charge Lésionnelle T2	0,71*	
Charge Lésionnelle T1	0,74*	0,99*
Charge Lésionnelle T2-FLAIR	0,79*	0,98*
FPC	-0,25	-0,35
Largeur du 3ème ventricule	0,13	0,09
Largeur des ventricules latéraux	0,26	0,51*

TAB. 7.2 – *Valeurs du coefficient de corrélation de Spearman entre les mesures de charges lésionnelles et le nombre de lésions donné par la même méthode et pour l'instant T_0.*

	Auto	T2	T1	T2-FLAIR
Charge Lésionnelle vs Nombre	0,47*	0,90*	0,80*	0,87*

TAB. 7.3 – *Valeurs du coefficient de corrélation de Spearman entre les valeurs d'EDSS et celles de charges lésionnelles (automatique ou manuelle en T2) et de fraction parenchymateuse cérébrale (FPC) pour l'instant T_0.*

	Charge Lésionnelle Auto	Charge Lésionnelle T2	FPC
EDSS	0,06	0,15	-0.16

rence significative entre le nombre de lésions pour les différentes méthodes (à l'inverse de la mesure de charge lésionnelle).

Pour cet instant, les valeurs de FPC se situent dans l'intervalle $[87; 93]$, de la largeur du troisième ventricule dans $[3; 8]$, de la largeur des ventricules latéraux dans $[16; 25]$ (un point aberrant à 28,8 correspondant au sujet SCHIS) et les valeurs d'EDSS dans $[0; 3, 5]$ (un point aberrant à 5 correspondant au sujet GUENA).

La table 7.4 donne les coefficients de corrélation (Pearson ou Spearman selon le cas) entre les différentes mesures et les mesures de charges lésionnelles (automatique ou manuelle en T2) pour l'instant T_1. Les corrélations trouvées à l'instant T_0 sont toujours présentes à l'instant T_1.

La table 7.5 montre qu'il y a toujours une corrélation significative entre le volume des lésions et leur nombre, excepté pour la méthode automatique.

La table 7.6 montre que l'EDSS n'est pas corrélé significativement avec une autre mesure.

7.1.1.3 Instant T_2

La base de données contient 7 sujets avec des IRM au troisième instant, T_2, de l'étude. Les boîtes à moustaches de la figure 7.2 donnent la distribution des différentes mesures à cet instant. Pour cet instant, les distributions des mesures sont similaires aux instants précédents. Seules les valeurs correspondantes au sujet SCHIS

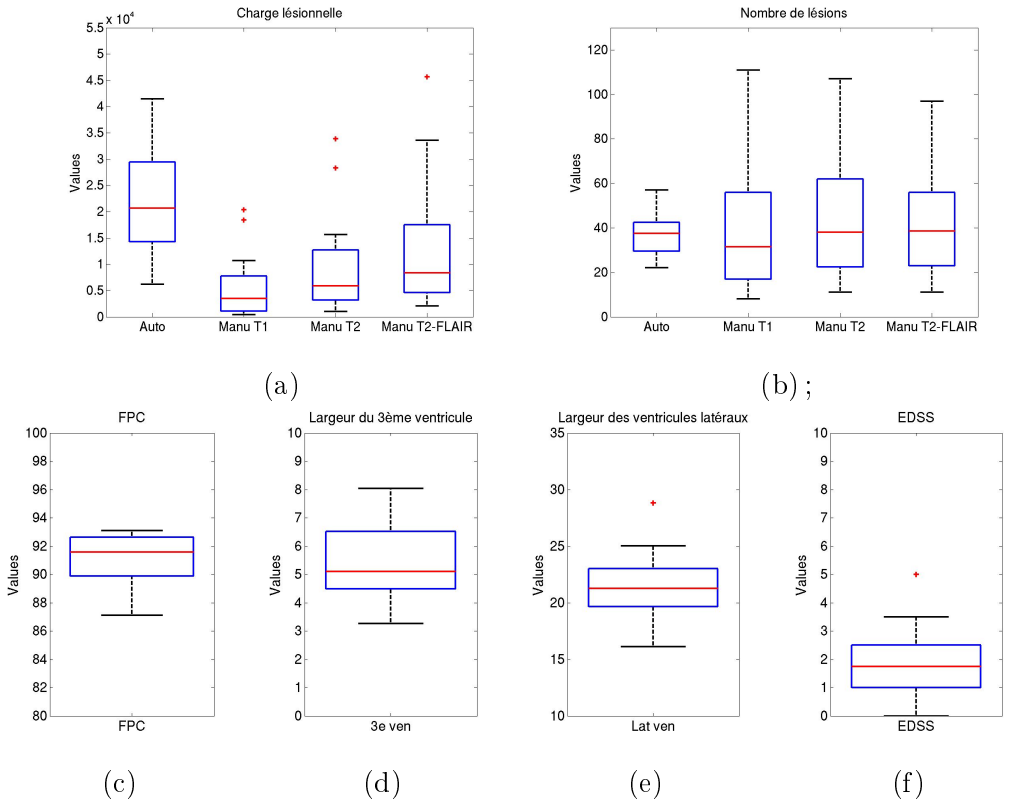


FIG. 7.2 – *Boîtes à moustaches des valeurs : (a) de charge lésionnelle, (b) du nombre des lésions, (c) de la fraction parenchymateuse cérébrale (FPC), (d) de la largeur du 3ème ventricule, (e) de la largeur des ventricules latéraux et (f) du score EDSS pour l'instant* T_1.

TAB. 7.4 – *Valeurs du coefficient de corrélation (de Pearson ou de Spearman selon le cas) entre les différentes mesures et les mesures de charges lésionnelles (automatique ou manuelle en T2) pour l'instant T_1.*

	Charge Lésionnelle Auto	Charge Lésionnelle T2
Charge Lésionnelle T2	0,76*	
Charge Lésionnelle T1	0,77*	0,97*
Charge Lésionnelle T2-FLAIR	0,74*	0,99*
FPC	-0,17	-0,22
Largeur du 3ème ventricule	-0,29	0,07
Largeur des ventricules latéraux	0,30	0,62*

TAB. 7.5 – *Valeurs du coefficient de corrélation de Spearman entre les mesures de charges lésionnelles et le nombre de lésions donné par la même méthode et pour l'instant T_1.*

	Auto	T2	T1	T2-FLAIR
Charge Lésionnelle vs Nombre	0,11	0,93*	0,88*	0,87*

TAB. 7.6 – *Valeurs du coefficient de corrélation de Spearman entre les valeurs d'EDSS et celles de charges lésionnelles (automatique ou manuelle en T2) et de fraction parenchymateuse cérébrale (FPC) pour l'instant T_1.*

	Charge Lésionnelle Auto	Charge Lésionnelle T2	FPC
EDSS	0,36	0,33	-0.13

restent comme points aberrants (le sujet DELEM n'a pas de troisième instant).

De même que pour les instants précédents, une analyse ANOVA à un facteur ne donne pas de différence significative entre le nombre de lésions pour les différentes méthodes.

Pour cet instant, les valeurs de FPC se situent dans l'intervalle $[84; 93]$, de la largeur du troisième ventricule dans $[4; 8]$, des ventricules latéraux dans $[17; 30]$ et les valeurs d'EDSS dans $[0; 5]$.

La table 7.7 donne les coefficients de corrélation (Pearson ou Spearman selon le cas) entre les différentes mesures et les mesures de charges lésionnelles (automatique ou manuelle en T2) pour l'instant T_2. Les corrélations entre la charge lésionnelle automatique et celles mesurées manuellement restent élevées mais non significatives. De même la corrélation entre la largeur des ventricules latéraux et la charge lésionnelle T2 n'est plus significative. Il sera intéressant de voir si ces absences de corrélations persistent avec plus de sujets. En effet, ces corrélations sont ici faites entre 7 valeurs.

La table 7.8 donne les coefficients de corrélation (Pearson ou Spearman selon le cas) entre les mesures de charges lésionnelles et le nombre de lésions donné par la même méthode et pour l'instant T_2. La table 7.8 montre que seules les corrélations entre le volume des lésions et leur nombre des séquences T2 et T1 restent significatives. De même, il faudra effectuer cette analyse avec plus de sujets.

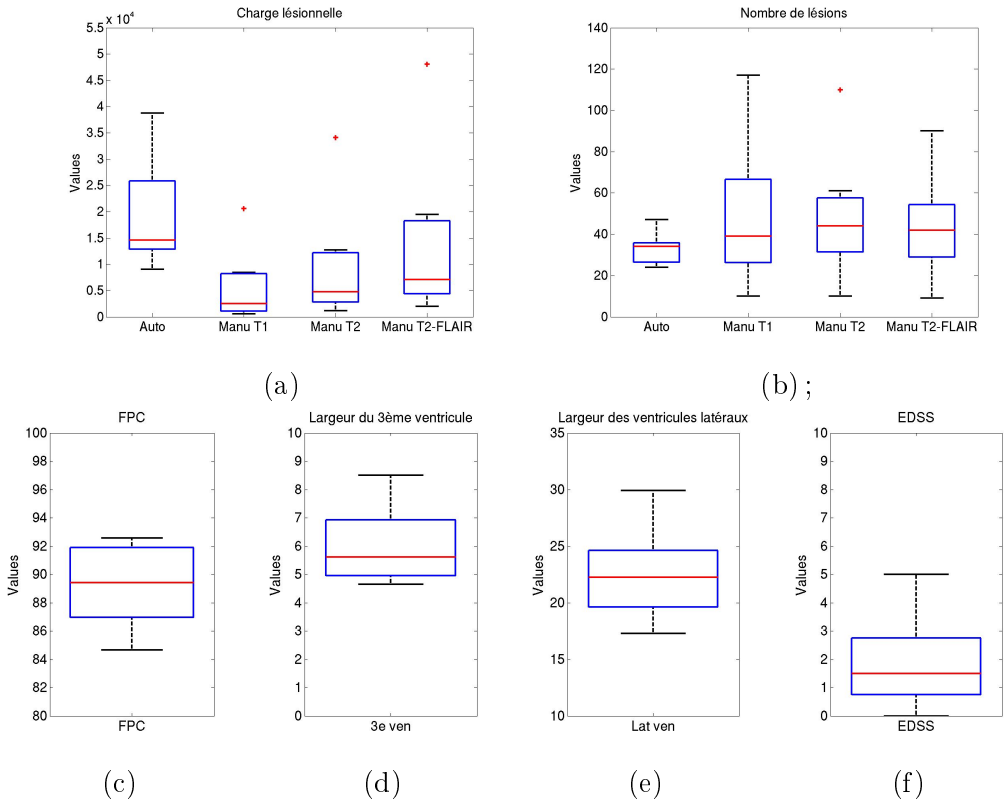


FIG. 7.3 – *Boîtes à moustaches des valeurs : (a) de charge lésionnelle, (b) du nombre des lésions, (c) de la fraction parenchymateuse cérébrale (FPC), (d) de la largeur du 3ème ventricule, (e) de la largeur des ventricules latéraux et (f) du score EDSS pour l'instant* T_2.

La table 7.9 montre que l'EDSS n'est corrélé significativement avec aucune mesure.

TAB. 7.7 – *Valeurs du coefficient de corrélation de Spearman entre les mesures de charges lésionnelles et le nombre de lésions donné par la même méthode et pour l'instant T_2.*

	Charge lésionnelle auto	Charge lésionnelle T2
Charge Lésionnelle T2	0,84	
Charge Lésionnelle T1	0,90	0.98*
Charge Lésionnelle T2-FLAIR	0,87	0,99*
FPC	-0,19	0,14
Largeur du 3ème ventricule	0,25	0,49
Largeur des ventricules latéraux	0,07	0,40

TAB. 7.8 – *Valeurs du coefficient de corrélation de Spearman entre les mesures de charges lésionnelles et le nombre de lésions donné par la même méthode et pour l'instant T_2.*

	Auto	T2	T1	T2-FLAIR
Charge Lésionnelle vs Nombre	0.46	0.94*	0.94*	0.88

TAB. 7.9 – *Valeurs du coefficient de corrélation de Spearman entre les valeurs d'EDSS et celles de charges lésionnelles (automatique ou manuelle en T2) et de fraction parenchymateuse cérébrale (FPC) pour l'instant T_2.*

	Charge Lésionnelle Auto	Charge Lésionnelle T2	FPC
EDSS (N=5)	0,90	0,90	-0.60

7.1.1.4 Instant T_3

La base de données contient trois sujets avec des IRM au quatrième instant, T_3, de l'étude. Les boîtes à moustaches de la figure 7.4 donnent la distribution des différentes mesures à cet instant.

Pour cet instant, les distributions des mesures sont similaires aux instants précédents. Les valeurs de FPC se situent dans l'intervalle $[86; 92]$, de la largeur du troisième ventricule dans $[5; 7]$, de la largeur des ventricules latéraux dans $[17; 23]$ et le score EDSS est égal à 1 pour les deux sujets où il est connu.

N'ayant que trois sujets pour cet instant, je n'ai pas fait l'analyse de corrélation (résultats non significatifs).

7.1.1.5 Discussion

L'analyse de la base, instant par instant, a permis de confirmer son homogénéité. Ces sujets sont tous atteints d'une SEP rémittente et ont un score EDSS assez faible. Cette homogénéité se retrouve dans les mesures de charge lésionnelle et d'atrophie. Pour ces mesures, le caractère homogène se retrouve sur les différents

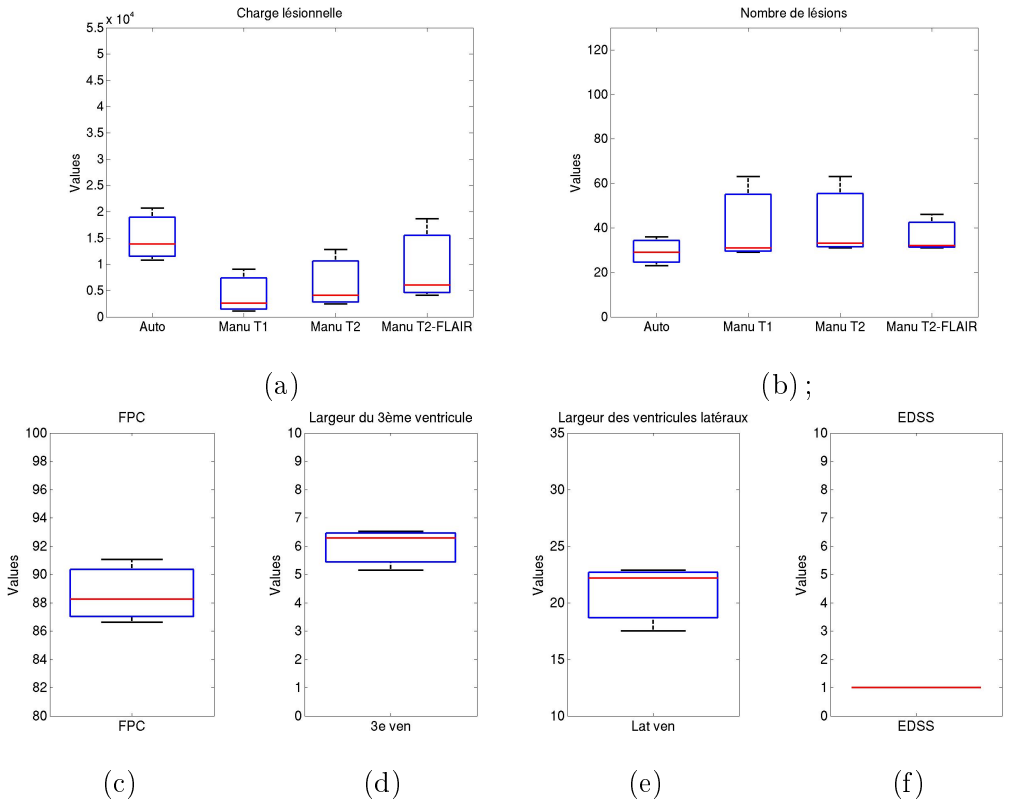


FIG. 7.4 – *Boîtes à moustaches des valeurs : (a) de charge lésionnelle, (b) du nombre des lésions, (c) de la fraction parenchymateuse cérébrale (FPC), (d) de la largeur du 3ème ventricule, (e) de la largeur des ventricules latéraux et (f) du score EDSS pour l'instant* T_3.

instants. Il semble toutefois impossible de dire que la cohorte évolue, elle aussi, de manière homogène. Premièrement, les différents intervalles de valeurs de mesures donnés à chaque instant ne sont pas comparables parce qu'ils ont été obtenus sur un nombre de sujets différents. Deuxièmement, le nombre de sujets pour les troisième et quatrième instants reste très faible. Enfin, certains sujets (par exemple niçois) sont toujours avec le même traitement de fond alors que d'autres sont passés sous immunosuppresseurs. La prédiction de l'évolution des mesures (qu'elle soit naturelle ou sous traitement) semble donc difficile à faire.

Il existe cependant une certaine cohérence des résultats dans la mesure où les points aberrants correspondent aux mêmes sujets. Ainsi SCHIS et DELEM correspondent aux points aberrants de charge lésionnelle pour les divers instants. De même, on retrouve sur ces instants la largeur des ventricules latéraux de SCHIS comme étant supérieure à celles des autres sujets.

Cette analyse montre aussi que le volume de lésions mesuré automatiquement est corrélé aux volumes mesurés manuellement sur les deux premiers instants. Cette corrélation existe aussi au troisième instant mais n'est pas significative (peut-être à cause d'un nombre de données trop faible).

En revanche, comme cela avait déjà été trouvé [Ge 2000, McFarland 2002], il n'y a pas de corrélation entre les mesures d'atrophie et la charge lésionnelle. Pour les deux premiers instants, la charge lésionnelle T2 est toutefois corrélée significativement ($\alpha < 0,05$) avec la largeur des ventricules latéraux. Je ne dispose pour l'instant d'aucune explication à ce phénomène.

L'absence de corrélation entre le score EDSS et les diverses mesures est conforme à la littérature. Concernant l'atrophie, l'absence de ce lien semble d'autant plus logique que dans le cas de sujets atteints de SEP rémittente, cette corrélation se fait entre un processus permanent et irréversible (l'atrophie) et un score physique, qui, dans cette forme de SEP est voué à fluctuer énormément en fonction des poussées et revient souvent à un état de base proche de 0. Il reste qu'il était intéressant de valider l'absence de ce lien sur notre base.

7.1.2 Analyse dynamique

Dans cette section, je propose de regarder les résultats de manière dynamique, c'est-à-dire d'étudier et de comparer les pourcentages de variation (PV) entre l'instant T_0 et l'instant considéré. De même que dans la section 6.4.1, ce pourcentage est obtenu par le calcul de Δ, entre l'instant de mesure, IM, et l'instant de référence, IR (équation 7.1).

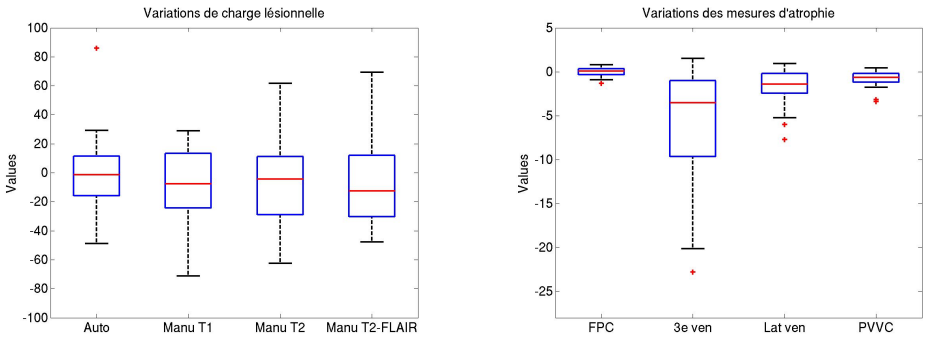
$$\Delta = 100 \times \frac{\text{Mesure(IM)} - \text{Mesure(IR)}}{\text{Mesure(IR)}} \tag{7.1}$$

Au vu des résultats de l'étude précédente, j'ai décidé d'analyser ici les variations des mesures de charges lésionnelles et d'atrophie. Les valeurs du pourcentage de variation de volume cérébral donné par SIENA ont été rajoutées. Les valeurs du

coefficient de corrélation (de Pearson ou de Spearman selon le cas (cf. chapitre 6)) entre ces diverses mesures seront aussi calculées.

7.1.2.1 Variations $T_0 \rightarrow T_1$, sur 16 sujets

La figure 7.5 (a) montre les distributions des variations de charge lésionnelle pour les différentes mesures. Celles-ci sont similaires (pas de différence significative montrée par une analyse ANOVA à un facteur). Un point aberrant est présent pour la méthode automatique (sujet : CALCE). De même, la figure 7.5 (b) montre les distributions des variations de mesures d'atrophie pour les différentes méthodes. Plusieurs points aberrants sont présents (sujet HUBSO, CELFR (2 fois) et PAUCH (2 fois)), ainsi qu'un point aberrant (non représenté ici) correspondant à une variation de -85% pour la largeur du 3ème ventricule (sujet CALCE). Cette très forte variation correspond sûrement à une erreur de mesure. Une analyse ANOVA à un facteur donne une différence significative entre les distributions.



(a) (b)

FIG. 7.5 – *Boîtes à moustaches des valeurs de variations : (a) de charge lésionnelle, (b) de mesures d'atrophie entre T_0 et T_1.*

La table 7.10 donne les coefficients de corrélation (Pearson ou Spearman selon le cas) entre le pourcentage de variation des mesures de charges lésionnelles (automatique ou manuelle en T2) et le pourcentage de variation des autres mesures. Ce tableau montre que les variations de charge lésionnelle T2 sont corrélées avec les variations des autres charges lésionnelles mesurées manuellement. Cette corrélation est par contre inexistante avec les variations de charge lésionnelle mesurée automatiquement. Paradoxalement, il existe une corrélation significative entre le pourcentage de variation du volume cérébral (PVVC) donné par SIENA et le PV de charge lésionnelle mesurée automatiquement.

La table 7.11 donne les coefficients de corrélation (Pearson ou Spearman selon le cas) entre les mesures automatiques d'atrophie et les autres mesures. Les corrélations entre le PV de la FPC mesurée par LCLASS et les PV des mesures manuelles (déjà mises en évidence dans la section 6.4.3) sont ici retrouvées. Le PVVC donné par SIENA n'est lui corrélé avec aucune autre mesure.

TAB. 7.10 – *Valeurs du coefficient de corrélation (de Pearson ou de Spearman selon le cas) entre le pourcentage de variation des mesures de charges lésionnelles (automatique ou manuelle en T2) et le pourcentage de variation des autres mesures dans l'intervalle T_0 à T_1.*

	PV Charge Lésionnelle Auto	PV Charge Lésionnelle T2
PV Charge Lésionnelle T2	-0,12	
PV Charge Lésionnelle T1	-0,11	0.78*
PV Charge Lésionnelle T2-FLAIR	-0,06	0,93*
PV FPC LCLASS	0,14	-0,23
PV Largeur du 3ème ventricule	-0,17	0,02
PV Largeur des ventricules latéraux	-0,16	-0,34
PV VC SIENA	0,63*	-0,15

TAB. 7.11 – *Valeurs du coefficient de corrélation (de Pearson ou de Spearman selon le cas) entre les mesures automatiques d'atrophie et les autres mesures dans l'intervalle T_0 à T_1.*

	PV FPC LCLASS	PV VC SIENA
PV Largeur du 3ème ventricule	0.51*	0,14
PV Largeur des ventricules latéraux	0.68*	0,23
PV VC SIENA	0,28	

7.1.2.2 Variations $T_0 \rightarrow T_2$, sur 7 sujets

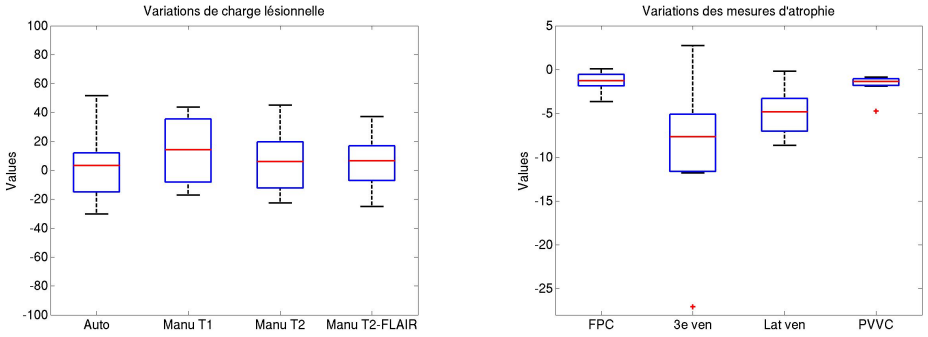
La figure 7.6 (a) montre les distributions des variations de charge lésionnelle pour les différentes mesures. De même, la figure 7.6 (b) montre les distributions des variations de mesures d'atrophie pour les différentes méthodes. Le sujet CELFR correspond au point aberrant de variation de la largeur du troisième ventricule. Le dernier point aberrant (SIENA PVVC) correspond à SCHIS.

Comme pour l'instant précédent, une analyse ANOVA à un facteur ne révèle aucune différence entre les différentes méthodes de mesure de la charge lésionnelle contrairement aux mesures d'atrophie où une différence existe.

Aucune des valeurs des tables 7.12 (coefficients de corrélation (Pearson ou Spearman selon le cas) entre le pourcentage de variation des mesures de charges lésionnelles (automatique ou manuelle en T2) et le pourcentage de variation des autres mesures) et 7.13 (coefficients de corrélation (Pearson ou Spearman selon le cas) entre les mesures d'atrophie) ne sont significatives. Le nombre de sujets peu élevé explique peut-être cela.

7.1.2.3 Variations $T_0 \rightarrow T_3$, sur 3 sujets

La figure 7.7 (a) montre les distributions des variations de charge lésionnelle pour les différentes mesures. De même, la figure 7.7 (b) montre les distributions des variations de mesures d'atrophie pour les différentes méthodes.



(a) (b) ;

FIG. 7.6 – *Boîtes à moustaches des valeurs de variations : (a) de charge lésionnelle, (b) de mesures d'atrophie entre T_0 et T_2.*

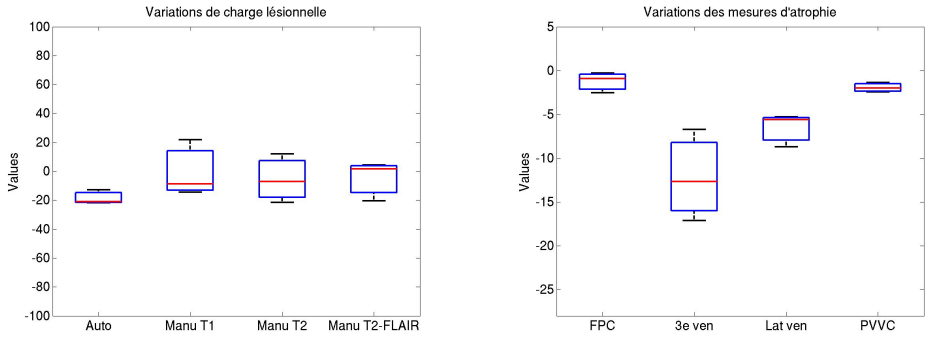
TAB. 7.12 – *Valeurs du coefficient de corrélation (de Pearson ou de Spearman selon le cas) entre le pourcentage de variation des mesures de charges lésionnelles (automatique ou manuelle en T2) et le pourcentage de variation des autres mesures dans l'intervalle T_0 à T_2.*

	PV Charge Lésionnelle Auto	PV Charge Lésionnelle T2
PV Charge Lésionnelle T2	-0,27	
PV Charge Lésionnelle T1	-0,24	0.65
PV Charge Lésionnelle T2-FLAIR	-0,12	0,84
PV FPC LCLASS	-0,12	-0,52
PV Largeur du 3ème ventricule	-0,68	-0,39
PV Largeur des ventricules latéraux	-0,46	-0,29
PV VC SIENA	0,46	-0,64

TAB. 7.13 – *Valeurs du coefficient de corrélation (de Pearson ou de Spearman selon le cas) entre les mesures automatiques d'atrophie et les autres mesures dans l'intervalle T_0 à T_2.*

	PV FPC LCLASS	PV VC SIENA
PV Largeur du 3ème ventricule	0.75	0,18
PV Largeur des ventricules latéraux	0.54	-0,39
PV VC SIENA	0,25	

Comme pour les instants précédents, une analyse ANOVA à un facteur ne révèle aucune différence entre les différentes méthodes de mesure de la charge lésionnelle contrairement aux mesures d'atrophie où une différence existe.



(a) (b) ;

FIG. 7.7 – *Boîtes à moustaches des valeurs de variations : (a) de charge lésionnelle, (b) de mesures d'atrophie entre T_0 et T_3.*

De même que dans la section 7.1.1, n'ayant que trois sujets ayant des images à l'instant T_3, je n'ai pas fait l'analyse de corrélation dans ce cas (résultats non significatifs).

7.1.2.4 Discussion

L'analyse des variations des résultats est un argument supplémentaire à l'hypothèse d'une évolution globale similaire de la base. En effet, peu de variations correspondent à des points aberrants (excepté principalement pour les sujets CALCE, CELFR et PAUCH).

Pour les différents instants, les variations de charges lésionnelles montrées par les diverses mesures sont similaires même si des variations, notamment de valeurs médianes, existent.

Excepté entre les mesures manuelles, les variations de charge lésionnelle des différentes mesures ne sont pas corrélées significativement entre elles. Concernant les mesures d'atrophie, la corrélation entre les variations de FPC données par LCLASS et les variations de mesures manuelles a été confirmée. En revanche, les variations mesurées par SIENA ne se corrèlent avec aucune autre mesure excepté les variations de charge lésionnelle mesurées automatiquement. Je ne vois pas d'explication à ce dernier point.

Ces corrélations ne sont plus constatées pour les variations entre T_0 et T_2 ou T_0 et T_1 mais cela peut s'expliquer par le faible nombre de sujets disponibles pour ces instants.

7.1.3 Analyse par sujet

Les sections précédentes ont permis de prouver une certaine homogénéité de la base et d'étudier la présence ou l'absence de corrélation entre les différentes mesures. Toutefois, ces analyses ne permettent pas de suivre une évolution des mesures dans le temps. En outre, ces évolutions peuvent dépendre de phénomènes cliniques [Bar-Zohar 2008]. Dans le but de chercher des explications cliniques aux résultats obtenus, une étude de l'évolution des mesures sujet par sujet est nécessaire.

D'un point de vue pratique cette section donne les courbes d'évolution d'atrophie entre les instants et l'instant d'origine, T_0. On a ensuite cherché à expliquer ces courbes en fonction de celle de la charge lésionnelle en T2 (mesure de référence manuelle) et d'effets cliniques dont la présence a été confirmée ou infirmée par Mikael Cohen, interne en neurologie, avec l'aide du dossier clinique du sujet.

La charge lésionnelle en T2 à l'instant T_0 est exprimée en trois catégories : faible (valeur à T_0 inférieure au premier quartile des charges lésionnelles (CL) T2 à ce même instant), moyenne (valeur à T_0 comprise entre le premier et le troisième quartile des CL T2 à ce même instant), forte (valeur à T_0 supérieure au troisième quartile des CL T2 à ce même instant).

7.1.3.1 Effets cliniques possibles

Poussées : Il s'agit d'une période de survenue ou d'aggravation de signes neurologiques durant plus de 24 heures et séparée de la précédente d'au moins un mois. Celle-ci peut correspondre à une augmentation de la CL. Cependant, la corrélation clinico-radiologique entre la survenue d'un évènement clinique (une poussée) et l'évolution de la charge lésionnelle reste difficile à prouver. Par ailleurs, notre étude ne prend en compte que la charge lésionnelle encéphalique et non les plaques médullaires.

Augmentation de la charge lésionnelle (CL) : Le volume des lésions peut augmenter. Il peut alors y avoir un œdème focal (par exemple si une plaque devient œdémateuse ou si elle prend le gadolinium). On peut alors supposer que cette infiltration œdémateuse du tissu cérébral peut "artificiellement" induire une augmentation du volume cérébral et ainsi fausser la mesure automatique de l'atrophie.

Syndrome pseudo grippal (SPG) : L'interféron peut être responsable (au début du traitement) d'une "réaction inflammatoire" qui se traduit cliniquement par un SPG (le sujet rapporte un état fébrile, des douleurs articulaires et musculaires diffuses). Ce syndrome est le principal effet indésirable de l'interféron. Celui-ci répond très bien au traitement par paracétamol. Il faut du reste préciser que certains sujets n'ont pas de SPG mais déclarent prendre du paracétamol de façon systématique, masquant peut être ainsi un SPG. Le SPG produirait un œdème diffus au niveau cérébral. Cet effet pourrait donc produire une augmentation du volume cérébral même si dans les faits un phénomène d'atrophie reste présent. Le volume

cérébral peut augmenter alors qu'il peut exister une diminution diffuse de la densité des tissus. La présence de ce syndrome et son intensité varient énormément d'un sujet à l'autre. Dans notre étude cet effet se produit principalement entre T_0 et T_1. Dans de rares cas, il se prolonge au-delà de T_1. En cas d'absence du syndrome entre T_0 et T_1, il ne devrait pas apparaître après, sauf en cas de changement de la dose du traitement.

Changement de dose de traitement (22 ou 44 microgrammes) : L'augmentation de la dose du traitement peut induire un nouveau SPG. Il doit toutefois être noté que cet effet est rare parce que la dose n'est augmentée que si le sujet supportait plutôt bien le traitement avec une dose plus faible. Un sujet peut donc présenter un SPG entre d'autres instants que T_0 et T_1.

7.1.3.2 Sujets avec 2 instants

Sujet ALLJE

La table 7.14 présente les résultats des mesures pour le sujet ALLJE. La charge lésionnelle diminue fortement (-62%). Les mesures automatiques d'atrophie montrent une augmentation du volume cérébral alors que les mesures manuelles non. Cette augmentation de volume peut s'expliquer par la présence d'un syndrome pseudo grippal. Dans ce cas, on peut supposer que la diminution de charge lésionnelle n'impacte pas les courbes d'atrophie car la charge lésionnelle est faible.

