



Distance entre distributions : application à l'imagerie médicale et à l'aéronautique

Sana Rebbah

► To cite this version:

Sana Rebbah. Distance entre distributions : application à l'imagerie médicale et à l'aéronautique. Traitement du signal et de l'image [eess.SP]. Université Paul Sabatier - Toulouse III, 2019. Français. NNT : 2019TOU30124 . tel-02798518

HAL Id: tel-02798518

<https://theses.hal.science/tel-02798518>

Submitted on 5 Jun 2020

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

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



THÈSE

En vue de l'obtention du DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Université Toulouse 3 - Paul Sabatier

Présentée et soutenue par

Sana REBBAH

Le 30 septembre 2019

**Distance entre distributions : application à l'Imagerie Médicale
et à l'Aéronautique**

Ecole doctorale : **AA - Aéronautique, Astronautique**

Spécialité : **Radiophysique et Imagerie Médicale**

Unité de recherche :

INSERM ToNIC-Toulouse NeuroImaging Center (UMR 1214)

École Nationale de l'Aviation Civile

Thèse dirigée par

**Isabelle BERRY
Daniel DELAHAYE
Pierre MARÉCHAL**

Jury

Pr. Jean-philippe Ranjeva, Rapporteur et président

Pr. Nidhal Carla Bouaynaya, Rapporteuse

Dr. Christian Barillot, Examineur

Pr. Isabelle Berry, Directrice de thèse

Pr. Daniel Delahaye, Directeur de thèse

Pr. Pierre Maréchal, Co-directeur de thèse

Remerciements

Je commence par remercier Pr. Jean-Philippe Ranjeva, Pr. Nidhal Bouaynaya et Dr. Christian Barillot d'avoir accepté de faire partie de mon jury de thèse. J'ai beaucoup apprécié notre échange lors de la soutenance.

Merci au Pr. Isabelle Berry à qui je tiens à exprimer ma profonde gratitude, pour avoir encadré et dirigé mes recherches. Je vous remercie pour votre soutien et votre appui tout au long de cette thèse. Je m'estime honorée et chanceuse de vous avoir eue comme directrice de thèse. Merci aussi au Pr. Daniel Delahaye et Pr. Pierre Maréchal pour leur co-encadrement. Merci à toi Daniel pour ta constante bonne humeur et tes précieux conseils (et désolée pour les footing !). Merci à vous aussi Pr. Maréchal, nos échanges ont toujours été intéressants.

Un particulier (et très grand) merci à Stéphane Puechmorel et Florence Nicol, chercheurs à l'ENAC, sans qui cette thèse aurait été bien différente. Merci Stéphane d'avoir été là du début jusqu'à la fin, d'avoir partagé ton temps et tes idées avec moi, mais aussi d'avoir participé à la rédaction de la partie sur la géométrie de l'information. Merci Florence pour ton soutien, ta disponibilité, tes corrections et tes mots d'encouragement dans les moments les plus délicats.

Cette thèse a été effectuée entre INSERM ToNIC et ENAC, je voudrais donc remercier Pr. Pierre Payoux et Pr. Patrick Sénac d'avoir accepté de m'accueillir dans ces unités de recherche.

Je remercie tous mes amis de l'ENAC. Tout d'abord, à Sana(2) pour les nombreuses et longues discussions. Cette dernière année aurait été bien triste sans toi. Merci à mes amis de Chine : Maji, Shangrong, Ying, Victor et Bug ; grâce à vous j'ai appris deux mots en chinois : *xié xié* et *Wo ai ni*, ce qui est parfait car c'est exactement ce que j'ai envie de vous dire. Je pense aussi à Alex, Serge, Thuhuyen, Evgenii, Micha, Moctar, Alice, Isa, Vincent, Florian... Sans oublier Philippe et Gaby, je vous dois une fin de thèse plus décontractée à base de "courtes" pauses et de blind-test. Merci à vous tous, je vous souhaite le meilleur pour la suite !

Mes remerciements s'adressent également à toute l'unité ToNIC. Merci pour les sourires et les encouragements. Une pensée à Florent Aubry et Nicolas Chauveau, mes anciens encadrants de stage qui n'ont jamais hésité à m'aider durant la thèse.

Je tiens également à remercier toutes les personnes qui ont contribué, d'une façon ou d'une autre, à la réussite de cette thèse et à rendre ce passage agréable.

Enfin, je remercie encore une fois le soutien inconditionnel de ma famille, plus particulièrement ma mère, mon frère Habib et ma meilleure amie Kana durant ces trois dernières années, mais aussi tout au long de mes études.

Et puis à toi, je te dédie cette thèse. J'espère que tu es fier de ta fille.

Table des matières

Introduction	13
I État de l'art	17
1 Neuroimagerie : analyse d'images quantitatives	19
1.1 IRM et maladies neurodégénératives	19
1.1.1 Les modalités d'IRM	20
1.1.2 L'IRM et la maladie d'Alzheimer	27
1.1.3 L'IRM et la Sclérose en plaques	31
1.2 Les différentes approches d'analyse d'images	40
1.2.1 Analyse par régions d'intérêt	40
1.2.2 Analyse voxel à voxel	42
1.2.3 Analyse d'histogrammes	43
1.3 Analyse d'histogrammes en imagerie cérébrale	45
1.3.1 Production et analyse d'histogrammes	45
1.3.2 Les applications cliniques	49
2 Aéronautique : les retards dans le transport aérien	55
2.1 Définition des retards : types et causes	55
2.1.1 Notion de retard	55
2.1.2 Les causes des retards	56
2.2 Mesure et caractérisation des retards	57
2.2.1 Les retards moyens	58
2.2.2 Modélisation des retards par des distributions	58
II Partie expérimentale	61
3 Les métriques d'histogrammes	63
3.1 Description de l'approche	63
3.1.1 Production d'histogrammes	64
3.1.2 Définition des métriques d'histogrammes	64
3.1.3 Classification basée sur les métriques d'histogrammes	66
3.2 Application 1 : classification des populations de la maladie d'Alzheimer	68
3.2.1 Matériels et méthodes	69
3.2.2 Résultats	81
3.2.3 Discussion	84
3.3 Application 2 : essai clinique de Sclérose en plaques	87

3.3.1	Matériels	89
3.3.2	Méthodes	93
3.3.3	Résultats	96
3.3.4	Discussion et limites	101
4	Distance entre histogrammes	103
4.1	Description de l'approche	103
4.1.1	Contexte	103
4.1.2	Notion de distance	104
4.2	Application : classification de la population de Sclérose en plaques	107
4.2.1	Matériels et méthodes	108
4.2.2	Résultats	111
4.2.3	Discussion	112
5	Distribution de probabilité et géométrie de l'information	115
5.1	État de l'art sur la géométrie de l'information	116
5.1.1	Contexte : Comparer deux distributions	116
5.1.2	Géométrie différentielle	121
5.2	Application 1 : classification de la population de la maladie d'Alzheimer	127
5.2.1	Conception mathématique	127
5.2.2	Application aux données de la maladie d'Alzheimer	134
5.3	Application 2 : caractérisation des retards dans le transport aérien	141
5.3.1	Contexte	141
5.3.2	Conception mathématique	142
5.3.3	Application : clustering des retards aéroportuaires	148
III	Discussion générale	153
	Références	157
	Annexes	181
	Contributions scientifiques	191
	Résumés	195

Table des figures

1.1	IRM pondérées en T1 (avec ou sans gadolinium), T2 et FLAIR . . .	21
1.2	Imagerie de diffusion : DWI, ADC et FA	23
1.3	Représentation du tenseur de diffusion	23
1.4	Principe du transfert d'aimantation	25
1.5	Imagerie de transfert d'aimantation : La carte MTR	25
1.6	La spectroscopie par résonance magnétique - Spectre RMN	26
1.7	Physiopathologie de la SEP	32
1.8	Les différentes formes cliniques de la SEP	33
1.9	IRM conventionnelle en SEP (T1 avec et sans Gd, T1, T2 et FLAIR	36
1.10	Spectroscopie par résonance magnétique et sclérose en plaques	39
1.11	Représentation des différentes approches d'analyses d'images quantitatives	40
1.12	Histogramme MTR de la SGAN partiellement et totalement normalisé	48
1.13	Histogrammes MTR des différentes formes de la SEP	51
1.14	Histogrammes de l'imagerie de diffusion (DM et FA) des patients atteints de SEP	52
2.1	Les causes des retards par an, en pourcentage du nombre total de minutes de retard	57
2.2	Modélisation des retards de départ en utilisant un modèle log-normal, gamma et Rayleigh (d'après Novianingsih and Hadiani (2014))	59
3.1	Schéma résumé de l'approche basée sur les métriques d'histogrammes	64
3.2	Schéma représentant les différentes étape de prétraitement des volumes IRM	72
3.3	Trois géométries possibles pour le ruban de substance grise dans le cerveau	73
3.4	Surfaces équipotentiels entre S et S'	74
3.5	Cartes d'épaisseur corticale	75
3.6	Anatomie des surfaces corticales : Gyri et sillons	75
3.7	Surface corticale interne et externe	76
3.8	Carte des courbures corticales	77
3.9	Schéma des support vector machines	79
3.10	Histogramme extraits des images paramétriques d'EC et de courbure corticale	82
3.11	Plan d'expérience de l'étude IRM ancillaire à l'essai MS-SPI	89
3.12	Les différents biomarqueurs IRM utilisés dans l'essai clinique MS-SPI	91
3.13	Boxplot à M00, M12 et M24 de l'EC (mm), ADC, FA et MTR.	97

3.14 Le pourcentage de variance expliquée pour chaque dimension . . .	98
3.15 Représentation des individus - AFM	99
3.16 Représentation des groupes - AFM	99
3.17 Dendrogramme clustering HCPC	100
3.18 Évaluation du clustering HCPC	100
4.1 Schéma résumé de l'approche basée sur les distances entre histogrammes.	108
4.2 Illustration de l'algorithme des k plus proches voisins	110
4.3 Illustration de l'algorithme des k plus proches voisins basé sur les histogrammes	110
5.1 Moyenne de gaussiennes : Comparaison de la distance euclidienne et la distance de Wasserstein	117
5.2 Chemin de distributions : Comparaison de la distance euclidienne et la distance de Wasserstein	118
5.3 L'exemple de la sphère.	119
5.4 Distributions normales univariées et leurs représentations dans le demi-plan $(\mu; \sigma)$	120
5.5 Surface plongée de dimension 2	121
5.6 Approche basée sur la géométrie de l'information dans le cadre de la classification de la population Alzheimer	134
5.7 Estimation de l'épaisseur corticale moyenne par ROI	138
5.8 Distribution des retards de départ d'un aéroport selon trois modèles : Gamma, Weibull et exponentiel	150

Liste des tableaux

1.1	Description des séquences IRM conventionnelles	21
1.2	Les critères 2017 de McDonald pour le diagnostic de la SEP	34
1.3	Les trois principales méthodes d'analyse d'images quantitatives	44
3.1	Comparaison des méthodes d'extraction de caractéristiques et de sélection de caractéristiques	68
3.2	La population ADNI étudiée	70
3.3	Les paramètres de l'acquisition MPRAGE	71
3.4	Matrice de confusion	81
3.5	Résultats du test MANOVA sur les métriques des histogrammes	83
3.6	Les paramètres du noyau RBF pour chaque groupe de classification SVM	84
3.7	Résultats de la classification SVM de la population ADNI	84
3.8	Performances de classification AD/HC de notre approche avec d'autres approches	86
3.9	Performances de classification sMCI/pMCI de notre approche avec d'autres approches	87
3.10	Performances de classification HC/pMCI de notre approche avec d'autres approches	87
3.11	Performances de classification HC/sMCI de notre approche avec d'autres approches	87
3.12	Performances de classification AD/pMCI de notre approche avec d'autres approches	87
3.13	Performances de classification AD/sMCI de notre approche avec d'autres approches	87
3.14	Information démographique de la population de notre étude	90
3.15	Paramètres d'acquisition des différentes séquences utilisées dans notre étude	91
3.16	Les différentes séquences IRM utilisées dans notre étude et les biomarqueurs extraits	91
3.17	Tableau représentatif des données de notre étude sur l'essai MS-SPI	94
3.18	Résultats de l'évolution des biomarqueurs (%), en un an (M00→M12).	96
3.19	Résultats de l'évolution des biomarqueurs (%), en deux ans (M00→M24).	96
3.20	Coefficients RV entre les groupes et l'AFM	100
4.1	Information démographique de la population de notre étude (approche 2)	108

4.2	Résultat de la classification des kNN basé sur les histogrammes . .	112
5.1	Caractéristiques démographiques et cliniques de la population étudiée	135
5.2	Résultat du clustering k-médoides AD/HC suivant 4 scénarios classiques.	141
5.3	Résultat du clustering k-medoids basé sur les distributions suivant différents modèles et mesures de dissimilarité	141
5.4	Premier résultat du clustering k-médoides des aéroports	151
5.5	Second résultat du clustering k-médoides des aéroports	151
5.6	Les variables optimales, sélectionner à l'aide du gain d'information	183
5.7	Les formules de 45 différentes mesures de dissimilarité	186
5.8	Définition des 22 régions d'intérêt du cortex. Ces 22 ROI correspondent à 11 zones de l'hémisphère droit ("R") et les 11 zones correspondantes de l'hémisphère gauche ("L").	187
5.9	Les variables les plus discriminantes pour la classification AD/HC .	188

Abréviations

10CV	10-fold cross-validation
ACP	Analyse en Composantes Principales
ADC	Apparent Diffusion Coefficient
AD	Alzheimer's disease
BTS	Bureau of Transportation Statistics
CIS	Clinically isolated syndrome
DKL	Divergence de Kullback-Leibler
DM	Diffusivité moyenne
DWI	Diffusion Weighted Imaging
EC	Épaisseur corticale
EVP	Effet de Volume Partiel
FAA	Federal Aviation Administration
FA	Fraction d'Anisotropie
FLAIR	Fluid Attenuated Inversion Recovery
Gd	Gadolinium
GI	Gain d'information
Gyrex	Courbure corticale externe au niveau des gyri
GyrInt	Courbure corticale interne au niveau des gyri
H-kNN	K-plus proches voisins basé sur les histogrammes
IRMq	Imagerie par Résonance Magnétique quantitative
IRM	Imagerie par Résonance Magnétique
kNN	K plus proches voisins
LCS	Liquide Cérébro-Spinal

MA	Maladie d'Alzheimer
MCI	Mild Cognitive Impairment
MRS	Magnetic Resonance Spectroscopy
MTI	Magnetization Transfer Imaging
MTR	Ratio de transfert d'aimantation
NABT	Normal Appearing Brain Tissue
PD	Densité de Protons
pMCI	Progressive Mild Cognitive Impairment
RBF	Radial basis function
ROI	Région d'intérêt
SBAN	Substance Blanche d'Apparence Normale
SB	Substance Blanche
SCI	Syndrome cliniquement isolé
SEP-PP	Sclérose en plaques progressive primaire
SEP-RR	Sclérose en plaques rémittente-récurrente
SEP-SP	Sclérose en plaques secondairement progressive
SEP	Sclérose en plaques
SGAN	Substance Grise d'Apparence Normale
SG	Substance Grise
SilExt	Courbure corticale externe au niveau des sillons
SilInt	Courbure corticale interne au niveau des sillons
sMCI	Stable Mild Cognitive Impairment
sMRI	Structural Magnetic Resonance Imaging
SNC	Système nerveux central
SPP	Sclérose en plaques progressive
TEP	Tomographie par émission de positons
TE	Temps d'écho
TR	Temps de répétition

Introduction générale

Contexte

En neuroimagerie, l'IRM est un des outils les plus informatifs concernant la prise en charge de patients atteints de troubles neurologiques telles que la maladie d'Alzheimer et la sclérose en plaques, tant pour le diagnostic, le pronostic, le suivi et l'évaluation de la réponse à un traitement. Afin d'exploiter les images provenant d'IRM et d'en extraire de précieuses informations, plusieurs méthodes d'analyse d'images quantitatives ont été développées. Parmi ces dernières, on retrouve l'analyse par régions d'intérêt et l'analyse voxel à voxel. Le principe est de mesurer un paramètre IRM dans une zone spécifique du cerveau ou dans la plus petite entité de l'image, le voxel, puis d'effectuer des tests statistiques entre deux groupes de sujets distincts. La statistique utilisée pour décrire le paramètre IRM est majoritairement la moyenne. Toutefois, d'autres études ont constaté la limite de cette statistique et ont proposé une autre approche d'analyse d'images, l'analyse d'histogrammes, qui est tout particulièrement utilisée dans l'étude de maladies marquées par des modifications diffuses, telle que la sclérose en plaques. Après avoir extrait les histogrammes des images quantitatives, des mesures locales sont dérivées de ces histogrammes telles que la moyenne, la médiane, les quantiles, skewness... Mais, il s'agit encore de résumés et toute l'information incluse dans le biomarqueur n'est pas exploitée.

De plus, nous faisons l'analogie avec l'aéronautique dans le cadre de l'étude des retards. Les retards des vols sur les plates-formes aéroportuaires représentent un indicateur de performance particulièrement important et contribuent de manière significative à la perception du fonctionnement de l'ensemble du système par les voyageurs. L'analyse actuellement réalisée se situe au niveau macroscopique et fournit un indicateur de retard moyen, sans tenir compte des mécanismes intermédiaires pouvant conduire au retard final. De plus, les retards observés sont souvent la conséquence d'un retard de vols précédents, qui a conduit à définir la notion de multiplicateur de retard, indicateur de la tendance à la propagation des retards. Aucun de ces outils d'analyse n'est pleinement satisfaisant et nécessite d'être plus précis pour permettre une réelle orientation sur les moyens de réduire ces retards. Ainsi, dans cette thèse, l'idée est de remplacer les statistiques agrégées telle que la moyenne par un modèle statistique paramétrique plus complet : les distributions de probabilité. Ainsi, que cela soit dans le contexte de neuroimagerie ou pour étudier des retards de vol, l'utilisation d'estimateur comme la moyenne et l'écart type pour résumer les distributions ne parviendra pas à caractériser le degré d'interdépendance des retards de vol ou l'étendue des dommages sous-jacents dans le cadre des maladies neurologiques.

Contributions

Dans ce travail de thèse nous proposons différentes manières de prendre en compte toute l'information incluse dans un marqueur à travers sa distribution et non uniquement une mesure de tendance centrale. Ainsi, les principales contributions de cette thèse sont :

- Mettre en avant l'utilisation des distributions à travers deux approches non-paramétriques et une approche paramétrique.
- Proposer, développer et appliquer les outils de la géométrie de l'information dans le cadre de l'imagerie médicale et de l'aéronautique.