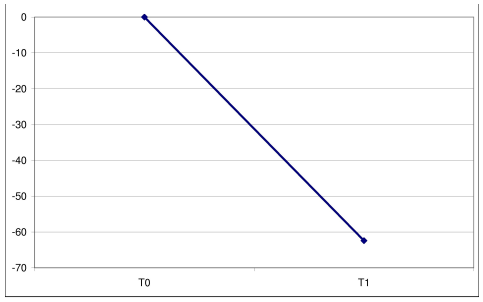
Sujet CALCE

La table 7.15 présente les résultats des mesures pour le sujet CALCE. La courbe de variation de la largeur du troisième ventricule n'est pas tracée ici car elle correspond à une augmentation de largeur du ventricule (= une diminution du volume cérébral) de 85% (cette dynamique empêcherait de voir l'évolution des autres courbes). La charge lésionnelle diminue fortement (-38%). Comme pour ALLJE, les mesures automatiques d'atrophie montrent une augmentation du volume cérébral alors que les mesures manuelles non. Cette augmentation de volume peut s'expliquer par la présence d'un syndrome pseudo grippal. Celui-ci causerait une augmentation du volume cérébral qui serait minorée par la diminution de la charge lésionnelle (forte). Il est à noter que le syndrome pseudo grippal est apparu ici lors d'un changement de dose de traitement.

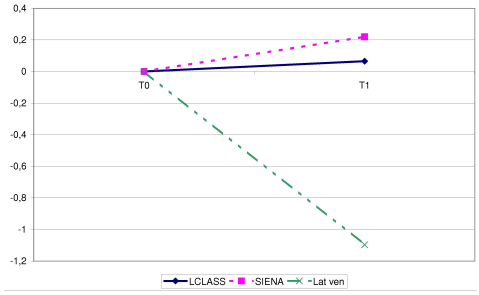
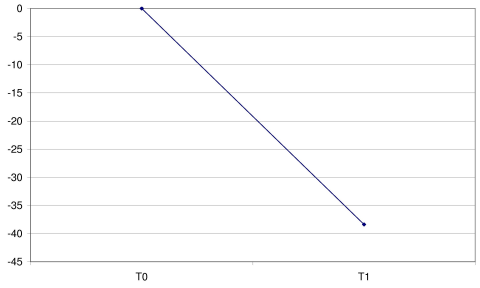
Sujet DAZIN

Les résultats des mesures pour le sujet DAZIN sont donnés dans la table 7.16. LCLASS et les mesures manuelles indiquent une augmentation du volume cérébral. Celui-ci semble se justifier pleinement par une augmentation de la charge lésionnelle ainsi que par la présence d'un syndrome pseudo grippal. Pourtant, SIENA ne semble pas être sensible à ce phénomène.

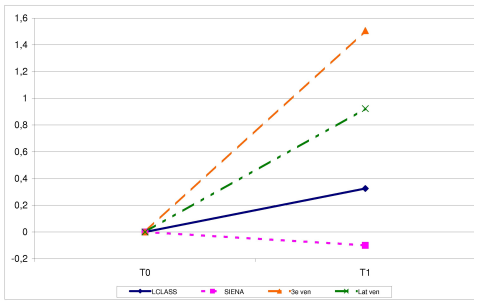
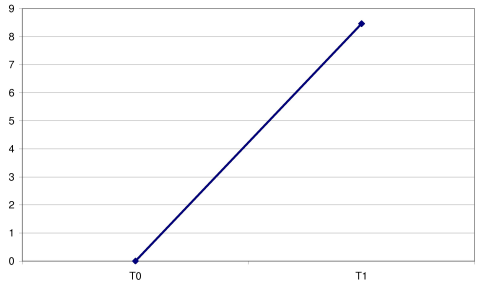
TAB. 7.14 – *Données du sujet ALLJE.*

	
Évolution des mesures d'atrophie	Évolution de la charge lésionnelle en T2
Charge lésionnelle T2 à T_0 : Faible	
Confirmation clinique d'un syndrome pseudo grippal : Positive	

TAB. 7.15 – *Données du sujet CALCE.*

	
Évolution des mesures d'atrophie	Évolution de la charge lésionnelle en T2
Charge lésionnelle T2 à T_0 : Forte	
Confirmation clinique d'un syndrome pseudo grippal : Positive	

TAB. 7.16 – *Données du sujet DAZIN.*

	
Évolution des mesures d'atrophie	Évolution de la charge lésionnelle en T2
Charge lésionnelle T2 à T_0 : Moyenne	
Confirmation clinique d'un syndrome pseudo grippal : Positive	

Sujet GIULA

La table 7.17 présente les résultats des mesures pour le sujet GIULA. Dans ce cas, l'atrophie mesurée par SIENA semble classique (proche des -0,6%). Ceci semble se confirmer par les mesures manuelles, une diminution de la charge lésionnelle et une absence de syndrome pseudo grippal. Étrangement, LCLASS indique une augmentation du volume cérébral. On peut ici supposer une erreur de mesure. Il serait aussi intéressant de voir si cette évolution persiste avec l'ajout du prochain instant. Je rappelle que LCLASS dépend du nombre d'instants traités (cf. section 6.4.5).

Sujet GUENA

Les résultats des mesures pour le sujet GUENA sont donnés dans la table 7.18. LCLASS et les mesures manuelles détectent une atrophie relativement faible. Celle-ci pourrait s'expliquer par l'augmentation de 4.3% d'une forte charge lésionnelle. Une atrophie plus forte serait masquée par cette augmentation. Dans ce cas, SIENA semble par contre ne pas être sensible à cette variation de charge lésionnelle et détecte une atrophie forte (-1,77%).

7.1.3.3 Sujets avec 3 instants

Sujet CELFR

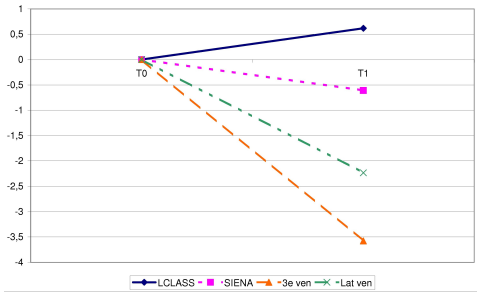
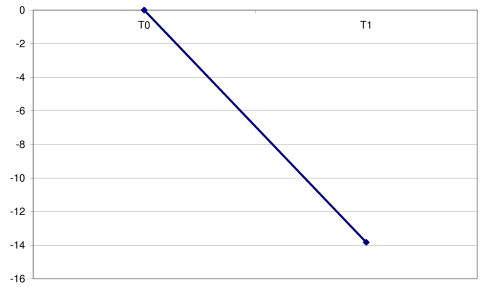
Les résultats des mesures pour le sujet CELFR sont donnés dans la table 7.19.

Variations entre l'instant T_0 et l'instant T_1 : La diminution du volume cérébral est assez faible pour SIENA (-0,23%) et classique pour LCLASS (-0,56%). Au vu de l'atrophie révélée par les mesures manuelles, on peut toutefois supposer qu'il y a en fait une atrophie plus forte. Les résultats des mesures automatiques seraient minorés par l'augmentation (très forte +61,8%) de la charge lésionnelle (même si celle-ci est faible à T_0). Cette augmentation peut ne pas avoir d'effet sur les mesures manuelles si, par exemple, les nouvelles lésions ne sont pas proches des ventricules.

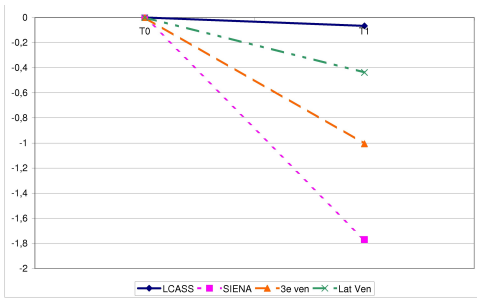
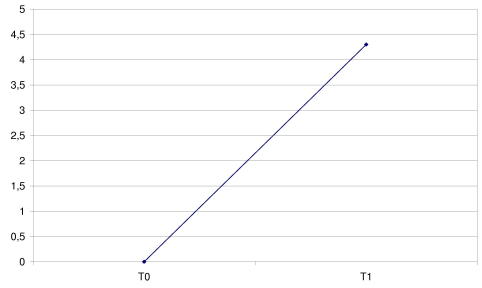
Variations entre l'instant T_1 et l'instant T_2 : Les méthodes automatiques indiquent une atrophie plus forte alors que les taux de variation des largeurs des ventricules diminuent. On peut supposer que le phénomène d'atrophie globale ralentit et que l'augmentation constatée par les méthodes manuelles ne reflète que la diminution de la charge lésionnelle.

Ne sachant pas si un syndrome pseudo grippal s'est produit (patient de Clermont-Ferrand pour lequel nous n'avons pas accès aux données cliniques), cette analyse est entièrement hypothétique.

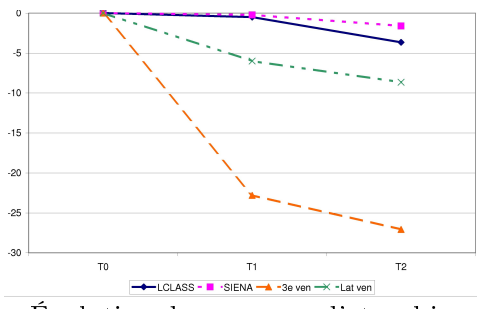
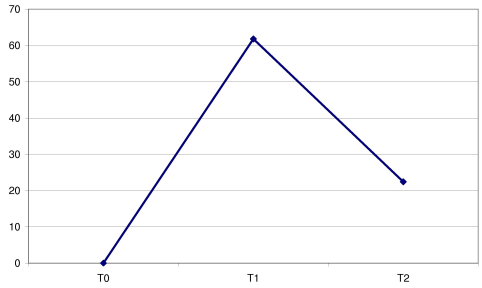
TAB. 7.17 – *Données du sujet GIULA.*

	
Évolution des mesures d'atrophie	Évolution de la charge lésionnelle en T2
Charge lésionnelle T2 à T_0 : Moyenne	
Confirmation clinique d'un syndrome pseudo grippal : Négative	

TAB. 7.18 – *Données du sujet GUENA.*

	
Évolution des mesures d'atrophie	Évolution de la charge lésionnelle en T2
Charge lésionnelle T2 à T_0 : Forte	
Confirmation clinique d'un syndrome pseudo grippal : Négative	

TAB. 7.19 – *Données du sujet CELFR.*

	
Évolution des mesures d'atrophie	Évolution de la charge lésionnelle en T2
Charge lésionnelle T2 à T_0 : Faible	
Confirmation clinique d'un syndrome pseudo grippal : Inconnue	

Sujet HUBSO

La table 7.20 présente les résultats des mesures pour le sujet HUBSO.

Variations entre l'instant T_0 et l'instant T_1 : Une atrophie assez forte est constatée (LCLASS : -1,31% ; SIENA : -1,04%). Celle-ci peut s'expliquer par la diminution de la charge lésionnelle et l'absence de syndrome pseudo grippal.

Variations entre l'instant T_1 et l'instant T_2 : Excepté SIENA, tous les marqueurs montrent une augmentation du volume cérébral. Celle-ci peut s'expliquer par l'augmentation de la charge lésionnelle qui d'ailleurs chez ce patient s'est accompagnée d'un évènement clinique authentifié (une poussée). Pendant cet intervalle, la dose de traitement a aussi été changée. Il y a eu au début une consommation de paracétamol en raison de symptômes certes de faible intensité mais pouvant aussi expliquer l'augmentation du volume cérébral constatée.

Sujet SAIVE

Les résultats des mesures pour le sujet SAIVE sont donnés dans la table 7.21.

Variations entre l'instant T_0 et l'instant T_1 : SIENA et les mesures manuelles indiquent une diminution du volume cérébral. En revanche, LCLASS indiquerait une faible augmentation (+0,17%). Cette dernière pourrait s'expliquer par l'augmentation de la charge lésionnelle. Toutefois, il est difficile de conclure sur ces hypothèses parce qu'on ignore s'il y a eu un syndrome pseudo grippal (sujet provenant de Montpellier).

Variations entre l'instant T_1 et l'instant T_2 : Tous les marqueurs donnent une diminution du volume cérébral. Cette diminution semble logique car la charge lésionnelle diminue aussi.

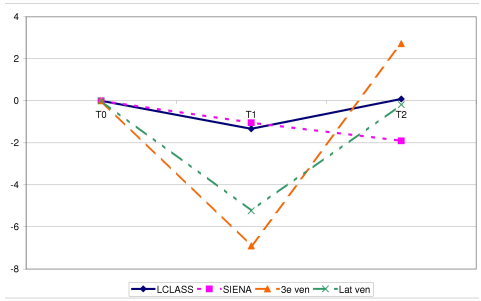
Sujet SCHIS

La table 7.22 présente les résultats des mesures pour le sujet SCHIS.

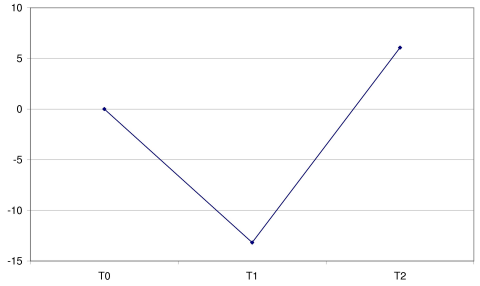
Variations entre l'instant T_0 et l'instant T_1 : SIENA mesure une forte atrophie (-1,22%) alors que l'évolution de la largeur des ventricules latéraux indique une atrophie plus faible. LCLASS et la largeur du troisième ventricule mesurent même une augmentation du volume cérébral. En absence de syndrome pseudo grippal, on peut supposer que cette augmentation est due à la forte augmentation (+44%) de la charge lésionnelle (qui était déjà forte). Celle-ci masquerait l'atrophie réelle. Comme pour le sujet GUENA, les résultats de SIENA semblent ici ne pas dépendre de l'évolution de la charge lésionnelle.

Variations entre l'instant T_1 et l'instant T_2 : Pendant cet intervalle, la charge lésionnelle se stabilise. Toutes les mesures d'atrophie indiquent une forte atrophie.

TAB. 7.20 – *Données du sujet HUBSO.*



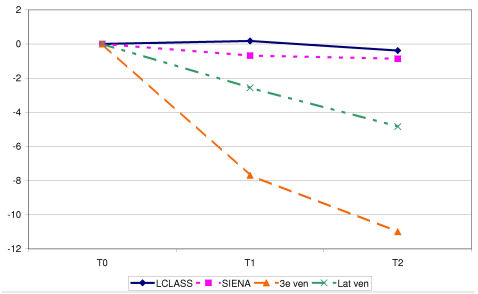
Évolution des mesures d'atrophie



Évolution de la charge lésionnelle en T2

Charge lésionnelle T2 à T_0	:	Moyenne
Confirmation clinique d'un syndrome pseudo grippal	:	Négative entre T_0 et T_1, Positive entre T_1 et T_2.

TAB. 7.21 – *Données du sujet SAIVE.*



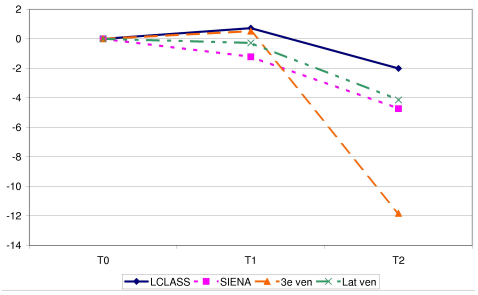
Évolution des mesures d'atrophie



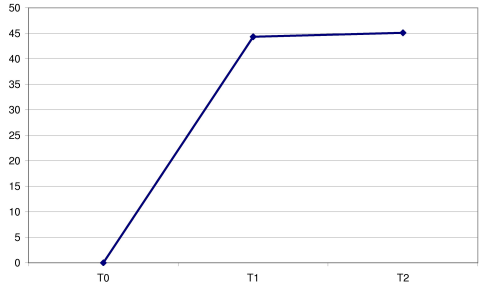
Évolution de la charge lésionnelle en T2

Charge lésionnelle T2 à T_0	:	Moyenne
Confirmation clinique d'un syndrome pseudo grippal	:	Inconnue

TAB. 7.22 – *Données du sujet SCHIS.*



Évolution des mesures d'atrophie



Évolution de la charge lésionnelle en T2

Charge lésionnelle T2 à T_0	:	Forte
Confirmation clinique d'un syndrome pseudo grippal	:	Négative

7.1.3.4 Sujets avec 4 instants

Sujet FRECO

La table 7.23 présente les résultats des mesures pour le sujet FRECO.

Variations entre l'instant T_0 et l'instant T_1 : Une faible atrophie est révélée par LCLASS et SIENA (-0.34% pour les deux). Ces mesures peuvent en fait refléter une atrophie plus forte (montrée par les mesures manuelles) mais masquée par l'augmentation de la charge lésionnelle.

Variations entre l'instant T_1 et l'instant T_2 : Les marqueurs mesurent une atrophie plus forte. Celle-ci peut s'expliquer par la diminution de la charge lésionnelle.

Variations entre l'instant T_2 et l'instant T_3 : Excepté SIENA, les marqueurs d'atrophie semblent se stabiliser. On peut supposer une stabilisation de l'état du sujet (stabilisation de la charge lésionnelle, retour de l'EDSS à 1).

Dans ce cas, LCLASS suit l'évolution des mesures manuelles, contrairement à SIENA qui ne le fait pas.

Sujet LICPA

Les résultats des mesures pour le sujet LICPA sont donnés dans la table 7.24.

Variations entre l'instant T_0 et l'instant T_1 : LCLASS et la mesure de la largeur du troisième ventricule montrent une très légère augmentation (LCLASS : +0,15%, 3ème ven. : +0.93%) du volume cérébral. Les autres mesures en donnent une faible diminution (<0,15%). Dans le même temps, la charge lésionnelle (qui était déjà faible) diminue. La seule explication semble donc une augmentation du volume cérébral due à un syndrome pseudo grippal. La présence de ce dernier a été confirmée cliniquement.

Variations entre l'instant T_1 et l'instant T_2 : Les mesures d'atrophie révèlent une forte atrophie malgré une augmentation de la charge lésionnelle. Cette observation pourrait s'expliquer par la disparition du syndrome pseudo grippal dont les effets s'amenderaient en révèlant ainsi l'atrophie sous-jacente.

Variations entre l'instant T_2 et l'instant T_3 : SIENA et les variations de la largeur des ventricules latéraux montrent une diminution de l'atrophie. LCLASS montre un gain de volume cérébral. Ceci semble correspondre à une stabilisation de la charge lésionnelle. En revanche, la mesure de largeur du troisième ventricule montre une atrophie très forte. On peut supposer dans ce cas une erreur de mesure.

Sujet VELMI

Les résultats des mesures pour le sujet VELMI sont donnés dans la table 7.25.

Variations entre l'instant T_0 et l'instant T_1 : SIENA et LCLASS reflètent une atrophie légèrement faible (-0,31%) malgré une augmentation de la charge lésionnelle. On peut donc ici supposer que l'atrophie réelle est en fait plus forte mais qu'elle est masquée par cette augmentation de charge lésionnelle. Pour ce sujet, l'absence ou la présence d'un syndrome pseudo grippal est incertaine car celui-ci prenait du paracétamol de manière préventive.

Variations entre l'instant T_1 et l'instant T_2 : Toutes les mesures indiquent une atrophie plus forte entre ces deux instants qui pourrait s'expliquer par la diminution de la charge lésionnelle.

Variations entre l'instant T_2 et l'instant T_3 : L'atrophie reste forte malgré une stabilisation de la charge lésionnelle.

Pour ce sujet, on peut noter qu'entre T_0 et T_2, les courbes d'atrophie de LCLASS et de SIENA se superposent parfaitement.

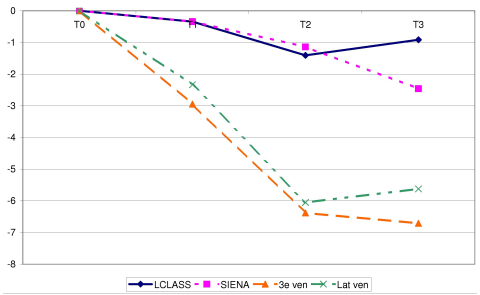
7.1.3.5 Discussion

Dans la majorité des cas, les courbes d'atrophie sont expliquées par des données cliniques. Cette analyse basée sur la formulation d'hypothèses (confirmées ou infirmées ensuite par les données cliniques) montre bien la complexité du phénomène d'atrophie et donc la difficulté à le mesurer. Cette complexité s'illustre par l'exemple d'une éponge gorgée d'eau : le volume n'est pas le même que celui de la même éponge sèche et pourtant il y a toujours la même quantité d'éponge. Chercher à mesurer une masse sur des observations visuelles est donc très délicat. Pour cette raison, je préfère parler d'évaluation de l'atrophie plutôt que de mesure d'atrophie.

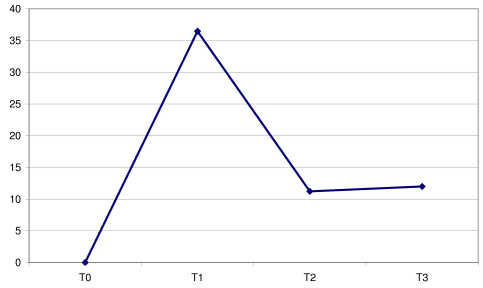
De plus, les différentes méthodes d'évaluation de l'atrophie sont basées sur des principes différents. Par exemple, la largeur du 3ème ventricule est une mesure locale qui certes est corrélée avec l'atrophie globale dans la majorité des cas mais qui peut sur certains sujets ne pas l'être. De même, l'atrophie globale n'est pas non plus forcément liée à l'atrophie des ventricules latéraux. Concernant les méthodes automatiques d'évaluation globale LCLASS et SIENA, on peut également supposer que celles-ci ne mesurent pas exactement les mêmes phénomènes. Par exemple, LCLASS rejette des voxels qu'elle ne sait pas classer et est appliquée dans un masque du cerveau différent de celui de SIENA.

En conséquence, l'hypothèse qu'une méthode peut être meilleure dans certains cas qu'une autre (et inversement) a du sens. Par exemple, LCLASS, en contradiction avec les autres mesures, semble détecter une augmentation du volume cérébral pour GIULA or celle-ci ne semble avoir aucune justification clinique. Ceci tend donc à montrer une erreur de mesure de la part de LCLASS. SIENA, au contraire des

TAB. 7.23 – *Données du sujet FRECO.*



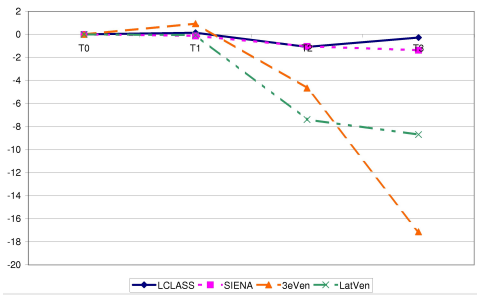
Évolution des mesures d'atrophie



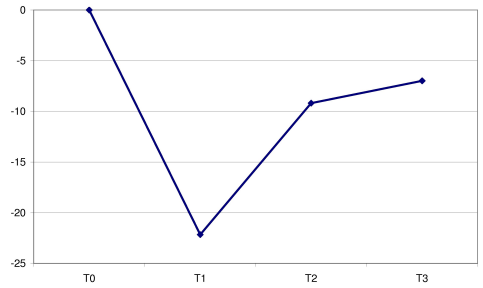
Évolution de la charge lésionnelle en T2

Charge lésionnelle T2 à T_0	:	Moyenne
Confirmation clinique d'un syndrome pseudo grippal	:	Négative

TAB. 7.24 – *Données du sujet LICPA.*



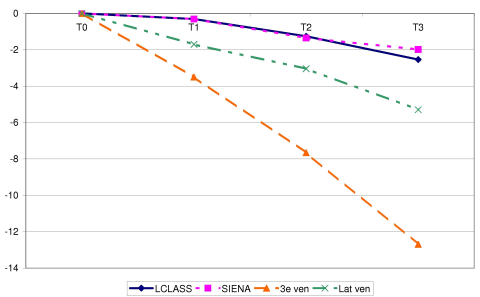
Évolution des mesures d'atrophie



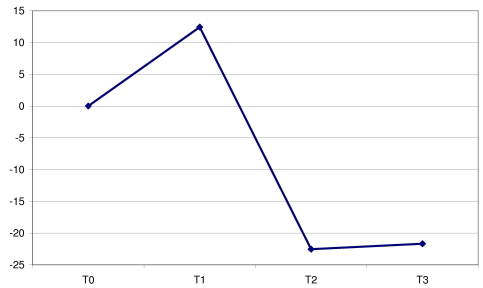
Évolution de la charge lésionnelle en T2

Charge lésionnelle T2 à T_0	:	Faible
Confirmation clinique d'un syndrome pseudo grippal	:	Positive

TAB. 7.25 – *Données du sujet VELMI.*



Évolution des mesures d'atrophie



Évolution de la charge lésionnelle en T2

Charge lésionnelle T2 à T_0	:	Moyenne
Confirmation clinique d'un syndrome pseudo grippal	:	Incertaine

autres méthodes, ne détecte pas une augmentation du volume cérébral (justifié cliniquement) pour DAZIN et HUBSO (par exemple). Cette observation semble induire que des erreurs de mesures peuvent aussi avoir lieu avec cette méthode.

Le problème est alors de définir les cas plus favorables à telle ou telle méthode en fonction de paramètres tels que la charge lésionnelle, l'EDSS, le type de sclérose en plaques, la localisation des lésions, des scores cognitifs, *etc.*

Dans notre étude, les sujets peuvent être triés en quatre groupes :

- syndrome pseudo grippal avéré et augmentation de la charge lésionnelle (DAZIN),
- syndrome pseudo grippal avéré et diminution de la charge lésionnelle (CALCE, ALLJE, LICPA),
- absence de syndrome pseudo grippal et augmentation de la charge lésionnelle (GUENA, VELMI, SCHIS, FRECO),
- absence de syndrome pseudo grippal et diminution de la charge lésionnelle (GIULA, HUBSO).

Les courbes d'atrophie des sujets de chaque groupe semblent se ressembler (mise à part quelques exceptions). Cependant, le faible nombre de données dont nous disposons nous empêche de réaliser de tels modèles.

7.1.4 Conclusion

L'analyse de la base a prouvé son homogénéité. Elle a aussi permis de chercher des corrélations entre les différentes mesures. Premièrement, les résultats de cette étude sont similaires à la littérature (*i.e.* l'EDSS ne se corrèle ni avec la charge lésionnelle ni avec l'atrophie). Deuxièmement, les résultats de la mesure automatique de charge lésionnelle se corrèlent avec la charge lésionnelle des mesures manuelles. Enfin, comme présenté dans la section 6.4.3, les résultats de LCLASS se corrèlent à ceux des mesures de largeurs de ventricules, contrairement à ceux de SIENA.

Toutefois, excepté sur certains sujets, les résultats de LCLASS et SIENA sont similaires. Il reste cependant très difficile de démontrer la supériorité d'une méthode par rapport à une autre. En effet, celles-ci peuvent être des mesures locales ou globales et les différences même de leur principe font qu'elles ne mesurent pas forcément les mêmes effets.

A cause du faible nombre de données de notre base, les résultats de cette analyse doivent être validés sur une base de plus grande importance. Cependant, par ces travaux, je propose un cadre complet pour l'analyse d'une telle base. Les méthodes proposées et utilisées sont entièrement automatiques et donc facilement appliquables sur un grand volume de données. Les résultats préliminaires sont encourageants. Cette étude sera aussi utilisée par Mikael Cohen pour sa thèse de médecine. Ce dernier la complétera en cherchant des corrélations entre les mesures obtenues et des données cliniques (telles que le *Multiple Sclerosis Functional Composite Score* qui se corrèle normalement mieux aux mesures d'atrophie [Cutter 1999, Fischer 1999, Rudick 2002]).

7.2 Intégration et mise à disposition des outils d'analyse

Cette section vise à présenter brièvement les outils logiciels développés pendant cette thèse. Les sections ne s'attarderont pas sur l'utilisation des divers outils mais décriront les possibilités de visualisation, de mesures et de calculs que ces derniers offrent. Le but étant de pouvoir les utiliser pour mener à bien d'autres travaux de recherche.

7.2.1 SepINRIA

Nos partenaires neurologues ont exprimé le besoin d'un outil logiciel permettant de visualiser, de segmenter et d'analyser les images IRM de leurs patients. De notre côté, un outil permettant d'obtenir des mesures de référence auxquelles les résultats de nos algorithmes pourront être confrontés était aussi recherché.

Dans ce domaine, peu de logiciels sont disponibles. On peut citer par exemple QBrain[1], Medimax[2] ou JIM[3] mais ces logiciels sont le plus souvent payants ou non diffusés en dehors de leur laboratoire d'origine. Le besoin de créer un nouveau logiciel dédié à l'analyse d'IRM de patients atteints de SEP s'est donc fait ressentir.

7.2.1.1 Introduction

Pour répondre à ces besoins, j'ai donc créé un logiciel gratuit dédié à l'analyse d'images IRM de patients atteints de SEP : SepINRIA. Son but est de fournir aux cliniciens des outils permettant d'analyser les images de manière plus rapide et quantitative.

Voici quelques points du cahier des charges le concernant :

- interface logicielle simple et conviviale,
- gestion des images par une base de données,
- visualisation d'images en 2D et en 3D,
- comparaison de diverses séquences IRM précédemment recalées entre elles,
- mesure quantitative de la charge lésionnelle (manuellement ou automatiquement),
- comparaison quantitative de deux segmentations des lésions,
- évaluation de l'atrophie (manuellement ou automatiquement).

Le site internet de SepINRIA (figure 7.8) se situe à l'adresse suivante :

http ://www-sop.inria.fr/asclepios/software/SepINRIA

SepINRIA est disponible sur les plates-formes suivantes : Windows XP/Vista, Linux Fedora Core et MacOSX.

[1] http ://www.medis.nl/Products/QBrain.htm
[2] http ://www-ulpmed.u-strasbg.fr/ipb/tim/doku.php ?contenu=medimax/index.html
[3] http ://www.xinapse.com/



FIG. 7.8 – *Site internet de SepINRIA.*

7.2.1.2 Les fonctions du logiciel

SepINRIA est divisé en diverses fonctionnalités (blocs logiciels). Le passage d'une fonctionnalité à une autre se fait depuis le menu "Functions". Lorsqu'une fonctionnalité est chargée, le bloc logiciel auquel elle est liée est montré dans la fenêtre principale de SepINRIA, remplaçant ainsi la fonctionnalité précédemment chargée.

Voici la liste des fonctionnalités de SepINRIA :

- **Lésion Segmentation Edition (LSE)**, segmentation manuelle ou semi-automatique des lésions (par exemple, segmentation de référence réalisée par un expert). Si plusieurs séquences sont disponibles, les images peuvent être recalées sur le T2 et visualisées simultanément pour aider la segmentation. Lors de la segmentation divers labels peuvent être associés aux lésions (par exemple en fonction de leur sous-type ou de leur emplacement). Les segmentations peuvent être sauvées et visualisées en 2D ou en 3D. Le nombre de lésions correspondant à chaque label ainsi que leur volume peuvent être calculés. Un rapport complet de ces différentes charges lésionnelles peut être imprimé. L'algorithme semi-automatique a été présenté dans la section 5.2. Ses paramètres ont été ici fixés empiriquement.
- **Automatic Lesion Segmentation (ALS)**, segmentation automatique des lésions depuis quatre séquences IRM : Dual Spin Echo T2-PD, T1, T2-FLAIR. La méthode implémentée dans SepINRIA est pour le moment la version PEMS-G de l'algorithme de segmentation (cf. chapitre 5). Il est envisagé de la remplacer dès que possible par la version PEMS-COMBI, donnant de meilleurs résultats et permettant une segmentation des lésions du cortex et du cervelet.
- **Image or Segmentation Comparison (ISC)**, recalage de deux images, calcul de l'image de différence ou fusion d'images (dans le but, par exemple, de visualiser l'évolution entre deux images de même modalité mais acquises à des instants différents). Cette fonction permet aussi de comparer des segmentations (segmentation(s) automatique(s) versus segmentation(s) de référence) entre elles, qualitativement (calcul de l'image de différence) ou quantitativement (calcul des valeurs des critères présentés dans la section 5.1.5).
- **Brain Atrophy Evaluation (BAE)**, évaluation manuelle ou automatique de l'atrophie. En mode manuel, l'utilisateur réalise des mesures linéaires

(largeurs du cerveau, des ventricules latéraux ou du troisième ventricule) à l'aide des outils présentés dans la section 6.2. En mode automatique, la fraction parenchymateuse cérébrale est calculée pour chacun des instants disponibles à l'aide d'une version simplifiée de LCLASS (cf. section 6.3). Cette simplification réside dans le fait que le masque du cerveau n'est obtenu qu'à partir de EBM et non de la méthode composée (cf. section 4.2).

Dans toutes ces fonctionnalités, les images (DICOM ou autres) peuvent être importées/exportées dans/de la base de données et visualisées en 2D et/ou en 3D.

Pour faciliter son utilisation, j'ai réalisé une documentation utilisateur de SepINRIA. Celle-ci est téléchargeable à partir du site internet de SepINRIA.

7.2.1.3 Aspects de programmation

SepINRIA est basé sur diverses librairies (figure 7.9). ITK[4] et Mips[5] sont utilisées pour le traitement des images. La première est disponible en ligne et est utilisée principalement pour les conversions de format d'image. La seconde est une librairie interne à Asclepios où j'ai rajouté les algorithmes que j'ai développés et qui sont présents dans les chapitres 3, 4, 5 et 6.

La visualisation est assurée par les librairies VTK[6] et vtkINRIA3D[7]. L'interface graphique est basée sur wxWidgets[8]. Enfin, le cadre global (passage d'une fonctionnalité à l'autre) est une réutilisation du cadre du logiciel MedINRIA[9].

7.2.1.4 Conclusion

SepINRIA répond au besoin d'un logiciel dédié à l'analyse d'IRM de patients atteints de SEP. Il permet d'obtenir des mesures de charge lésionnelle ou d'évaluer l'atrophie. Dans le cas des mesures manuelles effectuées par un expert, les résultats obtenus peuvent servir de référence pour valider des algorithmes automatiques. Les algorithmes automatiques présentés dans les chapitres précédents sont aussi présents dans SepINRIA. Le but de leur insertion est de permettre leur comparaison avec d'autres algorithmes (qui eux ne sont pas disponibles). Du reste, SepINRIA offre le cadre de cette comparaison en permettant un calcul automatique de critères (présentés dans la section 5.1.5.2) à partir des segmentations obtenues.

D'un point de vue pratique, SepINRIA est téléchargeable en ligne[10] et référencé sur plusieurs sites regroupant des logiciels gratuits (plus de 3200 téléchargements). Son code a été déposé à l'agence pour la protection des programmes[11]. Enfin, une

[4] http ://www.itk.org/
[5] http ://www-sop.inria.fr/asclepios/software.php
[6] http ://www.vtk.org/
[7] http ://www-sop.inria.fr/asclepios/software.php
[8] http ://www.wxwidgets.org/
[9] http ://www-sop.inria.fr/asclepios/software.php
[10] http ://www-sop.inria.fr/asclepios/software/SepINRIA
[11] Inter Deposit Digital Number : IDDN.FR.001.130005.000.S.P.2008.000.31230

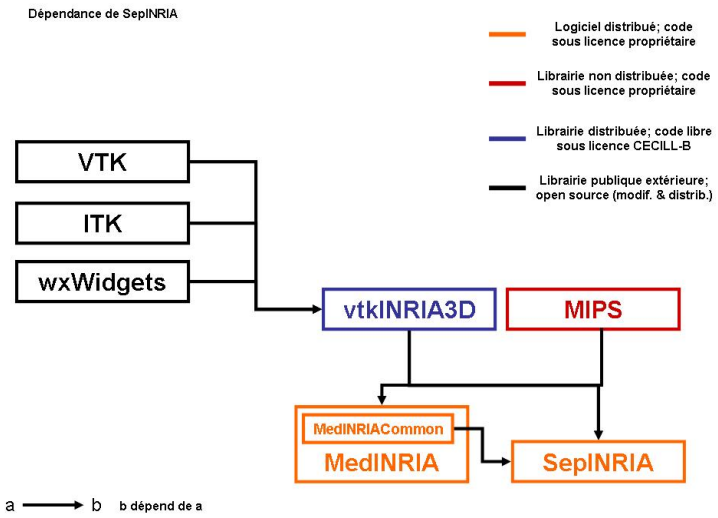


FIG. 7.9 – Diagramme des dépendances de SepINRIA

communauté d'utilisateurs commence à se former. Je citerai en premier lieu Mikael Cohen qui a pu effectuer toutes les mesures manuelles (qui nous ont servi de référence et dont il se servira pour sa thèse de médecine) grâce à SepINRIA. Des utilisateurs externes existent aussi. Concernant mes travaux sur SepINRIA, j'ai aussi reçu le prix Etienne Roullet du club francophone de la sclérose en plaques aux journées de neurologie de langue française. Ce club regroupe les professionnels de terrain (neurologues et autres) impliqués dans la prise en charge de la SEP.

SepINRIA est un premier pas vers la mise à disposition d'algorithmes de traitement d'IRM de patients atteints de SEP. En permettant de placer deux IRM dans le même repère, ce logiciel peut déjà rendre des services dans le contexte clinique pour comparer visuellement les images. Les calculs automatiques de la charge lésionnelle et de la fraction parenchymateuse cérébrale restent pour le moment du ressort de la recherche mais leur mise à disposition rend une évolution possible.

7.2.2 Web-services exécutés sur une grille de calcul

L'utilisation d'algorithmes de traitement d'images médicales nécessite de plus en plus de capacité de calcul. Tout d'abord, la complexité de ces algorithmes a augmenté. De même, la taille et la résolution des images font que celles-ci contiennent un nombre de voxels de plus en plus conséquent. Enfin, les bases de données d'images à traiter contiennent davantage de sujets et d'instants. Les utilisateurs et chercheurs se tournent donc de plus en plus vers l'exécution de calculs parallèles.

Cette section est un résumé du travail de développement réalisé en collaboration avec Erik Pernod, lors de son stage. Ce travail s'inscrit dans le cadre du projet NeuroLOG[12] qui vise à créer un intergiciel (logiciel intermédiaire entre diverses

[12] http ://neurolog.polytech.unice.fr/

applications reliées sur un réseau) dans le but de fédérer des données médicales, des données composées (médicales + autres), des algorithmes et de partager des ressources informatiques via une grille de calcul [Montagnat 2008].

Une présentation complète du projet NeuroLOG ainsi qu'une description des divers aspects techniques de ce travail sont données dans l'annexe D. Celle-ci présente aussi les sept web-services et la chaîne de traitement mise en place par notre travail.

Ce travail a plusieurs enjeux. Premièrement, il va permettre d'utiliser les différents parallélismes pour traiter les données : parallélisme de flux (deux services indépendants l'un de l'autre peuvent être exécutés en même temps), parallélisme de données (même instruction exécutée simultanément sur différentes données), parallélisme de services (dans le cas de deux services S1 et S2, séquenciellement liés, lorsque S1 a fini de traiter les données du patient A, S2 commence son traitement sur les données de A et S1 peut commencer à traiter les données du patient B). Ceci permettra une analyse plus rapide d'une importante base de données.

Deuxièmement, ce traitement plus rapide de la base de données permettra de tester différents jeux de paramètres pour la chaîne de traitement. On saura alors avec davantage de précision l'impact d'un paramètre sur le résultat final. Cette connaissance aidera grandement à automatiser totalement les algorithmes et à les rendre davantage robustes (puisqu'on pourra choisir le meilleur jeu de paramètres) dans le but de fournir au sein de SepINRIA (par exemple) les outils les plus fiables, robustes et précis.

Enfin, la mise à disposition de ces web-services et de la chaîne de traitement via la grille permettra à d'autres équipes de recherche de tester nos algorithmes. On pourra aussi remplacer chaque service par un service similaire développé par une autre équipe. Par exemple, on pourra comparer les résultats finaux de la chaîne de traitement présentée ici à ceux d'une chaîne de traitement similaire, excepté pour le service de segmentation du cerveau qui aurait été remplacé par un autre (développé par une autre équipe de recherche). Il découlera de cette analyse une multitude de nouvelles approches (méthodes) qu'il faudra tester et comparer. Cette analyse ne serait pas possible sans la mise en commun de ces ressources via la grille.

Chapitre 8

Conclusion et perspectives

Sommaire

8.1 Synthèse des contributions 160
8.1.1 La chaîne de traitement, EMA 160
8.1.2 Segmentation des lésions 161
8.1.3 Évaluation de l'atrophie 162
8.1.4 Cadre d'analyse d'IRM de patients atteints de SEP : application à la base QUALICORE 162
8.2 Perspectives 163
8.2.1 Segmentation des lésions du tronc et sous classification 163
8.2.2 Perfectionnement et évaluations locales d'atrophie 163
8.2.3 Analyses complémentaires de la base QUALICORE et d'autres bases de données 164
8.2.4 Intensifier l'intégration dans le domaine médical des outils développés 164
8.3 Publications de l'auteur 165
8.3.1 Articles de journaux 165
8.3.2 Articles de congrès (avec actes et comité de sélection) 165
8.3.3 Conférences sur résumé 166

Organisation du chapitre : Ce chapitre commence par une synthèse des contributions de cette thèse (cf. section 8.1). Puis, des perspectives découlant de ces travaux sont présentées dans la section 8.2. Pour finir, la liste de mes publications est donnée (cf. section 8.3).

8.1 Synthèse des contributions

Le sujet de cette thèse vise à produire un cadre permettant l'analyse d'images IRM de patients atteints de sclérose en plaques pour l'aide au diagnostic et à la prise de décisions thérapeutiques. Cependant, ce sujet est vaste et regroupe de nombreux domaines et notions. Ces travaux ont commencé par mettre en place un socle commun aux différentes méthodes : la chaîne de traitement EMA, visant à segmenter les différents tissus cérébraux. Puis, 2 axes ont été developpés : mesures de la charge lésionnelle et évaluations de l'atrophie. Dans chaque axe, l'analyse des mesures possibles a été menée par le raisonnement suivant :

- état de l'art des méthodes existantes et des critères de validation,
- amélioration ou développement de nouvelles méthodes semi-automatiques (permettant d'obtenir des mesures de référence par un expert),
- développement de nouvelles méthodes entièrement automatiques,
- comparaison des méthodes automatiques (développées ou de la littérature) entre elles et avec les mesures de reférence (analyse à la fois sur des données tests et sur les images de la base QUALICORE),
- choix d'une ou de plusieurs mesures pour l'analyse finale des sujets de la base QUALICORE.

Pour finir, ces travaux de thèse ont porté sur l'analyse de la base de données QUALICORE et sur l'intégration des algorithmes développés dans le logiciel SepINRIA. Des web-services et une chaîne de traitement ont aussi été mis en place sur une grille de calcul, ouvrant des perspectives de mise en commun d'algorithmes et de données ainsi qu'un traitement plus rapide de ces données.

8.1.1 La chaîne de traitement, EMA

Cette chaîne permet de segmenter les différents compartiments du cerveau. Elle reste proche de celle qui avait été mise en place par G. Dugas dans sa thèse [Dugas-Phocion 2006] (modèle de volume partiel, gestion des points aberrants) mais ses limites sont maintenant connues et maîtrisées. Par exemple, une solution a été proposée dans le cas de la non-inversibilité de la matrice de covariance. Les effets du paramètre fixant le pourcentage de points considérés pour calculer les paramètres des classes ont aussi été étudiés. EMA se base sur les informations apportées par trois séquences IRM (T1, T2, DP) et non seulement deux (T2, DP).

L'étape d'extraction du cerveau a été changée. Ceci s'est fait par une étude comparative de différentes méthodes disponibles sous forme de logiciels par l'algorithme STAPLE [Warfield 2004]. Ce dernier a aussi été utilisé pour mettre en place un cadre permettant de combiner des segmentations entre elles (une nouvelle segmentation

binaire est obtenue à partir de plusieurs autres segmentations). Parmi toutes les méthodes ou combinaisons de méthodes, une combinaison de trois méthodes a été retenue. Lors de l'ensemble des travaux présentés dans ce manuscrit, aucun échec de cette méthode combinée n'a été constaté.

Enfin, cette chaîne de traitement peut diviser le masque du cerveau en trois sous-régions d'intérêt (cervelet, cortex, tronc) et une version, EMAL, prenant en compte les informations longitudinales a aussi été développée (cf. section 8.1.3).

8.1.2 Segmentation des lésions

De très nombreuses méthodes de segmentation des lésions de SEP sont présentes dans la littérature. Toutefois, elles ne sont pas ou peu disponibles sous forme de logiciel et sont rarement comparées. Le choix de l'une de ces méthodes pour analyser une base de données était donc impossible. Les travaux de cette thèse se basent donc sur la méthode proposée par G. Dugas, appelée ici PEMS-G, mais en proposent deux nouvelles versions.

Premièrement, un algorithme semi-automatique a été développé pour permettre une réalisation rapide de segmentations de référence par un expert. Le caractère générique de cet algorithme a en outre été mis en avant par les travaux de stage de master recherche du radiologue S. Novellas. Lors de ces travaux, l'algorithme semi-automatique proposé a été utilisé pour segmenter les lymphomes médiastinaux (proliférations cellulaires malignes localisées dans la région de la cage thoracique située entre les deux poumons).

Deuxièmement, deux nouvelles versions de PEMS-G ont été proposées. Tout d'abord, celles-ci utilisent la nouvelle chaîne de traitement EMA. Ensuite, un seuil est calculé automatiquement à partir des segmentations obtenues. Dans le cas de la première méthode (segmentant les lésions du cortex) PEMS-COR, ce seuil est calculé à partir de la séquence T2-FLAIR avant d'y être appliqué. Dans le cas de la seconde la séquence DP est utilisée. Pour finir, de nouveaux post-traitements sont appliqués sur les segmentations des lésions obtenues. L'union des résultats de PEMS-COR et PEMS-CER donne la méthode PEMS-COMBI.

Troisièmement, ces méthodes ont été comparées. Tout d'abord, PEMS-COR a été comparée à PEMS-G sur les données test du challenge MICCAI de segmentation. Cette comparaison a montré la supériorité de PEMS-COR sur ces données. PEMS-COR a donc été utilisée pour participer au challenge. Sur l'ensemble de ce dernier PEMS-COR est arrivée première des neuf méthodes en course. Ensuite, PEMS-COR, PEMS-G et PEMS-COMBI ont été comparées sur les données de la base QUALICORE par rapport aux segmentations manuelles réalisées par un expert. Cette étude montre une supériorité des résultats de PEMS-COMBI par rapport aux autres méthodes. PEMS-COMBI a donc été choisie pour analyser la base de données.

8.1.3 Évaluation de l'atrophie

Comme dans le cas des méthodes segmentant les lésions, de nombreuses méthodes d'évaluation de l'atrophie existent. Dans ce cas, certaines sont aussi disponibles sous forme de logiciel. Cependant, leur validation et leur comparaison est un sujet délicat.

Le premier axe de recherche développé ici est de permettre une sélection automatique de la coupe sur laquelle effectuer des mesures manuelles linéaires. Cette automatisation vise à augmenter la reproductibilité de ces méthodes manuelles.

Ensuite, deux méthodes automatiques ont été développées. Celles-ci évaluent l'atrophie, en calculant la fraction parenchymateuse cérébrale (FPC) aux différents instants. La première méthode, CLASS, se sert des segmentations obtenues par la chaîne de traitement EMA pour calculer le volume des divers compartiments cérébraux. Cette méthode se différencie des méthodes pré-existantes par une classification utilisant un modèle de volumes partiels SG/LCR. Cette classification est aussi réalisée sur l'ensemble du cerveau, excepté en des points considérés aberrants. La seconde méthode, LCLASS, utilise les segmentations calculées par EMAL (incluant les informations longitudinales). Dans ce cas, l'algorithme de classification EM se sert des informations apportées par les voxels des images des différents instants pour calculer les paramètres des classes.

CLASS et LCLASS ont été comparées entre elles, avec les mesures manuelles, et avec d'autres méthodes (SIENA et SIENAX). Cette comparaison a été réalisée sur des données simulées et sur des données réelles. Sur ces deux types de données les résultats de CLASS ne sont pas réalistes. Par contre, sur les données simulées LCLASS se corrèle très fortement avec l'atrophie simulée ainsi que SIENA et la mesure de volume cérébral normalisé (VCN) donnée par SIENAX. Sur les données réelles les résultats de cette étude sont moins contrastés car la vérité terrain n'est pas connue. LCLASS se corrèle bien aux mesures manuelles contrairement à SIENA mais les résultats de SIENA sont plus réalistes. LCLASS et SIENA ont donc été choisies pour analyser la base de données.

8.1.4 Cadre d'analyse d'IRM de patients atteints de SEP : application à la base QUALICORE

Les algorithmes développés et les études comparatives menées ont permis de sélectionner des méthodes entièrement automatiques pour analyser une base d'IRM de patients atteints de SEP. Dans notre cas, ce cadre a été utilisé pour analyser la base de données QUALICORE. Cette analyse a prouvé l'homogénéité de la base. Des corrélations significatives entre les mesures automatiques et les mesures manuelles ont aussi été trouvées. Des justifications cliniques ont aussi été avancées pour expliquer les courbes d'évolution des différentes mesures.

Les méthodes qui ont permis cette analyse ont été intégrées au logiciel SepINRIA. Ce dernier se présente comme une plate-forme regroupant les différents algorithmes nécessaires à l'analyse d'images de patients atteints de SEP. Il reste pour le moment

du ressort de la recherche mais il peut déjà rendre des services dans le contexte clinique avec par exemple le recalage d'images. La création de web-services sur une grille de calcul ouvre aussi la voie à une mise en commun de ressources (données, algorithmes) entre plusieurs laboratoires, à la création de nouvelles chaînes de traitement basées sur les algorithmes provenant de laboratoires différents et surtout à une analyse rapide d'un important volume de données en utilisant la puissance de calcul de la grille.

8.2 Perspectives

Comme présenté dans la section précédente, ces travaux de thèse ont permis de faire le point sur les techniques actuelles de mesure de la charge lésionnelle et d'évaluation de l'atrophie. Je pense que ces travaux ont aussi ouvert différentes perspectives.

8.2.1 Segmentation des lésions du tronc et sous classification

La méthode employée pour segmenter les lésions du cervelet semble pouvoir s'appliquer au tronc. Toutefois, notre base de données contenait moins d'images avec de telles lésions. Une étude plus poussée sur l'extraction des lésions de cette région est donc nécessaire.

Une fois un masque complet des lésions obtenu, il serait intéressant de pouvoir classer les lésions dans les différents sous-types présentés dans le chapitre 5. Ceci semble pouvoir se faire à l'aide de la table 5.1.

8.2.2 Perfectionnement et évaluations locales d'atrophie

Les méthodes d'évaluation de l'atrophie, CLASS et LCLASS, proposées dans ce document ne prennent pas en compte certains points qu'elles considèrent comme aberrants. Notre hypothèse est que ces points représentent majoritairement des lésions. Cependant, selon les cliniciens ces lésions doivent être incluses dans le volume cérébral lors du calcul de la fraction parenchymateuse (FCP). L'utilisation de PEMS-COMBI (méthode développée dans ces travaux et union de PEMS-COR et PEMS-CER) permettrait de générer ce masque des lésions. Il resterait alors à effectuer une nouvelle classification des tissus cérébraux dans le masque du cerveau privé des voxels correspondants aux lésions. En effet, il semble plausible que la classification fournira de meilleurs résultats si les voxels correspondants aux lésions ne sont pas utilisés. Ces hypothèses restent bien sûr à vérifier.

Le chapitre 4 a présenté une méthode de segmentation du cerveau en trois régions d'intérêt. L'application de ces masques sur les segmentations des compartiments cérébraux (obtenues par la méthode de classification) peut permettre de connaître le volume des différents compartiments dans ces régions d'intérêt. Le calcul d'une FPC locale est donc possible. Cependant, ces mesures semblent encore plus difficiles à valider que la mesure globale actuelle.

Un autre point concernant ces mesures est de savoir s'il vaut mieux :

- réaliser cette classification en premier (de manière globale sur tout le cerveau) puis appliquer les masques des régions d'intérêt aux résultats de cette classification,
- ou s'il vaut mieux, appliquer les masques des régions d'intérêt en premier et réaliser trois classifications, une dans chaque région.

Une étude préliminaire que j'ai effectuée sur les images de Brainweb favorise la première solution. Il semble que le fait de réaliser la classification sur plus de points, permet un meilleur calcul des paramètres des classes. Toutefois, ces résultats doivent être confirmés par des études plus poussées. Enfin, il serait intéressant d'étudier l'impact de l'utilisation de champs de Markov sur les résultats de la chaîne de traitement EMA.

8.2.3 Analyses complémentaires de la base QUALICORE et d'autres bases de données

L'une des perspectives les plus encourageantes pourrait sortir de la thèse de médecine de Mikael Cohen. Celui-ci va comparer les différentes mesures à des données cliniques. Il sera donc intéressant de voir si des corrélations existent entre les mesures ou si des explications plus précises peuvent être apportées aux évolutions des paramètres IRM mesurés.

Dans le même ordre d'idée, l'utilisation du cadre d'analyse développé sur une plus grande base de données permettra aussi de valider nos résultats et potentiellement de trouver des modèles d'évolution de l'atrophie en fonction de la présence d'un syndrome pseudo-grippal et des variations de charge lésionnelle (cf. section 7.1.3.5).

Les différentes méthodes présentées dans ces travaux ont été développées pour segmenter les plaques ou évaluer l'atrophie cérébrale sur des images IRM de sujets atteints de sclérose en plaques. Cependant, d'autres maladies neuro-dégénératives (ainsi que tout simplement le vieillissement) induisent des phénomènes similaires (apparition de plaques et atrophie). Il serait donc intéressant d'étudier la portabilité des méthodes développées à ces cas.

8.2.4 Intensifier l'intégration dans le domaine médical des outils développés

De nombreux outils de segmentation et d'évaluation de l'atrophie sont présents dans la littérature. Cependant, peu sont disponibles gratuitement (au moins académiquement) sous forme de logiciels ou sous forme de web-services d'une grille de calculs. Quitte à sembler utopiste, ceci est, à mon avis, l'un des freins à leurs usages.

Premièrement, le fait de rendre ces algorithmes disponibles permet de les comparer à d'autres. Ceci est à mon sens un des enjeux prioritaires notamment concernant la segmentation des lésions, domaine regorgeant de méthodes de segmentation difficilement comparables. De plus, chacune de ces méthodes peut dépendre de la

résolution des images. Les tester sur un nombre maximum de cas semble donc une priorité pour mieux définir les limites de chacune d'elles.

Deuxièmement, il est particulièrement intéressant de tester ces logiciels dans le domaine clinique. Ces tests permettent de mieux définir les attentes de nos partenaires cliniques (pouvant fixer nos prochains axes de recherche). Ces logiciels peuvent également se révéler utiles dans ce cadre clinique. Par exemple, le recalage d'images permet de suivre plus facilement les évolutions pouvant se produire entre deux séquences IRM espacées dans le temps.

8.3 Publications de l'auteur

8.3.1 Articles de journaux

1. **J.C. Souplet**, C. Lebrun, S. Chanalet, N. Ayache, G. Malandain. Revue des approches de segmentation des lésions de sclérose en plaques dans les séquences conventionnelles IRM. Revue Neurologique (2008), Masson, Paris, 165(1) 7-14, 2009.
2. **J.C. Souplet**, C. Lebrun, N. Ayache, G. Malandain. Multiple Sclerosis Lesion Segmentation Methods : Review, Classification and Validation. Neuroimage, *Soumis*.

8.3.2 Articles de congrès (avec actes et comité de sélection)

1. **J.C. Souplet**, C. Lebrun, N. Ayache, and G. Malandain. An Automatic Segmentation of T2-FLAIR Multiple Sclerosis Lesions. In Proceedings of Multiple Sclerosis Lesion Segmentation Challenge Workshop (workshop associé à MICCAI), New York, NY, USA, septembre 2008.
2. E. Pernod, **J.C. Souplet**, J. Rojas Balderrama, D. Lingrand, and X. Pennec. Multiple Sclerosis Brain MRI Segmentation Workflow deployment on the EGEE Grid. In Proceedings of MICCAI-Grid Workshop (workshop associé à MICCAI), New York, NY, USA, septembre 2008.
3. **J.C. Souplet**, Christine Lebrun, Nicholas Ayache, and Grégoire Malandain. 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, mai 2008.
4. N. Toussaint, **J.C. Souplet** and P.Fillard. MedINRIA : Medical Image Navigation and Resarch Tool by INRIA. In Proceedings of Worskop on Interaction in medical image analysis and visualization (workshop associé à MICCAI), Brisbane, Australia, 2007.
5. **J.C. Souplet**, C. Leger, V. Eder. Representation of the LV 3D phase dispersion using gated blood pool tomography. In Proceedings of Computers in Cardiology, Pages : 419 - 422, Lyon, France, 25-28 septembre 2005, ISBN : 0-7803-9337-6.

8.3.3 Conférences sur résumé

1. J. Rojas Balderrama, D. Lingrand, E. Pernod, **J.C. Souplet**, X. Pennec, and J. Montagnat. NeuroLOG : neuroscience application workflows execution on the EGEE grid. In Proceedings of Enabling Grid for E-sciencE (EGEE) conference, Istanbul, Turkey, septembre 2008.
2. S. Novellas, G. Malandain, J. Costa, **J.C. Souplet**, N. Ayache, and P. Chevallier. Apport d'un algorithme de segmentation semi-automatique pour analyse des adénopathies médiastinales. In Proceedings of Journées Françaises de Radiologie, Paris, France, octobre 2008.
3. E. Pernod, **J.C. Souplet**, M. Cohen, N. Toussaint, C. Lebrun, and G. Malandain. SepINRIA v1.7.2 : Multiple Sclerosis Brain MRI visualisation, comparison and analysis Software. In Proceedings of World Congress for Treatment and Research in Multiple Sclerosis (WCTRIMS), Montreals, Canada, septembre 2008.
4. **J.C. Souplet**, C. Lebrun, and G. Malandain. SepINRIA v1.7.1 : A Free Software for Lesion Segmentation and Atrophy Evaluation dedicated to Multiple Sclerosis Brain MRI. In Journal of Neurology - 18th European Neurological Society Meeting, pages 208, Springer, Nice, France, juin 2008.
5. **J.C. Souplet**, C. Lebrun, and G. Malandain. SepINRIA : Un logiciel gratuit dédié à l'analyse d'Images de Résonance Magnétique cérébrales de patients atteints de sclérose en Plaques. In Proceedings of Journées de Neurologie de Langue Française, Nice, France, avril 2008.
6. **J.C. Souplet**, P.Fillard, N. Toussaint, C. Lebrun, G. Malandain. SepINRIA : A software to analyse Multiple Sclerosis Brain MRI. In Proceedings of 60th American Academy of Neurology Annual Meeting, Chicago, USA, Avril 2008.
7. **J.C. Souplet**, C. Lebrun, P. Clavelou, W. Camu, S. Chanalet, N. Ayache and G. Malandain. A Comparative Study of Atrophy Measurements in Multiple Sclerosis. In Proceedings of 17th Franco-UK scientific meeting of ARSEP (François Lhermitte Conferences), Paris, France, mai 2008.
8. **J.C. Souplet**, C. Lebrun, P. Clavelou, W. Camu, S. Chanalet, N. Ayache and G. Malandain. A New Atrophy Measurement Method dedicated to Multiple Sclerosis. In Proceedings of 60th American Academy of Neurology Annual Meeting, Chicago, USA, avril 2008.
9. **J.C. Souplet**, C. Lebrun, P. Clavelou, W. Camu, S. Chanalet, N. Ayache and G. Malandain. A comparative study of skull stripping methods in relapsing-remitting multiple sclerosis : Emergence of a new automatic segmentation algorithm. In Proceedings of Congress of the European Committee for Treatment and Research in Multiple Sclerosis (ECTRIMS), Prague, Czech Republic, octobre 2007.
10. **J.C. Souplet**, C. Lebrun, P. Fillard, N. Ayache, P. Clavelou, G. Malandain. SepINRIA : Un outil d'évaluation de la charge lésionnelle chez des patients

atteints de sclérose en plaques. In Revue Neurologique - Journées Internationales de la Société Française de Neurologie, Vol. 6-7, n° 163, page 725, juin 2007, Paris, France.

11. C. Lebrun, S. Chanalet, **J.C. Souplet**, N. Ayache, G. Malandain. Outils de pré-traitement de l'image et segmentation automatique de la charge lésionnelle. In Proceedings of Journées Internationales de la Société Française de Neurologie, Paris, France, juin 2007.

12. V. Eder, **J.C. Souplet**, C. Leger. ACS, Un nouvel algorithme de reconstruction des volumes ventriculaires et d'analyse 3D de la cinétique segmentaire et de l'asynchronisme en tomographie cavitaire. In Revue Médecine Nucléaire - 43ème Colloque de Médecine Nucléaire de Langue Française, Vol. 29, Nn° 11, page 695, Marseille, novembre 2005, ISSN : 0928-1258.

Annexe A

Tableaux de suivi de la base QUALICORE

Les tables A.1, A.2 et A.3 correspondent respectivement aux tables de suivi de la base de données pour le centre de Clermont-Ferrand, de Montpellier et de Nice.

Tab. A.1 – *Table de suivi de la base de données pour le CHU de Clermont-Ferrand.*

Clermont-Ferrand

Mise a jour :

08/08/2008

Pour chaque patient un examen est reçu et est complet si une date figure dans la case correspondante. Dans le cas où l'examen n'est pas complet la case est vide et rouge ; de plus il y a un commentaire dans la dernière case de la ligne. Lorsque un examen n'est pas reçu la case indique le nombre de jours restants avant que l'examen ne soit réalisé selon le protocole. Si ce nombre est entre MAX et MIN la case est orange. Si il est inférieur à MIN la case est rouge. NbJN est le nombre de jours recommandés par le protocole entre deux examens.

NbJN	MIN	MAX
365	-10	30

Nb : NbJN, MIN, MAX peuvent être changés.
Le tableau se mettra à jour.

DUMMY-RUN : OK

Centre	NUM	Id	T0	T1	T2	T3	T4	Commentaires
CLE	01							
CLE	02							
CLE	03							
CLE	04	CELFR	28/10/2005	13/10/2006	12/10/2007	12,00		T0 T1 TSE ???=??? T1GD, T1 Ok
CLE	05	DELEM	28/10/2005	27/10/2006	-338,00			T1 TSE ???=??? T1GD, T1 2 T1GD
CLE	06	LAMAN	25/11/2005	24/11/2006	-310,00			T1 2 T1GD
CLE	07	ROCCA	02/12/2005	01/12/2006	-303,00			T1 TSE ???=??? T1GD
CLE	08	PAUCH	30/12/2005	17/01/2007	-256,00			T1 TSE ???=??? T1GD + 2 T1GD
CLE	09	LENFL	27/01/2006					Sorti de l'étude
CLE	10	FAUTH	31/03/2006					Sorti de l'étude
CLE	11	VIGFR	30/03/2007	-184,00				
CLE	12							
CLE	13							
CLE	14							
CLE	15							

TAB. A.2 – *Table de suivi de la base de données pour le CHU de Montpellier.*

Montpellier

Mise a jour :

08/08/2008

Pour chaque patient un examen est reçu et est complet si une date figure dans la case correspondante. Dans le cas où l'examen n'est pas complet la case est vide et rouge ; de plus il y a un commentaire dans la dernière case de la ligne. Lorsque un examen n'est pas reçu la case indique le nombre de jours restants avant que l'examen ne soit réalisé selon le protocole. Si ce nombre est entre MAX et MIN la case est orange. Si il est inférieur à MIN la case est rouge. NbJN est le nombre de jours recommandés par le protocole entre deux examens.

NbJN	MIN	MAX
365	-10	30

Nb : NbJN, MIN, MAX peuvent être changés.
Le tableau se mettra à jour.

DUMMY-RUN : OK (=> RUARE à confirmer comme non sujet de l'étude)

Centre	NUM	Id	T0	T1	T2	T3	T4	Commentaires
MON	01	SAIVE	31/08/2006	18/06/2007	13/06/2008	257,00		T0: Ok T1:Ok
MON	02	ROQLA	25/06/2007	**-97,00**				T0: Ok
MON	03	CAKMA	22/10/2007	**22,00**				T0: Ok
MON	04							
MON	05							
MON	06							
MON	07							
MON	08							
MON	09							
MON	10							
MON	11							
MON	12							
MON	13							
MON	14							
MON	15							

TAB. A.3 – *Table de suivi de la base de données pour le CHU de Nice.*

Nice

Mise a jour :

08/08/2008

Pour chaque patient un examen est reçu et est complet si une date figure dans la case correspondante. Dans le cas où l'examen n'est pas complet la case est vide et rouge ; de plus il y a un commentaire dans la dernière case de la ligne. Lorsque un examen n'est pas reçu la case indique le nombre de jours restants avant que l'examen ne soit réalisé selon le protocole. Si ce nombre est entre MAX et MIN la case est orange. Si il est inférieur à MIN la case est rouge. NbJN est le nombre de jours recommandés par le protocole entre deux examens.

NbJN	MIN	MAX
365	-10	45

Nb : NbJN, MIN, MAX peuvent être changés.
Le tableau se mettra à jour.
T13G : Présence d'un T1 3D Gado
T1GD : Absence d'un T1 Gado

DUMMY-RUN : OK

Centre	NUM	Id	T0	T1	T2	T3	T4	Commentaires
NIC	01	DAZIN	10/09/2004	05/09/2005				Sorti de l'étude
NIC	02	LICPA	10/09/2004	16/09/2005	15/09/2006	14/09/2007	**-16,00**	
NIC	03	ZIAKA	13/09/2004					T0 : T13G + Sorti de l'étude
NIC	04	FRECO	01/10/2004	21/10/2005	13/10/2006	07/09/2007	**-23,00**	T0 : T13G
NIC	05	MARTH	21/01/2005					Sorti de l'étude
NIC	06	SCHIS	18/03/2005	24/02/2006	20/04/2007	**-163,00**		T0 : T1GD,T13G --- Ok
NIC	07	VELMI	18/02/2005	20/02/2006	09/03/2007	07/03/2008	159,00	T0 : T1GD,T13G --- Ok
NIC	08	GUENA	17/06/2005	23/06/2006	**-464,00**			
NIC	09	GIULA	07/10/2005	16/06/2006	**-471,00**			**t2 a 8 mois**
NIC	10	EDDTA	18/11/2005					T0 : T13G **T1 et T1bis: Inutilisable (Erreur de protocole)**
NIC	11	ALLJE	09/12/2005	06/10/2006	**-359,00**			T0 : T1GD,T13G; t0-->**t1=10 mois**
NIC	12	DRAJA	20/01/2006					T0 : T13G
NIC	13	CHADE	05/05/2006					T0 : T2FL a 36c, **T1 : Inutilisable (Erreur de protocole)**
NIC	14	CALCE	12/05/2006	25/05/2007	**-128,00**			T1: OK
NIC	15	DITMA						**T0 et T1 : Inutilisable (Erreur de protocole)**
NIC	16	HUBSO	23/06/2006	08/06/2007	30/05/2008	243,00		

Annexe B

Exemple de segmentations des lésions

B.1 Segmentation des lésions du cortex

Les figures suivantes illustrent la segmentation des lésions dans le cortex. Ceci est réalisé sur le sujet DELEM présentant de nombreuses lésions. La segmentation obtenue regroupe les différents types d'erreurs qui ont pu être constatés.