Ce travail a ainsi permis de :

- Classer les populations de la maladie d'Alzheimer d'une façon plus précise, obtenant une précision de 93.1%, en se basant sur l'utilisation des résumés des histogrammes de paramètres IRM (Rebbah et al., 2018b).
- Améliorer la classification des populations de sclérose en plaques suite à l'utilisation de diverses mesures de dissimilarité entre les histogrammes dans leur globalité et non entre les métriques de ces derniers (Rebbah et al., 2019a).
- Mettre en avant la possibilité de s'affranchir de toutes les variations individuelles des sujets et de déceler l'effet (ou l'absence d'effet) d'un médicament dans un essai thérapeutique (Rebbah et al., 2018a).
- Proposer et développer les outils de la géométrie différentielle qui sont plus naturels dans le cadre de l'utilisation des distributions de probabilité. En effet, l'histogramme n'est qu'une estimation trop grossière d'une distribution de probabilité. Ces outils ont permis de :
 - ⇒ Améliorer les performances de classification des populations de la maladie d'Alzheimer en modélisant les distributions d'épaisseur corticale par des lois gamma généralisées et en mesurant des distances géodésiques entre elles (Rebbah et al., 2019b).
 - ⇒ Mieux caractériser les retards aéroportuaires en les modélisant par des loi gamma et en mesurant des distances géodésiques entre elles.

Plan

Ce manuscrit s'organise en trois grandes parties qui présentent respectivement un état de l'art, nos contributions dans ce domaine et enfin une partie discussion.

- La première grande partie regroupe deux chapitres qui ont pour rôle d'introduire le travail de thèse dans le contexte scientifique international. Étant donné que la thèse a été effectuée entre l'INSERM et l'École Nationale de l'Aviation Civile, nous proposons un chapitre pour chaque domaine : la neuroimagerie et l'aéronautique. Le premier chapitre présente un état de l'art sur les différentes analyses d'images quantitatives utilisées, avec les détails de l'analyse d'histogrammes en imagerie cérébrale, tout en reprenant les notions concernant l'IRM et les maladies neurodégénératives. Le second chapitre a pour objectif d'exposer un bref état de l'art des retards dans le transport aérien. Nous présentons les causes de ces retards et la manière dans laquelle ces derniers sont représentés et mesurés dans la littérature.

- La deuxième partie est consacrée aux travaux réalisés au cours de cette thèse. On y retrouve trois chapitres : Les deux premiers proposent une approche

non-paramétrique et le troisième une approche paramétrique.

Le premier chapitre propose d'extraire des métriques des histogrammes des images paramétriques et de les intégrer dans des classifieurs afin de d'identifier différents groupes de patients. Deux applications sont proposées. La première porte sur la classification des patients atteints de troubles cognitifs légers et les patients atteints de la maladie d'Alzheimer. La deuxième sur un essai clinique de sclérose en plaques utilisant différentes modalités IRM.

Le second chapitre propose d'utiliser la totalité de l'information incluse dans les histogrammes en mesurant différentes distances entre ces derniers dans le cadre de la classification de la population de sclérose en plaques.

Le dernier chapitre présente l'approche paramétrique qui consiste à prendre en compte la distribution de probabilité et non une estimation tel que l'histogramme et cela en utilisant les outils de la géométrie de l'information. Nous exposons tout d'abord un état de l'art sur la géométrie non-euclidienne. Puis, nous proposons deux applications : la classification de la population Alzheimer et la caractérisation des retards dans le transport aérien.

- La dernière partie correspond à la discussion générale qui rappelle les principaux résultats apportés par ce travail de thèse, en présentant les limites et les perspectives associées.

Première partie

État de l'art

Chapitre 1

Neuroimagerie : analyse d'images quantitatives

Sommaire

1.1	IRM et maladies neurodégénératives	19
1.1.1	Les modalités d'IRM	20
1.1.2	L'IRM et la maladie d'Alzheimer	27
1.1.3	L'IRM et la Sclérose en plaques	31
1.2	Les différentes approches d'analyse d'images	40
1.2.1	Analyse par régions d'intérêt	40
1.2.2	Analyse voxel à voxel	42
1.2.3	Analyse d'histogrammes	43
1.3	Analyse d'histogrammes en imagerie cérébrale	45
1.3.1	Production et analyse d'histogrammes	45
1.3.2	Les applications cliniques	49

1.1 IRM et maladies neurodégénératives

L'imagerie par résonance magnétique (IRM) est devenue l'une des techniques d'imagerie les plus utilisées en recherche et en clinique. Elle est le premier choix pour imager morphologiquement le cerveau humain car elle est non-invasive et offre des images d'une résolution spatiale et temporelle relativement élevée. Le principal avantage de l'IRM par rapport à d'autres techniques d'imagerie telle que la tomographie par émission de positons (TEP) est sa très haute résolution spatiale in vivo, qui permet d'obtenir des informations anatomiques précises et claires. D'autre part, les champs magnétiques et les ondes électromagnétiques de basse énergie sont utilisés à la place des rayonnements ionisants, ainsi aucun dommage biologique n'est causé. Un autre avantage important est que l'IRM permet des études longitudinales, car elle ne repose pas sur l'utilisation d'isotopes radioactifs, ni de rayons X.

La prévalence des maladies neurodégénératives augmente du fait de l'augmentation du vieillissement de la population (Brayne and Miller, 2017), toutefois les mécanismes physiopathologiques à l'origine de ces maladies restent

obscurs. D'autres atteintes cérébrales telles que la sclérose en plaques surviennent plus précocement mais présentent des similarités morphologiques permettant d'en étudier le retentissement anatomique par les mêmes techniques

Cette partie résume la traduction sur l'imagerie de signes de neurodégénérescence communs ou spécifiques à la sclérose en plaques et à la maladie d'Alzheimer, et fournit un bref aperçu des techniques d'IRM utilisées pour étudier et suivre ces modifications morphologiques.

1.1.1 Les modalités d'IRM

Cette section présente brièvement les plus importantes modalités d'IRM, conventionnelles et non conventionnelles, appliquées dans le domaine de l'imagerie cérébrale.

1.1.1.1 IRM conventionnelle : T1, T2 et FLAIR

À ce jour, la majorité des images en clinique repose sur l'acquisition et l'interprétation du temps de relaxation (T1 et T2), dont le contraste dépend des propriétés intrinsèques du tissu ; sur le type d'acquisition (séquence d'imagerie et paramètres d'acquisition) et sur des facteurs matériels (géométrie de la bobine du récepteur radiofréquence (RF), sensibilité, positionnement du patient, chargement et gains de l'amplificateur de signal électronique).

Les images pondérées en T1 sont produites en utilisant des temps TE (temps d'écho) et TR (temps de répétition) courts en séquence d'écho de spin. Le contraste de l'image est principalement déterminé par les propriétés T1 du tissu. Inversement, les images pondérées en T2 sont produites en utilisant des temps TE et TR plus longs. Dans ces images, le contraste est principalement déterminé par les propriétés T2 du tissu. En général, les images pondérées en T1 et en T2 peuvent être facilement différenciées en regardant le liquide cérébro-spinal (LCS). Le LCS est sombre (hyposignal) pour les images pondérées en T1 et clair (hypersignal) pour les images pondérées en T2 (figure 1.1).

L'IRM pondérée en T1 peut également être réalisée en injectant du gadolinium (Gd) (figure 1.1). Le Gd est un agent de contraste paramagnétique non toxique. Lorsqu'il est injecté, il modifie l'intensité du signal en raccourcissant le temps de relaxation longitudinal T1 et transversal T2. Ainsi, le Gd produit un hypersignal sur les images pondérées en T1. Les images améliorées par le Gd sont particulièrement utiles pour examiner les structures vasculaires et la dégradation de la barrière hémato-encéphalique (e.g. tumeurs, abcès, inflammation en sclérose en plaques...).

Une troisième séquence couramment utilisée est la séquence FLAIR (Fluid Attenuated Inversion Recovery). La séquence FLAIR (figure 1.1) est basée sur la technique d'inversion-récupération. Elle est similaire à une image pondérée en T2, à la différence que l'on introduit un temps d'inversion TI annulant le signal du fait de son long T1. Ainsi, les anomalies restent en hypersignal mais le LCS est atténué et apparaît sombre sur l'image (hyposignal). Cette séquence est très sensible à la pathologie et facilite la différenciation entre le LCS et les anomalies (e.g. les lésions de la sclérose en plaques).

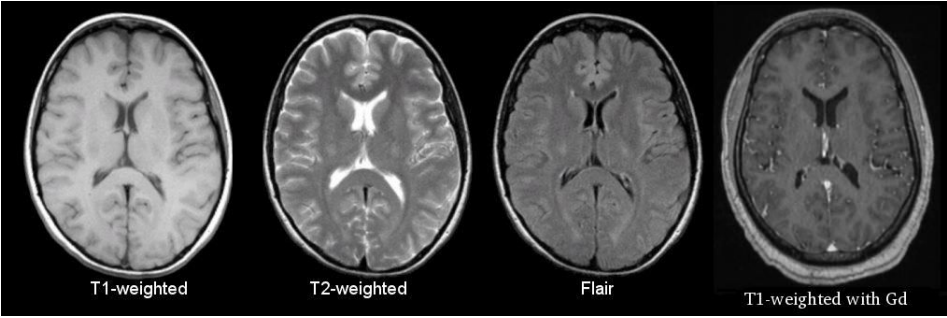


FIGURE 1.1 – Des IRM pondérées en T1 (avec et sans gadolinium), T2 et FLAIR (d’après [Preston \(2011\)](#), modifié)

Séquences IRM	Temps de répétition (TR en ms)	Temps d’écho (TE en ms)	Description
T1	500	14	SB : clair SG : gris LCS : sombre Graisse : très clair Lésions : sombre
T2	4000	90	SB : gris foncé SG : gris clair LCS : très clair Graisse : clair Lésions : très clair
FLAIR	9000	114	SB : gris foncé SG : gris clair LCS : sombre Graisse : clair Lésions : très clair

Tableau 1.1 – Les séquences d’IRM les plus courantes, leurs temps TR et TE approximatives et description de l’image

1.1.1.2 IRM non conventionnelle

Ces vingt dernières années, différentes techniques d’imagerie en IRM ont été développées et appliquées pour trouver de nouveaux marqueurs IRM, qui seraient plus proches des phénomènes physiopathologiques de la maladie et cela soit en s’intéressant aux mouvements des molécules d’eau (imagerie de diffusion), soit à l’organisation architecturale moléculaire (imagerie de transfert d’aimantation), soit par les modifications biochimiques (imagerie spectroscopique). Toutefois, ces méthodes ne sont pas recommandées en pratique clinique courante car elles nécessitent des protocoles d’acquisition IRM plus complexes et plus longs en termes de temps machine, et des procédures de post-traitements dédiés.

1.1.1.2.1 Imagerie de diffusion

La première image de diffusion chez l’homme a été réalisée par [Le Bihan and Breton](#) en 1985. Cette séquence apporte des information non disponibles avec les séquences d’IRM traditionnelles ou morphologiques. Elle permet d’évaluer le mouvement des molécules d’eau qui peut être modifié en cas de processus pathologique.

Le phénomène de diffusion reflète le mouvement aléatoire des molécules d’eau, appelé mouvement brownien. Au sein des tissus biologiques, le mouvement de l’eau est dirigé suivant l’architecture tissulaire qui l’entoure. En effet, au niveau du LCS, la diffusion est importante et isotrope (i.e. identique

dans toutes les directions), alors qu'au niveau de la substance blanche (SB) elle est faible et anisotrope (i.e. préférentielle dans une direction donnée). La diffusion de l'eau dans les fibres de la SB est privilégiée parallèlement aux fibres ($D_{//}$) et restreinte perpendiculairement aux fibres (D_{\perp}).

— DWI et ADC

L'acquisition d'une image pondérée en diffusion (DWI¹) se fait à l'aide d'une image sans gradient ($b^2=0$), c'est à dire pondérée en T2, notée S_0 et d'autant d'images de gradient de diffusion ($b>0$, notées S_i) que de directions d'acquisition. Le premier gradient introduit le déphasage des protons en fonction de leur position, tandis que le second gradient inverse les modifications apportées par le premier gradient. S'il y a des mouvements de protons, le second gradient ne pourra pas annuler complètement les changements induits par le premier gradient. En conséquence, il y aura une atténuation du signal. Cette perte de signal due au mouvement des particules est donnée par l'équation de [Stejskal and Tanner](#) :

$$S_i = S_0 e^{-b \cdot ADC_i} \quad (1.1)$$

où b représente le facteur de pondération en diffusion (sec/mm²). La pondération en diffusion de la séquence utilisée dépend de la performance des gradients. Plus b est élevé, plus la séquence est pondérée en diffusion, mais plus le rapport signal sur bruit sera faible,

ADC_i est le coefficient de diffusion apparent suivant la direction i . Il représente le degré de mobilité des molécules d'eau. Ainsi :

$$ADC_i = \frac{1}{b} \log \frac{S_0}{S_i} \quad (1.2)$$

Les cartes³ ADC (figure 1.2-b) sont dérivées d'au moins deux images DWI en utilisant l'équation de Stejskal-Tanner. Si la mobilité des molécules d'eau dans un tissu est faible, l'ADC diminue ce qui est représenté en hypersignal sur DWI et en hyposignal sur la carte ADC (figure 1.2 a et b) et au contraire si l'ADC augmente cela signifie qu'il y a une forte mobilité des molécules d'eau dans le tissu (hyposignal sur DWI et hypersignal sur la carte ADC).

— Tenseur de diffusion

À partir des cartes ADC, un tenseur de diffusion DTI (Diffusion Tensor Imaging) peut être calculé pour chaque voxel de l'image. Il est représenté graphiquement par une ellipsoïde (figure 1.3). Il correspond à une matrice D symétrique (3x3) :

$$D = \begin{pmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{yx} & D_{yy} & D_{yz} \\ D_{zx} & D_{zy} & D_{zz} \end{pmatrix} \quad (1.3)$$

1. DWI signifie "Diffusion Weighted Imaging"

2. b est le facteur de pondération en diffusion. Il dépend de l'amplitude du gradient, de la durée d'application et du temps entre les deux impulsions de gradient.

3. images paramétriques

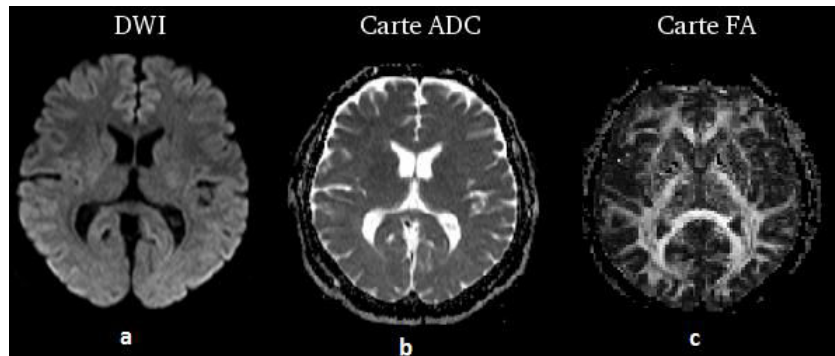


FIGURE 1.2 – Imagerie de diffusion. (a) IRM de diffusion DWI, (b) carte du coefficient apparent de diffusion (ADC), Carte de la fraction d'anisotropie (FA), d'après [MRIMASTER](#)

Les trois éléments diagonaux (D_{xx} , D_{yy} , D_{zz}) représentent les coefficients de diffusion mesurés le long de chacun des principaux axes (x , y et z). Les six termes non diagonaux (D_{xy} , D_{yz} , etc.) reflètent la corrélation des mouvements aléatoires entre chaque paire de directions principales.

La technique consiste à appliquer successivement des gradients de diffusion dans au moins six directions non-colinéaires, pour calculer les coefficients de diffusion correspondants à chacune de ces directions. La matrice du tenseur de diffusion est diagonalisable puisqu'elle est symétrique et définie positive :

$$D = \begin{pmatrix} D_{xx} & D_{xy} & D_{xz} \\ . & D_{yy} & D_{yz} \\ . & . & D_{zz} \end{pmatrix} = \begin{pmatrix} \epsilon_{1x} & \epsilon_{2x} & \epsilon_{3x} \\ \epsilon_{1y} & \epsilon_{2y} & \epsilon_{3y} \\ \epsilon_{1z} & \epsilon_{2z} & \epsilon_{3z} \end{pmatrix}^T \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} \begin{pmatrix} \begin{bmatrix} \epsilon_{1x} \\ \epsilon_{1y} \\ \epsilon_{1z} \end{bmatrix} & \begin{bmatrix} \epsilon_{2x} \\ \epsilon_{2y} \\ \epsilon_{2z} \end{bmatrix} & \begin{bmatrix} \epsilon_{3x} \\ \epsilon_{3y} \\ \epsilon_{3z} \end{bmatrix} \end{pmatrix} \quad (1.4)$$

Le tenseur est décomposé en valeurs propres λ_1 , λ_2 et λ_3 et en vecteurs propres ϵ_1 , ϵ_2 et ϵ_3 . Les valeurs propres sont classées dans l'ordre décroissant ($\lambda_1 > \lambda_2 > \lambda_3$) et donc ϵ_1 est représentatif de la direction principale de diffusion (i.e. la direction des fibres).

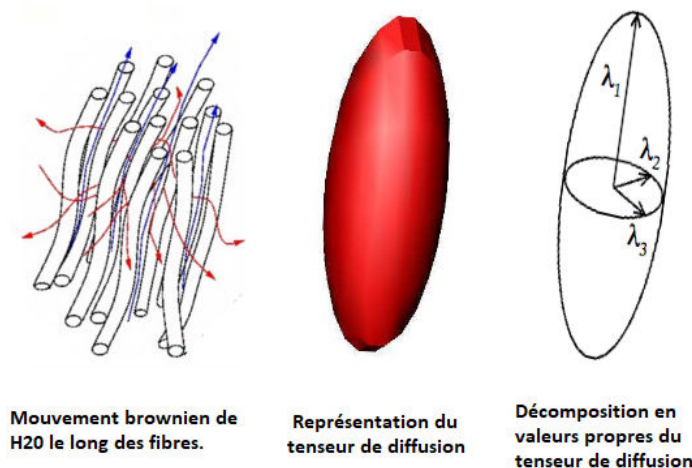


FIGURE 1.3 – Représentation du tenseur de diffusion (d'après ([Poupon, 1999](#)), modifié

— Les indices du tenseur de diffusion

À partir des données obtenues, on peut calculer et visualiser les différents indices reflétant l'amplitude moyenne de la diffusion (DM), le degré d'anisotropie (FA), ou encore la direction de la diffusion maximale au sein de

chaque voxel (tractographie).

-**La diffusivité moyenne** : La DM permet de quantifier l'amplitude moyenne des mouvements de diffusion. Elle indique si la diffusion est globalement (suivant toutes les directions d'acquisition) importante (DM élevée) ou restreinte (DM proche de 0)

$$DM = \frac{\lambda_1 + \lambda_2 + \lambda_3}{3} \quad (1.5)$$

-**Fraction d'anisotropie** : La FA caractérise le manque d'uniformité du phénomène de diffusion dans les différentes directions de l'espace. Une anisotropie proche de 1 signifie que la diffusion est privilégiée dans une direction et une anisotropie proche de 0 signifie que la diffusion est la même dans toutes les directions (i.e. un manque total d'orientation). Les faisceaux de fibres de substance blanche apparaissent ainsi en hypersignal sur la carte FA (figure 1.2-c).

$$FA = \sqrt{\frac{(\lambda_1 - \lambda_2)^2 + (\lambda_1 - \lambda_3)^2 + (\lambda_3 - \lambda_2)^2}{2(\lambda_1^2 + \lambda_2^2 + \lambda_3^2)}} \quad (1.6)$$

- **Diffusion axiale** : $D_{//}$ représente la diffusion suivant l'axe principal de diffusion. $D_{//} = \lambda_1$

- **Diffusion radiale** : D_{\perp} représente la diffusion perpendiculaire à l'axe principal de diffusion. $D_{\perp} = \frac{\lambda_2 + \lambda_3}{2}$

1.1.1.2.2 Imagerie de transfert d'aimantation

L'imagerie de transfert d'aimantation (MTI⁴), initialement proposée par [Wolff and Balaban](#) en 1989, a été largement utilisée pour étudier la microstructure tissulaire dans le cerveau. Le transfert d'aimantation offre des avantages qui se sont révélés utiles dans l'étude des troubles de la substance blanche, le développement du cerveau, le vieillissement et les maladies neurodégénératives.

Le signal IRM provient pour la plus grande part des protons des molécules d'eau libre. Il existe également de l'eau liée aux macromolécules (protéines, membranes, myéline) mais le signal de leurs protons n'est pas significatif car leur T_2 est très court. La technique de transfert d'aimantation est basée sur la quantification indirecte de changement d'états entre ces protons ([Hoa, 2007](#)). Une onde de radiofréquence sélective (MT-pulse) est appliquée⁵ afin de saturer les protons liés. Les échanges permanents entre eau libre et eau liée aux macromolécules, entraînent un transfert de la saturation des protons liés vers les protons libres. On observe donc une réduction du signal IRM (figure 1.4). Cette baisse relative du signal IRM est quantifiée par le taux de transfert d'aimantation MTR⁶ (figure 1.5) :

$$MTR = \frac{M_0 - M_S}{M_0} \times 100 \quad (1.7)$$

4. "Magnetization Transfer Imaging" en anglais

5. L'onde RF est décalée par rapport à la fréquence de résonance des protons libres. Environ 1500 Hz (1 à 2 kilohertz)

6. "Magnetization Transfert Ratio" en anglais

avec M_0 représente, pour chaque voxel, l'intensité du signal sans MT-pulse, M_s avec MT-pulse.

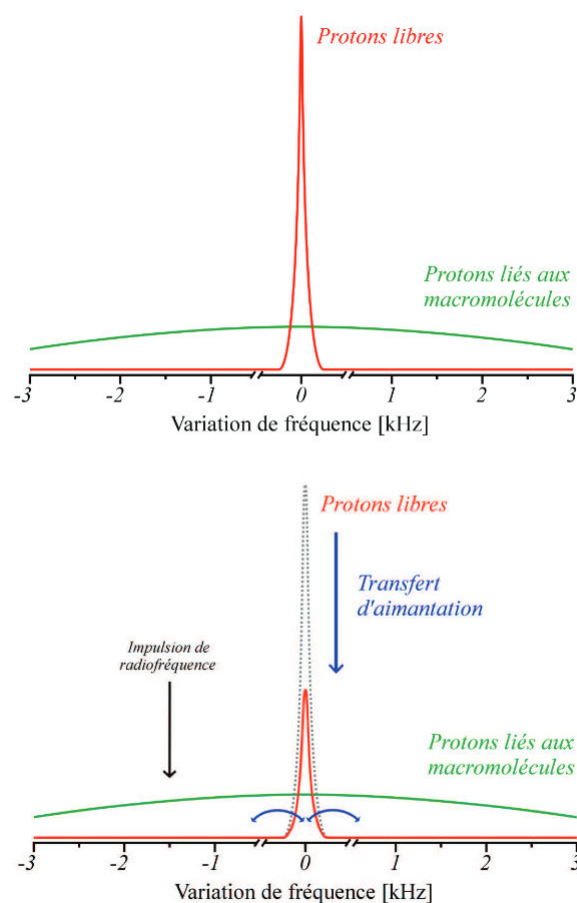


FIGURE 1.4 – Principe du transfert d'aimantation, d'après Pinson (2015).

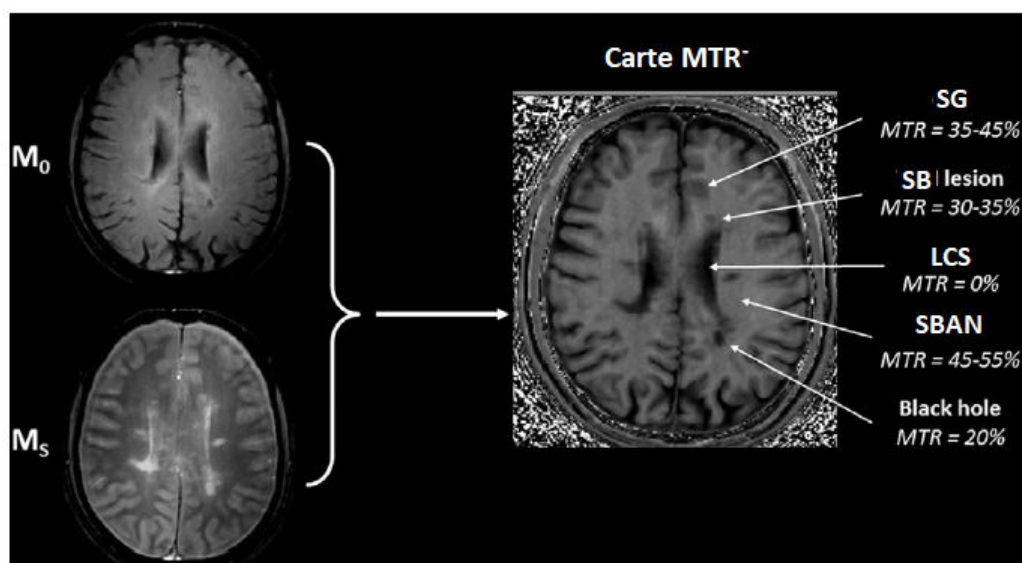


FIGURE 1.5 – Exemple d'application de l'imagerie de transfert d'aimantation, avant l'application de l'impulsion de saturation (M_0) et après impulsion (M_s). La cartographie du ratio de transfert d'aimantation (MTR) correspondante est alors calculée (image de droite), d'après Filippi et al. (2017), modifié.

Un faible MTR indique que la réduction du signal est inférieure à la normale

en raison d'une capacité réduite des macromolécules à échanger une aimantation avec les molécules d'eau environnantes. En conséquence, cet indice fournit une estimation de l'étendue de la perturbation tissulaire de la SEP, et le MTR est modifié lorsque la myéline ou d'autres structures cellulaires (par exemple, les neurones) sont endommagées. La myéline influe fortement sur la MTR, mais elle peut également être influencée par l'inflammation et la densité axonale.

Le MTR est considéré comme un marqueur de l'abondance de la myéline dans le cerveau. Des augmentations ou diminutions significatives du transfert d'aimantation au cours du temps peut donc être utilisé pour mesurer les processus de remyélinisation et de démyélinisation (Chen et al., 2008).

1.1.1.2.3 La spectroscopie par résonance magnétique

La spectroscopie par résonance magnétique (MRS⁷) est une technique d'imagerie permettant une étude du métabolisme cérébral. Son concept repose sur la détection de métabolites cérébraux, et l'étude des variations de leur concentration pour caractériser différentes pathologies. En pathologie cérébrale, on utilise essentiellement la spectroscopie du proton car on le retrouve en concentration importante dans le cerveau. Le signal RMN des protons alors obtenu fournit des renseignements sur les concentrations d'un certain nombre de métabolites. Les principaux métabolites détectables sont :

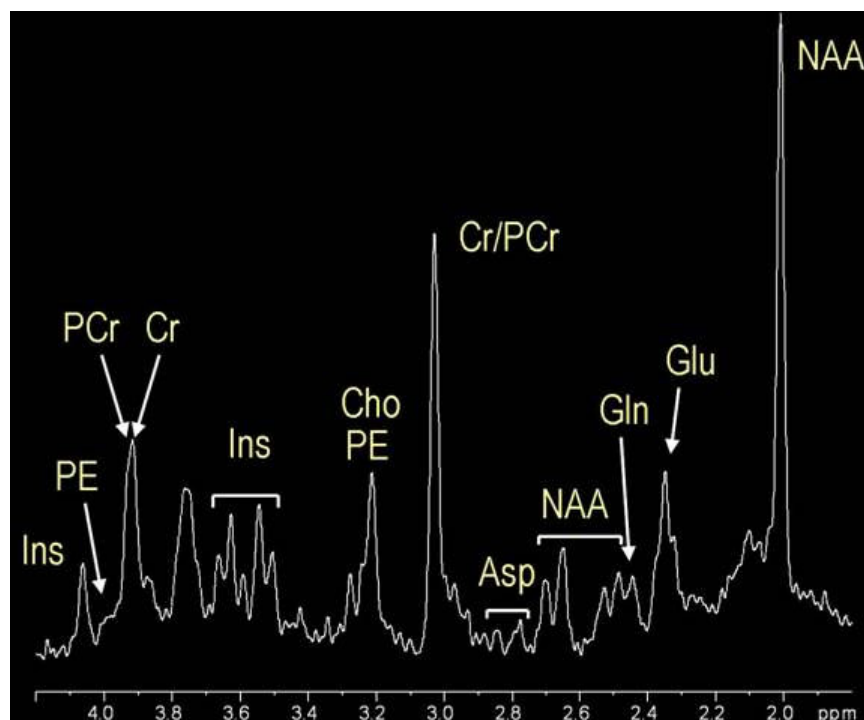


FIGURE 1.6 – Informations métaboliques obtenues à partir d'un spectre RMN du proton cérébral in vivo (source : <http://hirnforschung.kyb.mpg.de/en/methods/mrs.html>)

- Le N-Acétyl-Aspartate (NAA), marqueur de l'intégrité ou du fonctionnement axonal. Il diminue en cas de mort neuronales ;
- La Choline (Cho), marqueur du métabolisme phospholipidique et de la densité cellulaire. Elle augmente en cas d'inflammation ;

7. "Magnetic Resonance Spectroscopy" en anglais

- La Créatine (Cr), marqueur du métabolisme énergétique cellulaire ;
- Le Myoinositol (Ins), marqueur de l'inflammation et de la gliose ;
- Le Glutamate (Glu), neurotransmetteur, indicateur de neurotoxicité ;
- Le Lactate (Lac), indicateur de souffrance cellulaire liée à une ischémie, une hypoxie ou une réaction inflammatoire.

Pour plus de détails concernant les différentes modalités d'IRM, voir les ouvrages de [Hoa \(2007\)](#) et [Filippi \(2015\)](#).

1.1.2 L'IRM et la maladie d'Alzheimer

1.1.2.1 Généralités sur la maladie d'Alzheimer

La maladie d'Alzheimer (MA), initialement décrite par [Alzheimer](#) en 1906, est une maladie neurodégénérative qui est la cause la plus commune de démence chez les personnes âgées. Elle affecte le fonctionnement du système nerveux central par des dysfonctionnements génétiques ou métaboliques conduisant notamment à la mort de neurones. Cette dégénérescence du cortex cérébral s'aggrave au cours du temps entraînant une altération progressive des capacités cognitives. Au départ, la MA se manifeste par un déficit de la mémoire épisodique indiquant une atteinte des structures temporelles internes ([Hodges, 2006](#)), particulièrement dans la région du cerveau correspondant au siège de la mémoire à court terme, l'hippocampe. L'évolution de la maladie voit le processus neurodégénératif s'étendre à d'autres aires corticales entraînant secondairement une atteinte du langage et des fonctions instrumentales, des fonctions exécutives et des fonctions visio-spatiales, allant jusqu'à la perte d'autonomie qui signe le stade de la démence ([Ballard et al., 2011](#)).

1.1.2.1.1 Les différents stades d'évolution de la maladie

La MA est caractérisée par une altération progressive des capacités cognitives au cours de laquelle l'individu passe de sujet sain (vieillesse normale) vers la forme démentielle de la MA (MA probable) en passant par une phase de transition incluant des troubles cognitifs légers (stade MCI pour Mild Cognitive Impairment) :

- Stade MCI (trouble cognitif léger) : La notion de MCI a été introduite en 1999 par [Petersen et al.](#). Elle correspond à la phase de transition entre le vieillissement normal et la démence car ces troubles sont suffisamment limités pour ne pas impacter de façon significative la vie quotidienne des sujets. Tous les patients MCI ne convertissent pas vers la MA. Selon les cas, ils évoluent vers d'autres maladies ou restent stables.
- Stade MA probable : Pour qu'un patient soit déclaré atteint d'une démence de type Alzheimer, ou MA probable, il faut qu'il présente un déclin progressif et continu de ses capacités cognitives et qu'il présente des déficits dans au moins deux de ces domaines : altération de la mémoire, aphasie, apraxie, agnosie, altération des fonctions exécutives.

1.1.2.1.2 Diagnostic et biomarqueurs

Les études post-mortem ([Alzheimer, 1906](#); [Grignon et al., 1998](#)) sur des sujets Alzheimer ont mis en évidence les deux types de lésions cérébrales causant la perte neuronale et la dégradation des facultés cognitives : la dégénérescence neurofibrillaire (DNF) et les amas anormaux de fibrilles dans les neurones (plaques amyloïdes). Le diagnostic certain de la maladie d'Alzheimer ne peut être posé que post-mortem à partir d'une analyse histopathologique du cerveau révélant les deux types de lésions caractéristiques de la MA.

— Les critères du NINCDS-ADRDA⁸

Ces critères ([McKhann et al., 1984, 2011](#)) permettent de poser un diagnostic probable de la maladie. Ils s'appuient sur des tests cliniques et neuropsychologiques. Toutefois, ces critères possèdent des limites, en effet le phénomène de "réserve cognitive" ([Stern, 2002](#)) vient perturber le lien entre perte neuronale et altération cognitive car le cerveau de certains sujets, généralement ceux qui ont le plus haut niveau socioculturel, a la capacité de maintenir un fonctionnement normal malgré un certain nombre de déficits structurels.

— Les biomarqueurs de la MA

On distingue, parmi les principaux biomarqueurs de la maladie (d'après [Vanquin, 2015](#)) :

- Les biomarqueurs biologiques : détectés notamment dans le LCS, ils sont principalement les concentrations en protéines A β et protéines Tau ([Strozyk et al., 2003](#)).
- Les biomarqueurs de l'imagerie :
 - Les biomarqueurs des plaques amyloïdes, en imagerie TEP ;
 - Les biomarqueurs anatomiques et microstructuraux : principalement des biomarqueurs issus de l'IRM pondérée en T1, en T2 et en diffusion. Les biomarqueurs anatomiques donnent des informations quantitatives sur le volume des régions, leur épaisseur, leurs caractéristiques morphologiques ou leur composition ou, dans le cas du biomarqueur microstructurel, des informations quantitatives sur l'architecture fine du tissu neuronal ou sur l'intégrité du tissu à une échelle microscopique ;
 - Les biomarqueurs fonctionnels ; liés à l'activité cérébrale, ils sont notamment basés sur des variations de flux sanguin ; ce sont principalement des biomarqueurs issus de la TEP et de l'IRM fonctionnelle et de perfusion.

1.1.2.1.3 IRM structurelle et la maladie d'Alzheimer

La MA entraîne une perte neuronale que la neuroimagerie anatomique par IRM est capable de déceler et de mesurer indirectement en estimant l'atrophie induite par la MA ([Juottonen et al., 1999](#); [Bobinski et al., 2000](#)). Ainsi, la mesure de l'atrophie cérébrale serait un marqueur potentiel d'évolution depuis le stade MCI vers une MA. De plus, d'après [Smith \(2002\)](#) les déficits cognitifs sont plus liés à

8. Les critères du "National Institute of Neurological and Communicative Diseases and Stroke/Alzheimer's Disease and Related Disorders Association"

la perte de neurones qu'à la formation de DNF ou au dépôt de plaques amyloïdes. En effet, il semblerait que le nombre de neurones morts est bien plus élevé que le nombre de ceux qui contiennent des DNF (Killiany et al., 2002). De ce fait, la perte neuronale serait un meilleur marqueur de la pathologie que les techniques de mesure directe des DNF, par ailleurs plus coûteuses et moins accessibles (Mueller et al., 2006).

L'IRM structurelle (sMRI⁹) est importante pour le diagnostic différentiel de la MA en raison de sa capacité à visualiser des profils d'atrophie spécifiques dans le cerveau. Nous considérons ainsi que les biomarqueurs tels que, l'épaisseur corticale et la courbure corticale présenteraient un bon reflet des déficits cognitifs. L'application de sMRI a révélé un schéma spécifique d'amincissement cortical chez les adultes plus âgés qui semble être associé à la MA. Ce modèle, connu sous le nom de "signature" corticale de la MA (Dickerson et al., 2009a), comprend les régions médiales, inférieures et les zones du lobe temporal; le gyrus angulaire; le lobe frontal supérieur et inférieur; le lobule pariétal supérieur; le gyrus supramarginal; et le precuneus. Dans ces zones, l'amincissement est important chez les patients MA (Braak and Braak, 1991; Lerch and Evans, 2005) et chez les sujets présentant un risque de MA, les MCI (Bakkour et al., 2009; Morris, 2009; Dickerson et al., 2009b). Les individus ayant des signes de formes fibrillaires de l'amyloïde présentent un amincissement cortical dans ces régions par rapport aux individus ne présentant pas de signe amyloïde (Dickerson et al., 2009a, 2011). Pris ensemble, ces résultats ont conduit les chercheurs à conclure que cette signature corticale est un marqueur biologique valide et fiable de la MA (Bakkour et al., 2009, 2013; Dickerson and Wolk, 2013).