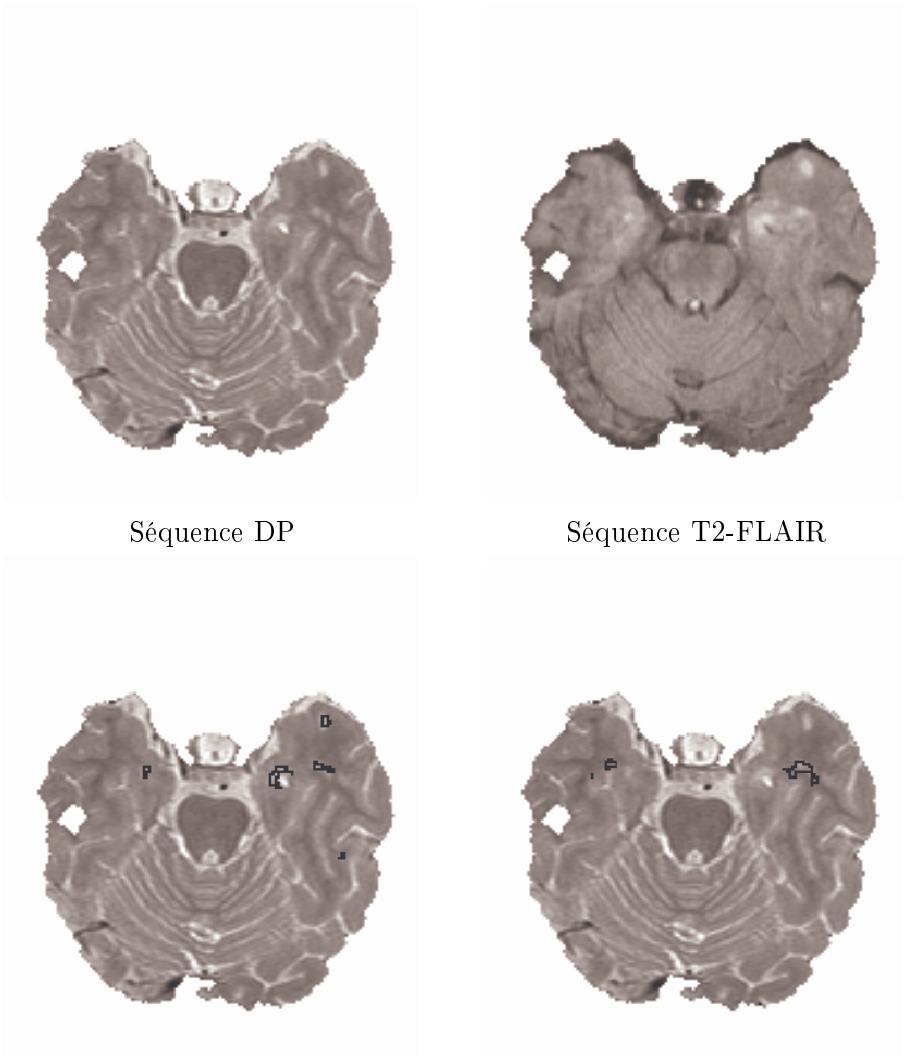


FIG. B.1 – *Présence de deux faux positifs et de deux faux négatifs.*

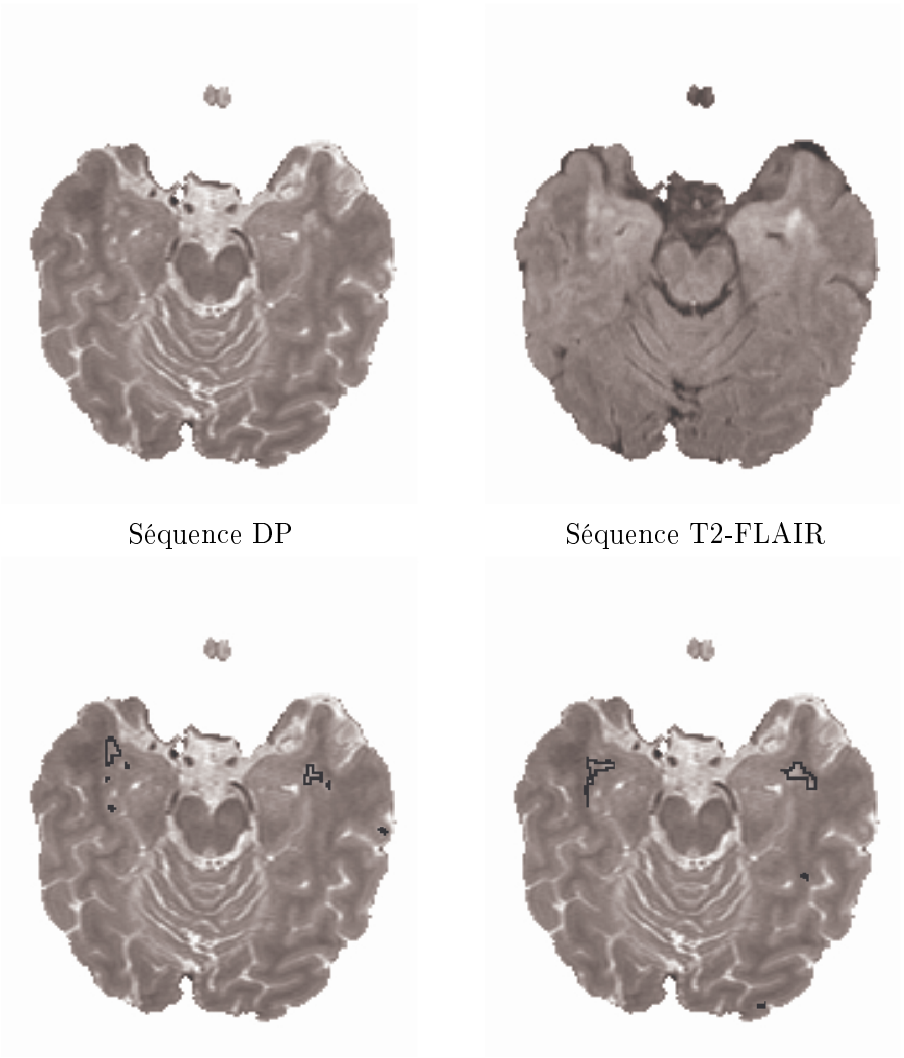


FIG. B.2 – *Non segmentation des petites lésions juxta-corticales.*

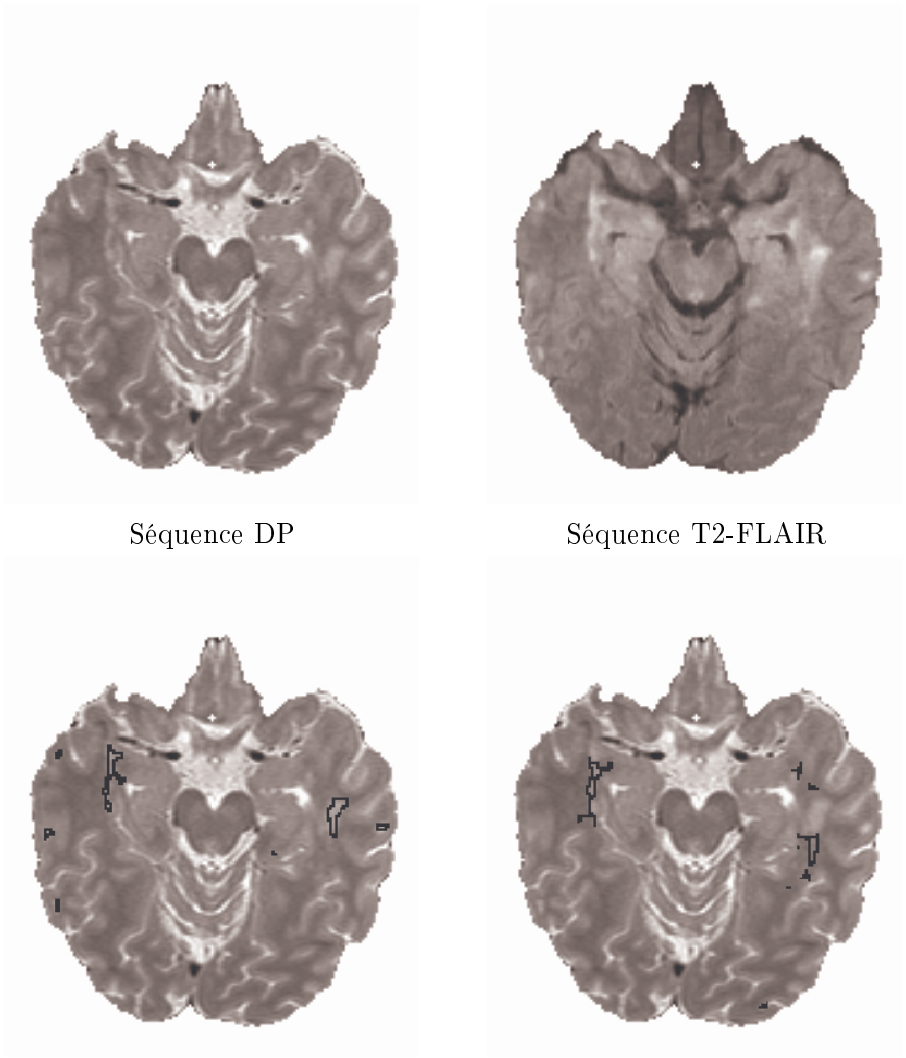


FIG. B.3 – *Non segmentation des petites lésions juxta-corticales.*

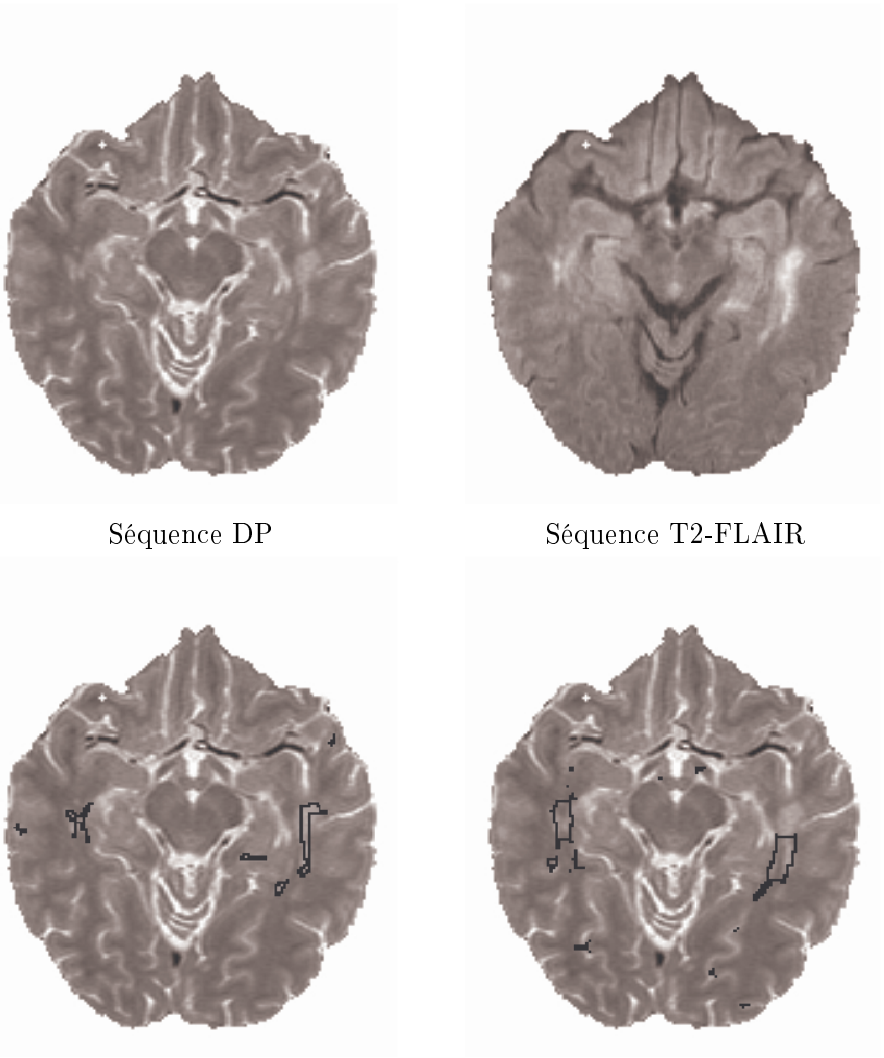


FIG. B.4 – *La sur-segmentation des lésions peut s'expliquer par la prise en compte de la SB "sale". Présence de quelques faux positifs. Le haut de la plus grosse lésion de droite n'est pas segmenté à cause de sa proximité avec la SG du cortex.*

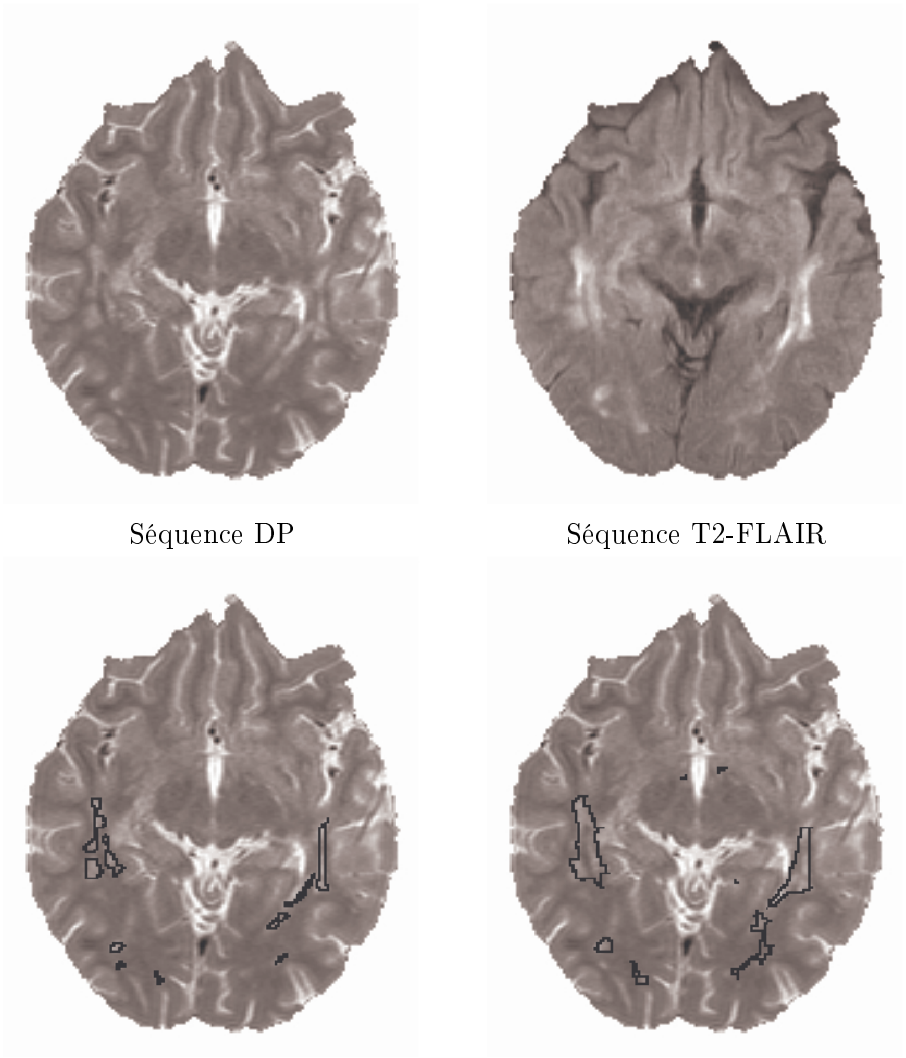


FIG. B.5 – *La sur-segmentation des lésions peut s'expliquer par la prise en compte de la SB "sale". Non détection d'une petite lésion juxta-corticale.*

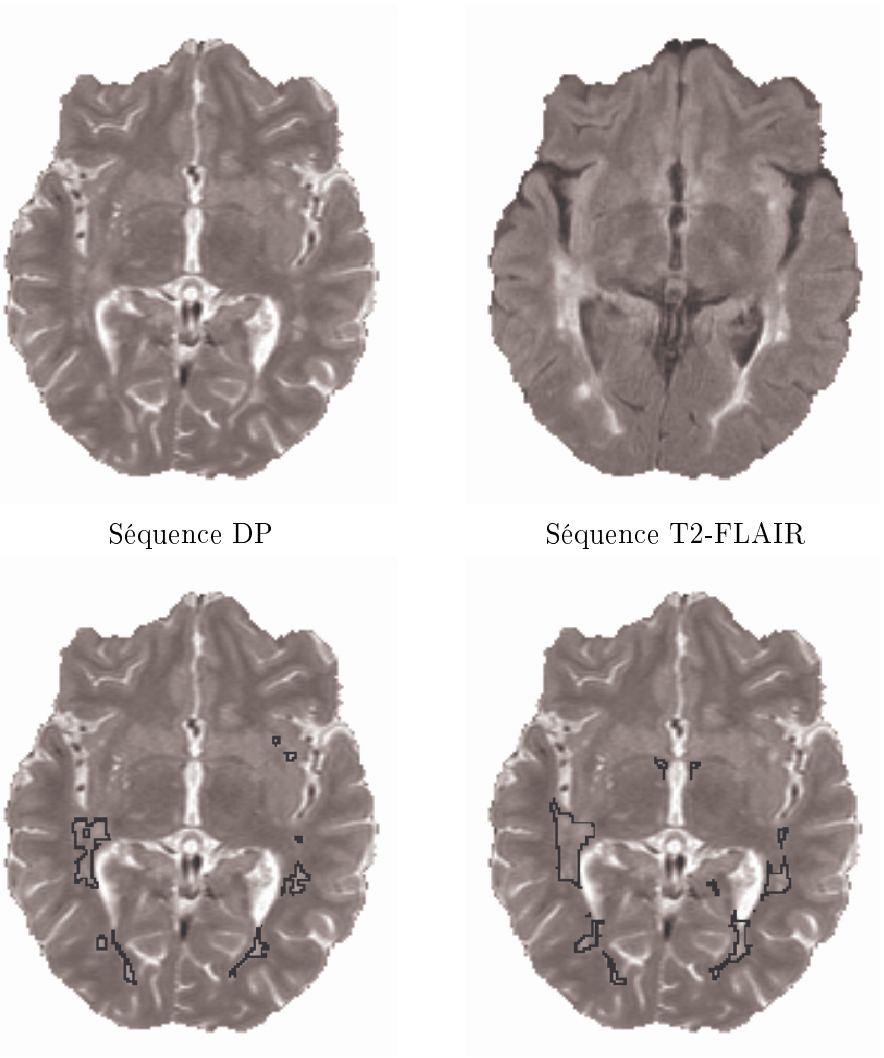


FIG. B.6 – *La sur-segmentation des lésions peut s'expliquer par la prise en compte de la SB "sale".*

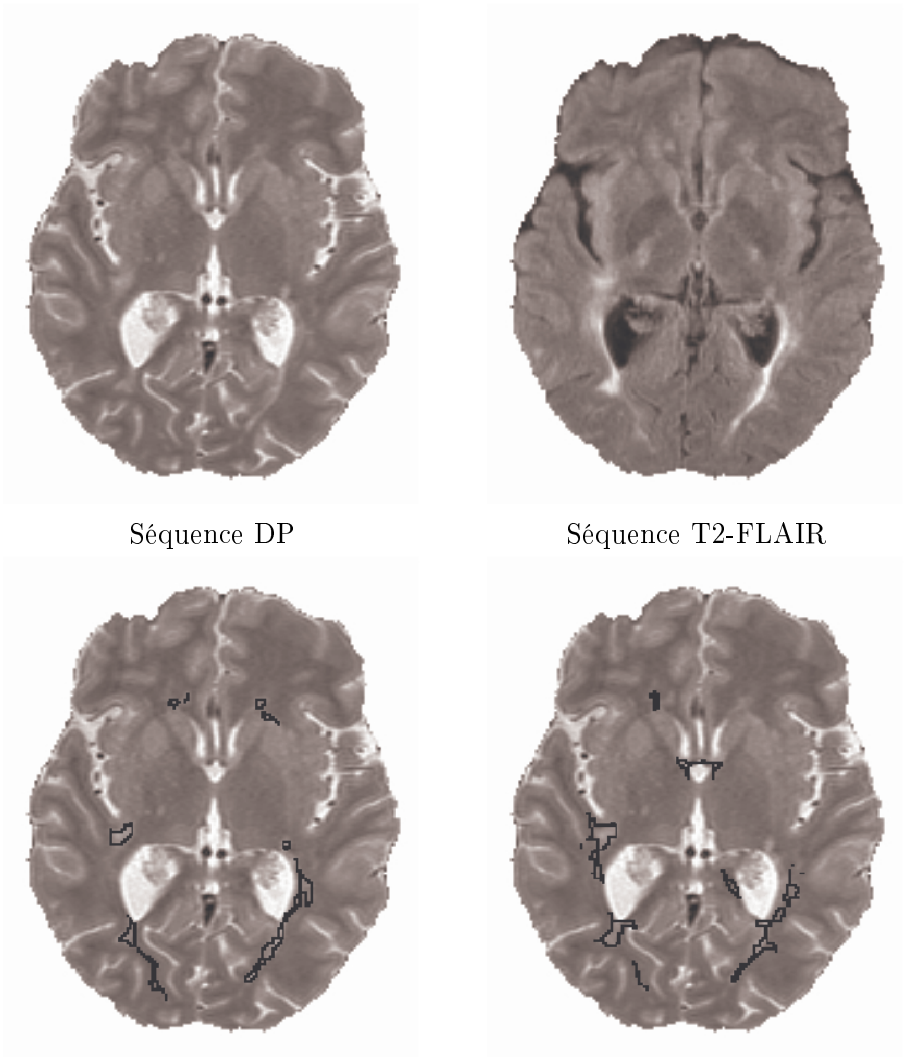


FIG. B.7 – *La sur-segmentation des lésions peut s'expliquer par la prise en compte de la SB "sale".*

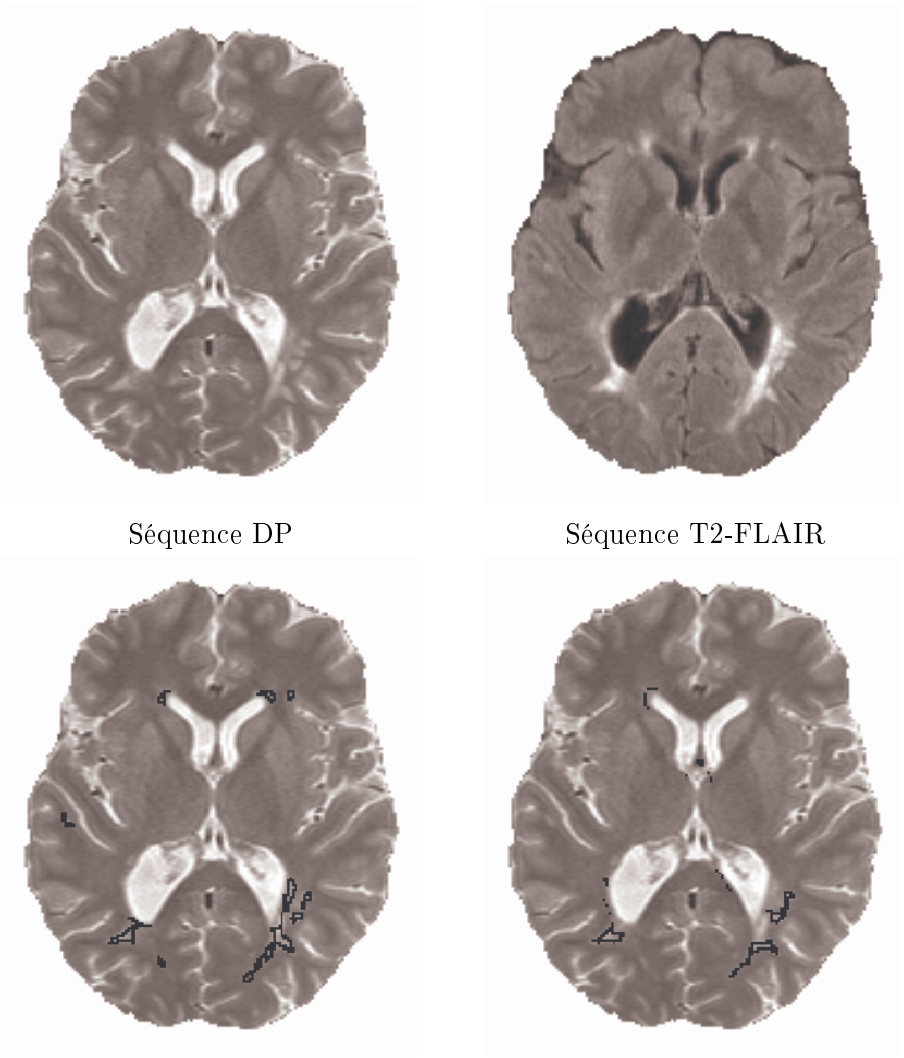


FIG. B.8 – *Non détection de lésions proches des ventricules. Dans ce cas, la proximité de ces lésions avec les ventricules les a exclues du masque de la SB supposée.*

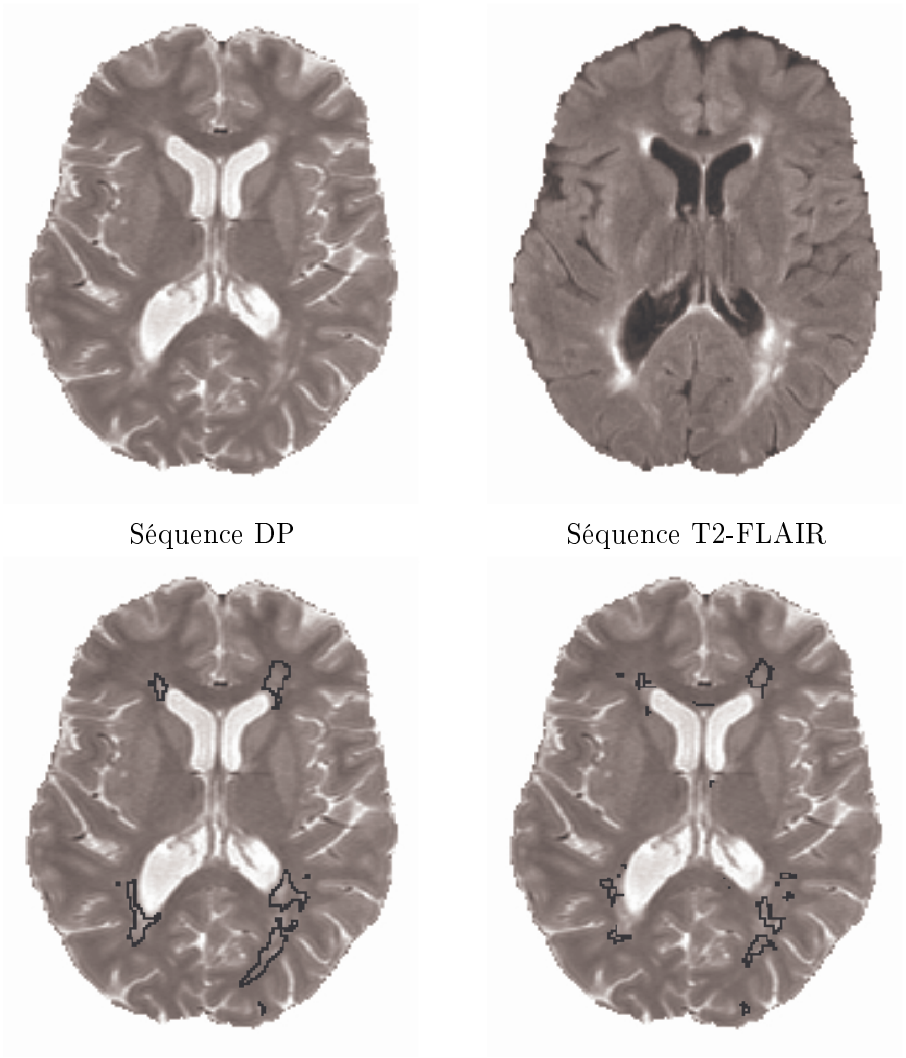


FIG. B.9 – *Non détection de lésions proches des ventricules. Dans ce cas, la proximité de ces lésions avec les ventricules les a exclues du masque de la SB supposée.*

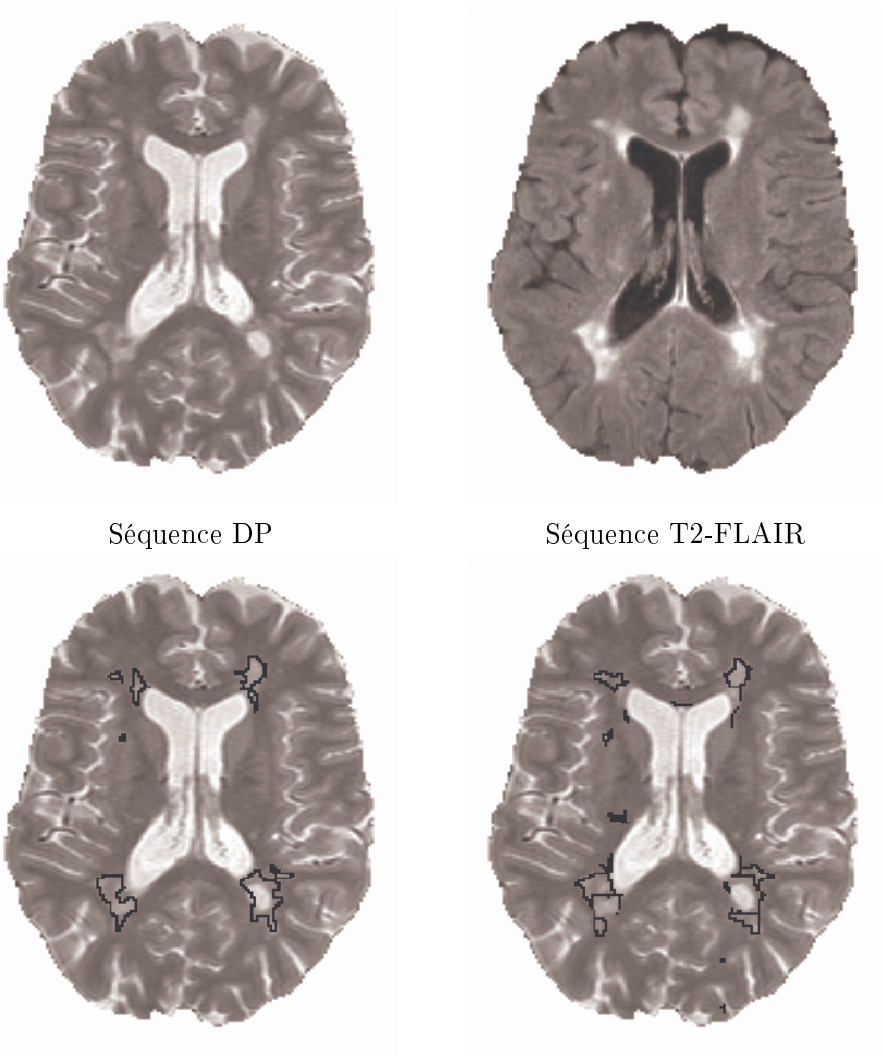


FIG. B.10 – *Non détection de lésions proches des ventricules. Dans ce cas, la proximité de ces lésions avec les ventricules les a exclues du masque de la SB supposée.*

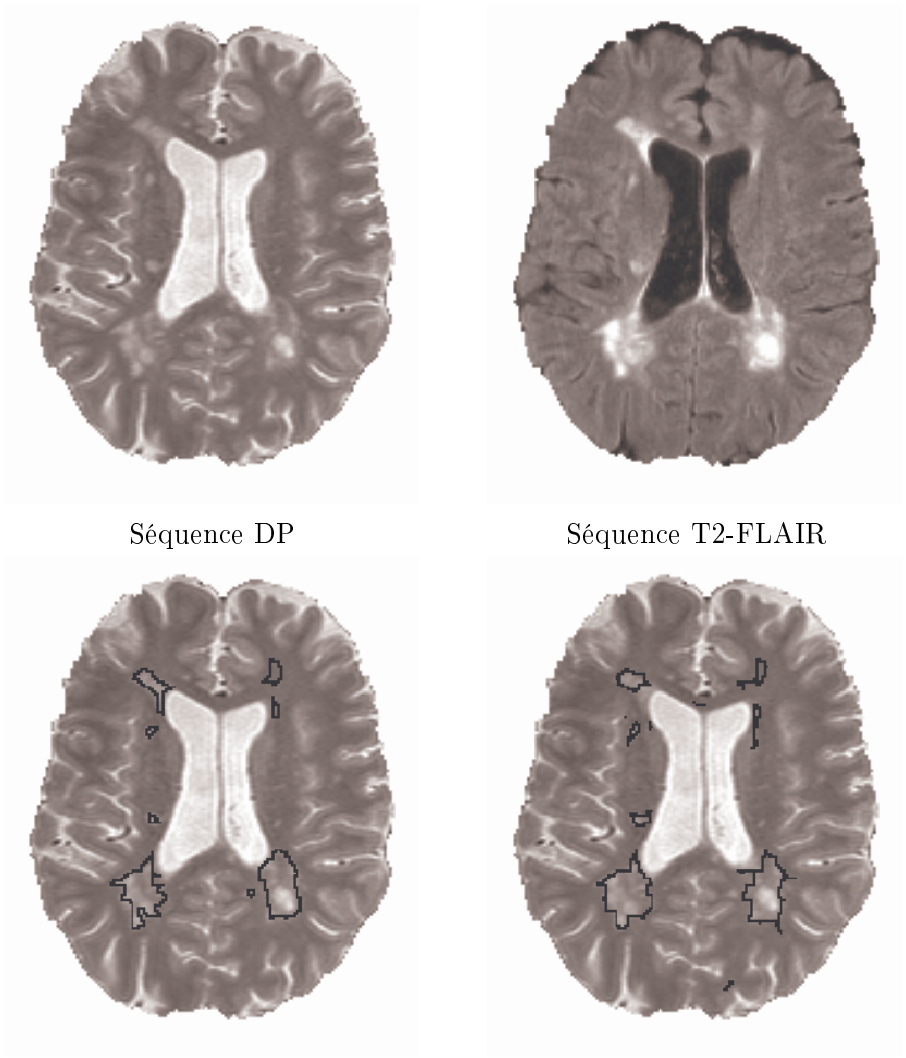


FIG. B.11 – *Légère sur-segmentation des lésions.*

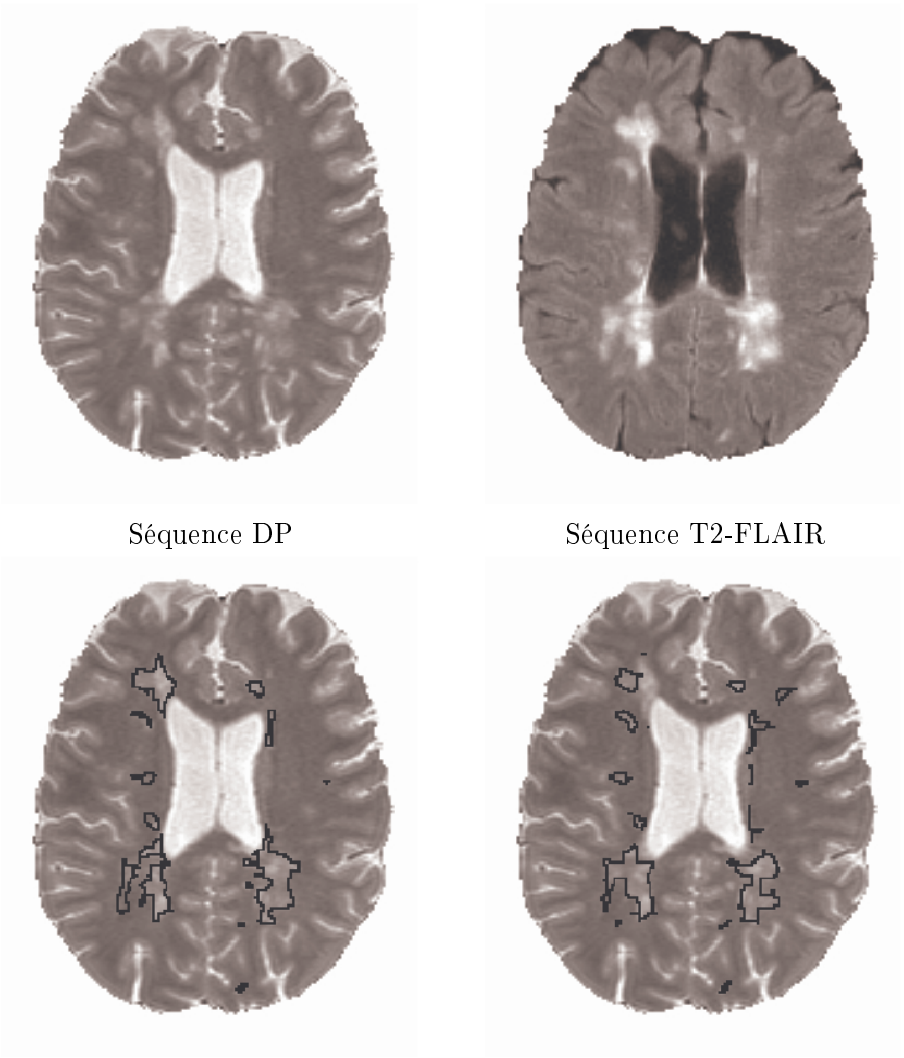


FIG. B.12 – *Légère sur-segmentation des lésions pouvant s'expliquer par la prise en compte de la SB "sale". Une lésion proche des ventricules n'est pas détectée car elle a été exclue du masque de la SB supposée.*