Pour plus de détails sur sMRI, notamment la mesure de l'épaisseur et la courbure corticale, voir les thèses de Querbes (2009) et de Vanquin (2015).

1.1.2.1.4 Imagerie de diffusion et la maladie d'Alzheimer

De nombreuses études DTI sur la MA ont révélé une augmentation de la diffusivité moyenne (DM) et une diminution de la fraction d'anisotropie (FA) dans la substance blanche, qui sont des indicateurs prometteurs de la MA (Amlie et al., 2014). Étant donné que DM et FA sont influencés par la diffusion de l'eau dans les plans radial et axial, les diffusivités radiale D_{\perp} et axiale $D_{//}$ pourraient elles-mêmes donner une représentation plus précise des lésions tissulaires de la substance blanche en MA (Song et al., 2002, 2004) et peuvent être utiles pour la détermination des stades de la maladie en décrivant différentes pathologies de la substance blanche (Acosta-Cabronero et al., 2012). Cependant, la nature précise des dommages causés par la substance blanche dans la maladie d'Alzheimer n'a pas été caractérisée et on ignore si elle survient avant ou après les dommages dans la substance grise (Nir et al., 2013). Les corrélations entre les changements de DTI dans la SG et les pathologies des tissus restent à étudier, mais il a été démontré que la diffusivité de la SG avait un pouvoir prédictif supérieur par rapport à la volumétrie (Weston et al., 2015).

9. "Structural Magnetic Resonance Imaging" en anglais.

1.1.2.1.5 Imagerie de transfert d'aimantation et la maladie d'Alzheimer

Un nombre croissant de publications ont utilisé avec succès MTI pour caractériser des troubles neurodégénératifs, y compris la MA ([Tambasco et al., 2015](#)). [Ropele et al. \(2012\)](#) montre que MTI permet l'évaluation des lésions cérébrales en cours, indépendamment de l'atrophie, et apparaît donc comme un marqueur précieux des modifications tissulaires liées à la maladie. Plusieurs études ont examiné le ratio de transfert d'aimantation (MTR) et ont constaté qu'il diminuait dans le cerveau entier ([Ridha et al., 2007](#); [Es et al., 2010](#)), zones corticales, lobes temporaux ([Bozzali et al., 2001](#)) et dans l'hippocampe ([Hanyu et al., 2000](#)). [van der Flier et al. \(2002\)](#) ont trouvé des amplitudes de pics réduites dans les histogrammes MTR chez les MCI et AD par rapport aux sujets sains, reflétant des dommages structurels au cerveau. [Mascalchi et al. \(2013\)](#) ont souligné une diminution significative du MTR dans l'hippocampe gauche, l'amygdale et le cortex temporal médial postérieur gauche de patients atteints de MA, mais aucune différence n'a été observée entre le MCI et les témoins.

1.1.2.1.6 Spectroscopie par résonance magnétique et la maladie d'Alzheimer

Trois décennies de recherche indiquent que la MRS est un marqueur d'imagerie biochimique potentiel dans la maladie d'Alzheimer. Elle permet une mesure régionale des métabolites, notamment le myo-inositol(mI), la choline(Cho), le N-acétyl aspartate(NAA) et la créatine(Cr). En 1992, [Klunk et al.](#) ont démontré une diminution du métabolite neuronal NAA des patients atteints de MA par rapport aux témoins. Le NAA est considéré comme un marqueur de l'intégrité neuronale. Plusieurs études ont examiné la capacité de la MRS à distinguer les patients AD des témoins, avec des résultats variables en fonction de la région anatomique analysée et des paramètres d'acquisition. La sensibilité était de 90% dans la région temporo-pariétale et inférieure à 57% dans SG du lobe pariétal. La spécificité atteint 95% dans le lobe occipital médial et 73% dans le cingulaire postérieur ([Shonk et al., 1995](#); [Zhu et al., 2006](#); [Fernando et al., 2005](#); [Martínez-Bisbal et al., 2004](#)). En outre, l'ajout du volume de l'hippocampe au MRS améliore la capacité de distinguer la MA ([Kantarci et al., 2002](#); [Schuff et al., 2002](#)). Les taux de NAA/Cr dans le MCI sont légèrement réduits mais diminuent à mesure que les patients MCI progressent vers la MA. De plus, une diminution du NAA/Cr chez les patients MCI prédit la progression vers la MA ([Kantarci et al., 2007](#); [Chao et al., 2005](#); [Metastasio et al., 2006](#)). Le rapport Cho/Cr est également utile pour déterminer la progression de MCI vers AD. Chez les patients MCI, une baisse de Cho/Cr prédit la stabilité alors qu'une augmentation prédit la conversion vers MA.

Voir la revue de [Graff-Radford and Kantarci \(2013\)](#) pour plus de détails concernant MRS et l'Alzheimer et pour plus de détails sur l'IRM en général et la maladie d'Alzheimer, voir la revue de [Promteangtrong et al. \(2015\)](#).

1.1.3 L'IRM et la Sclérose en plaques

La sclérose en plaques (SEP) est une maladie neurodégénérative, inflammatoire et démyélinisante du système nerveux central. Elle a été décrite pour la première fois en 1868 par le neurologue français . La maladie affecte principalement le sexe féminin (3 femme pour un 1 homme) et le jeune adulte (20 ans à 40 ans). Actuellement, elle touche plus de 110.000 cas en France et 2.3 millions dans le monde (d'après l'ARSEP). L'origine exacte de la SEP n'est pas encore confirmée, il s'agit d'une maladie multifactorielle, elle peut être due à des facteurs propres à l'individu (facteurs génétiques) mais aussi à des facteurs environnementaux (e.g. les sujets d'origine nord-européenne sont plus fréquemment touchés).

L'IRM est l'examen complémentaire le plus fréquemment utilisé, donnant une indication sur la nature des tissus et faisant notamment apparaître les plaques de démyélinisation, évocatrices de la SEP. Ainsi, l'IRM a un rôle prépondérant dans le diagnostic (van Walderveen et al., 1995; Tourbah and Berry, 2000), le suivi des patients (Filippi et al., 2002), mais aussi l'évaluation des nouvelles thérapeutiques (McFarland et al., 2002; Río et al., 2017).

1.1.3.1 Généralités sur la sclérose en plaques

1.1.3.1.1 Physiopathologie

La SEP est une maladie du système nerveux central. le SNC comprend la moelle spinale et l'encéphale, ce dernier est constitué de trois régions : le cerveau, le tronc cérébral et le cervelet. Le cerveau constitue la plus grande partie du SNC. Il renferme des cellules nerveuses responsables de la réception et la transmission de l'influx nerveux : Les neurones. Ils sont formés d'un corps cellulaire et d'un prolongement d'axone. L'axone est entouré d'une gaine : La myéline. Elle assure la protection et la nutrition de l'axone. Sa fonction principale est d'accélérer la conduction de l'influx nerveux.

La SEP est une maladie auto-immune : le système immunitaire reconnaît la myéline comme un agent étranger et l'attaque. Ce phénomène entraîne des lésions dispersées dans le système nerveux central. Ces lésions sont appelées plaques de démyélinisation. Ces plaques sont responsables d'une altération de la conduction nerveuse entraînant ainsi un handicap en fonction de la localisation des plaques, qui peut être suivi d'une dégénérescence axonale

Dans un premier temps, l'atteinte de la gaine de myéline est légère (les axones ne sont pas touchés) ce qui cause un ralentissement de l'influx nerveux. Suite à ces perturbations, deux possibilités, soit une remyélinisation plus ou moins complète, qui s'accompagne d'une cicatrisation, soit au contraire une aggravation de la démyélinisation, ce qui au fil du temps causera une destruction axonale (figure 1.7).

1.1.3.1.2 Clinique

La SEP est une maladie imprévisible dont l'évolution, polymorphe, est le reflet de l'interaction entre les poussées et la progression du handicap.

- Une poussée : La survenue de nouveaux symptômes d'installation subaiguë, qui

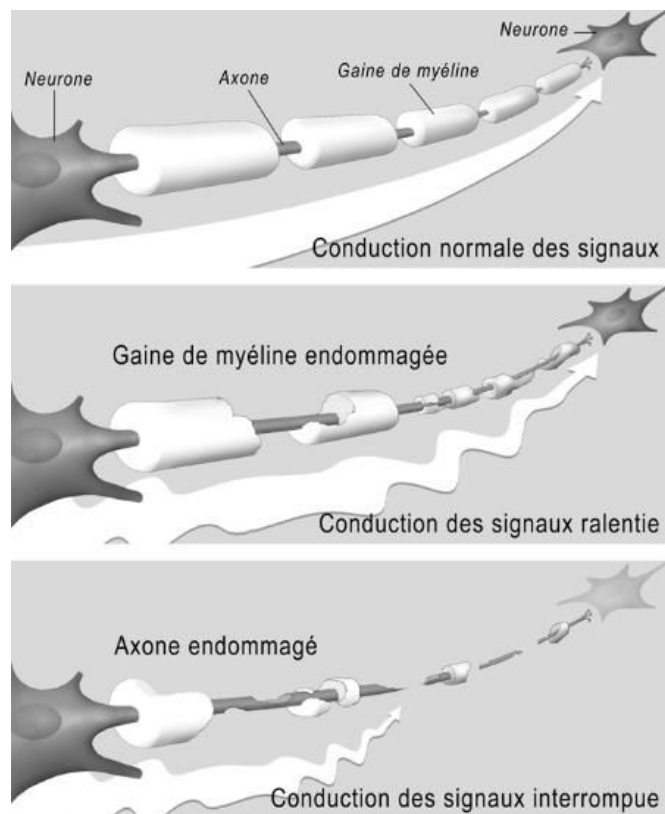


FIGURE 1.7 – L’atteinte de la gaine de myéline dans le cadre de la SEP

durent au moins 24heures, suivis souvent d’une phase de plateau puis une récupération qui peut être complète.

- La progression : Une aggravation continue des signes cliniques du patient, indépendamment des poussées, sur une période d’au moins six mois. La progression de la maladie est mesurée à l’aide d’une échelle : l’Expanded Disability Status Scale (EDSS) qui a été proposée par Kurtzke en 1983. Elle permet de décrire les déficits liés à la maladie et d’évaluer l’évolution du handicap. Elle repose sur un examen neurologique standardisé ou plusieurs paramètres sont évalués.

L’évolution de la maladie est très variable selon les personnes. Elle est classée selon trois formes évolutives principales (Weill and Batteux, 2003) figure 1.8 :

- **La SEP Rémittente-Récurrente (SEP-RR)** : Cette évolution est caractérisée par l’apparition de poussées suivies d’une phase de rétablissement (avec ou sans séquelles). Absence de progression entre les poussées. Cette forme concerne la plupart des patients SEP, environ 85% (Magy, 2009).
- **La SEP secondairement progressive (SEP-SP)** : Évolution naturelle de la forme SEP-RR. Après une phase initialement rémittente de 15 années en moyenne, la phase de poussées est suivie d’une phase d’accumulation progressive du handicap, à laquelle peuvent se surajouter des poussées.
- **La SEP progressive primaire (SEP-PP)** : Les patients font leur entrée dans la maladie par une progression continue du handicap, avec ou sans poussées ajoutée. Cette forme est plus grave et plus rare et touche 15% des patients.

En ce qui concerne les symptômes de la SEP, dû au polymorphisme clinique

de la maladie, aucun symptôme n'est spécifique de la SEP. Ils sont extrêmement variables d'un patient à l'autre : fatigue, douleurs musculaires, troubles oculaires, troubles moteur...

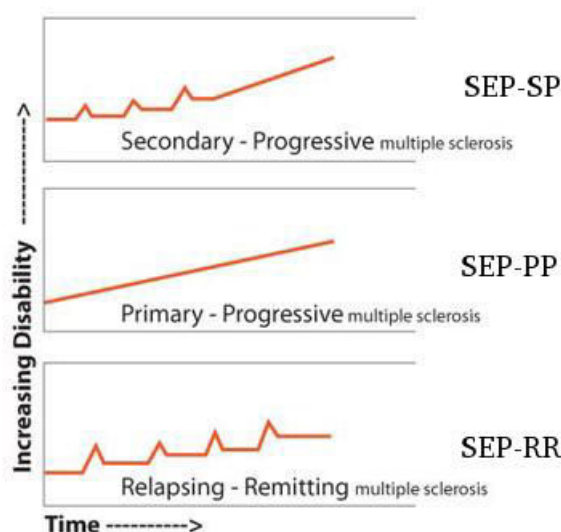


FIGURE 1.8 – Les trois formes cliniques de la SEP (source : <https://cnfs.ca/pathologies/sclerose-en-plaques>)

1.1.3.1.3 Diagnostic

A ce jour, il n'existe pas de test diagnostique spécifique à la SEP, ainsi le diagnostic repose sur un faisceau d'arguments cliniques, biologiques et radiologique. Il suppose aussi l'absence de toute autre affection susceptible d'expliquer les signes observés. Le diagnostic de la SEP repose sur les caractéristiques suivantes :

- **La dissémination dans le temps (DIT¹⁰)**, soit l'évolution de la maladie sous la forme de poussées et rémissions ou en imagerie, la démonstration de coexistence de plaques d'âges différents (apparition entre deux examens successifs et/ou prises de contraste limitées à certaines plaques).
- **La dissémination dans l'espace (DIS¹¹)** correspond à l'atteinte de plusieurs zones du SNC, de manière simultanée ou successive, soit démontrée cliniquement soit par imagerie ce qui revient à compter et localiser les plaques de démyélinisation

Cette notion de dissémination dans le temps et l'espace peut faire défaut au début de la maladie et dans les formes progressives. Ainsi, il est indispensable de passer des examens complémentaires :

- **Examen clinique**

La SEP étant une maladie polymorphe l'examen clinique est important. Il permet de mettre en évidence certains signes évocateurs tels que les antécédents personnels et familiaux du patient, la nature des signes cliniques

10. "dissemination in time" en anglais

11. "dissemination in space" en anglais

et leur évolution dans le temps.

— L'imagerie par résonance magnétique

L'IRM est l'examen complémentaire le plus approprié ; il occupe donc une place fondamentale dans le diagnostic. Il est d'ailleurs effectué dès le premier événement clinique. Les critères diagnostiques de la dissémination dans le temps et dans l'espace ont beaucoup évolué depuis McDonald ([McDonald et al., 2001](#); [Polman et al., 2005, 2011](#); [Thompson et al., 2018](#)). Sur le tableau 1.2 nous présentons les révisions 2017 de ces critères ([Thompson et al., 2018](#)).

Poussées	Lésions objectives	Données supplémentaires nécessaire pour le diagnostic de SEP
≥2 Poussées cliniques	≥2 lésions	Aucune
≥2 Poussées cliniques	1 lésion	DIS démontrée par une nouvelle poussée clinique impliquant un autre territoire du SNC OU par IRM
1 Poussée clinique	≥2 lésions	DIT démontrée par une nouvelle poussée clinique OU par IRM OU BOC spécifiques du LCR
1 Poussée clinique	1 lésion	DIS démontrée par une nouvelle poussée clinique impliquant un autre territoire du SNC OU par IRM DIT démontrée par une nouvelle poussée clinique OU par IRM OU BOC spécifiques du LCR

Tableau 1.2 – Les critères 2017 de McDonald pour le diagnostic de la SEP.

BOC : Bandes Oligoclonales, DIS : Dissémination dans l'espace, DIT : Dissémination dans le temps

— **L'inflammation au niveau du SNC** (Examen du liquide céphalo-rachidien LCS) : Lors d'un mécanisme inflammatoire, le LCS, prélevé par ponction lombaire, montre le plus souvent un syndrome inflammatoire : on retrouvera ainsi une augmentation du nombre de lymphocytes et une hyperprotéinorchie de profil assez spécifique.

1.1.3.1.4 Traitement

Aujourd'hui les médicaments permettent globalement de freiner la maladie mais pas de la guérir. L'efficacité thérapeutique se juge sur la fréquence des poussées et sur l'évolution du handicap neurologique.

- **Le traitement des poussées** : Combattre l'inflammation et raccourcir la durée des poussées (les corticoïdes).
- **Les traitements de fond** : Espacer les poussées, diminuer leur intensité et à plus long terme, ralentir la progression du handicap (immunomodulateurs et les immunosuppresseurs).
- **Le traitement des symptômes** : Diminuer l'intensité des symptômes ressentis par les patients.

1.1.3.1.5 Les essais cliniques en sclérose en plaques

Un essai clinique est une étude scientifique réalisée en thérapeutique médicale humaine pour évaluer l'efficacité et la tolérance d'une méthode diagnostique ou

d'un traitement. Pour garantir la sécurité des sujets et la rigueur scientifique, les essais cliniques comprennent quatre phases qui sont chacune destinées à recueillir des informations spécifiques sur le nouveau traitement.

Pour les patients atteints de la forme SEP-PP ou SEP-SP de la sclérose en plaques, malgré des différents phénotypes cliniques initiaux, la durée des périodes observées avant le franchissement de certaines étapes importantes du handicap, ainsi que l'âge auquel ces étapes sont atteintes, sont similaires. En conséquence, des directives récentes ont proposé de les regrouper en une seule entité dénommée "maladie progressive" (SPP). On estime que le taux de prévalence globale des patients atteints d'une maladie progressive correspond à au moins 40% de tous les patients atteints de sclérose en plaques.

La sclérose en plaques progressive (SPP) est décrite comme l'accumulation graduelle de lésions nerveuses et d'incapacités cliniques. A ce jour, contrairement à la forme rémittente-récurrente (SEP-RR) (Miller and Rhoades, 2012), aucun traitement efficace sur la SPP n'a été identifié et les résultats des essais cliniques restent majoritairement décevants (Miler AE et al., 2012 ; Wiendl H et al., 2009). Cela est dû au fait que les processus pathologiques sous-jacents qui la définissent restent flous et aucun biomarqueur n'a été totalement validés comme étant suffisamment sensible dans la forme SPP.

— La problématique de la SEP progressive dans les essais cliniques

Les patients qui ont reçu un diagnostic de SPP sont souvent frustrés par le nombre relativement bas d'essais cliniques sur la SPP (spécialement la primaire-progressive) par rapport au nombre important sur la SEP-RR. Les cliniciens et les chercheurs de la SEP partagent cette frustration et recherchent activement des moyens d'augmenter le nombre d'essais de traitements pour les SPP, en s'attaquant à plusieurs obstacles :

- Les médicaments actuellement utilisés pour traiter les formes récurrentes de la SEP (SEP-RR) ciblent principalement une inflammation dans le système nerveux central (SNC). Sachant que l'inflammation joue un rôle moins important dans la SPP que dans les formes récurrentes de la SEP, ces médicaments ne semblent pas être aussi efficaces dans la SPP, ce qui signifie que de nouvelles cibles de traitement doivent être identifiées.

- Dans la SPP, il y a un manque de marqueurs facilement identifiables qui puissent être mesurés dans les essais. Dans les essais concluants, les chercheurs ont examiné les marqueurs tels que le nombre de poussées et le nombre de nouvelles lésions à l'aide de l'IRM afin de déterminer si les personnes ayant reçu le traitement avaient un nombre inférieur de poussées et de nouvelles lésions par rapport à ceux qui ont reçu un placebo, mais dans le groupe SPP, ces mesures ne quantifient pas vraiment la progression du handicap.

- La progression du handicap dans les SPP peut être assez lente, ce qui rend difficile l'identification d'un effet du médicament sur la progression dans un essai de deux ou trois ans.

— Bilan sur les essais cliniques de la SEP progressive

Les études cliniques ont impliqué plus de 8500 sujets. Différentes catégories de traitement ont été étudiées, y compris les immunosuppresseurs classiques, les interférons bêta et les nouveaux immunomodulateurs. Malgré de grands espoirs,

les résultats des essais de phase III sont essentiellement négatifs. Cependant, on s'aperçoit que la méthodologie d'essai clinique moderne a tiré des leçons des 20 dernières années, on constate : un plus grand nombre de sujet, l'utilisation de biomarqueurs plus pertinents et des durées d'essai de phase III plus réalistes.

Les thérapies récentes de la SEP sont présentés dans les revues de [Coclitu et al. \(2016\)](#), [Correale et al. \(2017\)](#) et [De Angelis et al. \(2018\)](#).

1.1.3.2 IRM conventionnelle et sclérose en plaques

L'IRM conventionnelle est un outil très sensible pour détecter l'inflammation et pour visualiser les lésions et cela de façon très précoce. Elle consiste en l'utilisation de séquences en pondération T2, T1, FLAIR et séquences pondérées T1 après injection de Gadolinium.

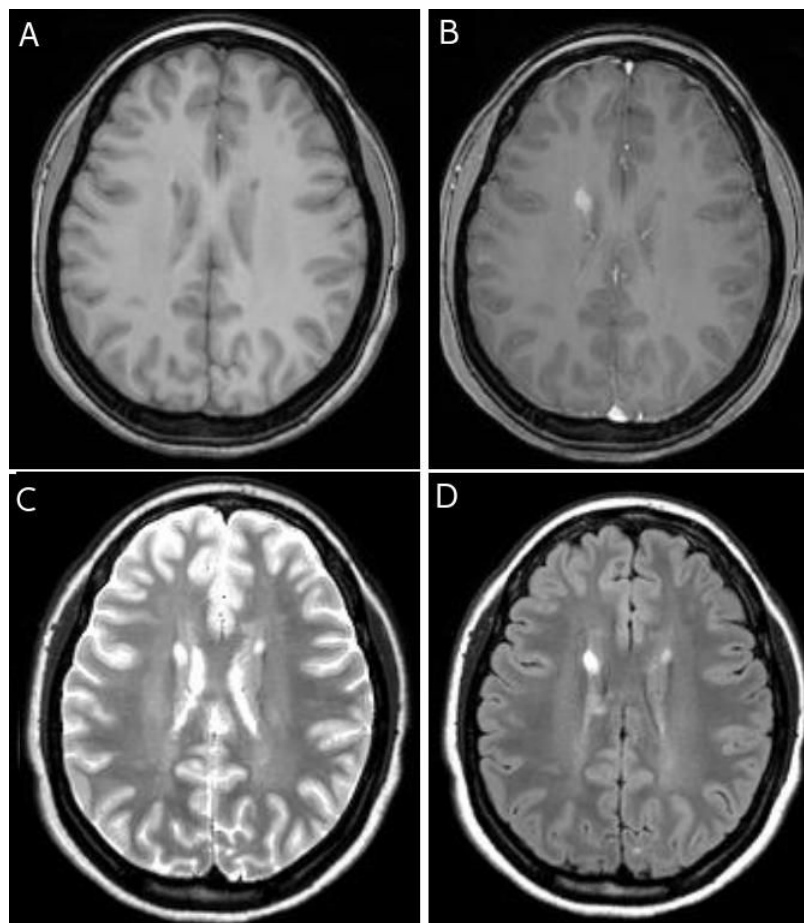


FIGURE 1.9 – IRM conventionnelle en SEP
(A) IRM pondérée en T1 sans injection Gd, lésions en iso- ou hypo-signal ; (B) IRM pondérée en T1 après injection Gd, lésions en hyper-signal ; (C) IRM pondérée en T2, lésions en hyper-signal ; (D) Séquence FLAIR, lésions en hyper-signal (acquisitions effectuées au sein du plateau IRM de l'INSERM ToNIC U1214, d'après [Dutilleul \(2015\)](#))

— IRM pondérée en T1

L'IRM pondérée en T1 permet de mettre en évidence les lésions anciennes et chroniques, qu'on appelle les "trous noirs" car elles apparaissent en hyposignal sur l'IRM pondérée en T1 (figure 1.9-A). Ces plaques représentent une destruction tissulaire sévère (lésions de démyélinisation et perte axonale). Ainsi l'IRM pondérée en T1 permet d'étudier l'ampleur des processus de réparation

tissulaire via l'étude de la persistance ou l'atténuation de ces hypo-intensités en T1.

la séquence pondérée en T1 avec injection de Gadolinium (figure 1.9-B) est une acquisition pondérée en T1 effectuée après l'injection par voie intraveineuse d'un produit de contraste qui est les chélates de Gadolinium (Gd). Les lésions rehaussées à l'injection du Gd représentent les lésions récentes et actives (inflammatoires). Ainsi cette séquence apporte une information sur l'activité inflammatoire et le caractère évolutif de la maladie.

D'autre part, l'IRM pondérée en T1 nous permet de mettre en évidence l'atrophie cérébrale. L'atrophie cérébrale est devenue un phénomène reconnu en SEP (Chard et al., 2004; Azevedo and Pelletier, 2016; Zivadinov et al., 2016). Elle est caractérisée par l'élargissement ventriculaire, la diminution du volume du corps calleux et par l'amincissement du ruban cortical.

C'est l'atrophie cérébrale qui est la plus fortement corrélée à l'évolution clinique de la SEP et permettrait de prédire la détérioration de la fonction motrice et cognitive. On sait que l'atrophie cérébrale se produit lors du vieillissement normal, avec une perte annuelle du volume cérébral d'environ 0.2 à 0.5% (Fjell et al., 2009) alors que chez les patients atteints de SEP, le taux d'atrophie est estimé entre 0,5 et 1,3% par an (Vågberg et al., 2013), probablement en raison d'une perte du volume de la SG. Généralement, l'atrophie cérébrale semble être plus importante chez les patients atteints de SEP progressive que chez ceux atteints de SEP rémittente-récurrente. Pourtant, plusieurs études ont montré qu'une perte importante de volume peut déjà se produire chez les patients atteints d'une SEP-RR récente (De Stefano et al., 2002; Chard et al., 2004) et chez les patients atteints de syndrome cliniquement isolé¹² (CIS). Plus récemment, une perte de volume du parenchyme cérébral de 0,4% par an a été proposée comme "le taux d'atrophie pathologique" pour définir les patients atteints de SEP (De Stefano et al., 2016) et présentait une forte corrélation avec l'EDSS. Compte tenu de sa corrélation avec le handicap, l'utilisation plus systématique des mesures d'atrophie cérébrale a été suggérée comme paramètre des études cliniques.

— IRM pondérée en T2 et séquence FLAIR

Ces séquences ont une grande sensibilité pour la détection des lésions de la SEP. Les plaques apparaissent en hypersignal et représentent soit des lésions évolutives inflammatoires et oedémateuses, soit des lésions anciennes et non actives (non-inflammatoires) qui témoignent d'une démyélinisation suivie d'une éventuelle perte axonale. Les hypo-intensités observées sur les séquences pondérées en T1 apparaissent sur les images pondérées en T2 (figure 1.9-C) ou en FLAIR en hyper-intensités (figure 1.9-D).

La séquence pondérée en T2 représente mal les lésions péri-ventriculaires et juxta-corticale qui ont souvent un signal similaire au liquide cérébrospinal (LCS) adjacent, d'où l'intérêt d'utiliser la séquence FLAIR, qui est une séquence pondérée en T2 mais sans la contribution du signal du LCS, ainsi les lésions sont

12. CIS est un premier épisode de démyélinisation d'origine inflammatoire, i.e. une poussée que l'on observe pour la première fois et qui représente pour la personne atteinte un risque accru que sa SEP se manifeste cliniquement complètement, c'est-à-dire selon les critères de dissémination temporelle et spatiale

en hypersignal et le LCS en hyposignal, ce qui permet une bonne différenciation et une meilleure visibilité.

— Paradoxe clinico-radiologique

Ces méthodes d'imagerie conventionnelles ont leurs limites, notamment du fait de la persistance d'un "paradoxe clinico-radiologique" (Barkhof, 2002). En effet les données issues de l'imagerie ne sont pas corrélées aux marqueurs cliniques. Cela est lié à la fois à un manque de spécificité de ces techniques permettant d'appréhender des processus physiopathologiques hétérogènes (inflammation, démyélinisation, atteinte axonale) et à l'insuffisance de leur sensibilité pour évaluer l'étendue des dommages tissulaires en dehors des lésions focales.

1.1.3.3 IRM de diffusion et sclérose en plaques

Dans la sclérose en plaques, les séquences de diffusion fournissent des informations sur les altérations dans le cerveau, la moelle épinière et le nerf optique. En particulier, il a été démontré que le DTI pouvait détecter des modifications pathologiques non visibles en IRM conventionnelle. Ces séquences ont permis de mettre en évidence des anomalies de la substance blanche dite d'apparence normale (SBAN) (Filippi et al., 2000a; Cercignani et al., 2000; Ge et al., 2004; Cassol et al., 2004; Ceccarelli et al., 2007).

Les lésions de SEP présentent des valeurs élevées de diffusivité (Castriota-Scanderbeg et al., 2002), une réduction de l'ADC y a également été rapportée (Tievsky et al., 1999). D'autre part, les études DTI ont démontré que la FA est réduite dans les lésions de SEP par rapport à la SBAN et au cerveau normal. Bien que le degré le plus élevé d'anomalies de diffusion ait été observé dans les lésions T1 hypo-intenses (Werring et al., 2000, 1999; Filippi et al., 2000a). Il ne semble pas y avoir de différences entre les lésions rehaussées par le Gd et non rehaussées (Droogan et al., 1999).

L'analyse en imagerie de diffusion de la SBAN a montré une diminution de la FA et une augmentation de la DM par rapport aux sujets contrôles (Werring et al., 1999; Bammer et al., 2000; Cercignani et al., 2001b). Cela s'interprète par le fait qu'il y ait un plus grand espace extracellulaire au niveau de la SBAN, dû à une démyélinisation, une perte axonale concourant à une certaine désorganisation progressive du tissu cérébral. De plus, il a été rapporté que les indices de diffusion dans la SBAN étaient en corrélation modérée avec le handicap clinique (Ciccarelli et al., 2001; Rovaris et al., 2002b). Cependant, il a été rapporté que les patients atteints de SEP primaire progressive SEP-PP ne présentaient aucune corrélation entre les paramètres de diffusion et le handicap. Par exemple, Filippi et al. (2001) n'ont trouvé aucune corrélation significative entre les mesures EDSS et de diffusion dans la SBAN ou dans les lésions, contrairement à ce qui a été observé dans les cas de SEP secondairement progressive (SEP-SP). Une autre étude sur les patients SEP-PP, utilisant l'analyse d'histogrammes des paramètres de diffusion, confirme ces résultats (Cercignani et al., 2001a).

Nous retrouvons ces observations pour la substance grise, il a été constaté (particulièrement dans les noyaux gris centraux) une plus grande diffusivité chez les patients atteints de SEP par rapport à des sujets témoins (Bozzali et al., 2002). Quant à la FA, elle est plus faible dans la SG que dans la SB et cela est liée

à la constitution anatomique même du tissu où il y a moins de myéline et où l'anisotropie est moins importante (Ciccarelli et al., 2001).

1.1.3.4 IRM de transfert d'aimantation et sclérose en plaques

Le MTR est considéré comme un marqueur de l'abondance de la myéline dans le cerveau. Des augmentations ou diminutions significatives du transfert d'aimantation au cours du temps peut donc être utilisé pour mesurer les processus de remyélinisation et démyélinisation respectivement (Chen et al., 2008).

Chez les patients SEP, des valeurs de MTR anormalement basses ont été trouvées dans la SBAN. Dousset et al. (1992) ont été les premiers à observer ce phénomène et plusieurs autres groupes l'ont reproduit (Hiehle et al., 1994; Pike et al., 1999). Les valeurs de MTR dans la SBAN et SGAN sont légèrement inférieures à celles du tissu cérébral normal et ont été attribuées à la présence de modifications microscopiques diffuses et étendues : l'existence d'œdème ou d'inflammation (diminution relative de la proportion d'eau liée) ainsi qu'une démyélinisation (Schmierer et al., 2004) et une perte axonale (van Waesberghe et al., 1999).

Ainsi, le MTI semble fournir de bons marqueurs pour le suivi de la SEP. En conséquence, plusieurs essais cliniques récents sur la SEP ont déjà incorporé le MTI afin d'évaluer l'impact d'un traitement sur la démyélinisation et la perte axonale. MTI a été utilisée dans les essais de phase II et de phase III de la SEP-RR et SEP-SP. Richert et al. (2001) n'ont pas trouvé de différence significative dans les valeurs de MTR de la SBAN avant et après le traitement interféron bêta-1b, ainsi que dans les paramètres dérivés des histogrammes du MTR du cerveau entier chez les patients atteints de SEP-RR (Richert et al., 1998).

1.1.3.5 Spectroscopie par résonance magnétique et sclérose en plaques

La MRS n'a pas encore trouvé de véritable application clinique dans la SEP et reste une technique utilisée en recherche uniquement.

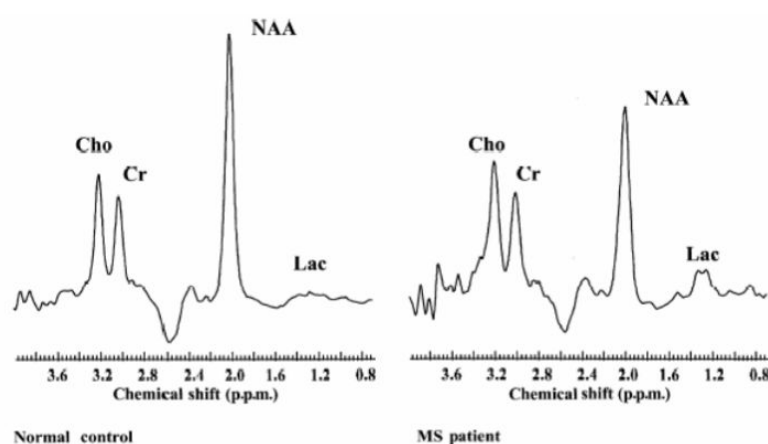


FIGURE 1.10 – Exemple de spectres obtenus chez un volontaire (spectre de gauche) et chez un patient porteur de SEP (à droite) Le rapport NAA/Cr est significativement diminué chez le patient par rapport au témoin (D'après De Stefano et al. (1998))

Elle a toutefois permis de recueillir des informations sur l'évolution naturelle

des plaques de démyélinisation, de révéler des anomalies métabolites dans la SBAN et de mieux comprendre les mécanismes des lésions cérébrales causées par la SEP.

La sclérose en plaques est marquée par une augmentation de la choline et une diminution du NAA (figure 1.10). Ces perturbations sont constatées dès le stade précoce de la maladie et dans la SBAN et la SGAN et démontrent l'activité inflammatoire, une démyélinisation et voire même une perte axonale (Filippi, 2001; Chard et al., 2002).

Pour une description plus approfondie de la sclérose en plaques, voir la thèse de Dutilleul (2015), la revue de Kaunzner and Gauthier (2017) et de Filippi et al. (2017).