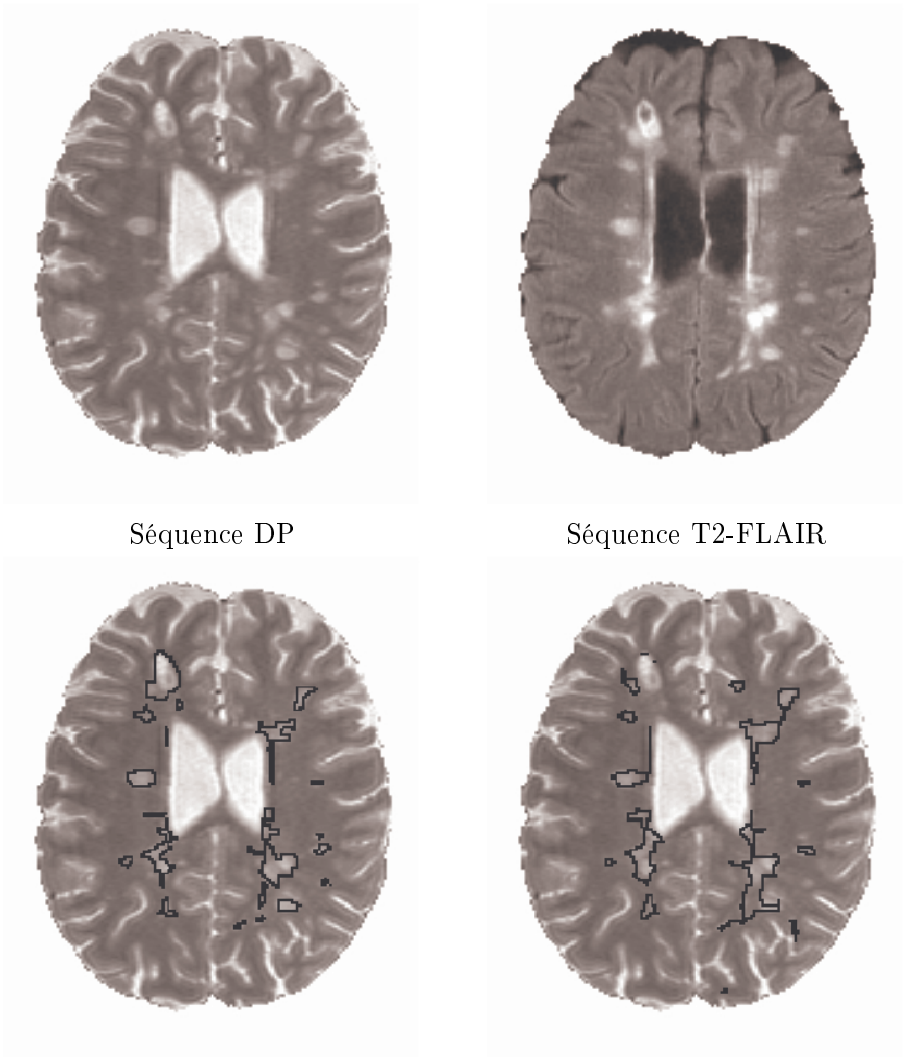


FIG. B.13 – *Légère sur-segmentation des lésions pouvant s'expliquer par la prise en compte de la SB "sale". Une lésion proche des ventricules n'est pas détectée car elle a été exclue du masque de la SB supposée.*

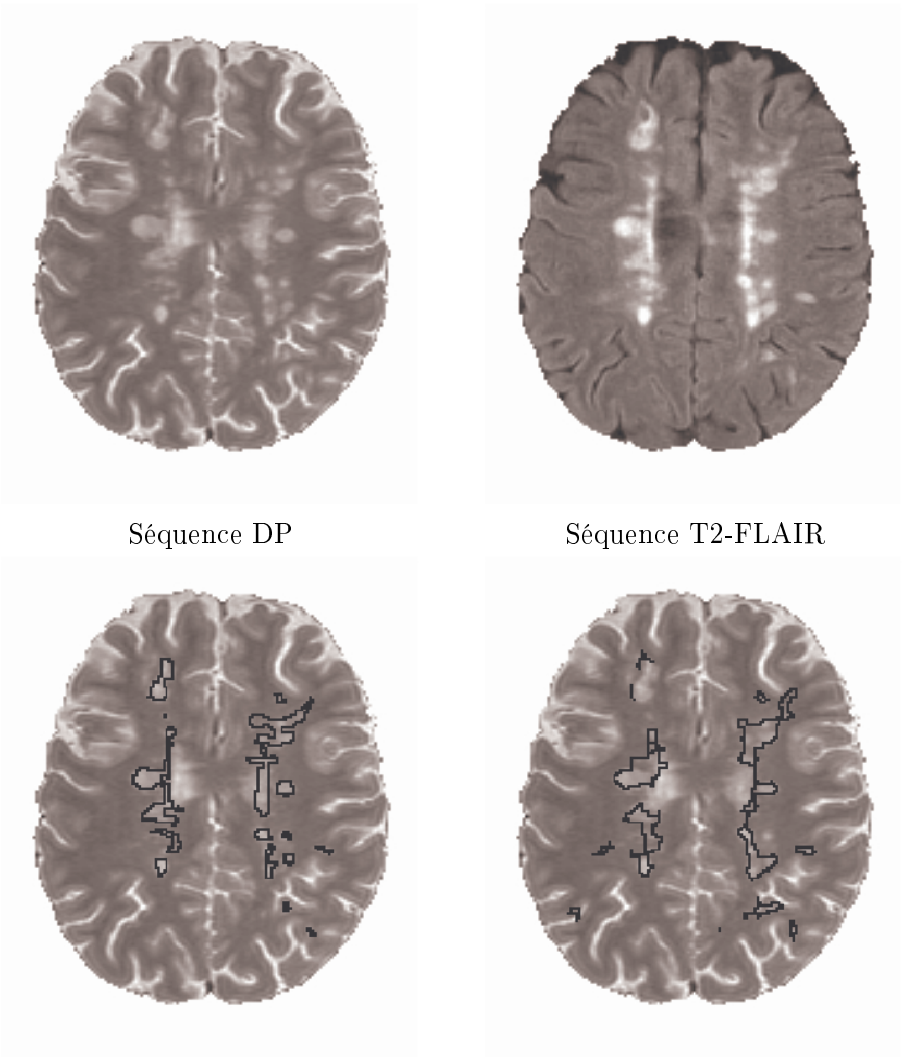


FIG. B.14 – *Légère sur-segmentation des lésions pouvant s'expliquer par la prise en compte de la SB "sale". Une lésion proche des ventricules n'est pas détectée car elle a été exclue du masque de la SB supposée.*

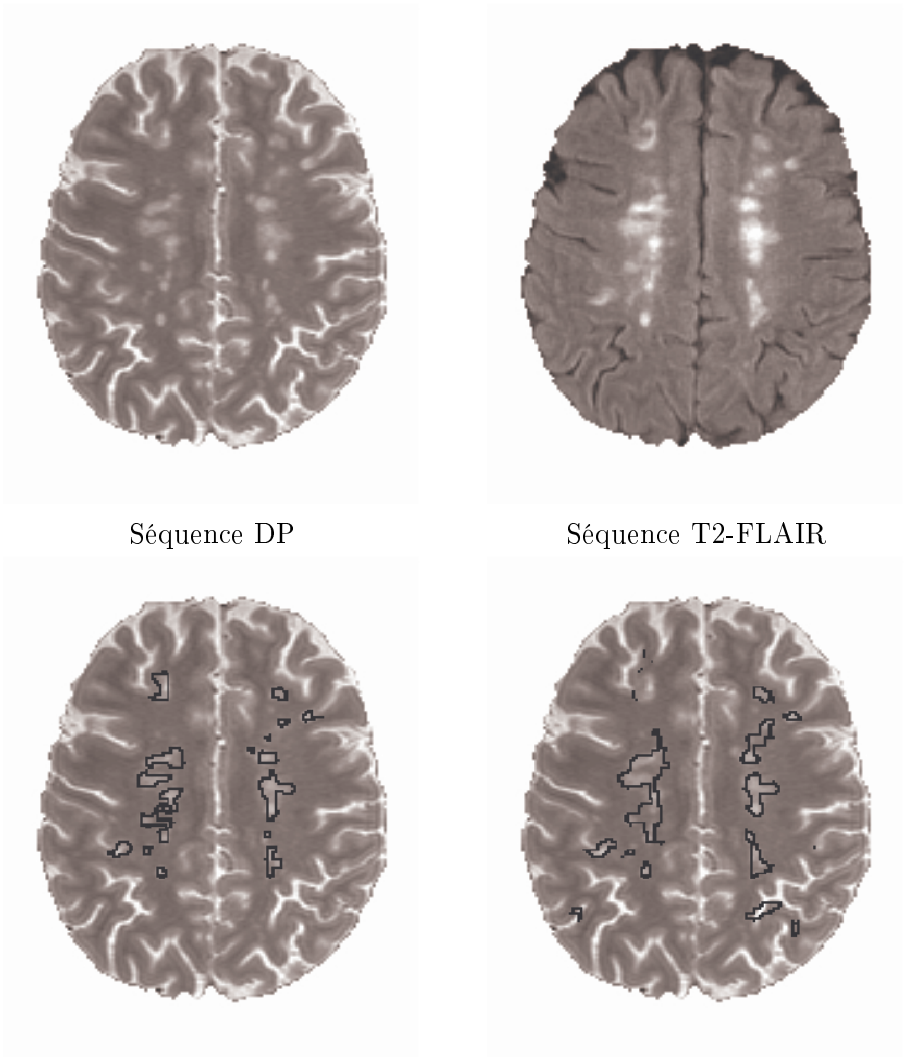


FIG. B.15 – *Légère sur-segmentation des lésions pouvant s'expliquer par la prise en compte de la SB "sale". Deux faux positifs sont présents en bas à droite (proximité avec le cortex).*

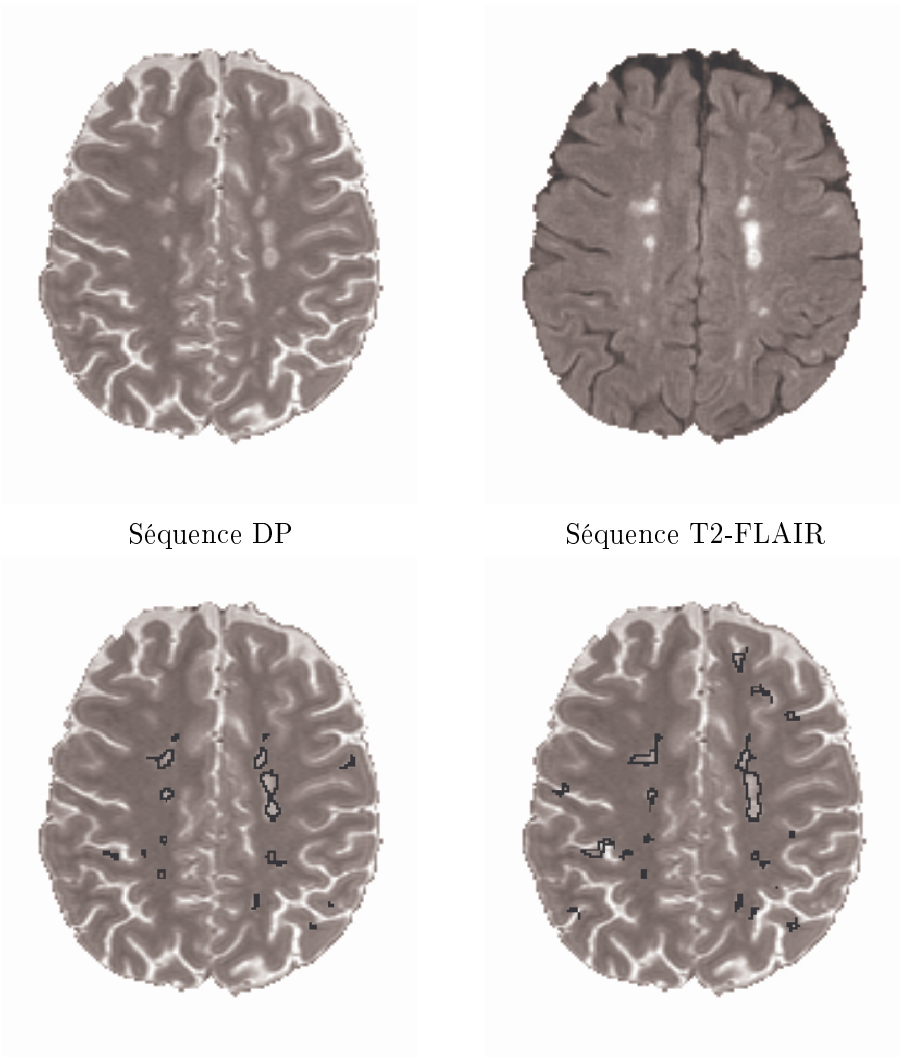


FIG. B.16 – *Légère sur-segmentation des lésions pouvant s'expliquer par la prise en compte de la SB "sale".*

B.2 Segmentation des lésions du cervelet

Les figures suivantes donnent un exemple de la segmentation obtenue dans la région du cervelet. Cette segmentation est illustrée sur le sujet DAZIN possèdant quatre lésions dans le cervelet (selon la segmentation manuelle).

Dans ce cas, il faut noter que les lésions du cervelet sont obtenues automatiquement à partir de la séquence DP, alors que la segmentation manuelle ne tient compte de cette séquence que si une lésion est visible en T2. La séquence DP est utilisée par l'expert uniquement pour confirmer ou infirmer une lésion supposée à partir du T2. Ceci pourrait donc expliquer une partie des faux positifs observés.

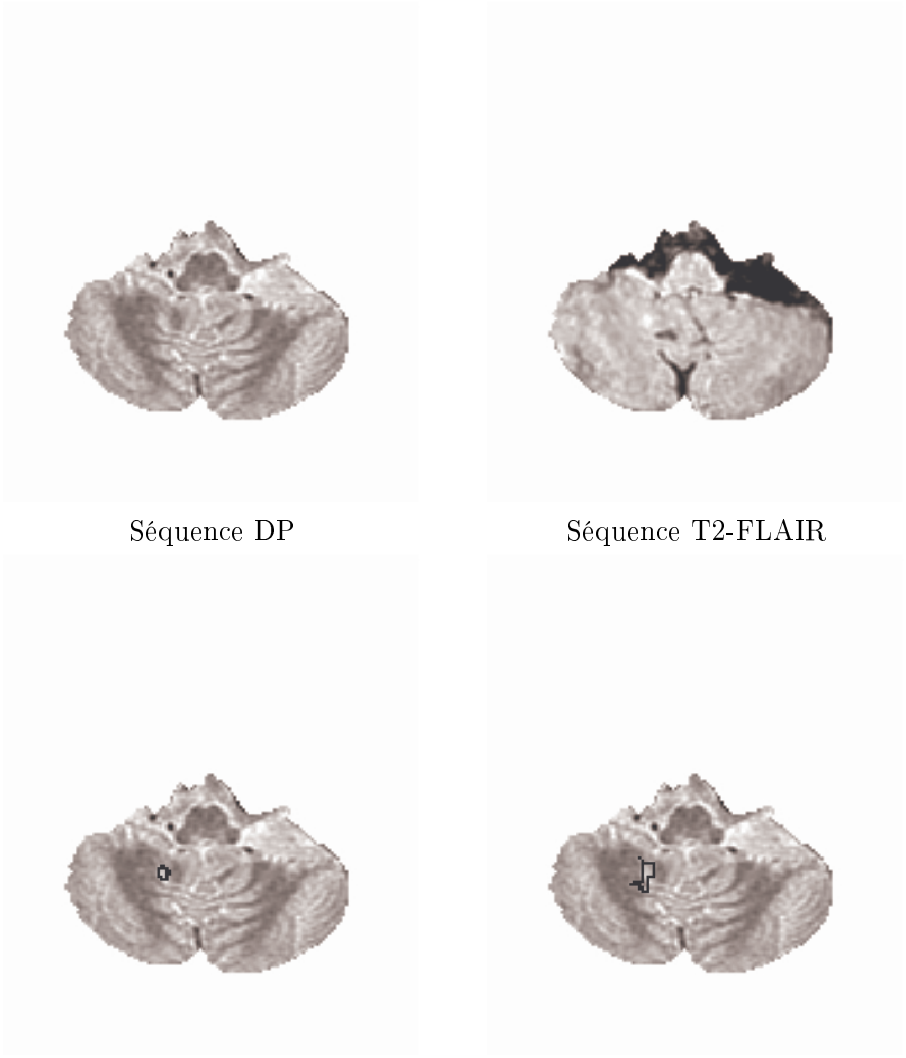


FIG. B.17 – *Légère sur-segmentation d'une lésion du cervelet par rapport à la segmentation manuelle.*

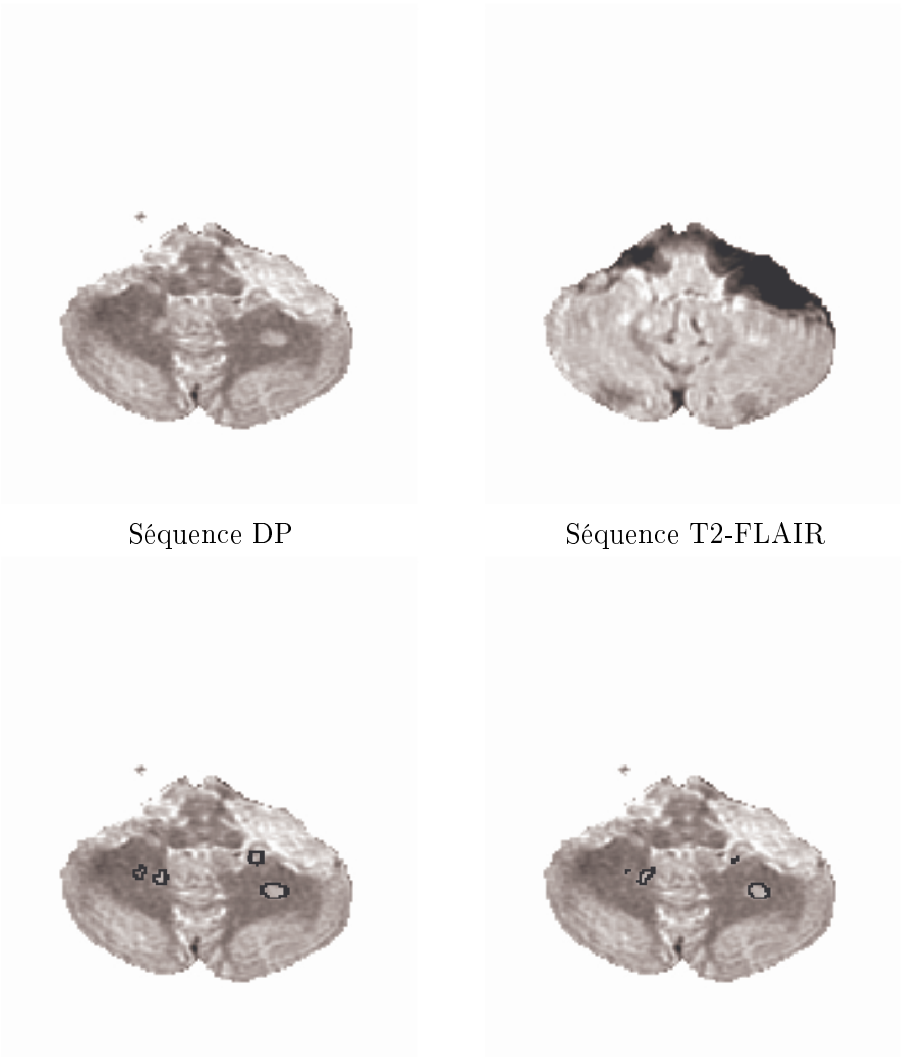


FIG. B.18 – *Les quatres lésions du cervelet sont bien détectées. Deux lésions sont sous-segmentées.*

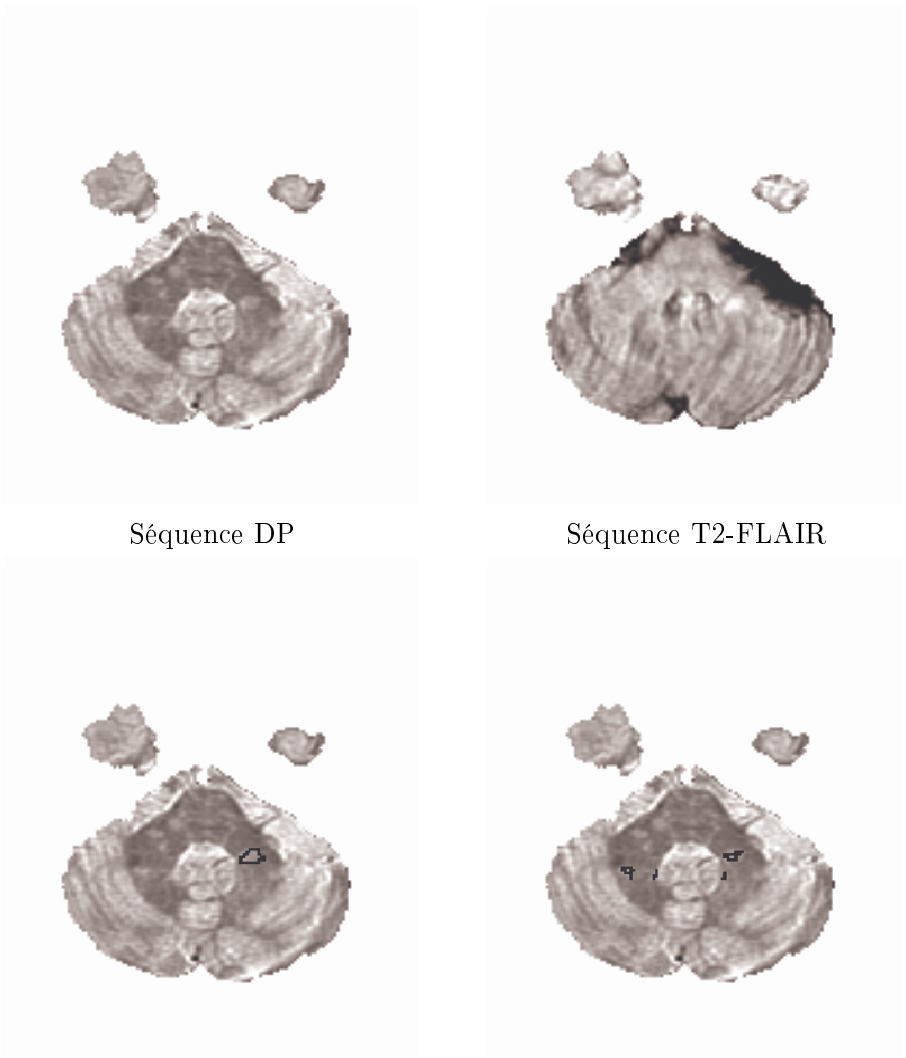
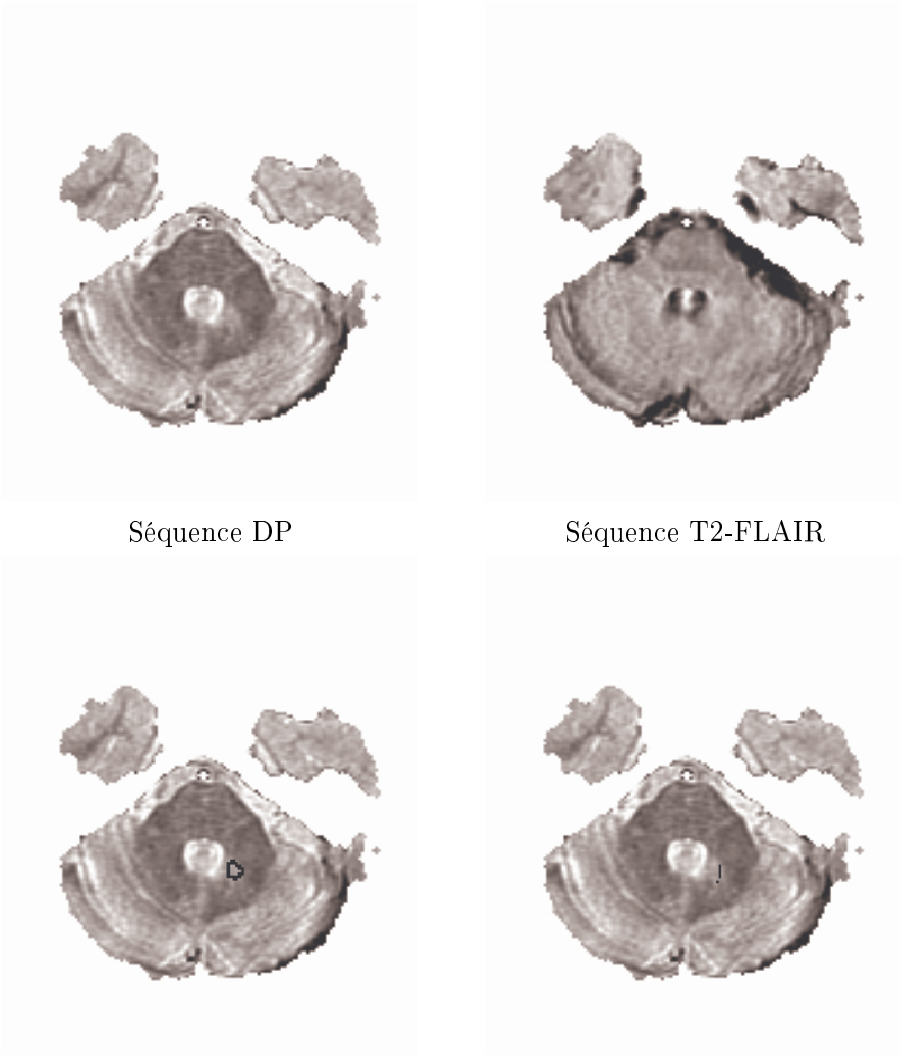


FIG. B.19 – *La plus grosse lésion de cervelet est détectée. Des faux positifs sont présents.*



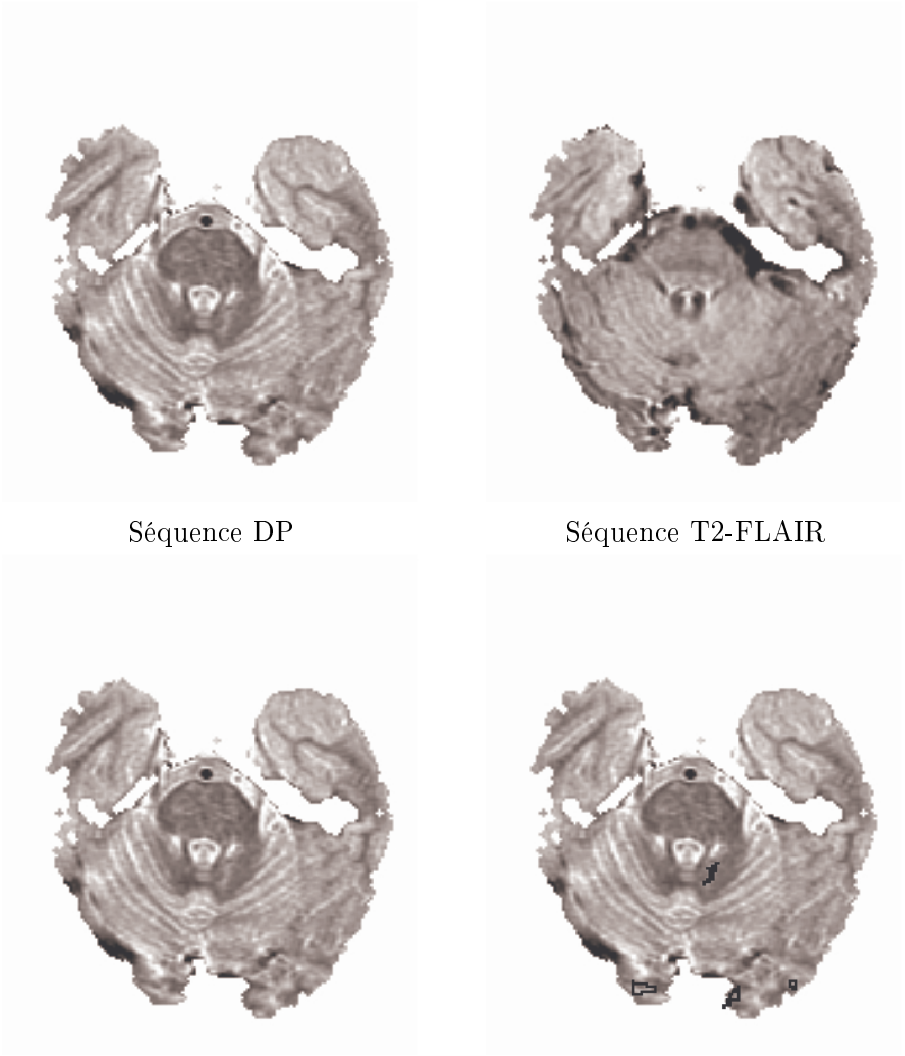
Séquence DP Séquence T2-FLAIR

Coupe 16

Segmentation manuelle PEMS-COMBI

Sujet : DAZIN

FIG. B.20 – *La lésion est détectée mais sous-segmentée.*



Séquence DP | Séquence T2-FLAIR

Coupe 18

Segmentation manuelle | PEMS-COMBI

Sujet : DAZIN

FIG. B.21 – *Présence d'un faux positif.*

Annexe C

Tableau récapitulatif des mesures effectuées sur les sujets de la base QUALICORE

La table C.1 présente les résultats des différentes mesures instant par instant sur les sujets de la base QUALICORE. La table C.2 donne le pourcentage de variation de ces mesures entre l'instant considéré et l'instant T_0.