1.2 Les différentes approches d'analyse d'images

L'un des challenges en imagerie médicale est le post-traitement et l'interprétation des images. Au cours des deux dernières décennies, un nombre croissant de techniques d'analyse d'images quantitative ont été développées, dont l'analyse d'histogrammes, de régions d'intérêt et l'analyse voxel-à-voxel. Chaque approche a ses avantages et ses inconvénients. Il n'existe pas de procédure standard en imagerie cérébrale, le choix de l'approche dépend de la problématique et de l'objectif de l'étude mais aussi des hypothèses formulées. Ces approches sont représentées sur la figure 1.11 et résumées dans le tableau 1.3

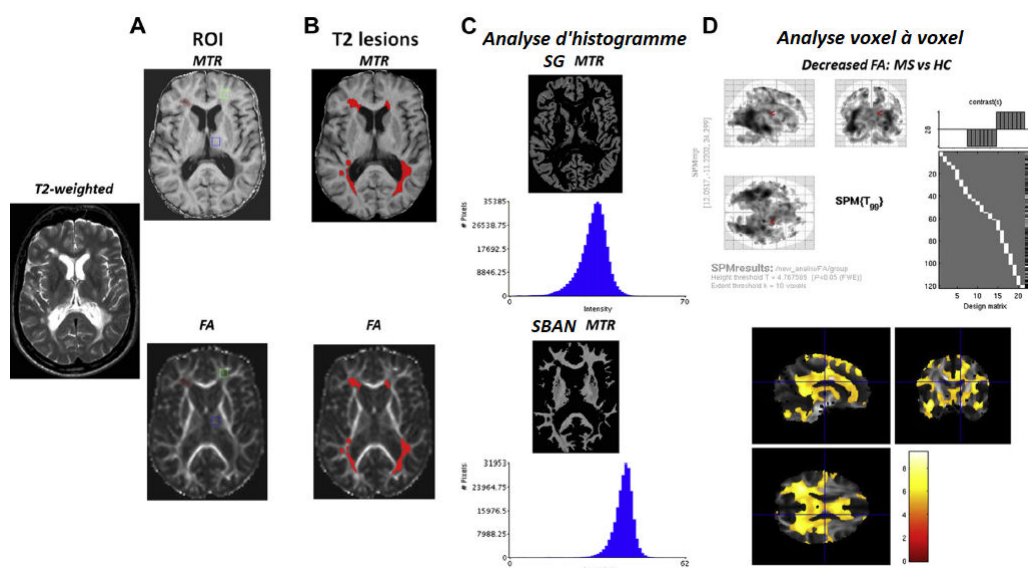


FIGURE 1.11 – Représentation schématique des différentes approches pouvant être appliquées à l'analyse des anomalies liées à la SEP sur les cartes de l'imagerie par transfert d'aimantation MTR et l'imagerie de diffusion FA. (A) Analyse par ROI avec sélection d'une lésion de la SB (carré rouge), une région de SBAN (carré vert) et une région de SG (carré bleu). (B) Analyse des lésions T2 sur les cartes MTR et FA. (C) Analyse d'histogrammes MTR de la SG et de la SBAN. (D) Analyse voxel à voxel. Comparaison des FA au niveau de voxel entre les patients SEP et sujets contrôles HC (d'après Filippi et al. (2017), modifié)

1.2.1 Analyse par régions d'intérêt

La technique la plus utilisée pour prendre en compte une information spatiale est l'approche par régions d'intérêt (ROI). L'étude est centrée sur une ou

plusieurs zones particulières du cerveau (e.g. des lésions visibles ou de grands volumes de substance blanche d'apparence normale), où les propriétés statistiques (e.g. valeurs moyennes) des intensités ou de paramètres IRM (e.g. diffusion, transfert d'aimantation...) à l'intérieur de ces régions d'intérêt sont comparées soit à une autre ROI de tissu sain chez le même patient, à travers le temps ou entre différentes populations de sujets. Une fois l'analyse par ROI choisie par l'examineur, ce dernier devra faire un certain nombre de choix, tels que la définition des régions, le choix de la forme et de la taille de ces ROI.

Le choix des ROI, étape cruciale pour la pertinence des résultats, peut être effectué de différentes manières, que l'on peut mettre en parallèle avec les différentes techniques de labellisation. Une des méthodes pour définir les ROI repose sur l'utilisation d'atlas, il est possible de normaliser les données dans un espace standard (e.g. espace de Talairach) mais cela a pour conséquence que les régions obtenues ne reposent plus sur les anatomies spécifiques de chaque sujet. Une autre méthode très répandue est la segmentation manuelle des ROI, elle n'a pas besoin de normalisation et permet ainsi de s'affranchir de la variabilité anatomique mais à de nombreux égards, elle est limitée et inadéquate. Comme il a été précisé par [Astrakas and Argyropoulou \(2010\)](#), il y a un fort risque d'erreur lorsque la région d'intérêt est un organe ou une structure aux bordures complexes. De plus, dans les études longitudinales, il est difficile de répéter exactement la même géométrie et de reproduire exactement les mêmes ROI d'une étude à l'autre. Enfin, dans le cas où il n'y a pas de connaissance à priori de la localisation de régions d'intérêt, le choix des ROI devient compliqué et augmenter le nombre de ROI n'est pas une bonne solution, car en ajoutant des régions non pertinentes, la spécificité spatiale et la puissance statistique sont réduites. Afin de remédier aux limites des méthodes manuelles, de nombreux algorithmes ont été conçus pour la segmentation semi-automatique et automatique des structures d'intérêt, ces derniers accélèrent le processus et améliorent la reproductibilité. Il peut s'agir de simples techniques de seuillage ou des méthodes de reconnaissance de formes plus complexes. Leurs applications cliniques rencontrent des obstacles. En effet, les plus avancés et les plus puissants d'entre eux requièrent des compétences particulières en mathématiques et en calcul et malgré leur variété, il n'existe pas d'algorithme de segmentation universel pouvant être appliqué de manière satisfaisante dans tous les cas.

Les ROI peuvent avoir une forme circulaire, ovale, carrée, rectangulaire ou irrégulière. Elles peuvent être définies dans une seule coupe ou s'étendre sur plusieurs coupes (l'ensemble est alors un volume d'intérêt [VOI]). Les ROI restent souvent grossières, de la taille des gyri, ce qui ne permet pas des détections focales précises. En ce qui concerne la taille des ROI, un compromis est fait entre la réduction du bruit (qui favorise les ROI de grande taille) et la réduction des erreurs d'estimation de volume partiel (qui favorise les ROI de petite taille).

En résumé, les principaux avantages de cette approche sont que les régions peuvent être choisies sur la base d'hypothèses à priori et peuvent être situées dans une partie spécifique du cerveau. Aussi, elle permet d'atténuer le problème des comparaisons multiples, dans lequel un espace de recherche de centaines de milliers de voxels est réduit à des zones plus petites et plus maniables, réduisant ainsi les seuils de correction des comparaisons multiples. D'autre part, cette approche peut passer à côté d'anomalies significatives dans les régions non

sélectionnées et dépend fortement de l'observateur et des indices anatomiques utilisés pour positionner les régions. Le processus de recalage et les problèmes méthodologiques associés doivent être soigneusement pris en compte lorsque des images de différentes modalités sont recalées afin de permettre le transfert de régions. Il est également nécessaire de repositionner les ROI exactement au même endroit dans les études longitudinales.

1.2.2 Analyse voxel à voxel

Une approche alternative à l'analyse par ROI est une méthode sans hypothèses *a priori* dans laquelle les régions avec une différence statistique significative sont automatiquement détectées entre les groupes : l'analyse voxel à voxel. Elle peut être considérée comme une analyse de régions d'intérêt où les ROI seraient représentées par les plus petites entités d'une image 3D, c'est à dire les voxels. La méthode repose sur des algorithmes automatisés, car il n'est pas humainement possible d'étudier les valeurs de milliers de voxels sur différentes images.

L'un des exemples les plus populaires de ce type d'analyse est la morphométrie voxel-à-voxel (VBM). La méthode VBM, développée par [Ashburner and Friston \(2000\)](#) et implémentée dans le logiciel [SPM](#), a pour objectif de détecter des différences significatives liés à des anomalies tissulaires au sein d'un groupe ou entre plusieurs groupes par des tests voxel à voxel. La VBM repose sur des mesures en chaque voxel de l'image de la concentration locale de tissu en termes de niveau de gris d'images T1, ou encore en termes de coefficient de diffusion dans l'imagerie de diffusion. Ces images sont lissées par un lissage gaussien. Puis, toutes les cartes obtenues pour les différents sujets sont mises en correspondance sur un espace commun (i.e. l'espace MNI). L'analyse des différences entre groupes repose sur des tests statistiques paramétriques de type test de Student réalisés voxel à voxel suivis d'une correction pour comparaisons multiples permettant d'éviter l'apparition de nombreux faux positifs. Les tests mettent ainsi en évidence les voxels pour lesquels il existe une différence significative de concentration entre groupes.

Cette technique est puissante pour effectuer des tests non biaisés sur tous les voxels et attirer l'attention sur les régions où d'autres études pourraient être fructueuses. Mais, bien que l'approche type voxel-à-voxel ait été utilisée avec succès en neuroimagerie pour identifier et caractériser les changements locaux liés aux maladies fonctionnelles et structurelles, elle reste très sensible à divers artefacts, notamment la distorsion des structures cérébrales, la classification erronée des types de tissus... Tous ces facteurs peuvent fausser l'analyse statistique et augmenter le risque de faux positifs. D'autre part, il s'agit d'une méthode d'analyse de groupe qui n'est pas conçue pour l'analyse individuelle des patients, il est donc difficile d'envisager un suivi longitudinal d'un patient donné. De plus, elle ne permet que la localisation des différences significatives (i.e. anomalies anatomiques) entre groupe et non leur quantification. En général, cette méthode a été critiquée pour son incapacité à détecter des changements structurels complexes sur le plan spatial, nécessitant des techniques alternatives.

1.2.3 Analyse d'histogrammes

L'analyse par histogramme est une technique quantitative utilisée dans plusieurs études en neuroimagerie, c'est une approche alternative aux ROI ([Law et al., 2007](#); [Young et al., 2007](#)). D'ailleurs, cette approche peut être considérée comme l'opposée de l'analyse par ROI car toute information spatiale est perdue. Elle est généralement appliquée dans des conditions qui imposent des changements globaux et généralisés, dans lesquelles choisir des ROI n'a pas grand intérêt (e.g. la sclérose en plaques). L'histogramme est la représentation graphique la plus couramment utilisée pour représenter la distribution de probabilité d'une variable quantitative en décrivant les fréquences dans certaines plages de valeurs. Des métriques peuvent être dérivées à partir de l'histogramme : moyenne, quantiles (dont la médiane), amplitude du pic, position du pic... Ces variables sont ensuite utilisées dans des tests statistiques, par exemple en comparant des patients et des sujets sains. L'analyse d'histogrammes permet de caractériser les paramètres IRM, tels que les paramètres de diffusion, dans tout le cerveau de manière entièrement automatisée. De plus, elle évite tout biais dû à la connaissance à priori des zones du cerveau susceptibles d'être affectées par la maladie. Cela évite également de devoir placer des ROI sur les images, spécialement dans les études longitudinales, où il faut repositionner les ROI exactement au même endroit sur les images. La section qui suit [1.3](#) détaillera plus amplement cette approche, qui a d'ailleurs inspiré ce travail de thèse.

1.2. LES DIFFÉRENTES APPROCHES D'ANALYSE D'IMAGES

Méthodes	Avantages	Inconvénients
Régions d'intérêt	<ul style="list-style-type: none"> - Cette étude se concentre sur des zones particulières du cerveau - Segmentation manuelle : Pas besoin de normalisation et ainsi s'affranchit de la variabilité anatomique - Segmentation semi-automatique et automatique : accélèrent le processus et améliorent la reproductibilité 	<ul style="list-style-type: none"> - Long processus - Connaissance à priori sur la localisation de la pathologie : Introduction d'un biais - Correction des comparaisons multiples nécessaire - Utilisation d'atlas : Les régions obtenues ne reposent pas sur les anatomies spécifiques de chaque sujet. - Segmentation manuelle : Fort risque d'erreur lorsque la ROI est un organe ou une structure aux bordures complexes. - Segmentation semi-automatique et automatique : Complexes ; Compétences particulières en mathématiques et en calcul requises.
Voxel-à-voxel	<ul style="list-style-type: none"> - En partie automatique - Relativement rapide - Tous les voxels sont testés d'une façon non biaisée 	<ul style="list-style-type: none"> - Très sensible à divers artefacts (Distorsion...) - VBM souffre d'un paramétrage complexe et ne permet que la localisation des différences significatives et non leur quantification - Correction des comparaisons multiples nécessaire
Histogrammes	<ul style="list-style-type: none"> - Aucune hypothèse à priori n'est nécessaire - Simple et rapide - Opérateur indépendant - Le repositionnement sur des analyses répétées est inutile - Très utile dans l'étude des maladies diffuses (e.g. Sclérose en plaques) 	<ul style="list-style-type: none"> - L'information spatiale est perdue - La recherche aujourd'hui n'a que superficiellement exploré cet outil

Tableau 1.3 – Les trois principales méthodes d'analyse d'images quantitatives

1.3 Analyse d'histogrammes en imagerie cérébrale

Les histogrammes représentant les valeurs des paramètres IRM mesurés dans tout le cerveau sont de plus en plus utilisés pour caractériser des maladies diffuses qui affectent de grandes parties du cerveau. L'application la plus courante des histogrammes est celle du paramètre MTR de l'imagerie de transfert d'aimantation dans l'étude de la Sclérose en plaques (première étude par [Buchem et al., 1996](#)). Cependant, cette approche est également applicable à tout paramètre IRM susceptible d'être affecté par des changements diffus dans le cerveau.

Dans cette section, on présentera comment générer les histogrammes des paramètres IRM, que peut-on extraire de ces derniers et enfin les différentes applications cliniques de l'analyse d'histogrammes.

1.3.1 Production et analyse d'histogrammes

Dans cette section, les principes de production d'histogrammes sont expliqués, avec entre autres des informations sur le choix du type de normalisation, le nombre adéquat de bins, ainsi que les métriques qui peuvent être extraites des histogrammes ([Tofts et al., 2004a](#)).

1.3.1.1 Carte paramétrique

Il y a une différence distincte entre les images provenant de l'IRM et les cartes paramétriques. Les images brutes contiennent des informations fondamentales spécifiques à la modalité. Les cartes paramétriques sont ultérieurement dérivées d'une série d'images à l'aide de modèles mathématiques et de mesures physiques (e.g. contenu métabolique, diffusion, relaxation de spin...).

Une carte paramétrique peut être calculée à partir de deux images (ou plus) du même tissu. Un exemple simple consisterait à collecter deux images de pondération T2 différentes. Le rapport de ces deux images ne dépend alors que du paramètre T2 du tissu et est indépendant des paramètres du scanner. En calculant ce rapport pour chaque pixel (ou voxel dans le cas d'une image 3D), il est possible de former une troisième matrice, ou carte paramétrique, qui a l'apparence d'une image cérébrale (les structures cérébrales peuvent être identifiées) mais est conceptuellement différente de l'image brute IRM. En effet, les valeurs des pixels ont désormais une signification numérique et physique (telle que la valeur de T2, en millisecondes, à chaque position du cerveau), et ne représentent pas juste l'intensité du signal. Quelques exemples :

- Une carte T2¹³ dérivée d'images pondérées en T2 avec différents temps d'écho. Le modèle mathématique est une fonction exponentielle qui décrit la relation inverse entre le temps d'écho et l'intensité du signal. T2 peut ensuite être calculé en ajustant le modèle de décroissance aux données.
- Une carte MTR (ratio de transfert d'aimantation) dérivée de deux images de densité de protons acquises avec et sans MT pulse. Sur la base des différences d'intensité du signal entre les deux acquisitions, une carte MTR peut être calculée.

13. relaxation transverse (en s)

- Carte ADC (coefficient de diffusion apparent) dérivée de deux images pondérées en diffusion avec différents gradients de diffusion (valeurs b). La carte obtenue est basée sur la pente de la courbe de décroissance du signal avec des valeurs b croissantes.

En résumé, le calcul des cartes paramétriques est basé sur une approche voxel à voxel. Le paramètre IRM est calculé pour chaque voxel, puis une carte paramétrique est synthétisée en remplaçant l'intensité du pixel par la valeur du paramètre calculé correspondant.

1.3.1.2 Quantification des histogrammes

Dans cette section, nous donnerons plus de détails sur la mise en œuvre pratique des caractéristiques de l'histogramme :

1.3.1.2.1 Acquisition des données

Les images doivent être acquises en tenant compte d'un certain nombre de points :

- Obtenir des valeurs précises du paramètre IRM en question. En effet, la non uniformité du champ B1 élargira l'histogramme (pour MTR et T1 mais pas pour la diffusion). Les erreurs non corrigées d'angle de bascule modifieront l'histogramme dans certains cas (e.g. MTR), ou encore de petits changements de la position du pic peuvent être importants, et il est donc important de faire attention aux techniques d'acquisition de données.
- De plus petits voxels donneront moins d'erreurs de volume partiel, mais cela est au détriment du bruit.
- Les cartes paramétriques doivent être stockées à une résolution suffisante.

1.3.1.2.2 Segmentation des images

Le processus de segmentation doit être reproductible, précis et aussi indépendant que possible d'un jugement à priori. L'une des plus grandes difficultés est de savoir comment traiter les voxels du volume partiel, e.g. ceux qui contiennent un mélange de cerveau et de LCS. En effet, un seul voxel peut contenir plusieurs types de tissus. Ce phénomène, appelé effet de volume partiel (EVP), complique le processus de segmentation. Ainsi, en raison de la complexité de l'anatomie du cerveau humain, l'EVP est un facteur important pour la quantification précise de la structure du cerveau. Plusieurs approches sont proposées dans la littérature, un critère trop strict rejeterait de grandes quantités de tissu cérébral susceptibles d'être biologiquement utiles pour évaluer l'effet de la maladie étudiée. Alors qu'un critère plus souple risquerait de fausser les résultats et leurs interprétations.

Pour plus d'information sur les différentes approches de segmentation en prenant compte l'EVP, voir la revue de [Tohka \(2014\)](#) et le rapport de [Grande-Barreto and del Pilar Gomez-Gil \(2017\)](#).

La segmentation doit être effectuée sur des images PD, T1 ou T2, dont la pondération est différente de celle utilisée pour générer la carte paramétrique. Par exemple, si un histogramme MTR est généré, le processus de segmentation ne devrait (idéalement) pas être réalisé sur cartes MTR (sinon une modification de

la valeur MTR modifierait la segmentation). Les pixels de faible valeur MTR, tels que ceux du LCS, doivent être supprimés par segmentation d'une image PD, T1 ou T2 et non d'une carte MTR. D'autre part, si le cerveau d'apparence normale est étudié, les lésions devront d'abord être identifiées et retirées (Tortorella et al., 2000).

1.3.1.2.3 Choix du nombre de bins

Un histogramme d'image indique le nombre d'occurrences de chaque valeur du paramètre IRM dans l'ensemble de données. Chaque ensemble d'occurrences est un "bin" de cet histogramme. Les bins sont aussi appelé les classes d'un histogramme.

"Combien de bins doit contenir l'histogramme?". Bien que l'histogramme soit la méthode la plus simple pour l'estimation de distribution de probabilité, il n'y a toujours pas de réponse définitive à cette question. De nombreux travaux antérieurs ont tendance à utiliser une valeur qui semble appropriée, sans justification statistique derrière cette analyse. Dans la littérature, il existe plusieurs suggestions de choix pour le nombre de classes, où N est le nombre d'observations (ou voxels) et K le nombre de bins :

- Visuel : faire varier le nombre de bins jusqu'à atteindre un compris qui faciliterait l'interprétation.
- Le choix le plus simple : $K = \sqrt{N}$
- La règle de Sturge : $K = 1 + 3.322 \log(N)$. Cette règle fonctionne bien dans le cas où les données sont continues et de distribution normale. Mais, pas dans le cas asymétrique.
- Règle de Scott : $3.49\sigma N^{1/3}$. Elle est basée sur l'écart-type σ des données.
- Règle de Freedman-Diaconis : $2(IQR)N^{1/3}$. Elle est basée sur l'écart interquartile IQR .

Réduire trop le nombre de bins dégradera considérablement l'information contenue dans l'image, ce qui signifiera la perte de caractéristiques distinctes. Les histogrammes ont tendance à prendre deux formes ; régulière ou irrégulière. Lorsqu'on discute du nombre de bins d'un histogramme, il est typique d'utiliser un histogramme régulier dans lequel la taille des bins est uniforme. Les histogrammes irréguliers, bien qu'ils puissent offrir une plus grande souplesse de classification, nécessiteraient une intervention de l'utilisateur, ce qui les rendraient inadéquats à une procédure automatisée. Ainsi, il n'existe pas de règle adaptée à toutes situations, parfois il serait plus approprié de sélectionner des méthodes qui reposent sur des observations expérimentales plutôt que sur une base statistique.

1.3.1.2.4 Lissage d'histogrammes

Pour éviter l'apparition de discontinuités (aussi appelées "spikes"¹⁴) sur l'histogramme des cartes paramétriques, les images doivent être aussi continues que possible. Le processus de de-spiking est important car l'amplitude ou la

14. Le processus s'appelle "image de-spiking"

position des pics d'histogrammes peuvent être complètement modifiés par des "spikes" indésirables, tandis que pour des analyses plus complexes, ces spikes peuvent provoquer de subtils changements dans les résultats qui peuvent passer inaperçus par le scientifique.

Les spikes d'histogrammes proviennent de diverses sources, telles que le bruit des images, des erreurs au niveau de l'acquisition ou du traitement des images. Ces spikes sont généralement atténués par l'application d'un filtre médian, ou par le choix de larges bins, ce qui lisse l'histogramme (Tozer and Tofts, 2003).

1.3.1.2.5 Normalisation des histogrammes

Après que les paramètres IRM soient collectées et les cartes paramétriques formées, un histogramme absolu est généré. Un histogramme absolu h^{vpb} ¹⁵ est un histogramme non normalisé. Il représente le nombre de voxels par unité du paramètre IRM (e.g. MTR, ADC...). La taille des classes de l'histogramme est choisie pour diviser la plage des valeurs du paramètre IRM en un nombre approprié d'intervalles, les bins. L'aire sous cette courbe est le nombre total de voxels (équivalent au volume total du cerveau). Étant donné que le volume cérébral varie considérablement d'un individu à l'autre, l'histogramme est généralement normalisé. Il existe deux méthodes de normalisation (figure 1.12) :

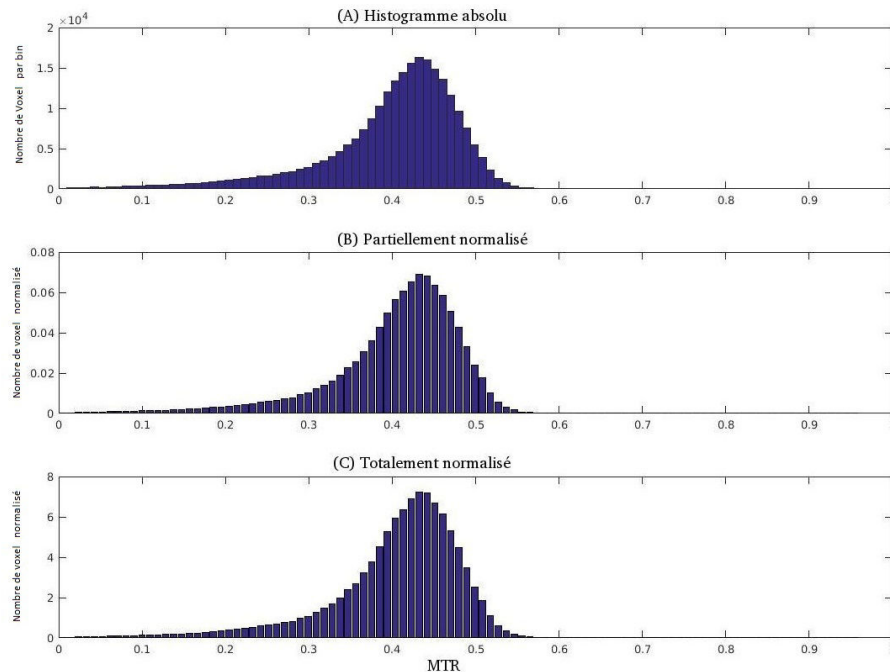


FIGURE 1.12 – Histogramme MTR de la SGAN partiellement et totalement normalisé.

(A) Histogramme absolu (non normalisé). (B) L'histogramme partiellement normalisé h^{np} a été calculé en divisant chaque valeur de l'histogramme par la somme totale des valeurs (sans prendre en compte la largeur des bins) ; la hauteur maximale de l'histogramme dépend de la largeur des bins (inversement proportionnelle). (C) L'histogramme totalement normalisé h^{nt} a été calculé en tenant compte de la largeur des bins ; il est indépendant de la largeur des bins.

- Normalisation partielle : Diviser toutes les valeurs de l'histogramme par le nombre total de voxels. Une distribution de fréquence est alors obtenue.

15. histogramme voxel per bin : histogramme absolu créé en comptant le nombre de voxels dans chaque bin

L'histogramme h^{np} est ainsi corrigé du volume cérébral mais reste dépendant de la taille des bins.

$$h^{np} = \frac{h_i^{vpb}}{\sum_i^N h_i^{vpb}} \quad (1.8)$$

- Normalisation totale : Diviser toutes les valeurs de l'histogramme par le nombre total de voxels et par la taille du bin Δ . L'aire sous la courbe est ainsi égal à un. L'histogramme h^{nt} est corrigé du volume cérébral et est indépendant de la taille des bins. Cette approche est plus adaptée dans le cas de l'étude d'histogrammes provenant de différents sites.

$$h^{nt} = \frac{h_i^{vpb}}{\Delta \sum_i^N h_i^{vpb}} \quad (1.9)$$

avec h^{np} histogramme partiellement normalisé, h^{nt} histogramme totalement normalisé, h_i^{vpb} histogramme absolu et Δ taille¹⁶ du bin.

1.3.1.3 Les métriques d'histogrammes

Les caractéristiques de l'histogramme peuvent être calculées pour réduire les informations de l'histogramme à quelques paramètres destinés à contenir les informations importantes dans l'histogramme. Les caractéristiques classiques appropriées aux histogrammes sont : L'amplitude et la position du pic, la valeur moyenne, skewness¹⁷, kurtosis¹⁸ et les quantiles (la médiane représente le 50ème quantile) du paramètre IRM. Ces résumés d'histogrammes sont définis en détails dans la section 3.1.2.

1.3.2 Les applications cliniques

L'analyse d'histogrammes permet de quantifier le changement pathologique diffus qui, dans les maladies neurologiques, peut fournir des informations utiles sur les mécanismes sous-jacent de la pathologie. La sclérose en plaques est un bon exemple où l'application de cette approche est appropriée. Mais également d'autres maladies, notamment la maladie d'Alzheimer, l'épilepsie, la neuropathie optique héréditaire de Leber ont été étudiées avec cette approche.

1.3.2.1 Les origines biologiques

Les histogrammes du cerveau entier sont sensibles aux changements diffus de grands volumes de tissus (par opposition aux modifications focales : "lésions"). Ils peuvent représenter tout paramètre IRM susceptible d'être modifié par la présence d'une maladie diffuse. Ils ont la capacité de détecter de petits changements dans la valeur du paramètre, en partie parce qu'un grand nombre de voxels sont regroupés (réduisant ainsi l'effet du bruit), et en partie parce que les problèmes de segmentation et de recalage sont moins cruciaux que dans l'analyse par ROI (Tofts et al., 2004a).

16. taille ou largeur du bin

17. correspond à l'asymétrie

18. coefficient d'aplatissement

Au niveau des grandes lésions, e.g. les tumeurs, l'analyse d'histogrammes peut être utile, en particulier si les lésions sont hétérogènes, car une simple valeur moyenne ne serait pas appropriée, bien que le résultat puisse être sensible à la procédure de segmentation.

La position du pic de l'histogramme correspond à la valeur paramétrique la plus répandue dans le tissu cérébral. Elle est étroitement liée à la valeur moyenne (bien que ce ne soit pas la même chose, en particulier pour les histogrammes asymétriques). Si le tissu normal est homogène, le pic peut être très bien définie. Un changement de la position du pic peut survenir si un grand nombre de voxels sont affectés; ou encore, le pic peut rester à la même position mais perdre en amplitude lorsqu'un volume important de tissu recueille des valeurs anormales du paramètre.

1.3.2.2 L'analyse d'histogrammes en sclérose en plaques

L'étude de la Sclérose en plaques repose non seulement sur l'aspect focale de la pathologie (lésions ou plaques de la SEP), mais également sur les altérations diffuses observables dans les tissus d'apparence normale (Evangelou et al., 2000). Les mesures quantitatives d'IRM, y compris le MTR, la diffusion et le temps de relaxation T1, sont sensibles à ces anomalies et c'est dans l'étude de ces altérations diffuses que l'analyse d'histogrammes a été utilisée.

Il a été démontré que le nombre et le volume des lésions de la SEP ne sont que modestement liés au handicap (Molyneux et al., 2001; Barkhof, 2002), ce qui suggère que ce sont les changements dans les tissus d'apparence normale qui refferment la majeure partie de l'information. L'analyse par histogramme est le moyen idéal de quantifier les changements dans ces tissus. Il n'est donc pas surprenant que cette vue d'ensemble montre une plus forte relation avec le handicap que les seules mesures du volume des lésions.

1.3.2.2.1 Histogramme MTR

L'analyse des histogrammes MTR fournit, à partir d'une simple procédure, de multiples métriques d'histogrammes qui sont influencées par des altérations macroscopiques et microscopiques, et pourraient être utilisées pour mesurer l'évolution de la SEP, que ce soit l'évolution naturelle de la maladie ou modifiée par un traitement lors d'un essai clinique.

De nombreuses études ont été réalisées à l'aide des histogrammes MTR. Certaines ont montré que les métriques de ces histogrammes présentaient une moyenne à forte corrélation avec le handicap physique et neuropsychologique (Filippi et al., 1999; Traboulsee et al., 2003). Ces corrélations se sont avérées plus fortes chez les patients atteints de SEP-RR et de SEP-SP que dans d'autres phénotypes de la maladie (Dehmeshki et al., 2001a; Kalkers et al., 2001). D'autres études ont mis en avant des corrélations avec les troubles cognitifs dans la SEP (Rovaris et al., 1998; van Buchem et al., 1998; Filippi et al., 2000b). L'étude de Rovaris et al. (2000) a révélé, à l'aide d'une régression multivariée, que les histogrammes MTR dérivés des tissus corticaux et sous-corticaux constituaient les seuls éléments associés aux troubles cognitifs.

Un certain nombre d'études ont examiné la capacité des histogrammes MTR à montrer des différences entre les formes cliniques de la SEP (Al-Radaideh et al.,

2015). Une étude (Tortorella et al., 2000) a révélé qu'il existait une différence entre les patients SEP-RR et les patients SEP-SP, l'amplitude du pic du tissu cérébral d'apparence normale étant plus faible chez ces derniers. Une vaste étude transversale (Filippi et al., 1999) a montré que les patients présentant un syndrome cliniquement isolé (SCI) et ceux présentant une SEP bénigne avaient un histogramme MTR similaire à celui des sujets sains. Les patients atteints de SEP-RR avaient des moyennes et des amplitudes de pic réduites, tandis qu'un changement similaire mais plus sévère était observé chez ceux atteints de la forme SEP-SP (figure 1.13).

Des études longitudinales ont montré que les métriques IRM étaient des marqueurs de l'évolution de la SEP. Il a été démontré par Santos et al. (2002) et Rovaris et al. (2003) que les métriques des histogrammes MTR de la SBAN permettent de prédire la progression du handicap clinique sur cinq ans. Quant aux métriques des histogrammes MTR de la SGAN, ils arrivent à prédire la progression du handicap sur trois ans (Khaleeli et al., 2008). L'étude de Agosta et al. (2006) quant à elle montre que les métriques MTR prédisent l'accumulation du handicap sur huit ans et suggère que les dommages au niveau de la SG sont l'un des facteurs clés associés à l'accumulation du handicap.

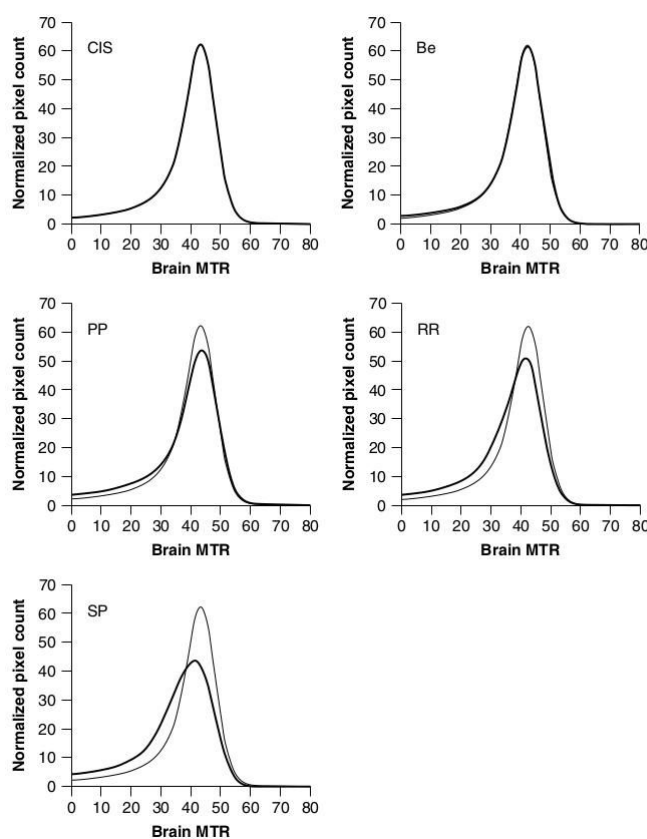


FIGURE 1.13 – Histogrammes MTR du cerveau entier de patients atteints de SEP (lignes épaisses) et de sujets sains (lignes fines). D'un point de vue général, on observe que chez les sujets SEP, le pic est diminué d'amplitude et dévié vers la gauche. La plus grande anomalie est observée sur l'histogramme des patients atteints de la plus sévère forme clinique de la maladie, la SEP-SP. Une partie de cette anomalie provient de lésions; alors que les plus petites anomalies sont observées au niveau des tissus d'apparence normale (d'après Filippi et al., 1999 et modifié par Tofts et al., 2004a).

Syndrome cliniquement isolé (CIS), SEP bénigne (Be), SEP progressive primaire (PP), SEP rémittente-récurrente (RR) et la SEP secondairement progressive (SP).

1.3.2.2.2 Histogramme de diffusion

L'analyse par histogramme a facilité l'étude de la SEP utilisant l'imagerie de diffusion (figure 1.14). Des variations de la DM et de la FA dans le cerveau entier et dans les tissus d'apparence normale ont été dévoilées à l'aide d'une analyse d'histogrammes de tous les principaux phénotypes cliniques de la SEP (Cercignani et al., 2000, 2001a; Nusbaum et al., 2000), y compris la SEP-PP (Rocca et al., 2003). L'étude de Yu et al. (2008) a montré que chez les patients atteints de SEP-RR, certaines mesures d'histogrammes de diffusion, DM et FA (principalement, de la SBAN) étaient significativement corrélées aux volumes des lésions T2 et la mesure clinique du handicap EDSS.

En utilisant l'analyse d'histogrammes, il a été démontré que les modifications de la DM impliquaient également la SGAN (et non juste SBAN) des patients atteints de SEP. Ces modifications ont tendance à être plus prononcées chez les patients présentant les formes évolutives de la maladie (Cercignani et al., 2001a; Bozzali et al., 2002; Rovaris et al., 2002a).

Les histogrammes de l'ADC montrant un décalage vers la droite et une amplitude de pic réduite chez les sujets atteints de SEP. Ce changement est lié de manière significative au handicap de la maladie, ce qui n'a pas réussi à être démontré avec une approche type ROI (Wilson et al., 2001). Tozer et al. (2006) ont utilisé une analyse en composantes principales et une analyse discriminante linéaire sur les histogrammes T1 et ADC de patients atteints de SEP et ont comparé les mesures de ces analyses aux mesures d'amplitude et de position du pic. Plus récemment, Bester et al. (2015) ont comparé le kurtosis moyen (MK)¹⁹, DM et FA entre les patients atteints de SEP et des sujets sains.

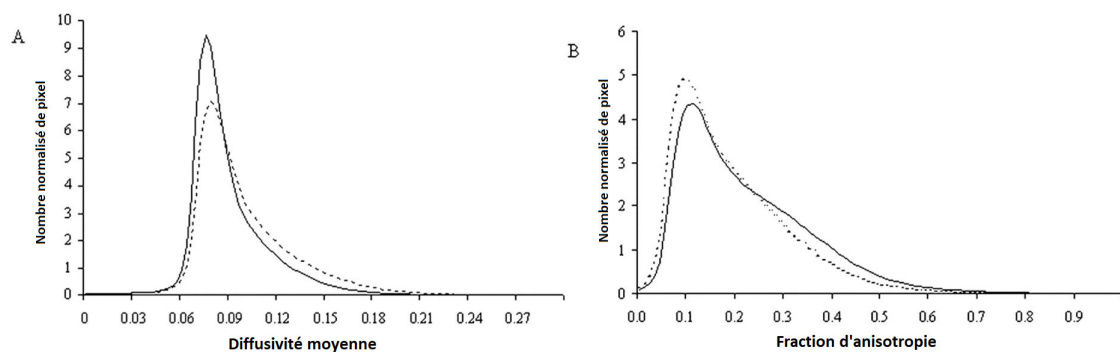


FIGURE 1.14 – Histogrammes des patients atteints de SEP (ligne en pointillée) et sujets contrôles (ligne en continue). (A) diffusivité moyenne, (B) Fraction d'anisotropie (d'après Cercignani et al., 2001b)

1.3.2.3 Autres maladies neurologiques

Nombreuses autres pathologies neurologiques ont été étudiées à l'aide d'histogrammes :

La maladie d'Alzheimer a été étudiée en utilisant à la fois la diffusivité moyenne et les histogrammes MTR (Bozzali et al., 2001). L'amplitude du pic

19. La diffusion non gaussienne peut être mesurée à l'aide du modèle Kurtosis de l'imagerie de diffusion (DKI), une modalité IRM récemment mise au point. En plus des métriques traditionnelles dérivées du DTI, elle permet de mesurer le kurtosis moyen (MK), indice de la complexité microstructurale des tissus.

pour les deux paramètres au niveau du cortex des patients AD était anormalement basse par rapport aux sujets témoins. Plus récemment, deux études se sont intéressées à l'utilisation des histogrammes pour prédire et classer la population Alzheimer. Ruiz et al. (2018) proposent un nouveau CAD²⁰. En entrée du système, ils utilisent des histogrammes en tant que représentation des images IRM, par la suite des algorithmes d'apprentissage automatique construisent des modèles très précis pour le diagnostic de la MA, permettant de prédire si l'image IRM d'entrée appartient à un sujet étiqueté comme patient atteint de MA. Giulietti et al. (2018) quand à eux, ont exploré la plus-value de l'analyse d'histogrammes appliquée aux mesures dérivées de l'imagerie par tenseur de diffusion pour détecter les différences de tissu cérébral entre les patients atteints de troubles cognitifs légers (MCI), des patients AD et les sujets sains. La diffusivité axiale de la SBAN pourrait être un marqueur précoce des modifications microstructurelles du tissu cérébral au cours de la MA et pourrait être utile pour évaluer la progression de la maladie.