TAB. C.1 – *Résultats des différentes mesures instant par instant sur les sujets de la base QUALICORE.*

Centre	ID	Instants	Volume des lésions				Nombre de lésions				BPF (LCLASS)	Largeur 3e ventricule	Largeur ventricules latéraux	EDSS
			Auto	Manu T1	Manu T1-Gd	Manu T2	Manu T2-FLAIR	Auto	Manu T1	Manu T1-Gd	Manu T2	Manu T2-FLAIR				
MON	SAIVE	T0	5985	1051	0	5489	7713	20	14	0	20	32	92,93	6,37	24,19	
MON	SAIVE	T1	6521	863	0	5889	8441	22	16	0	26	29	93,09	6,86	24,81	
MON	SAIVE	T2	9070	1216	0	4764	7056	24	26	0	39	38	92,57	7,07	25,36	
MON	ROQLA	T0	15421	2153	0	4381	7028	21	29	0	28	26	93,89	3,72	18,48	2
MON	CAKMA	T0	4859	1744	0	2795	3130	17	14	0	19	18	93,37	3,1	21,58	0
CLE	CELFR	T0	11538	520	0	932	1454	35	8	0	10	9	87,87	3,99	17,32	2
CLE	CELFR	T1	12969	670	37	1508	2463	55	10	1	13	13	87,42	4,9	18,36	2
CLE	CELFR	T2	13182	541	63	1141	1992	47	10	2	10	9	84,65	5,07	18,82	
CLE	DELEM	T0	35026	26443	652	45761	53082	33	103	1	95	80	88,66	4,1	24,44	0
CLE	DELEM	T1	35532	20336	0	28338	33590	29	105	0	74	66	88,67	4,29	25,01	
CLE	LAMAN	T0	20308	3932	0	6796	8552	44	30	0	37	39	92,78	3,26	18,74	0
CLE	LAMAN	T1	15996	3512	0	4910	6450	40	32	0	31	31	92,82	3,64	18,76	
CLE	ROCCA	T0	12118	3060	155	5026	7826	34	33	1	36	34	92,02	5,39	20,8	1
CLE	ROCCA	T1	6195	2127	0	3501	5022	26	39	0	49	44	92,74	5,52	20,79	1
CLE	PAUCH	T0	19085	1970	0	2668	4078	33	24	0	22	24	92,64	5,61	19,82	2
CLE	PAUCH	T1	12379	1322	0	2936	4224	40	16	0	21	21	91,77	6,74	21,35	2
CLE	LENFL	T0	4635	0	0	134	200	25	0	0	2	3	92,64	5,08	21,25	1,5
CLE	FAUTH	T0	17800	11572	141	18706	26155	18	32	3	30	17	91,74	2,87	19,86	1,5
CLE	VIGFR	T0	16139	9592	712	13606	18582	49	135	5	128	120	87,94	5,91	25,5	
NIC	DAZIN	T0	33534	6505	255	10539	16574	44	55	1	54	62	92,24	3,32	20,59	0
NIC	DAZIN	T1	33340	7009	0	11431	14512	57	50	0	70	72	92,54	3,27	20,4	
NIC	LICPA	T0	17508	1206	0	2590	4046	40	16	0	28	33	91,3	5,37	16,1	1
NIC	LICPA	T1	15694	897	0	2016	3528	36	18	0	21	20	91,43	5,32	16,11	1,5
NIC	LICPA	T2	14630	1058	0	2352	3918	34	27	0	29	26	90,31	5,62	17,29	1,5
NIC	LICPA	T3	13826	1102	0	2409	4113	23	29	0	31	31	91,05	6,29	17,5	1
NIC	ZIAKA	T0	40512	15387	0	22901	30962	34	53	0	47	45	89,71	4,89	23,62	3
NIC	FRECO	T0	26457	7399	0	11454	17835	24	84	0	92	60	87,42	6,11	20,98	1
NIC	FRECO	T1	25915	8527	0	15631	20817	22	62	0	54	46	87,12	6,29	21,47	2
NIC	FRECO	T2	27780	8444	0	12739	19469	31	72	0	61	52	86,19	6,5	22,25	2
NIC	FRECO	T3	20695	9018	0	12827	18620	29	63	0	63	46	86,62	6,52	22,16	1
NIC	MARTH	T0	33178	49377	568	61304	75179	52	99	6	87	77	91,07	6,91	23,11	1
NIC	SCHIS	T0	43410	14501	0	23480	40204	56	121	0	117	94	91,07	7,61	28,73	2,5
NIC	SCHIS	T1	41467	18455	256	33890	45647	39	101	2	87	83	91,73	7,57	28,81	3
NIC	SCHIS	T2	38739	20584	42	34070	48079	35	117	1	110	90	89,24	8,51	29,92	5
NIC	VELMI	T0	12364	3043	0	5255	7555	37	35	0	36	37	90,56	4,57	21,71	2
NIC	VELMI	T1	15984	2902	0	5908	8352	39	31	0	44	43	90,28	4,73	22,08	0
NIC	VELMI	T2	12779	2519	0	4071	5658	25	39	0	44	42	89,42	4,92	22,37	0
NIC	VELMI	T3	10786	2599	0	4116	6017	36	31	0	33	32	88,26	5,15	22,86	
NIC	GUENA	T0	22353	9556	0	13483	17199	32	93	0	100	100	89,7	7,95	20,55	5
NIC	GUENA	T1	21050	10678	79	14063	20576	45	111	1	107	97	89,64	8,03	20,64	5
NIC	GIULA	T0	19334	3466	0	5270	7476	31	31	0	36	36	92,44	4,47	22,38	2
NIC	GIULA	T1	21387	3397	0	4541	6054	30	30	0	32	34	93,01	4,63	22,88	1
NIC	EDDTA	T0	29545	12842	0	18591	28289	43	97	0	88	73	93,57	4,45	17,06	3
NIC	ALLJE	T0	19018	1406	0	2765	3880	40	19	0	25	24	89,81	5,04	18,78	1
NIC	ALLJE	T1	24387	403	0	1039	2032	57	8	0	11	11	90,13	5,09	18,9	1
NIC	DRAJA	T0	24509	6252	0	13392	18468	40	116	0	104	92	88,44	5,51	27,31	3
NIC	CHADE	T0	23242	6263	48	10690	16440	45	63	1	71	68	87,76	2,36	20,96	1,5
NIC	CALCE	T0	17710	6364	1506	12178	15750	41	24	5	14	14	91,86	2,36	20,96	2
NIC	CALCE	T1	32939	5012	59	7505	9816	34	23	1	24	25	91,92	4,36	21,19	3,5
NIC	HUBSO	T0	28771	5267	0	10036	13722	39	41	0	48	45	92,34	4,78	22	0
NIC	HUBSO	T1	20321	6244	0	8714	11556	34	50	0	49	44	91,11	5,11	23,15	0
NIC	HUBSO	T2	20048	7562	0	10645	14635	36	50	0	47	55	92,42	4,65	22,04	1

TAB. C.2 – *Pourcentage de variation des différentes mesures entre l'instant considéré et l'instant T_0 pour les sujets de la base QUALICORE.*

Centre	ID	Instant considéré par rapport à T0	Variation du volume des lésions					Variation du nombre de lésions					Variation de la BPF (LCLASS)	-Variation de la largeur 3e ventricule	-Variation de la largeur ventricules latéraux
			Auto	Manu T1	Manu T1-Gd	Manu T2	Manu T2-FLAIR	Auto	Manu T1	Manu T1-Gd	Manu T2	Manu T2-FLAIR			
MON	SAIVE	T1	8,96	-17,89	0,00	7,29	9,44	10,00	14,29	0,00	30,00	-9,38	0,17	-7,69	-2,56
MON	SAIVE	T2	51,55	15,70	0,00	-13,21	-8,52	20,00	85,71	0,00	95,00	18,75	-0,39	-10,99	-4,84
CLE	CELFR	T1	12,40	28,85	37,00	61,80	69,39	57,14	25,00	1,00	30,00	44,44	-0,51	-22,81	-6,00
CLE	CELFR	T2	14,25	4,04	63,00	22,42	37,00	34,29	25,00	2,00	0,00	0,00	-3,66	-27,07	-8,66
CLE	DELEM	T1	1,44	-23,09	-652,00	-38,07	-36,72	-12,12	1,94	-1,00	-22,11	-17,50	0,01	-4,63	-2,33
CLE	LAMAN	T1	-21,23	-10,68	0,00	-27,75	-24,58	-9,09	6,67	0,00	-16,22	-20,51	0,04	-11,66	-0,11
CLE	ROCCA	T1	-48,88	-30,49	-155,00	-30,34	-35,83	-23,53	18,18	-1,00	36,11	29,41	0,78	-2,41	0,05
CLE	PAUCH	T1	-35,14	-32,89	0,00	10,04	3,58	21,21	-33,33	0,00	-4,55	-12,50	-0,94	-20,14	-7,72
NIC	DAZIN	T1	-0,58	7,75	-255,00	8,46	-12,44	29,55	-9,09	-1,00	29,63	16,13	0,33	1,51	0,92
NIC	LICPA	T1	-10,36	-25,62	0,00	-22,16	-12,80	-10,00	12,50	0,00	-25,00	-39,39	0,14	0,93	-0,06
NIC	LICPA	T2	-16,44	-12,27	0,00	-9,19	-3,16	-15,00	68,75	0,00	3,57	-21,21	-1,08	-4,66	-7,39
NIC	LICPA	T3	-21,03	-8,62	0,00	-6,99	1,66	-42,50	81,25	0,00	10,71	-6,06	-0,27	-17,13	-8,70
NIC	FRECO	T1	-2,05	15,25	0,00	36,47	16,72	-8,33	-26,19	0,00	-41,30	-23,33	-0,34	-2,95	-2,34
NIC	FRECO	T2	5,00	14,12	0,00	11,22	9,16	29,17	-14,29	0,00	-33,70	-13,33	-1,41	-6,38	-6,05
NIC	FRECO	T3	-21,78	21,88	0,00	11,99	4,40	20,83	-25,00	0,00	-31,52	-23,33	-0,92	-6,71	-5,62
NIC	SCHIS	T1	-4,48	27,27	256,00	44,34	13,54	-30,36	-16,53	2,00	-25,64	-11,70	0,72	0,53	-0,28
NIC	SCHIS	T2	-10,76	41,95	42,00	45,10	19,59	-37,50	-3,31	1,00	-5,98	-4,26	-2,01	-11,83	-4,14
NIC	VELMI	T1	29,28	-4,63	0,00	12,43	10,55	5,41	-11,43	0,00	22,22	16,22	-0,31	-3,50	-1,70
NIC	VELMI	T2	3,36	-17,22	0,00	-22,53	-25,11	-32,43	11,43	0,00	22,22	13,51	-1,26	-7,66	-3,04
NIC	VELMI	T3	-12,76	-14,59	0,00	-21,67	-20,36	-2,70	-11,43	0,00	-8,33	-13,51	-2,54	-12,69	-5,30
NIC	GUENA	T1	-5,83	11,74	79,00	4,30	19,63	40,63	19,35	1,00	7,00	-3,00	-0,07	-1,01	-0,44
NIC	GIULA	T1	10,62	-1,99	0,00	-13,83	-19,02	-3,23	-3,23	0,00	-11,11	-5,56	0,62	-3,58	-2,23
NIC	ALLJE	T1	28,23	-71,34	0,00	-62,42	-47,63	42,50	-57,89	0,00	-56,00	-54,17	0,36	-0,99	-0,64
NIC	CALCE	T1	85,99	-21,24	-1447,00	-38,37	-37,68	-17,07	-4,17	-4,00	71,43	78,57	0,07	-84,75	-1,10
NIC	HUBSO	T1	-29,37	18,55	0,00	-13,17	-15,78	-12,82	21,95	0,00	2,08	-2,22	-1,33	-6,90	-5,23
NIC	HUBSO	T2	-30,32	43,57	0,00	6,07	6,65	-7,69	21,95	0,00	-2,08	22,22	0,09	2,72	-0,18

Annexe D

Projet NeuroLOG et Web-services

Cette annexe explique le contexte et les notions techniques des travaux réalisés en collaboration avec Erik Pernod. Les tenants et les aboutissants de ce travail sont donnés dans la section 7.2.2.

D.1 Le projet NeuroLOG

NeuroLOG[1] est un projet de 3 ans fondé par l'agence nationale de la recherche française sous le contrat numéro : ANR-06-TLOG-024. Ce projet vise à créer un intergiciel (logiciel intermédiaire entre diverses applications reliées sur un réseau) dans le but de fédérer des données médicales, des données composées (médicales + autres), des algorithmes et de partager des ressources informatiques via une grille de calcul [Montagnat 2008]. Il est particulièrement destiné aux applications portant sur les trois pathologies suivantes : sclérose en plaques, attaque cérébrale et développement de tumeurs. Concernant la SEP, le défi est de permettre un diagnostic précoce et d'aider à prédire l'évolution de la maladie. Ces applications utilisent différents algorithmes qui peuvent être partagés par les partenaires du projet (cf. figure D.1).



Fig. D.1 – *Partenaires du projet NeuroLOG.*

[1] http ://neurolog.polytech.unice.fr/

D.2 Aspects techniques

Cette section ne présente que brièvement les aspects techniques du découpage du développement de services sur la grille. Il présente aussi sommairement le découpage du pipeline de segmentation présenté dans le chapitre 3. Le lecteur intéressé par ces aspects peut se référer au rapport de stage d'Erik Pernod [Pernod 2008].

Grille de calcul : Une grille de calcul est un réseau de ressources (de calcul ou de stockage) connecté sous la forme topologique d'une grille [Foster 2001].

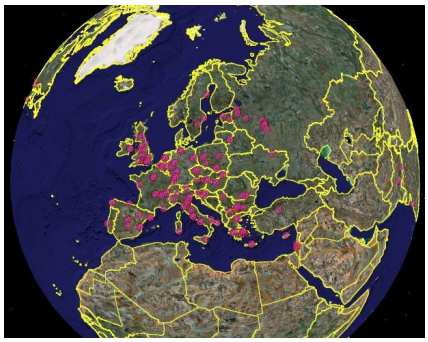


FIG. D.2 – *Partemaires du projet EGEE.*

Les travaux présentés dans la section 3.3.3.3 et les web-services présentés dans la section D.3 ont été réalisés sur la grille EGEE[1] (Enabling Grids for E-sciencE). EGEE est l'une des plus grandes infrastructures multidisciplinaires au monde. Elle regroupe 120 organisations (cf. figure D.2) allant du biomédical, à la fusion nucléaire et fournit des ressources de calculs à toute la communauté de recherche européenne.

Grille de calcul versus cluster (grappe) : Même si une grille et un cluster ont des buts similaires (*i.e.* fournir des ressources de stockage et de calcul sur un réseau et permettre de paralléliser des applications), il existe des différences entre eux. Les unités de calcul d'un cluster sont homogènes et localisées géographiquement sur un même site alors que les unités de calcul d'une grille sont inhomogènes et dispersées. L'homogénéité d'un cluster et son caractère géographique local permettent un meilleur contrôle de celui-ci et des connexions plus rapides mais il n'est partagé que par une communauté restreinte. À l'inverse, une grille permet un accès distant et transparent aux unités de calcul et aux bases de données mais l'inhomogénéité des éléments la composant peut être dangereuse en terme de sécurité des données et d'erreurs d'exécution dues à la différence d'environnement système.

Service : Un service est une fonctionnalité autonome. Celle-ci peut se représenter sous forme de boîte noire avec une liste d'entrées et de sorties. La boîte noire correspond à l'exécution d'un ou plusieurs logiciels mais la définition du service ne dépend pas de ceux-ci.

[1] EGEE, http://www.eu-egee.org

Architecture orientée service : Il s'agit d'une infrastructure informatique permettant aux services d'échanger des données entre eux. Les différents services peuvent être combinés et réutilisés pour former des chaînes de traitements.

Web-service : Les services prennent tout leur sens lorsqu'ils sont accessibles via un protocole standard d'internet (comme HTTP) qui est indépendant des plates-formes d'exécution et du langage de programmation. La description d'un web-service est réalisée dans un langage de description de services (dérivant du XML) qui a été standardisé par le W3C[2].

D.3 Web-services et chaîne de traitement

La chaîne de traitement de classification des tissus cérébraux (cf chapitre 3) doit être découpée en différents web-services. La description des liens entre ces web-services permet alors d'exécuter la chaîne de traitement sur la grille.

Le premier travail a été de détailler cette chaîne de traitement (cf. figure D.3). Il a ensuite été décidé de la découper en 7 services en fonction des rôles des divers algorithmes.

Nom	**Asclepios_Conversion_service**
Traitement	Conversion de format des données.
Description	Convertit une image d'un format vers un autre sans changer le contenu des données.

Nom	**Asclepios_Baladin_registration_service**
Traitement	Recalage de deux images (par une transformation rigide ou affine) [Ourselin 2000].
Description	Recale une image flottante sur une image de référence de manière rigide ou affine. Crée une matrice de transformation (4x4) et rééchantillonne après transformation l'image flottante dans le repère de l'image de référence.

Nom	**Asclepios_Resampling_service**
Traitement	Rééchantillonnage d'une image dans un autre repère.
Description	Rééchantillonne une image en fonction d'une transformation (obtenue lors d'un recalage) dans le même repère ou dans un nouveau repère (pouvant être spécifié).

[2]The World Wide Web Consortium, http://www.w3.org

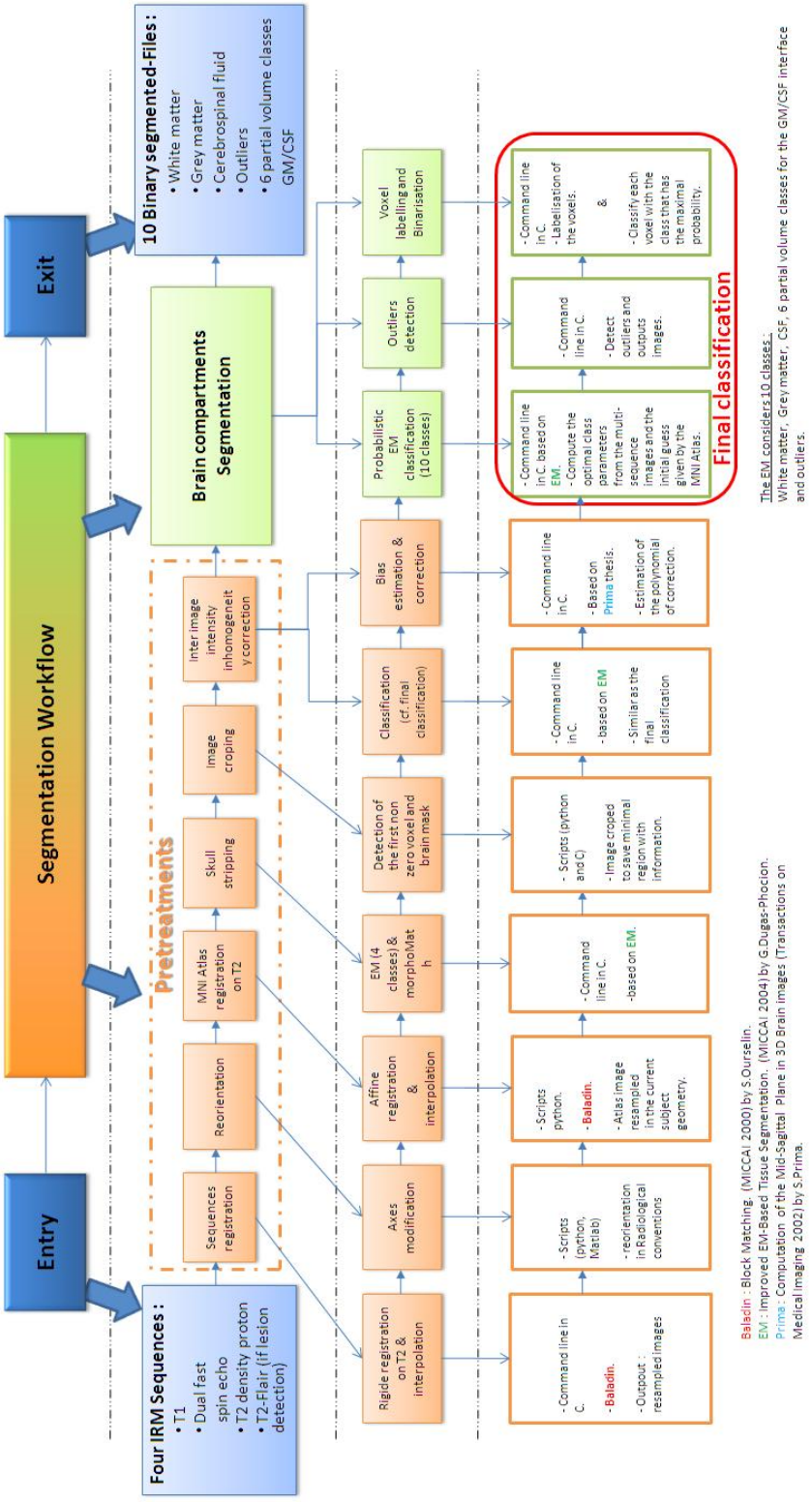


FIG. D.3 – *Première description de la chaîne de traitement de classification des tissus cerebraux.*

Nom	**Asclepios_EMBrainMask_service**
Traitement	Segmentation du cerveau [Dugas-Phocion 2004b].
Description	Segmente le cerveau grâce à l'utilisation d'un algorithme EM utilisant des informations *a priori* fournies par un atlas.

Nom	**Asclepios_EMBrainClassification_service**
Traitement	Classification des compartiments cérébraux [Dugas-Phocion 2004a].
Description	Réalise une classification des différents compartiments cérébraux grâce à un algorithme EM en fonction d'un masque du cerveau et en utilisant des informations *a priori* fournies par un atlas. Les cartes d'appartenances de chaque voxel aux différentes classes (SB, SG, LCR, 6 classes de volumes partiels SG/LCR) sont obtenues. Une segmentation binaire des points aberrants ne respectant pas le modèle de distributions des intensités est aussi fournie.

Nom	**Asclepios_Binarisation_service**
Traitement	Création des segmentations binaires à partir des cartes d'appartenances.
Description	Crée des segmentations binaires à partir des cartes d'appartenances. Le voxel appartient à la classe ayant le taux d'appartenance le plus élevé.

Nom	**Asclepios_Unbias_service**
Traitement	Estimation du champs de biais et calcul de l'image débiaisée [Prima 2001].
Description	À partir de segmentations binaires, un polynôme à appliquer pour corriger le biais est calculé, pour chaque séquence IRM. Ce polynôme est ensuite appliqué pour corriger le biais.

La description des liens entre ces web-services permet de définir la chaîne de traitement (cf. figure D.4).

On peut voir sur cette chaîne que les web-services sont paramétrisables (par exemple : différents types de recalages sont possibles). Ces services ont aussi été validés indépendamment et dans la chaîne de traitement. Pour cela, les résultats obtenus via la grille ont été comparés avec les résultats obtenus localement (sur une machine du projet) par les logiciels.

Les moyens de création de ces services et les difficultés rencontrées ne sont pas présentés ici. Le lecteur intéressé par ces aspects peut se référer à [Pernod 2008].

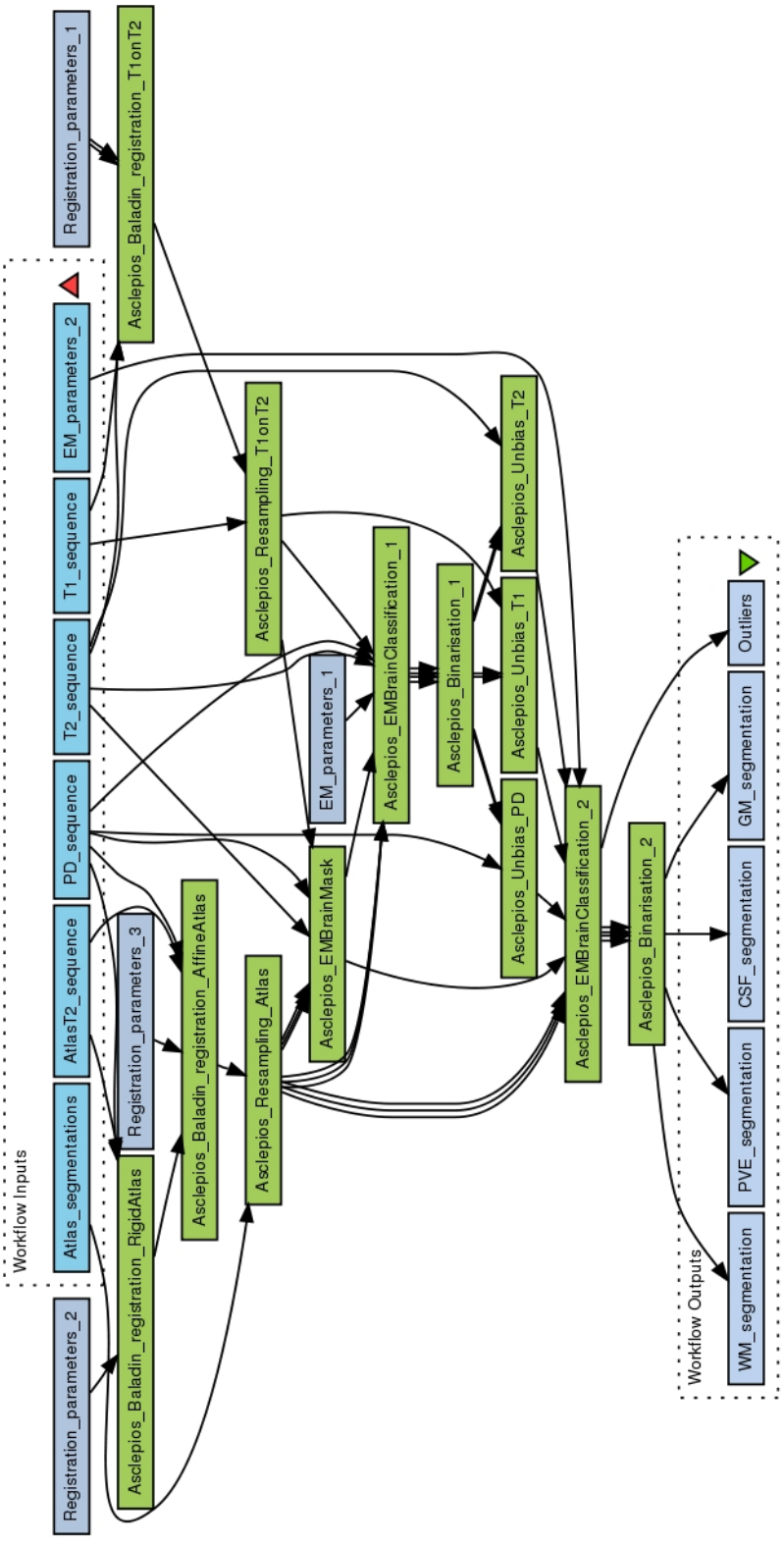


FIG. D.4 – *Définition de la chaîne de traitement à partir des web-services.*

Bibliographie

[Acar 2003] G. Acar, F. Idiman, E. Idiman, G. Kirkali, H. Cakmakci et S. Ozakbas. *Nitric oxide as an activity marker in multiple sclerosis.* Journal of Neurology, vol. 250, pages 588–592, 2003. 10

[Achiron 2002] Anat Achiron, Sebastien Gicquel, Shmuel Miron et Meir Faibel. *Brain MRI lesion load quantification in multiple sclerosis : a comparison between automated multispectral and semi-automated thresholding computer-assisted techniques.* Magn Reson Imaging, vol. 20, no. 10, pages 713–20, December 2002. 70

[Admasu 2003] F Admasu, S Al-Zubi, K Toennies, N Bodammer et H Hinrichs. *Segmentation of multiple sclerosis lesions from MR brain images using the principles of fuzzy-connectedness and artificial neuron networks.* Proceedings of 2003 International Conference on Image Processing, vol. 3, pages II–1081–4, September 2003. 35, 73, 74, 81

[Admiraal-Behloul 2005] F Admiraal-Behloul, DM van den Heuvel, H Olofsen, MJ van Osch, J van der Grond, M A van Buchem et JH Reiber. *Fully automatic segmentation of white matter hyperintensities in MR images of the elderly.* Neuroimage, vol. 28, no. 3, pages 607–17, November 2005. 36

[Aït-Ali 2005] L S Aït-Ali, S Prima, P Hellier, B Carsin, G Edan et C Barillot. *STREM : a robust multidimensional parametric method to segment MS lesions in MRI.* In Med Image Comput Comput Assist Interv, volume 3749 of *LNCS*, pages 409–16. Springer, 2005. 74, 79, 81, 115

[Aït-Ali 2006] Laure Soraya Aït-Ali. *Analyse spatio-temporelle pour le suivi de structures évolutives en imagerie cérébrale multi-séquences.* Thèse de sciences, Université de Rennes 1, 2006. 37, 45

[Akselrod-Ballin 2006] Ayelet Akselrod-Ballin, Meirav Galun, Ronen Basri, Achi Brandt, Moshe John Gomori, Massimo Filippi et Paula Valsasina. *An Integrated Segmentation and Classification Approach Applied to Multiple Sclerosis Analysis.* In 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, volume 1, pages 1122–1129, 2006. 70, 81

[Alonge 2001] F Alonge, E Ardizzone et R Pirrone. *Neural classification of multiple sclerosis lesions in MR images.* International Journal of Knowledge-Based Intelligent Engineering Systems, vol. 5, no. 4, pages 228–233, 2001. 35, 36, 73, 81

[Altman 1994] D G Altman et J M Bland. *Diagnostic tests 2 : Predictive values.* BMJ, vol. 309, no. 6947, page 102, July 1994. 79

[Amato 2007] Maria Pia Amato et Emilio Portaccio. *Clinical outcome measures in multiple sclerosis.* J Neurol Sci, vol. 259, no. 1-2, pages 118–22, August 2007. 10

[Anbeek 2004a] Petronella Anbeek, Koen L Vincken, MJP van Osch, RHC Bisschops et Jeroen van der Grond. *Automatic segmentation of different-sized white matter lesions by voxel probability estimation.* Med Image Anal, vol. 8, no. 3, pages 205–15, September 2004. 35, 79

[Anbeek 2004b] Petronella Anbeek, Koen L Vincken, MJP van Osch, Robertus H C Bisschops et Jeroen van der Grond. *Probabilistic segmentation of white matter lesions in MR imaging.* Neuroimage, vol. 21, no. 3, pages 1037–44, March 2004. 35, 79

[Anderson 2006] Valerie M Anderson, Nick C Fox et David H Miller. *Magnetic resonance imaging measures of brain atrophy in multiple sclerosis.* J Magn Reson Imaging, vol. 23, no. 5, pages 605–18, May 2006. 107, 110

[Ardizzone 1999] E. Ardizzone et R. Pirrone. *An Architecture for the Recognition and Classification of Multiple Sclerosis Lesions in MR Images.* In Intelligent Data Analysis in Medicine and Pharmacology, pages 1–12, 1999. 73, 81

[Ashton 2003] Edward A Ashton, Chihiro Takahashi, Michel J Berg, Andrew Goodman, Saara Totterman et Sven Ekholm. *Accuracy and reproducibility of manual and semiautomated quantification of MS lesions by MRI.* J Magn Reson Imaging, vol. 17, no. 3, pages 300–8, March 2003. 67, 77

[Bakshi 2005] R Bakshi. *Magnetic resonance imaging advances in multiple sclerosis.* J Neuroimaging, vol. 15, no. 4 Suppl, page 5, 2005. 67

[Bar-Zohar 2008] D Bar-Zohar, F Agosta, D Goldstaub et M Filippi. *Magnetic resonance imaging metrics and their correlation with clinical outcomes in multiple sclerosis : a review of the literature and future perspectives.* Mult Scler, vol. 14, no. 6, pages 719–27, July 2008. 143

[Barkhof 1997] F Barkhof, M Filippi, D H Miller, P Scheltens, A Campi, C H Polman, G Comi, H J Ader, N Losseff et J Valk. *Comparison of MRI criteria at first presentation to predict conversion to clinically definite multiple sclerosis.* Brain, vol. 120 (Pt 11), pages 2059–69, November 1997. 16, 67

[Barkhof 1998] F J Barkhof, M Elton, J Lindeboom, M W Tas, W F Schmidt, O R Hommes, C H Polman, A Kok et J Valk. *Functional correlates of callosal atrophy in relapsing-remitting multiple sclerosis patients. A preliminary MRI study.* J Neurol, vol. 245, no. 3, pages 153–8, March 1998. 108

[Barkhof 1999] F Barkhof. *MRI in multiple sclerosis : correlation with expanded disability status scale (EDSS).* Mult Scler, vol. 5, no. 4, pages 283–6, August 1999. 17

[Barkhof 2001] F Barkhof, J H van Waesberghe, M Filippi, T Yousry, D H Miller, D Hahn, A J Thompson, L Kappos, P Brex, C Pozzilli et C H Polman. *T1 hypointense lesions in secondary progressive multiple sclerosis : effect of interferon beta-1b treatment.* Brain, vol. 124, no. Pt 7, pages 1396–402, July 2001. 68

[Barkhof 2004] Frederik Barkhof. *Assessing treatment effects on axonal loss–evidence from MRI monitored clinical trials.* J Neurol, vol. 251 Suppl 4, September 2004. 20

[Benedict 2004] Ralph H B Benedict, RH Benedict, Bianca Weinstock-Guttman, Inna Fishman, Jitendra Sharma, Christopher W Tjoa et Rohit Bakshi. *Prediction of neuropsychological impairment in multiple sclerosis : comparison of conventional magnetic resonance imaging measures of atrophy and lesion burden.* Arch Neurol, vol. 61, no. 2, pages 226–30, February 2004. 110, 112

[Bermel 2002] Robert A Bermel, Rohit Bakshi, Christopher Tjoa, Srinivas R Puli et Lawrence Jacobs. *Bicaudate ratio as a magnetic resonance imaging marker of brain atrophy in multiple sclerosis.* Arch Neurol, vol. 59, no. 2, pages 275–80, February 2002. 108, 110, 113

[Bermel 2006] Robert A Bermel et Rohit Bakshi. *The measurement and clinical relevance of brain atrophy in multiple sclerosis.* Lancet Neurol, vol. 5, no. 2, pages 158–70, February 2006. 107

[Beyer 2008] Thomas Beyer, Markus Weigert, Harald H Quick, Uwe Pietrzyk, Florian Vogt, Christoph Palm, Gerald Antoch, Stefan P Muller et Andreas Bockisch. *MR-based attenuation correction for torso-PET/MR imaging : pitfalls in mapping MR to CT data.* Eur J Nucl Med Mol Imaging, vol. 35, no. 6, pages 1142–6, June 2008. 32

[Bezdek 1999] J. C. Bezdek, J. Keller, R. Krishnapuram et N. R. Pal. Fuzzy Models and Algorithms for Pattern Recognition and Image Processing. Kluwer Academic Publisher, 1999. 36

[Bloch 1946] F. Bloch. *Nuclear Induction.* Physical review, vol. 70, pages 460–474, 1946. 13

[Bondiau 2005] Pierre-Yves Bondiau, Grégoire Malandain, Stéphane Chanalet, Pierre-Yves Marcy, Jean-Louis Habrand, François Fauchon, Philippe Paquis, Adel Courdi, Olivier Commowick, Isabelle Rutten et Nicholas Ayache. *Atlas-based automatic segmentation of MR images : validation study on the brainstem in radiotherapy context.* Int J Radiat Oncol Biol Phys, vol. 61, no. 1, pages 289–98, January 2005. 35, 62

[Borman 2004] Sean Borman. The expectation maximization algorithm – a short tutorial. http ://www.seanborman.com/publications/EM_algorithm.pdf, Juillet 2004. 37

[Bosc 2003] Marcel Bosc, Fabrice Heitz, Jean Paul Armspach, Izzie Namer, Daniel Gounot et Lucien Rumbach. *Automatic change detection in multimodal serial MRI : application to multiple sclerosis lesion evolution.* Neuroimage, vol. 20, no. 2, pages 643–56, October 2003. 72

[Boudraa 2000] A O Boudraa, S M Dehak, Y M Zhu, C Pachai, Y G Bao et J Grimaud. *Automated segmentation of multiple sclerosis lesions in multispectral MR imaging using fuzzy clustering.* Comput Biol Med, vol. 30, no. 1, pages 23–40, January 2000. 36, 75, 76, 81

[Brex 2002] Peter A Brex, Olga Ciccarelli, Jonathon I O'Riordan, Michael Sailer, Alan J Thompson et David H Miller. *A longitudinal study of abnormalities on MRI and disability from multiple sclerosis.* N Engl J Med, vol. 346, no. 3, pages 158–64, January 2002. 19, 67

[Bricq 2008] S. Bricq, C. Collet et J.-P. Armspach. *Lesions detection on 3D brain MRI using trimmed likelihood estimator and probabilistic atlas.* In Fifth IEEE International Symposium on Biomedical Imaging ISBI'08, Paris, France, may 2008. 74, 81

[Brown 1992] L.G. Brown. *A Survey of Image Registration Techniques.* ACM Computing Surveys, vol. 24, no. 4, pages 325–376, Décembre 1992. 29, 30

[Bruneteau 2005] G. Bruneteau, R. Guillevin, A. Tourbah, C. Papeix, J. Chiras et C. Pierrot-Deseilligny. *Contribution of proton magnetic resonance spectroscopy to the diagnosis of Balo's concentric sclerosis.* Rev Neurol (Paris), vol. 161, no. 4, pages 455–8, April 2005. 20

[Buckle 2005] Guy J Buckle. *Functional magnetic resonance imaging and multiple sclerosis : the evidence for neuronal plasticity.* J Neuroimaging, vol. 15, no. 4 Suppl, page 82, 2005. 68

[Butzkueven 2008] H Butzkueven, S C Kolbe, D J Jolley, J Y Brown, M J Cook, I A van der Mei, P S Groom, J Carey, J Eckholdt, J P Rubio, B V Taylor, P J Mitchell, G F Egan et T J Kilpatrick. *Validation of linear cerebral atrophy markers in multiple sclerosis.* J Clin Neurosci, vol. 15, no. 2, pages 130–7, February 2008. 108, 112, 125

[Camara 2006] O. Camara, M. Schweiger, R.I. Scahill, W.R. Crum, B.I. Sneller, J.A. Schnabel, G.R. Ridgway, D.M. Cash, D.L.G. Hill et N.C. Fox. *Phenomenological model of diffuse global and regional atrophy using finite-element methods.* IEEE TMI, vol. 25, no. 11, pages 1417–30, 2006. 117, 119, 120

[Camara 2007] Oscar Camara, Rachael I Scahill, Julia A Schnabel, William R Crum, Gerard R Ridgway, Derek L G Hill, DL Hill et Nick C Fox. *Accuracy assessment of global and local atrophy measurement techniques with realistic simulated longitudinal data.* In Int Conf Med Image Comput Comput Assist Interv, volume 4792 of *LNCS*, pages 785–92. Springer, 2007. 121

[Charcot 1898] J.M. Charcot. Leçons sur les maladies du système nerveux, volume 1. Bourneville, 1898. 8

[Charil 2003] Arnaud Charil, Alex P Zijdenbos, Jonathan Taylor, Cyrus Boelman, Keith J Worsley, Alan C Evans et Alain Dagher. *Statistical mapping analysis of lesion location and neurological disability in multiple sclerosis : application to 452 patient data sets.* Neuroimage, vol. 19, no. 3, pages 532–44, July 2003. 75, 81

[Coles 1999] A J Coles, M G Wing, P Molyneux, A Paolillo, C M Davie, G Hale, D Miller, H Waldmann et A Compston. *Monoclonal antibody treatment exposes three mechanisms underlying the clinical course of multiple sclerosis.* Ann Neurol, vol. 46, no. 3, pages 296–304, September 1999. 20

[Collins 1998] D.L. Collins, A.P. Zijdenbos, V. Kollokian, J.G. Sled, N.J. Kabani, C.J. Holmes et A.C. Evans. *Design and Construction of a Realistic Digital Brain Phantom.* IEEE Transactions on Medical Imaging, vol. 17, no. 3, pages 463–468, June 1998. 78

[Collins 2001] D. Louis Collins, Johan Montagnat, Alex P. Zijdenbos, Alan C. Evans et Douglas L. Arnold. *Automated Estimation of Brain Volume in Multiple Sclerosis with BICCR.* In IPMI '01 : Proceedings of the 17th International Conference on Information Processing in Medical Imaging, pages 141–147, London, UK, 2001. Springer-Verlag. 108, 118

[Commowick 2006] Olivier Commowick, Vincent Arsigny, Jimena Costa, Nicholas Ayache et Grégoire Malandain. *An Efficient Locally Affine Framework for the Registration of Anatomical Structures.* In Proceedings of the Third IEEE International Symposium on Biomedical Imaging (ISBI 2006), pages 478–481, Crystal Gateway Marriott, Arlington, Virginia, USA, April 2006. 60

[Commowick 2007] Olivier Commowick. *Création et Utilisation d'Atlas Anatomiques Numériques pour la Radiothérapie (Design and Use of Anatomical Atlases for Radiotherapy).* Thèse de sciences (phd thesis), Université de Nice – Sophia-Antipolis, February 2007. 56, 60

[Commowick 2008] Olivier Commowick, Vincent Grégoire et Grégoire Malandain. *Atlas-Based Delineation of Lymph Node Levels in Head and Neck Computed Tomography Images.* Radiotherapy Oncology, vol. 87, no. 2, pages 281–289, 2008. 35

[Crum 2003] W R Crum, L D Griffin, D L G Hill et D J Hawkes. *Zen and the art of medical image registration : correspondence, homology, and quality.* Neuroimage, vol. 20, no. 3, pages 1425–37, Novembre 2003. 30

[Crum 2006] W.R. Crum, O. Camara et D.L.G. Hill. *Generalized overlap measures for evaluation and validation in medical image analysis.* IEEE Trans Med Imaging, vol. 25, no. 11, pages 1451–61, 2006. 116

[Cutter 1999] G R Cutter, M L Baier, R A Rudick, D L Cookfair, J S Fischer, J Petkau, K Syndulko, B G Weinshenker, J P Antel, C Confavreux, G W Ellison, F Lublin, A E Miller, S M Rao, S Reingold, A Thompson et E Willoughby. *Development of a multiple sclerosis functional composite as a clinical trial outcome measure.* Brain, vol. 122 (Pt 5), pages 871–82, May 1999. 20, 153

[Dastidar 1999] P Dastidar, T Heinonen, T Lehtimaki, M Ukkonen, J Peltola, T Erila, E Laasonen et I Elovaara. *Volumes of brain atrophy and plaques correlated with neurological disability in secondary progressive multiple sclerosis.* J Neurol Sci, vol. 165, no. 1, pages 36–42, May 1999. 19, 67, 108

[Datta 2006] Sushmita Datta, Balasrinivasa Rao Sajja, Renjie He, Jerry S Wolinsky, Rakesh K Gupta et Ponnada A Narayana. *Segmentation and quantification of black holes in multiple sclerosis.* Neuroimage, vol. 29, no. 2, pages 467–74, January 2006. 70, 71, 79, 80, 81, 82

[Datta 2007] Sushmita Datta, Balasrinivasa Rao Sajja, Renjie He, Rakesh K Gupta, Jerry S Wolinsky et Ponnada A Narayana. *Segmentation of gadolinium-enhanced lesions on MRI in multiple sclerosis.* J Magn Reson Imaging, vol. 25, no. 5, pages 932–7, May 2007. 70, 71

[Dellaert 2002] Frank Dellaert. *The Expectation Maximization Algorithm.* Rapport technique, College of Computing, Georgia Institute of Technology, February 2002. http ://www.cc.gatech.edu/ dellaert/em-paper.pdf. 37

[Dempster 1977] A. Dempster, N. Laird et D. Rubin. *Maximum likelihood for incomplete data via the EM algorithm.* Journal of the Royal Statistical Society, vol. 39, no. 1, pages 1–38, 1977. 36, 37

[Dice 1945] L.R. Dice. *Measures of the amount of ecologic association between species.* Ecology, vol. 26, no. 3, pages 297–302, 1945. 79, 116

[Duan 2008] Y Duan, P G Hildenbrand, M P Sampat, D F Tate, I Csapo, B Moraal, R Bakshi, F Barkhof, D S Meier et C R G Guttmann. *Segmentation of subtraction images for the measurement of lesion change in multiple sclerosis.* AJNR Am J Neuroradiol, vol. 29, no. 2, pages 340–6, February 2008. 72

[Dugas-Phocion 2004a] G. Dugas-Phocion, M. Ángel G. Ballester, G. Malandain, C. Lebrun et N. Ayache. *Improved EM-Based Tissue Segmentation and Partial Volume Effect Quantification in Multi-Sequence Brain MRI.* In Proc. of MICCAI'04. Springer LNCS 3216, 2004. 36, 40, 74, 81, 118, 201

[Dugas-Phocion 2004b] Guillaume Dugas-Phocion, Miguel Ángel González Ballester, Christine Lebrun, Stéphane Chanalet, Caroline Bensa, Grégoire Malandain et Nicholas Ayache. *Hierarchical Segmentation of Multiple Sclerosis Lesions in Multi-Sequence MRI.* In International Symposium on Biomedical Imaging : From Nano to Macro (ISBI'04), Arlington, VA, USA, April 2004. IEEE. 36, 55, 66, 74, 201

[Dugas-Phocion 2006] Guillaume Dugas-Phocion. *Segmentation d'IRM Cérébrales Multi-Séquences et Application à la Sclérose en Plaques.* Thèse de sciences, École des Mines de Paris, 2006. ix, x, 3, 4, 12, 22, 24, 37, 39, 40, 45, 47, 55, 56, 78, 83, 87, 160

[Dunn 1973] J.C. Dunn. *A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters.* Journal of Cybernetics, vol. 3, pages 32–57, 1973. 36

[Evans 1992] A. C. Evans, D. L. Collins et B. Milner. *An MRI-based Stereotactic Brain Atlas from 250 Young Normal Subjects.* In 22nd Symposium of the Society for Neuroscience, page 408, Anaheim, 1992. 39

[Evans 1993] A. C. Evans, D. L. Collins, S. R. Mills, E. D. Brown, R. L. Kelly et T. M. Peters. *3D statistical neuroanatomical models from 305 MRI volumes.* In IEEE-Nuclear Science Symposium and Medical Imaging Conference, pages 1813–1817, 1993. 39

[Fennema-Notestine 2006] Christine Fennema-Notestine, I Burak Ozyurt, Camellia P Clark, Shaunna Morris, Amanda Bischoff-Grethe, Mark W Bondi,

Terry L Jernigan, Bruce Fischl, Florent Segonne, David W Shattuck, Richard M Leahy, David E Rex, Arthur W Toga, Kelly H Zou et Gregory G Brown. *Quantitative evaluation of automated skull-stripping methods applied to contemporary and legacy images : effects of diagnosis, bias correction, and slice location.* Hum Brain Mapp, vol. 27, no. 2, pages 99–113, February 2006. 52, 54, 55, 56, 58

[Fenster 2005] Aaron Fenster et Bernard Chiu. *Evaluation of Segmentation algorithms for Medical Imaging.* In IEEE Engineering in Medicine and Biology, pages 7186–7189, 2005. 78

[Filippi 2001] M Filippi. *Linking structural, metabolic and functional changes in multiple sclerosis.* Eur J Neurol, vol. 8, no. 4, pages 291–7, July 2001. 34

[Filippi 2003] M Filippi et M A Rocca. *MRI aspects of the "inflammatory phase" of multiple sclerosis.* Neurol Sci, vol. 24 Suppl 5, December 2003. 19

[Filippi 2004a] Massimo Filippi, Marco Rovaris, Matilde Inglese, Frederik Barkhof, Nicola De Stefano, Steve Smith et Giancarlo Comi. *Interferon beta-1a for brain tissue loss in patients at presentation with syndromes suggestive of multiple sclerosis : a randomised, double-blind, placebo-controlled trial.* Lancet, vol. 364, no. 9444, pages 1489–96, October 2004. 10, 122, 125

[Filippi 2004b] Massimo Filippi, Marco Rovaris, Matilde Inglese, Frederik Barkhof, Nicola De Stefano, Steve Smith et Giancarlo Comi. *Interferon beta-1a for brain tissue loss in patients at presentation with syndromes suggestive of multiple sclerosis : a randomised, double-blind, placebo-controlled trial.* Lancet, vol. 364, no. 9444, pages 1489–96, October 2004. 67

[Filippi 2004c] Massimo Filippi, Marco Rovaris et Maria A Rocca. *Imaging primary progressive multiple sclerosis : the contribution of structural, metabolic, and functional MRI techniques.* Mult Scler, vol. 10 Suppl 1, June 2004. 20

[Filippi 2005] Massimo Filippi et Maria Assunta Rocca. *MRI evidence for multiple sclerosis as a diffuse disease of the central nervous system.* J Neurol, vol. 252 Suppl 5, October 2005. 20, 82

[Fischer 1999] J S Fischer, R A Rudick, G R Cutter et S C Reingold. *The Multiple Sclerosis Functional Composite Measure (MSFC) : an integrated approach to MS clinical outcome assessment. National MS Society Clinical Outcomes Assessment Task Force.* Mult Scler, vol. 5, no. 4, pages 244–50, August 1999. 20, 153

[Fisher 2000] E Fisher, R A Rudick, G Cutter, M Baier, D Miller, B Weinstock-Guttman, M K Mass, D S Dougherty et N A Simonian. *Relationship between brain atrophy and disability : an 8-year follow-up study of multiple sclerosis patients.* Mult Scler, vol. 6, no. 6, pages 373–7, December 2000. 20

[Flandin 2004] G. Flandin. *Utilisation d'informations géométriques pour l'analyse statistique des données d'IRM fonctionnelle.* PhD thesis, Université de Nice-Sophia Antipolis, Avril 2004. 37, 38

[Foster 2001] Ian Foster, Carl Kesselman et Steven Tuecke. *The Anatomy of the Grid : Enabling Scalable Virtual Organizations.* International Jounral of Supercomputer Applications, vol. 15, no. 3, 2001. 198

[Freeborough 1997] P A Freeborough et N C Fox. *The boundary shift integral : an accurate and robust measure of cerebral volume changes from registered repeat MRI.* IEEE Trans Med Imaging, vol. 16, no. 5, pages 623–9, October 1997. 109

[Furby 2008] J Furby, T Hayton, V Anderson, D Altmann, R Brenner, J Chataway, RA Hughes, KJ Smith, DH Miller et R Kapoor. *Magnetic resonance imaging measures of brain and spinal cord atrophy correlate with clinical impairment in secondary progressive multiple sclerosis.* Mult Scler, July 2008. 20, 108

[Garcia-Lorenzo 2008] D. Garcia-Lorenzo, S. Prima, D.L. Collins, D.L. Arnold, S. Morissey et C. Barillot. *Combining Robust Expectation Maximization and Mean Shift algorithms for Multiple Sclerosis Brain Segmentation.* In Proceedings of MICCAI workshop on Medical Image Analysis on Multiple Sclerosis (validation and methodological issues), September 2008. 74, 81

[Ge 2000] Y Ge, R I Grossman, J K Udupa, L Wei, L J Mannon, M Polansky et D L Kolson. *Brain atrophy in relapsing-remitting multiple sclerosis and secondary progressive multiple sclerosis : longitudinal quantitative analysis.* Radiology, vol. 214, no. 3, pages 665–70, March 2000. 138

[Ge 2002] Yulin Ge, Robert I Grossman, James S Babb, Marcie L Rabin, Lois J Mannon et Dennis L Kolson. *Age-related total gray matter and white matter changes in normal adult brain. Part I : volumetric MR imaging analysis.* AJNR Am J Neuroradiol, vol. 23, no. 8, pages 1327–33, September 2002. 108

[Ge 2003] Yulin Ge, Robert I Grossman, James S Babb, Juan He et Lois J Mannon. *Dirty-appearing white matter in multiple sclerosis : volumetric MR imaging and magnetization transfer ratio histogram analysis.* AJNR Am J Neuroradiol, vol. 24, no. 10, pages 1935–40, Nov-Dec 2003. 20

[Ge 2006] Y Ge. *Multiple sclerosis : the role of MR imaging.* AJNR Am J Neuroradiol, vol. 27, no. 6, pages 1165–76, Jun-Jul 2006. 67

[Gerig 2000] G Gerig, D Welti, C R Guttmann, A C Colchester et G Szekely. *Exploring the discrimination power of the time domain for segmentation and characterization of active lesions in serial MR data.* Med Image Anal, vol. 4, no. 1, pages 31–42, March 2000. 72, 81

[Gholipour 2007] Ali Gholipour, Nasser Kehtarnavaz, Richard Briggs, Michael Devous et Kaundinya Gopinath. *Brain functional localization : a survey of image registration techniques.* IEEE Trans Med Imaging, vol. 26, no. 4, pages 427–51, April 2007. 30

[Giugni 1997] E Giugni, C Pozzilli, S Bastianello, C Gasperini, A Paolillo, T Koudriavtseva, M Frontoni, D Farina et L Bozzao. *MRI measures and their*

relations with clinical disability in relapsing-remitting and secondary progressive multiple sclerosis. Mult Scler, vol. 3, no. 4, pages 221–5, August 1997. 19

[Godement 2005] Roger Godement. Cours d'algèbre. Hermann Éditeurs, 2005. 44

[Goldberg-Zimring 2005] Daniel Goldberg-Zimring, Andrea U J Mewes, AU Mewes, Mahnaz Maddah et Simon K Warfield. *Diffusion tensor magnetic resonance imaging in multiple sclerosis.* J Neuroimaging, vol. 15, no. 4 Suppl, page 68, 2005. 68

[González Ballester 1999] M. A. González Ballester. *Morphometric Analysis of Brain Structures in MRI.* PhD thesis, University of Oxford, 1999. 40

[Gonzalez-Toledo 2006] Eduardo Gonzalez-Toledo, Roger E Kelley et Alireza Minagar. *Role of magnetic resonance spectroscopy in diagnosis and management of multiple sclerosis.* Neurol Res, vol. 28, no. 3, pages 280–3, April 2006. 68

[Grimaud 1996] J Grimaud, M Lai, J Thorpe, P Adeleine, L Wang, G J Barker, D L Plummer, P S Tofts, W I McDonald et D H Miller. *Quantification of MRI lesion load in multiple sclerosis : a comparison of three computer-assisted techniques.* Magn Reson Imaging, vol. 14, no. 5, pages 495–505, 1996. 34, 67, 70, 77

[Grimaud 2001] J Grimaud, N Pageot et M Rovaris. *Parallels between clinical aspects and MRI in multiple sclerosis.* Rev Neurol (Paris), vol. 157, no. 8-9 Pt 1, pages 884–90, September 2001. 17, 67

[Gudbjartsson 1995] H Gudbjartsson et S Patz. *The Rician distribution of noisy MRI data.* Magnetic resonance in medicine, vol. 34, no. 6, pages 910–914, 1995. 37

[Guttmann 1999] C R Guttmann, R Kikinis, M C Anderson, M Jakab, S K Warfield, R J Killiany, H L Weiner et F A Jolesz. *Quantitative follow-up of patients with multiple sclerosis using MRI : reproducibility.* J Magn Reson Imaging, vol. 9, no. 4, pages 509–18, April 1999. 76

[Guttmann 2000] C R Guttmann, S.K. Warfield, A. Guimond, R. Kikinis, F.A. Jolesz et H. Weiner. *Discrimination of Relapsing and Progressive MS through MRI Estimates of Brain Parenchymal Fraction and Lesion Burden.* In 52nd American Academy of Neurology Annual Meeting, Avril 2000. 108

[Hajnal 2001] J. V. Hajnal, D. L. G. Hill et D. Hawkes, editeurs. Medical image registration. D. J. Hawkes, 2001. 30

[Hardmeier 2005] M Hardmeier, S Wagenpfeil, P Freitag, E Fisher, R A Rudick, M Kooijmans, M Clanet, E W Radue et L Kappos. *Rate of brain atrophy in relapsing MS decreases during treatment with IFN beta-1a.* Neurology, vol. 64, no. 2, pages 236–40, January 2005. 126

[Hartley 1958] H. O. Hartley. *Maximum Likelihood Estimation from Incomplete Data.* Biometrics, vol. 14, no. 2, pages 174–194, 1958. 36

[Hartley 2006] S W Hartley, A I Scher, ESC Korf, L R White et L J Launer. *Analysis and validation of automated skull stripping tools : a validation study based on 296 MR images from the Honolulu Asia aging study.* Neuroimage, vol. 30, no. 4, pages 1179–86, May 2006. 52, 56, 58

[He 2000] Renjie He et Ponnada A. Narayana. *Detection and Delineation of Multiple Sclerosis Lesions in Gadolinium-Enhanced 3D T1-Weighted MRI Data.* In Proceedings of the 13th IEEE Symposium on Computer-Based Medical Systems, page 201, Washington, DC, USA, 2000. IEEE Computer Society. 70, 71, 80

[He 2002] Renjie He et Ponnada A Narayana. *Automatic delineation of Gd enhancements on magnetic resonance images in multiple sclerosis.* Med Phys, vol. 29, no. 7, pages 1536–46, July 2002. 70, 71, 81, 82

[Hill 2001] D L Hill, P G Batchelor, M Holden et D J Hawkes. *Medical image registration.* Phys Med Biol, vol. 46, no. 3, Mars 2001. 29

[Hogben 2007] Leslie Hogben, editeur. Handbook of Linear Algebra. Discrete Mathematics and its Applications. Chapman & Hall/CRC, 2007. 44

[Horakova 2008] D Horakova, J L Cox, E Havrdova, S Hussein, O Dolezal, D Cookfair, M G Dwyer, Z Seidl, N Bergsland, M Vaneckova et R Zivadinov. *Evolution of different MRI measures in patients with active relapsing-remitting multiple sclerosis over 2 and 5 years : a case-control study.* J Neurol Neurosurg Psychiatry, vol. 79, no. 4, pages 407–14, April 2008. 108

[Horsfield 2007] Mark A Horsfield, Rohit Bakshi, Marco Rovaris, Mara A Rocca, Venkata S R Dandamudi, Paola Valsasina, Elda Judica, Fulvio Lucchini, Charles R G Guttmann, Maria Pia Sormani et Massimo Filippi. *Incorporating domain knowledge into the fuzzy connectedness framework : application to brain lesion volume estimation in multiple sclerosis.* IEEE Trans Med Imaging, vol. 26, no. 12, pages 1670–80, December 2007. 70, 71

[Hu 2007] Shiyan Hu et D Louis Collins. *Joint level-set shape modeling and appearance modeling for brain structure segmentation.* Neuroimage, vol. 36, no. 3, pages 672–83, July 2007. 35

[Inglese 2004] Matilde Inglese, Yulin Ge, Massimo Filippi, Andrea Falini, Robert I Grossman et Oded Gonen. *Indirect evidence for early widespread gray matter involvement in relapsing-remitting multiple sclerosis.* Neuroimage, vol. 21, no. 4, pages 1825–9, April 2004. 82

[Jaccard 1912] P. Jaccard. *The distribution of flora in the alpine zone.* New Phytologist, vol. 11, pages 37–50, 1912. 79

[Johnston 1996] B Johnston, MS Atkins, B Mackiewich et M Anderson. *Segmentation of multiple sclerosis lesions in intensity corrected multispectral MRI.* IEEE Trans Med Imaging, vol. 15, no. 2, pages 154–169, 1996. 76, 81

[Juengling 2003] Freimut D Juengling et Jan Kassubek. *Standardized calculation of brain parenchymal fraction : an approach to objective assessment of cerebral*

atrophy. AJNR Am J Neuroradiol, vol. 24, no. 7, pages 1492–3, August 2003. 108

[Kamber 1995] Micheline Kamber, Rajjan Shinghal, D.L. Collins, Gordon S. Francis et Alan C. Evans. *Model-Based 3D Segmentation of Multiple Sclerosis Lesions in Magnetic Resonance Brain Images.* IEEE Trans Med Imaging, vol. 14, no. 3, pages 442–453, September 1995. 73, 81, 86

[Kappos 2006] L Kappos, C H Polman, M S Freedman, G Edan, H P Hartung, D H Miller, X Montalban, F Barkhof, L Bauer, P Jakobs, C Pohl et R Sandbrink. *Treatment with interferon beta-1b delays conversion to clinically definite and McDonald MS in patients with clinically isolated syndromes.* Neurology, vol. 67, no. 7, pages 1242–9, October 2006. 19, 67

[Kapur 1996] T Kapur, WEL Grimson, WM Wells et R Kikinis. *Segmentation of brain tissue from magnetic resonance images.* Medical Image Analysis, vol. 1, pages 109–128, 1996. 55

[Karacali 2006] Bilge Karacali et Christos Davatzikos. *Simulation of tissue atrophy using a topology preserving transformation model.* IEEE Trans Med Imaging, vol. 25, no. 5, pages 649–52, May 2006. 117

[Kassubek 2003] Jan Kassubek, Hayrettin Tumani, Daniel Ecker, Anja Kurt, Albert C Ludolph et Freimut D Juengling. *Age-related brain parenchymal fraction is significantly decreased in young multiple sclerosis patients : a quantitative MRI study.* Neuroreport, vol. 14, no. 3, pages 427–30, March 2003. 108

[Kawa 2007] Jacek Kawa et Ewa Pietka. *Kernelized fuzzy c-means method in fast segmentation of demyelination plaques in multiple sclerosis.* Conf Proc IEEE Eng Med Biol Soc, vol. 2007, pages 5616–9, 2007. 36

[Kikinis 1999] R Kikinis, C R Guttmann, D Metcalf, WM Wells, G J Ettinger, H L Weiner et F A Jolesz. *Quantitative follow-up of patients with multiple sclerosis using MRI : technical aspects.* J Magn Reson Imaging, vol. 9, no. 4, pages 519–30, April 1999. 76, 77, 81, 82

[Kindermann 1980] R. Kindermann et J.L. Snell. Markov Random Fields and Their Applications. AMS Books, 1980. 36

[Korteweg 2006] Tijmen Korteweg, Mar Tintore, Bernard Uitdehaag, Alex Rovira, Jette Frederiksen, David Miller, Kryshani Fernando, Massimo Filippi, Federica Agosta, Maria Rocca, Franz Fazekas, Christian Enzinger, Paul Matthews, Allyson Parry, Chris Polman, Xavier Montalban et Frederik Barkhof. *MRI criteria for dissemination in space in patients with clinically isolated syndromes : a multicentre follow-up study.* Lancet Neurol, vol. 5, no. 3, pages 221–7, March 2006. 19, 67, 107

[Krishnan 1998] K Krishnan et MS Atkins. *Segmentation of multiple sclerosis lesions in MRI - An image analysis approach.* Proceedings of the SPIE-Medical Imaging 1998, vol. 3338, pages 1106–1116, February 1998. 75, 76, 77, 81

[Kurtzke 1983] J F Kurtzke. *Rating neurologic impairment in multiple sclerosis : an expanded disability status scale (EDSS).* Neurology, vol. 33, no. 11, pages 1444–52, November 1983. 10, 17

[Lauterbur 1973] P.C. Lauterbur. *Image formation by induced local interactions : Examples employing nuclear magnetic resonance.* Nature, vol. 242, pages 190–191, 1973. 12

[Learned-Miller 2005] E. G. Learned-Miller et P. Ahammad. *Joint MRI Bias Removal using Entropy Minimization Across Images.* Neural Information Processing System 17, pages 761–768, 2005. 32

[Lebrun 2004] C Lebrun, D Rey, S Chanalet, V Bourg, C Bensa, M Chatel, N Ayache et G Malandain. *Intérêt du recalage automatique des IRM cérébrales dans le suivi : Application à la sclérose en plaques.* Rev Neurol, vol. 160, no. 8-9, pages 805–10, 2004. 30

[Lecoeur 2007] J. Lecoeur et C. Barillot. *Segmentation d'images cérébrales : État de l'art.* Rapport technique 6306, INRIA, September 2007. 34

[Lecoeur 2008] J. Lecoeur, S. Morissey, J.-C. Ferré, D. Arnold, D. Collins et C. Barillot. *Multiple Sclerosis Lesions Segmentation using Spectral Gradient and Graph Cuts.* In Proceedings of MICCAI workshop on Medical Image Analysis on Multiple Sclerosis (validation and methodological issues), September 2008. 75, 81

[Leemput 2001] Koen Van Leemput, Frederik Maes, Dirk Vandermeulen et Paul Suetens. *A Statistical Framework for Partial Volume Segmentation.* Rapport technique, Katholieke Universiteit Leuven - Center for Processing Speech, April 2001. 36

[Lester 1999] Hava Lester et Simon R. Arridge. *A survey of hierarchical non-linear medical image registration.* Pattern Recognition, vol. 32, no. 1, pages 129–149, 1999. 30

[Li 2003] Lihong Li, Xiang Li, Hongbing Lu, Wei Huang, C Christodoulou, A Tudorica, LB Krupp et Zhengrong Liang. *MRI volumetric analysis of multiple sclerosis : methodology and validation.* IEEE Transactions on Nuclear Science, vol. 50, no. 2(5), pages 1686–1692, October 2003. 76, 81

[Li 2006] D K B Li, U Held, J Petkau, M Daumer, F Barkhof, F Fazekas, J A Frank, L Kappos, D H Miller, J H Simon, J S Wolinsky et M Filippi. *MRI T2 lesion burden in multiple sclerosis : a plateauing relationship with clinical disability.* Neurology, vol. 66, no. 9, pages 1384–9, May 2006. 19

[Liang 1999] Z.P. Liang et P.C. Lauterbur. Principles of Magnetic Resonance Imaging. Biomedical Engineering. IEEE press, 1999. 13

[Lin 2003] X Lin, L D Blumhardt et C S Constantinescu. *The relationship of brain and cervical cord volume to disability in clinical subtypes of multiple sclerosis : a three-dimensional MRI study.* Acta Neurol Scand, vol. 108, no. 6, pages 401–6, Décembre 2003. 20

[Lin 2008] X Lin, C R Tench, P S Morgan et C S Constantinescu. *Use of combined conventional and quantitative MRI to quantify pathology related to cognitive impairment in multiple sclerosis.* J Neurol Neurosurg Psychiatry, vol. 79, no. 4, pages 437–41, April 2008. 108

[Losseff 1996] N A Losseff, L Wang, H M Lai, D S Yoo, M L Gawne-Cain, W I McDonald, D H Miller et A J Thompson. *Progressive cerebral atrophy in multiple sclerosis. A serial MRI study.* Brain, vol. 119 (Pt 6), pages 2009–19, December 1996. 108

[Lublin 2005] Fred Lublin. *History of modern multiple sclerosis therapy.* J Neurol, vol. 252 Suppl 3, pages III/3–III/9, September 2005. 11

[Maintz 1998] J B Maintz et M A Viergever. *A survey of medical image registration.* Med Image Anal, vol. 2, no. 1, pages 1–36, Mars 1998. 30

[Mangin 1998] J.-F. Mangin, O. Coulon et V. Frouin. *Robust brain segmentation using histogram scale-space analysis and mathematical morphology.* In W. M. Wells, A. Colchester et S. Delp, editeurs, Proc. 1st MICCAI, LNCS-1496, pages 1230–1241, MIT, Boston, Oct. 1998. Springer Verlag. 51

[Mangin 2000] J. F. Mangin. *Entropy minimization for automatic correction of intensity non uniformity.* In MMBIA (Math. Methods in Biomed. Image Analysis), pages 162–169. IEEE Press, 2000. 32

[Mazziotta 2001] J Mazziotta, A Toga, A Evans, P Fox, J Lancaster, K Zilles, R Woods, T Paus, G Simpson, B Pike, C Holmes, L Collins, P Thompson, D MacDonald, M Iacoboni, T Schormann, K Amunts, N Palomero-Gallagher, S Geyer, L Parsons, K Narr, N Kabani, G Le Goualher, D Boomsma, T Cannon, R Kawashima et B Mazoyer. *A probabilistic atlas and reference system for the human brain : International Consortium for Brain Mapping (ICBM).* Philos Trans R Soc Lond B Biol Sci, vol. 356, no. 1412, pages 1293–322, August 2001. 108

[McDonald 2001] W I McDonald, A Compston, G Edan, D Goodkin, H P Hartung, F D Lublin, H F McFarland, D W Paty, C H Polman, S C Reingold, M Sandberg-Wollheim, W Sibley, A Thompson, S van den Noort, B Y Weinshenker et J S Wolinsky. *Recommended diagnostic criteria for multiple sclerosis : guidelines from the International Panel on the diagnosis of multiple sclerosis.* Ann Neurol, vol. 50, no. 1, pages 121–7, July 2001. 16, 67

[McFarland 2002] H F McFarland, F Barkhof, J Antel et D H Miller. *The role of MRI as a surrogate outcome measure in multiple sclerosis.* Mult Scler, vol. 8, no. 1, pages 40–51, February 2002. 138

[Meier 2003] Dominik S Meier et Charles R Guttmann. *Time-series analysis of MRI intensity patterns in multiple sclerosis.* Neuroimage, vol. 20, no. 2, pages 1193–209, October 2003. 72

[Meier 2006] Dominik S Meier et Charles R Guttmann. *MRI time series modeling of MS lesion development.* Neuroimage, vol. 32, no. 2, pages 531–7, August 2006. 72

[Miller 2002] David H Miller, Frederik Barkhof, Joseph A Frank, Geoffrey J M Parker, GJ Parker et Alan J Thompson. *Measurement of atrophy in multiple sclerosis : pathological basis, methodological aspects and clinical relevance.* Brain, vol. 125, no. Pt 8, pages 1676–95, August 2002. 107

[Milles 2007] Julien Milles, Yue Min Zhu, Gerard Gimenez, CRG Guttmann et Isabelle E Magnin. *MRI intensity nonuniformity correction using simultaneously spatial and gray-level histogram information.* Comput Med Imaging Graph, vol. 31, no. 2, pages 81–90, March 2007. 32

[Molyneux 1998] PD Molyneux, L Wang, M Lai, G J, PS Tofts, IF Moseley et D H. *Quantitative techniques for lesion load measurement in multiple sclerosis : an assessment of the global threshold technique after non uniformity and histogram matching corrections.* Eur J Neurol, vol. 5, no. 1, pages 55–60, January 1998. 32

[Montagnat 2008] Johan Montagnat, Alban Gaignard, Diane Lingrand, Javier Rojas Balderrama, Philippe Collet et Philippe Lahire. *NeuroLOG : a community-driven middleware design.* In HealthGrid, Chicago, Juin 2008. IOS Press. 158, 197

[Narayana 1988] P. A. Narayana, W. W. Brey, M. V. Kulkarni et C. L. Sievenpiper. *Compensation for surface coil sensitivity variation in magnetic resonance imaging.* Magn Reson Imaging, vol. 6, no. 3, pages 271–4, May-Jun 1988. 30

[Narayana 2005] Ponnada A Narayana. *Magnetic resonance spectroscopy in the monitoring of multiple sclerosis.* J Neuroimaging, vol. 15, no. 4 Suppl, page 46, 2005. 68

[Niessen 1997] W. J. Niessen. *Multiscale Medical Image Analysis.* PhD thesis, University of Utrecht, 1997. 40

[Nijeholt 1998] G J Nijeholt, M A van Walderveen, J A Castelijns, J H van Waesberghe, C Polman, P Scheltens, P F Rosier, P J Jongen et F Barkhof. *Brain and spinal cord abnormalities in multiple sclerosis. Correlation between MRI parameters, clinical subtypes and symptoms.* Brain, vol. 121 (Pt 4), pages 687–97, April 1998. 19, 67

[Novellas 2007] Sébastien Novellas. Segmentation semi-automatique des lymphomes médiastinaux en tomodensitométrie. Rapport de master, Master IGMMV, Université de Nice - Sophia-Antipolis, 2007. 83

[O'Riordan 1998] J I O'Riordan, M Gawne Cain, A Coles, L Wang, D A Compston, P Tofts et D H Miller. *T1 hypointense lesion load in secondary progressive multiple sclerosis : a comparison of pre versus post contrast loads and of manual versus semi automated threshold techniques for lesion segmentation.* Mult Scler, vol. 4, no. 5, pages 408–12, October 1998. 17, 19

[Ourselin 2000] S. Ourselin, A. Roche, S. Prima et N. Ayache. *Block Matching : A General Framework to Improve Robustness of Rigid Registration of Medical Images.* In Proc. of MICCAI'00, pages 557–566. Springer, 2000. 29, 30, 47, 199

[Parodi 2002] Roberto C Parodi, Francesco Sardanelli, Paolo Renzetti, Elisabetta Rosso, Caterina Losacco, Alessandra Ferrari, Fabrizio Levrero, Alberto Pilot, Matilde Inglese et Giovanni L Mancardi. *Growing Region Segmentation Software (GRES) for quantitative magnetic resonance imaging of multiple sclerosis : intra- and inter-observer agreement variability : a comparison with manual contouring method.* Eur Radiol, vol. 12, no. 4, pages 866–71, April 2002. 77

[Paty 2001] D W Paty et D K Li. *Interferon beta-1b is effective in relapsing-remitting multiple sclerosis. II. MRI analysis results of a multicenter, randomized, double-blind, placebo-controlled trial.* Neurology, vol. 57, no. 12 Suppl 5, December 2001. 68

[Pernod 2008] Erik Pernod. Interface development for medical imaging applications in the framework of the multiple sclerosis disease. Rapport de master, Institut des Sciences de l'Ingénieur de Toulon et du Var, 2008. 41, 198, 201

[Pham 1999] D. L. Pham et J. L. Prince. *Adaptive fuzzy segmentation of magnetic resonance images.* IEEE Trans Med Imaging, vol. 18, no. 9, pages 737–52, September 1999. 36

[Pham 2000] D. L. Pham, C. Xu et J. L. Prince. *Current methods in medical image segmentation.* Annu Rev Biomed Eng, vol. 2, pages 315–37, 2000. 34

[Pitiot 2003] A. Pitiot. *Segmentation automatique des structures cérébrales s'appuyant sur des connaissances explicites.* Thèse de sciences, École des mines de Paris, November 2003. 35

[Polman 2005] C.H. Polman, S.C. Reingold, G. Edan, M. Filippi, H-P. Hartung, L. Kappos, F.D. Lublin, L.M. Metz, H.F. McFarland, P.W. O'Connor, M. Sandberg-Wollheim, A.J. Thompson, B.G. Weinshenker et J.S. Wolinsky. *Diagnostic criteria for multiple sclerosis : 2005 revisions to the "Mc Donald Criteria".* Ann Neurol, vol. 58, no. 6, pages 840–6, 2005. 17, 67, 107

[Poser 1983] C M Poser, D W Paty, L Scheinberg, W I McDonald, F A Davis, G C Ebers, K P Johnson, W A Sibley, D H Silberberg et W W Tourtellotte. *New diagnostic criteria for multiple sclerosis : guidelines for research protocols.* Ann Neurol, vol. 13, no. 3, pages 227–31, March 1983. 16

[Prima 2001] S. Prima, N. Ayache, Tom Barrick et Neil Roberts. *Maximum Likelihood Estimation of the Bias Field in MR Brain Images : Investigating Different Modelings of the Imaging Process.* In Proc. of MICCAI'01, pages 811–819. Springer LNCS 2208, 2001. 32, 47, 201

[Prima 2002] S. Prima, N. Ayache, A. Janke, S. Francis, D. Arnold et L. Collins. *Statistical Analysis of Longitudinal MRI Data : Application for detection of Disease Activity in MS.* In Takeyoshi Dohi et Ron Kikinis, editeurs, Medical Image Computing and Computer-Assisted Intervention (MICCAI'02), volume 2488 of *LNCS*, pages 363–371, Tokyo, September 2002. Springer. 72

[Prima 2003] S. Prima, D.L. Arnold et D.L. Collins. *Multivariate statistics for detection of MS activity in serial multimodal MR images.* In R.E. Ellis et T.M.

Peters, editeurs, 6th International Conference on Medical Image Computing and Computer-Assisted Intervention, MICCAI'2003, volume 2878 of *LNCS*, pages 663–670, Montreal, Canada, November 2003. Springer. 32, 72, 81

[Quarantelli 2003] Mario Quarantelli, Andrea Ciarmiello, Vincenzo Brescia Morra, Giuseppe Orefice, Michele Larobina, Roberta Lanzillo, Vittorio Schiavone, Elena Salvatore, Bruno Alfano et Arturo Brunetti. *Brain tissue volume changes in relapsing-remitting multiple sclerosis : correlation with lesion load.* Neuroimage, vol. 18, no. 2, pages 360–6, February 2003. 108

[Rashid 2007] W Rashid, G R Davies, D T Chard, C M Griffin, D R Altmann, A J Thompson et D H Miller. *Relationship of triple dose contrast enhanced lesions with clinical measures and brain atrophy in early relapsing-remitting multiple sclerosis : a two-year longitudinal study.* Mult Scler, vol. 13, no. 2, pages 178–85, March 2007. 19

[Raudys 1998] S Raudys et R P W Duin. *Expected classification error of the Fisher linear classifier with pseudo-inverse covariance matrix.* Pattern Recogn. Lett., vol. 19, no. 5-6, pages 385–392, 1998. 44

[Rey 2001a] D. Rey, J. Stoeckel, G. Malandain et N. Ayache. *Using SPM to Detect Evolving MS Lesions.* In 4th Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI'01), volume 2208 of *LNCS*, pages 1232–1234, Utrecht, The Netherlands, October 2001. 72

[Rey 2001b] David Rey, Jonathan Stoeckel, Grégoire Malandain et Nicholas Ayache. *A Spatio-temporal Model-based Statistical Approach to Detect Evolving Multiple Sclerosis Lesions.* In IEEE Workshop on Mathematical Methods in Biomedical Image Analysis, MMBIA'01, Kauia, Hawai, USA, December 2001. 72

[Rey 2002a] D. Rey. *Détection et quantification automatiques de processus évolutifs dans les images médicales tridimensionnelles : application à la sclérose en plaques.* PhD thesis, University of Sophia-Antipolis, october 2002. 3, 32

[Rey 2002b] David Rey, Gérard Subsol, Hervé Delingette et Nicholas Ayache. *Automatic Detection and Segmentation of Evolving Processes in 3D Medical Images : Application to Multiple Sclerosis.* Medical Image Analysis, vol. 6, no. 2, pages 163–179, June 2002. 72, 81

[Rio 2008] J Rio, A Rovira, M Tintore, E Huerga, C Nos, N Tellez, C Tur, M Comabella et X Montalban. *Relationship between MRI lesion activity and response to IFN-beta in relapsing-remitting multiple sclerosis patients.* Mult Scler, vol. 14, no. 4, pages 479–484, May 2008. 19

[Roche 2000] A. Roche, G. Malandain et N. Ayache. *Unifying Maximum Likelihood Approaches in Medical Image Registration.* International Journal of Imaging Systems and Technology : Special Issue on 3D Imaging, vol. 11, no. 1, pages 71–80, 2000. 29

[Roche 2003] Alexis Roche. EM algorithm and variants : an informal tutorial. ftp ://ftp.cea.fr/pub/dsv/madic/publis/Roche_em.pdf, 2003. 37

[Rousson 2004] M. Rousson, N. Paragios et R. Deriche. *Implicit Active Shape Models for 3D Segmentation in MR Imaging.* In MICCAI (1), volume 3216 of *Lecture Notes in Computer Science*, pages 209–216. Springer, 2004. 35

[Rovaris 1997] M Rovaris, M Filippi, G Calori, M Rodegher, A Campi, B Colombo et G Comi. *Intra-observer reproducibility in measuring new putative MR markers of demyelination and axonal loss in multiple sclerosis : a comparison with conventional T2-weighted images.* J Neurol, vol. 244, no. 4, pages 266–70, April 1997. 34

[Rudick 1999] R A Rudick, E Fisher, J C Lee, J Simon et L Jacobs. *Use of the brain parenchymal fraction to measure whole brain atrophy in relapsing-remitting MS. Multiple Sclerosis Collaborative Research Group.* Neurology, vol. 53, no. 8, pages 1698–704, November 1999. 108

[Rudick 2000] R A Rudick, E Fisher, J C Lee, J T Duda et J Simon. *Brain atrophy in relapsing multiple sclerosis : relationship to relapses, EDSS, and treatment with interferon beta-1a.* Mult Scler, vol. 6, no. 6, pages 365–72, December 2000. 20

[Rudick 2002] R A Rudick, G Cutter et S Reingold. *The multiple sclerosis functional composite : a new clinical outcome measure for multiple sderosis trials.* Mult Scler, vol. 8, no. 5, pages 359–65, October 2002. 20, 153

[Saint-Jean 2001] Christophe Saint-Jean. *Classification paramétrique robuste partiellement supervisée en reconnaissance des formes.* Thèse de sciences, Université de la Rochelle, dec 2001. 39

[Sajja 2006] Balasrinivasa Rao Sajja, Sushmita Datta, Renjie He, Meghana Mehta, Rakesh K Gupta, Jerry S Wolinsky et Ponnada A Narayana. *Unified approach for multiple sclerosis lesion segmentation on brain MRI.* Ann Biomed Eng, vol. 34, no. 1, pages 142–51, January 2006. 71, 76, 81

[Samarasekera 1997] S Samarasekera, J K Udupa, Y Miki, L Wei et R I Grossman. *A new computer-assisted method for the quantification of enhancing lesions in multiple sclerosis.* J Comput Assist Tomogr, vol. 21, no. 1, pages 145–51, Jan-Feb 1997. 70, 71, 80, 81, 82

[Sandor 1991] T Sandor, D Metcalf et Y J Kim. *Segmentation of brain CT images using the concept of region growing.* Int J Biomed Comput, vol. 29, no. 2, pages 133–47, November 1991. 34

[Sandor 1997] S Sandor et R Leahy. *Surface-based labeling of cortical anatomy using a deformable atlas.* IEEE Trans Med Imaging, vol. 16, no. 1, pages 41–54, February 1997. 33, 51

[Segonne 2004] F Segonne, A M Dale, E Busa, M Glessner, D Salat, H K Hahn et B Fischl. *A hybrid approach to the skull stripping problem in MRI.* Neuroimage, vol. 22, no. 3, pages 1060–75, July 2004. 33, 51, 53

[Shahar 2004] A. Shahar et H. Greenspan. *Probabilistic Spatial-Temporal Segmentation of Multiple Sclerosis Lesions.* In Proceedings to the Computer Vision

Approaches to Medical Image Analysis Workshop (CVAMIA), Prague, 2004. 70, 71, 81, 82

[Sharma 2004] Jitendra Sharma, Michael P Sanfilipo, RH Benedict, Bianca Weinstock-Guttman, FE Munschauer et Rohit Bakshi. *Whole-brain atrophy in multiple sclerosis measured by automated versus semiautomated MR imaging segmentation.* AJNR Am J Neuroradiol, vol. 25, no. 6, pages 985–96, Jun-Jul 2004. 108, 125

[Sharma 2008] S. Sharma, V. Noblet, F. Rousseau, F. Heitz, L. Rumbach et J.-P. Armspach. *Evaluation of Brain Atrophy Estimation Algorithms using Simulated Ground-Truth Data.* In MICCAI 2008 - Workshop on Medical Image Analysis on Multiple Sclerosis, 2008. 117, 121

[Shattuck 2001] D W Shattuck, S R Sandor-Leahy, K A Schaper, D A Rottenberg et R M Leahy. *Magnetic resonance image tissue classification using a partial volume model.* Neuroimage, vol. 13, no. 5, pages 856–76, May 2001. 33, 51

[Sijbers 1998] J. Sijbers, A. J. den Dekker, P. Scheunders et D. Van Dyck. *Maximum-likelihood estimation of Rician distribution parameters.* IEEE Trans Med Imaging, vol. 17, no. 3, pages 357–61, June 1998. 37

[Simon 1987] J H Simon, R B Schiffer, R A Rudick et R M Herndon. *Quantitative determination of MS-induced corpus callosum atrophy in vivo using MR imaging.* AJNR Am J Neuroradiol, vol. 8, no. 4, pages 599–604, Jul-Aug 1987. 108

[Simon 1998] J H Simon, L D Jacobs, M Campion, K Wende, N Simonian, D L Cookfair, R A Rudick, R M Herndon, J R Richert, A M Salazar, J J Alam, J S Fischer, D E Goodkin, C V Granger, M Lajaunie, A L Martens-Davidson, M Meyer, J Sheeder, K Choi, A L Scherzinger, D M Bartoszak, D N Bourdette, J Braiman, C M Brownscheidle et R H Whitham. *Magnetic resonance studies of intramuscular interferon beta-1a for relapsing multiple sclerosis. The Multiple Sclerosis Collaborative Research Group.* Ann Neurol, vol. 43, no. 1, pages 79–87, January 1998. 17, 19

[Simon 1999] J H Simon, L D Jacobs, M K Campion, R A Rudick, D L Cookfair, R M Herndon, J R Richert, A M Salazar, J S Fischer, D E Goodkin, N Simonian, M Lajaunie, D E Miller, K Wende, A Martens-Davidson, R P Kinkel, FE Munschauer et C M Brownscheidle. *A longitudinal study of brain atrophy in relapsing multiple sclerosis. The Multiple Sclerosis Collaborative Research Group (MSCRG).* Neurology, vol. 53, no. 1, pages 139–48, July 1999. 108

[Simon 2006] Jack H Simon. *Update on multiple sclerosis.* Magn Reson Imaging Clin N Am, vol. 14, no. 2, pages 203–24, May 2006. 19

[Sled 1998a] J. G. Sled et G. B. Pike. *Standing-wave and RF penetration artifacts caused by elliptic geometry : an electrodynamic analysis of MRI.* IEEE Trans. Med. Imaging, vol. 17, no. 4, pages 653–662, 1998. 32

[Sled 1998b] J. G. Sled, A. P. Zijdenbos et A. C. Evans. *A Nonparametric Method for Automatic Correction of Intensity Nonuniformity in MRI Data.* IEEE Trans Med Imaging, vol. 17, no. 1, pages 87–97, 1998. 32

[Smith 2002a] S.M. Smith, Y. Zhang, M. Jenkinson, J. Chen, P.M. Matthews, A. Federico et N. De Stefano. *Accurate, robust, and automated longitudinal and cross-sectional brain change analysis.* Neuroimage, vol. 17, no. 1, pages 479–89, 2002. 108, 110, 126

[Smith 2002b] Stephen M Smith. *Fast robust automated brain extraction.* Hum Brain Mapp, vol. 17, no. 3, pages 143–55, November 2002. 33, 52, 53

[Smith 2004] Stephen M Smith, Mark Jenkinson, Mark W Woolrich, Christian F Beckmann, Timothy E J Behrens, TE Behrens, Heidi Johansen-Berg, Peter R Bannister, Marilena De Luca, Ivana Drobnjak, David E Flitney, Rami K Niazy, James Saunders, John Vickers, Yongyue Zhang, Nicola De Stefano, J Michael Brady et Paul M Matthews. *Advances in functional and structural MR image analysis and implementation as FSL.* Neuroimage, vol. 23 Suppl 1, 2004. 117

[Song 2007] Tao Song, Mo M Jamshidi, Roland R Lee et Mingxiong Huang. *A modified probabilistic neural network for partial volume segmentation in brain MR image.* IEEE Trans Neural Netw, vol. 18, no. 5, pages 1424–32, September 2007. 35

[Spearman 1904] C Spearman. *The Proof and Measurement of Association between Two Things.* The American Journal of Psychology, vol. 15, pages 72–101, 1904. 119

[Stefano 2007] N. De Stefano, M. Battaglini et S.M. Smith. *Measuring brain atrophy in multiple sclerosis.* J Neuroimaging, vol. 17 Suppl 1, page 10, 2007. 107, 108, 113

[Studholme 2006] Colin Studholme, Corina Drapaca, Bistra Iordanova et Valerie Cardenas. *Deformation-based mapping of volume change from serial brain MRI in the presence of local tissue contrast change.* IEEE Trans Med Imaging, vol. 25, no. 5, pages 626–39, May 2006. 117

[Styner 2000] M Styner, C Brechbuhler, G Szekely et G Gerig. *Parametric estimate of intensity inhomogeneities applied to MRI.* IEEE Trans Med Imaging, vol. 19, no. 3, pages 153–65, March 2000. 32

[Suri 2002a] J. Suri, S. Singh et L. Reden. *Computer Vision and Pattern Recognition Techniques for 2-D and 3-D MR Cerebral Cortical Segmentation (Part I) : A State-of-the-Art Review.* Pattern Analysis & Applications, vol. 5, no. 1, pages 46–76, May 2002. 34

[Suri 2002b] J. Suri, S. Singh et L. Reden. *Computer Vision and Pattern Recognition Techniques for 2-D and 3-D MR Cerebral Cortical Segmentation (Part II) : A State-of-the-Art Review.* Pattern Analysis & Applications, vol. 5, no. 1, pages 77–98, May 2002. 34

[Thirion 1999] J.-P. Thirion et G. Calmon. *Deformation Analysis to Detect and Quantify Active Lesions in Three-Dimensional Medical Image Sequences.* IEEE Transactions on Medical Imaging, vol. 18, no. 5, pages 429–441, 1999. 72, 81

[Tintore 2000] M Tintore, A Rovira, M J Martinez, J Rio, P Diaz-Villoslada, L Brieva, C Borras, E Grive, J Capellades et X Montalban. *Isolated demyelinating syndromes : comparison of different MR imaging criteria to predict conversion to clinically definite multiple sclerosis.* AJNR Am J Neuroradiol, vol. 21, no. 4, pages 702–6, April 2000. 16, 67

[Tobie 2007] Mathieu Tobie, François Rousseau, Yvan Narboux, Lucien Rumbach, Fabrice Heitz et Jean-Paul Armspach. *Quantitative Evaluation of Brain Atrophy measurement using SIENA in Multiple Sclerosis.* In 16th Meeting of ARSEP, Paris, France, Mars 2007. 126

[Tourbah 2001] A. Tourbah. *IRM et sclérose en plaques.* Neurologies, vol. 4, no. 33, pages 270–274, September 2001. 17

[Truyen 1996] L Truyen, J H van Waesberghe, M A van Walderveen, B W van Oosten, C H Polman, O R Hommes, H J Ader et F Barkhof. *Accumulation of hypointense lesions ("black holes") on T1 spin-echo MRI correlates with disease progression in multiple sclerosis.* Neurology, vol. 47, no. 6, pages 1469–76, December 1996. 19

[Turner 2001] B Turner, N Ramli, L D Blumhardt et T Jaspan. *Ventricular enlargement in multiple sclerosis : a comparison of three-dimensional and linear MRI estimates.* Neuroradiology, vol. 43, no. 8, pages 608–14, August 2001. 108, 125

[Udupa 1997] J.K. Udupa, L. Wei, S. Samarasekera, Y. Miki, M.A. van Buchem et R.I. Grossman. *Multiple Sclerosis Lesion Quantification Using Fuzzy-Connectedness Principles.* IEEE Trans Med Imaging, vol. 16, no. 5, pages 598–609, 1997. 73, 74, 81

[Udupa 2006] JK Udupa, VR Leblanc, Y Zhuge, C Imielinska, H Schmidt, LM Currie, BE Hirsch et J Woodburn. *A framework for evaluating image segmentation algorithms.* Computerized Medical Imaging and Graphics, vol. 20, no. 1, pages 75–87, 2006. 78

[Van den Elsen 1993] P.A. Van den Elsen, E.J.D. Pol et M.A. Viergever. *Medical Image Matching - A Review with Classification.* IEEE Engineering in Medicine and Biology, vol. 12, no. 4, pages 26–39, Mars 1993. 30

[Van Leemput 1999a] K Van Leemput, F Maes, D Vandermeulen et P Suetens. *Automated model-based bias field correction of MR images of the brain.* IEEE Trans Med Imaging, vol. 18, no. 10, pages 885–96, October 1999. 32, 36, 74

[Van Leemput 1999b] K Van Leemput, F Maes, D Vandermeulen et P Suetens. *Automated model-based tissue classification of MR images of the brain.* IEEE Trans Med Imaging, vol. 18, no. 10, pages 897–908, October 1999. 36, 39, 74

[Van Leemput 2001] K Van Leemput, F Maes, D Vandermeulen, A Colchester et P Suetens. *Automated segmentation of multiple sclerosis lesions by model outlier detection.* IEEE Trans Med Imaging, vol. 20, no. 8, pages 677–88, August 2001. 36, 45, 74, 81

[van Walderveen 1999] M A van Walderveen, L Truyen, B W van Oosten, J A Castelijns, G J Lycklama a Nijeholt, J H van Waesberghe, C Polman et F Barkhof. *Development of hypointense lesions on T1-weighted spin-echo magnetic resonance images in multiple sclerosis : relation to inflammatory activity.* Arch Neurol, vol. 56, no. 3, pages 345–51, March 1999. 19

[van Walderveen 2001] M A van Walderveen, G J Lycklama A Nijeholt, H J Ader, P J Jongen, C H Polman, J A Castelijns et F Barkhof. *Hypointense lesions on T1-weighted spin-echo magnetic resonance imaging : relation to clinical characteristics in subgroups of patients with multiple sclerosis.* Arch Neurol, vol. 58, no. 1, pages 76–81, January 2001. 19, 67

[Wang 1998a] L Wang, H M Lai, G J Barker, D H Miller et P S Tofts. *Correction for variations in MRI scanner sensitivity in brain studies with histogram matching.* Magn Reson Med, vol. 39, no. 2, pages 322–7, February 1998. 32

[Wang 1998b] Y Wang, T Adaly, SY Kung et Z Szabo. *Quantification and Segmentation of Brain Tissues from MR Images : A Probabilistic Neural Network Approach.* IEEE Trans Image Process, vol. 7, no. 8, pages 1165–1181, August 1998. 35

[Ward 1999] B. Douglas Ward. *Intracranial Segmentation.* Rapport technique, Biophysics Research Institute Medical College of Wisconsin, June 1999. 33, 51, 54

[Warfield 1995] S Warfield, J Dengler, J Zaers, CRG Guttmann, WM Wells, GJ Ettinger, J Hiller et R Kikinis. *Automatic Identification of Grey Matter Structures from MRI to Improve the Segmentation of White matter Lesions.* Jounal of Image Guided Surgery, vol. 1, no. 6, pages 326–338, 1995. 73, 74, 81, 82, 86

[Warfield 2000] S K Warfield, M Kaus, F A Jolesz et R Kikinis. *Adaptive, template moderated, spatially varying statistical classification.* Med Image Anal, vol. 4, no. 1, pages 43–55, March 2000. 76

[Warfield 2004] Simon K Warfield, Kelly H Zou et William M Wells. *Simultaneous truth and performance level estimation (STAPLE) : an algorithm for the validation of image segmentation.* IEEE Transactions on Medical Imaging, vol. 23, no. 7, pages 903–921, July 2004. 4, 33, 56, 77, 79, 160

[Wei 2002] X Wei, SK Warfield, KH Zou, Y Wu, X Li, A Guimond, JP Mugler, RR Benson, L Wolfson, HL Weiner et CR Guttmann. *Quantitative analysis of MRI signal abnormalities of brain white matter with high reproducibility and accuracy.* J Magn Reson Imaging, vol. 15, no. 2, pages 203–9, February 2002. 76, 77, 81

[Wells 1995] W. M. Wells, W. E. L. Grimson, R. Kikinis et F. A. Jolesz. *Adaptive Segmentation of MRI Data.* In CVRMed '95 : Proceedings of the First International Conference on Computer Vision, Virtual Reality and Robotics in Medicine, volume 905 of *Lecture Notes in Computer Science*, pages 59–69, London, UK, 1995. Springer-Verlag. 36

[Welti 2001] D Welti, G Gerig, Ernst-Wilhelm Radu, Ludwig Kappos et G Szekely. *Spatio-temporal Segmentation of Active Multiple Sclerosis Lesions in Serial MRI Data.* In 17th International Conference, IPMI 2001 Proceedings, volume 2082 of *LNCS*, page 438, CA, USA, June 2001. Springer. 72, 81

[Wu 2006] Y Wu, SK Warfield, IL Tan, WM Wells, DS Meier, RA van Schijndel, F Barkhof et CR Guttmann. *Automated segmentation of multiple sclerosis lesion subtypes with multichannel MRI.* Neuroimage, vol. 32, no. 3, pages 1205–15, September 2006. 35, 76, 80, 81

[Yang 2004] F Yang, T Jiang, W Zhu et F Kruggel. *White Matter Lesion Segmentation from Volumetric MR Images.* Medical Imaging and Augmented Reality Springer, pages 113–120, 2004. 73, 74, 81

[Yeo 2008] Boon Thye Thomas Yeo, Tom Vercauteren, Pierre Fillard, Xavier Pennec, Polina Golland, Nicholas Ayache et Olivier Clatz. *DTI Registration with Exact Finite-Strain Differential.* In Proceedings of the IEEE International Symposium on Biomedical Imaging : From Nano to Macro (ISBI'08), Paris, France, Mai 2008. IEEE. 30

[Yu 1999] O Yu, Y Mauss, G Zollner, I J Namer et J Chambron. *Distinct patterns of active and non-active plaques using texture analysis on brain NMR images in multiple sclerosis patients : preliminary results.* Magn Reson Imaging, vol. 17, no. 9, pages 1261–7, November 1999. 80

[Zaidi 2006] Habib Zaidi. *Medical image segmentation : quo vadis.* Comput Methods Programs Biomed, vol. 84, no. 2-3, pages 63–5, December 2006. 34

[Zhang 2001] Y Zhang, M Brady et S Smith. *Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm.* IEEE Trans Med Imaging, vol. 20, no. 1, pages 45–57, January 2001. 36, 76, 108, 110

[Zhu 2003] H. Zhu et O. Basir. *Automated brain tissue segmentation and MS lesion detection using fuzzy and evidential reasoning.* In 10th IEEE Int. Conf. Electronics, Circuits and Systems, 2003. 73, 81

[Zhuang 2006] Audrey H Zhuang, Daniel J Valentino et Arthur W Toga. *Skull-stripping magnetic resonance brain images using a model-based level set.* Neuroimage, vol. 32, no. 1, pages 79–92, August 2006. 33, 51

[Zijdenbos 1994] A.P. Zijdenbos, B.M. Dawant, R.A. Margolin et A.C. Palmer. *Morphometric analysis of white matter lesions in MR images : method and validation.* IEEE Trans Med Imaging, vol. 13, no. 4, pages 716–724, December 1994. 75, 76, 79

[Zijdenbos 1998] Alex Zijdenbos, Reza Forghani et Alan Evans. *Automatic Quantification of MS Lesions in 3D MRI Brain Data Sets : Validation of INSECT*. In MICCAI'1998, volume 1496 of *LNCS*, pages 439–448, Boston USA, October 1998. Springer. 67, 75, 77

[Zijdenbos 2002] Alex P Zijdenbos, Reza Forghani et Alan C Evans. *Automatic pipeline analysis of 3-D MRI data for clinical trials : application to multiple sclerosis*. IEEE Trans Med Imaging, vol. 21, no. 10, pages 1280–91, October 2002. 75, 81

[Zitova 2003] Barbara Zitova et Jan Flusser. *Image registration methods : a survey*. Image and Vision Computing, vol. 21, no. 11, pages 977–1000, October 2003. 30

[Zivadinov 2004] Robert Zivadinov et Rohit Bakshi. *Role of MRI in multiple sclerosis I : inflammation and lesions*. Front Biosci, vol. 9, pages 665–83, January 2004. 20

[Zivadinov 2005] Robert Zivadinov, Attilio Grop, Jitendra Sharma, Alessio Bratina, Christopher W Tjoa, Michael Dwyer et Marino Zorzon. *Reproducibility and accuracy of quantitative magnetic resonance imaging techniques of whole-brain atrophy measurement in multiple sclerosis*. J Neuroimaging, vol. 15, no. 1, pages 27–36, January 2005. 110

[Zou 2004a] KH Zou, SK Warfield, A Bharatha, CM Tempany, MR Kaus, SJ Haker, WM Wells, FA Jolesz et R Kikinis. *Statistical validation of image segmentation quality based on a spatial overlap index*. Acad Radiol, vol. 11, no. 2, pages 178–89, February 2004. 79

[Zou 2004b] KH Zou, WM Wells, R Kikinis et SK Warfield. *Three validation metrics for automated probabilistic image segmentation of brain tumours*. Stat Med, vol. 23, no. 8, pages 1259–82, April 2004. 79

Résumé : L'objectif de cette thèse est d'analyser une base de données d'images de résonance magnétique (IRM) cérébrales de patients atteints de sclérose en plaques (SEP). Dans ce but, deux marqueurs IRM ont été choisis (la charge lésionnelle et l'atrophie). Ceux-ci peuvent être évalués manuellement par le clinicien. Cependant, ces mesures sont fastidieuses et sujettes à une grande variabilité intra- et inter-experts. L'enjeu de cette thèse est donc d'automatiser ces mesures.

Cette automatisation commence par l'amélioration et/ou la proposition de nouveaux algorithmes semi-automatiques. Ceux-ci permettent d'obtenir plus rapidement et de manière contrôlée par l'expert des mesures de référence auxquelles pourront être confrontés les résultats de méthodes automatiques. Concernant ces dernières, de nouvelles méthodes ont été proposées. Leurs réalisations se décomposent en trois parties.

Tout d'abord, une chaîne de traitement visant à segmenter les différents compartiments (substance blanche, substance grise et liquide céphalo-rachidien) a été mise en place. Ensuite, les segmentations produites par cette chaîne ont été utilisées pour calculer automatiquement un seuil à appliquer sur une séquence IRM pour obtenir une segmentation des lésions. Cette approche est reprise de la littérature mais deux nouvelles versions en sont ici proposées. La première segmente les lésions du cortex, la seconde celles du cervelet. Enfin, les segmentations obtenues par la chaîne de traitement sont utilisées pour calculer la fraction parenchymateuse cérébrale et ainsi évaluer l'atrophie. Pour ce calcul, deux versions de l'algorithme sont proposées. Contrairement à la première, la seconde prend en compte les informations longitudinales dans le calcul. Ces méthodes ont été comparées aux autres méthodes automatiques existantes et disponibles sur internet ainsi qu'aux mesures de référence obtenues par un expert.

À partir de ces travaux, la base de données à étudier est ensuite analysée selon trois axes : statique, dynamique et par sujet. Les algorithmes développés ont été implémentés dans un nouveau logiciel dédié à l'analyse de patients atteints de SEP : SepINRIA.

Mots clefs : Imagerie médicale, Analyse d'image, Sclérose en Plaques, IRM, Segmentation, Atrophie, Biomarqueurs.

Abstract : The aim of this PhD thesis is to analyse a database of multiple sclerosis (MS) brain magnetic resonance images (MRI). For this purpose, two biomarkers have been selected (lesion load and brain atrophy). They can be evaluated manually by the physician. However, manual measurements are a fastidious task and are subject to inter- and intra- expert variabilities. Consequently, this PhD thesis described new methods to perform these measurements.

Firstly, semi-automatic algorithms have been improved. Thanks to them, an expert can faster realise different measurements of reference. Secondly, different fully-automatic methods have been proposed.

The first method is a pipeline allowing to segment the different brain compartments (white matter, gray matter and cerebro-spinal fluid). The second method uses these segmentations to automatically compute a threshold that has to be applied on a MRI sequence to obtain the MS lesions segmentation. This approach has already been described in the literature. However, this PhD thesis work proposes two new versions of this method. The first segments the MS lesion of the cortex and the second of the cerebellum. The last method uses the segmentations obtained by the first one to compute the brain parenchymal fraction and to evaluate the atrophy. Two versions of this method are proposed. At the opposite of the first version, the second take into consideration longitudinal information.

These methods have been compared to other automatic methods and to the measurements of reference performed by the expert. Thanks to these methods, our database has been studied and analysed. Finally, the developed algorithms have been included into a software dedicated to MS brain MRI analysis : SepINRIA.

Keywords : Medical imaging, image processing, Multiple Sclerosis, MRI, Segmentation, Atrophy, Biomarkers.