Kang et al. (2011) ont mesuré les différences entre les quantiles des histogrammes ADC des données de volume de la tumeur afin de déterminer le degré de gliome et d'évaluer les performances diagnostiques des cartes ADC à deux différentes valeurs de b.

L'épilepsie frontale à crises nocturnes (NFLE) a été étudiée avec l'histogramme MTR et DM (Ferini-Strambi et al., 2000). L'amplitude du pic pour les deux paramètres étaient significativement réduites pour les sujets NFLE. Il a été suggéré que les résultats indiquent une anomalie diffuse et généralisée.

Inglese et al. (2001) ont étudié la neuropathie optique héréditaire de Leber (LHON) à l'aide de l'imagerie de transfert d'aimantation et l'imagerie de diffusion. Une baisse significative de l'amplitude du pic de l'histogramme MTR a été observée avec une tendance à la hausse des valeurs moyennes de DM chez les patients atteints de LHON.

Plusieurs études ont effectué une analyse plus complexe des données d'histogrammes. Goodyear et al. (2015) ont étudié les modifications du nerf optique en étudiant les moyennes de la DM, la FA, la diffusivité axiale (DA) et de la diffusivité radiale (DR), ainsi que les variations du skewness de ces paramètres chez les patients atteints de névrite optique. Wagner et al. (2016) ont caractérisé les tumeurs cérébelleuses de l'enfant en utilisant le 25ème et 75ème quantile et le skewness de FA, DM, DA et DR. Steffen-Smith et al. (2014) ont analysé l'histogramme du DM par le biais de l'écart type, du skewness, amplitude et position du pic. Les valeurs de DM ont également été ajustées à l'aide d'un modèle de mélange de deux gaussiennes.

Pour plus de détails sur l'analyse d'images quantitatives et l'analyse d'histogrammes, voir l'ouvrage de Filippi et al. (2005), la 1ère édition de l'ouvrage de Tofts et al. (2004b) et sa 2ème édition par Cercignani et al. (2018).

20. "Computer-aided diagnosis" ce qui signifie Diagnostic assisté par ordinateur

Chapitre 2

Aéronautique : les retards dans le transport aérien

Sommaire

2.1 Définition des retards : types et causes	55
2.1.1 Notion de retard	55
2.1.2 Les causes des retards	56
2.2 Mesure et caractérisation des retards	57
2.2.1 Les retards moyens	58
2.2.2 Modélisation des retards par des distributions	58

Les retards de vol sont un problème majeur aux États-Unis et en Europe. Ils sont l'un des indicateurs de performance les plus utilisés de tous les systèmes de transport. Les autorités nationales de réglementation disposent d'une multitude d'indicateurs liés aux seuils de tolérance pour les retards de vol. En effet, le retard de vol est un sujet essentiel dans le contexte des systèmes de transport aérien. En 2017, 44.4% des vols en Europe sont partis avec un retard supérieur à 5 minutes et 38.5% sont arrivés avec un retard supérieur à 5 minutes (d'après [EUROCONTROL](#)). Aux États-Unis, il représente 27.0% des vols au départ et 27.8% des vols à l'arrivée (d'après [BTS](#)).

Dans cette section nous présenterons un bref état de l'art des retards dans le transport aérien.

2.1 Définition des retards : types et causes

2.1.1 Notion de retard

Le terme "retard" est assez simple et généralement appliqué lorsqu'un événement se produit plus tard que prévu. Cependant, le retard est utilisé ou interprété différemment par les différentes parties prenantes impliquées dans les études de planification aéroportuaire, les analyses de performance opérationnelle des compagnies aériennes et le public. En général, les retards dans l'aviation peuvent décrire l'une des deux situations suivantes :

- Les retards en temps réel, souvent comparés à l'horaire de vol. Pour les vols réels -actuels ou historiques- les retards sont souvent mesurés en tant qu'horaires

réels par rapport aux horaires prévus ou planifiés. L'horaire peut faire référence aux heures figurant dans un plan de vol ou dans un horaire publié des compagnies aériennes.

- Estimations mathématiques ou calculées (à l'aide de modèles analytiques ou de simulations) pour la planification, souvent comparées à des temps de parcours sans entrave, nominaux ou optimaux. Les analystes utilisent souvent des outils de simulation informatique ou d'autres procédures analytiques pour évaluer les retards et les économies réalisées.

2.1.2 Les causes des retards

Les retards de vol sont des phénomènes complexes, car ils peuvent être provoqués par plusieurs causes. Deux agences gouvernementales tiennent des statistiques de trafic aérien aux États-Unis. Le Bureau of Transportation Statistics (BTS) compile des données sur les retards dans l'intérêt des passagers. Ils définissent un vol retardé lorsque l'avion ne libère pas son stationnement moins de 15 minutes après l'heure de départ prévue. La Federal Aviation Administration (FAA) est plus intéressée par les retards indiquant des inefficacités de mouvement de surface et enregistrera un retard lorsqu'un avion a besoin de 15 minutes ou plus longtemps que l'heure normale de sortie ("taxi-out") ou d'arrivée ("taxi-in") (Mueller and Chatterji, 2002).

Depuis Juin 2003, les compagnies aériennes qui communiquent des données on-time déclarent également les causes des retards et des annulations au BTS. Les compagnies aériennes signalent les causes de retard dans les grandes catégories créées par le Comité consultatif sur les notifications on-time des transporteurs aériens. Ces catégories sont définies ci-dessous (et sur la figure 2.1) :

- **Transporteur aérien** : la cause de l'annulation ou du retard est due aux contrôles effectués (problèmes de maintenance ou d'équipage, nettoyage des avions, chargement des bagages, ravitaillement en carburant, etc.).
- **Conditions météorologiques extrêmes** : conditions météorologiques significatives (réelles ou prévues) qui retardent ou empêchent l'exécution d'un vol tel qu'une tornade, un blizzard ou un ouragan.
- **Gestion du trafic aérien** : retards et annulations imputables au système d'aviation national faisant référence à un large éventail de conditions, telles que des conditions météorologiques peu extrêmes, les opérations aéroportuaires, un trafic lourd et le contrôle du trafic aérien.
- **Arrivée tardive des avions** : un vol précédent avec le même avion est arrivé en retard, ce qui a retardé le vol actuel.
- **Sécurité** : retards ou annulations causées par l'évacuation d'un terminal ou d'un hall, le réembarquement d'un avion en raison d'une violation de la sécurité, un équipement de contrôle inopérant et / ou des files d'attente de plus de 29 minutes dans les zones de contrôle.

D'après la figure 2.1, la cause de retard la plus importante est de loin "réactionnaire", ce qui fait référence à l'arrivée tardive de l'avion entrant (ou du vol en correspondance/ équipage/ cargo). Cela représente environ 40% du total des retards en Europe ou cinq minutes par vol en moyenne en 2018. Cela explique également pourquoi les passagers sont plus susceptibles d'être retardés

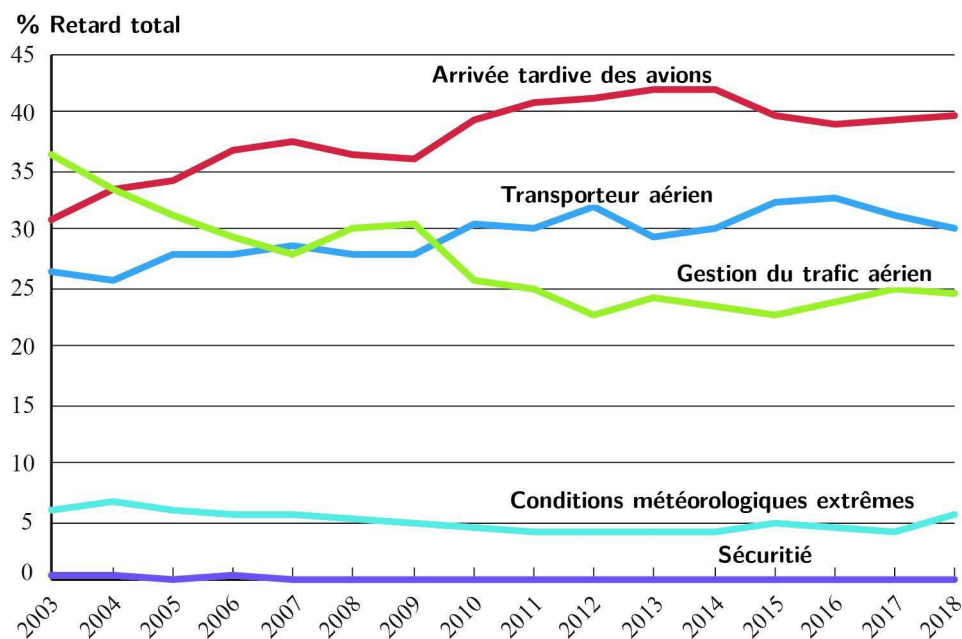


FIGURE 2.1 – La cause des retards par an, en pourcentage du retard total (d'après BTS)

s'ils voyagent plus tard dans la journée plutôt que le matin. Cela signifie également que si les retards non réactionnaires sont minimisés (appelés "retards primaires"), en particulier ceux en début de journée, cela profiterait non seulement à ce vol, mais également aux vols suivants. L'élimination d'un retard sur les deux premiers vols matinaux d'un avion, en particulier, pourrait profiter à tous les passagers utilisant cet avion au cours des 15 prochaines heures.

Les compagnies aériennes pourraient en théorie réduire les retards réactionnaires en permettant un délai beaucoup plus long entre les vols. Cependant, cela augmenterait considérablement les coûts pour les compagnies aériennes, qui s'efforçaient d'optimiser les rotations en minimisant les temps de connexions. Les avions et les équipages ne génèrent des revenus que lorsqu'ils se trouvent dans les airs. Par conséquent, disposer de beaucoup de temps de latence inutile afin d'éviter les retards intermittents se traduirait par des tarifs plus élevés. Comme mentionné ci-dessus, les aéroports souhaitent également des délais d'exécution courts afin de maximiser l'utilisation de leurs portes/stands.

2.2 Mesure et caractérisation des retards

De nombreux retards opérationnels (arrivées ou départs tardifs) résultent de problèmes indépendants de la volonté des aéroports (conditions météorologiques, planification des vols, etc.) ou de la gestion du trafic aérien (e.g. contraintes d'espace aérien, contrôle de la circulation aérienne, tempêtes dans d'autres régions). Dans certains cas, les aéroports collectent des données sur les retards et sont en mesure de montrer que peu de retards sont dus à l'aéroport actuel, mais sont plutôt dus à des contraintes en amont ou en aval. Les retards opérationnels sont généralement mesurés à l'aide des bases de données de la FAA ou en comparant les heures réelles aux horaires des compagnies aériennes.

2.2.1 Les retards moyens

Les aéroports se concentrent généralement sur les retards moyens. En effet, la métrique standard actuelle pour mesurer les retards dans un aéroport est le retard moyen par opération. Il est convenu que cette mesure n'est pas adéquate et manque d'informations. Cependant, dans un grand aéroport, il existe également un consensus sur le fait que :

- les retards moyens inférieurs à 5 minutes sont tolérables,
- les retards moyens supérieurs à 10 minutes posent problème,
- les retards moyens supérieurs à 20 minutes indiquent que l'aéroport subit des problèmes de congestion très importants, au point de ne plus pouvoir fonctionner en raison de la saturation du trafic.

Bien que les retards moyens puissent convenir à la planification des aéroports, ils ne sont généralement pas assez détaillés pour permettre aux compagnies aériennes de prendre des décisions. De plus, sur la scène publique, annoncer qu'un aéroport a un retard moyen de 10 minutes peut sembler pas être "grave", alors qu'il coûte très cher aux compagnies aériennes. Exprimer des retards de plusieurs millions de dollars en temps perdu et en coûts de fonctionnement est généralement plus efficace. De plus, les retards annuels moyens peuvent ne pas constituer des mesures efficaces pour certains aéroports présentant un trafic événementiel important ou une grande diversité de vols au cours de différentes saisons (e.g. l'été, les stations de ski...).

Pour décrire les retards, d'autres mesures sont parfois utilisées :

- Retards horaire moyens pour chaque configuration vent/météo,
- Retards aux heures de pointe,
- 95ème quantiles (i.e. que 95% des vols ont une valeur de retard inférieure à la valeur indiquée ; à l'inverse, 5% des vols sont estimés/simulés de manière à subir un retard supérieur à cette valeur),
- Retard maximum, i.e. le temps maximum au cours duquel un avion est retardé.

2.2.2 Modélisation des retards par des distributions

De nombreuses réflexions concernant l'optimisation des retards ont été menées ces dernières années afin de réduire les impacts économiques, environnementaux et sociaux des divers moyens de transports utilisés. Dans le transport ferroviaire, une étude réalisée par [Zieger et al. en 2018](#) a souligné l'importance du choix de la meilleure distribution probabiliste pour modéliser des retards. En s'appuyant sur une modélisation de la propagation des retards par une distribution mathématique des "buffer times" elle-même basée sur la méthode de Monte-Carlo, il a été établi que le choix d'une distribution mathématique peut avoir une influence sur l'estimation des retards. Dès lors, il devient nécessaire de choisir une distribution mathématique adaptée aux retards dans le transport aérien. Nombreux auteurs ont étudié les modèles de distributions des retards de vol. [Lan et al. \(2006\)](#) ont modélisé le retard d'arrivée à l'aide d'une distribution log-normale. [Novianingsih and Hadiani \(2014\)](#) ont

modélisé les retards au départ à l'aide du modèle log-normal, gamma et Rayleigh (figure 2.2). Schaefer et al. (2005) ont modélisé les retards de vol et les retards au sol en utilisant les distributions gamma, Erlang et bêta. Mueller and Chatterji (2002) ont modélisé le retard de départ en utilisant la distribution de Poisson et la distribution normale pour le retard d'arrivée. Ils ont ajusté moyenne et écart-type via la méthode des moindres carrés pour améliorer le modèle. Les retards résiduels aléatoires ont été estimés par Tu et al. (2008) à l'aide de distributions de mélanges de gaussiennes.

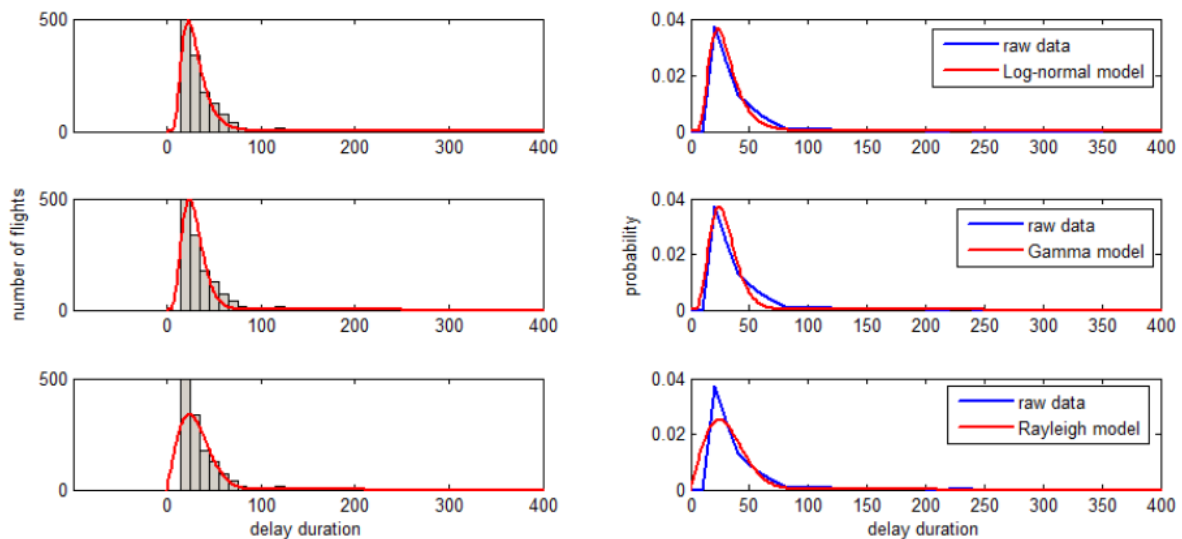


FIGURE 2.2 – Modélisation des retards de départ en utilisant un modèle log-normal, gamma et Rayleigh (d'après Novianingsih and Hadiani (2014))

Deuxième partie

Partie expérimentale

Chapitre 3

Les métriques d'histogrammes

Sommaire

3.1 Description de l'approche	63
3.1.1 Production d'histogrammes	64
3.1.2 Définition des métriques d'histogrammes	64
3.1.3 Classification basée sur les métriques d'histogrammes	66
3.2 Application 1 : classification des populations de la maladie d'Alzheimer	68
3.2.1 Matériels et méthodes	69
3.2.2 Résultats	81
3.2.3 Discussion	84
3.3 Application 2 : essai clinique de Sclérose en plaques	87
3.3.1 Matériels	89
3.3.2 Méthodes	93
3.3.3 Résultats	96
3.3.4 Discussion et limites	101

3.1 Description de l'approche

Dans ce chapitre, nous allons décrire une première approche pour l'analyse des histogrammes d'images quantitatives. Cette méthode consiste à construire des histogrammes et d'en extraire des descripteurs locaux tels que la moyenne, les quantiles, skewness etc... Ces métriques d'histogrammes sont utilisées comme entrées dans un algorithme de classification (supervisé ou non) afin d'identifier des populations. La figure 3.1 présente un schéma résumé de l'approche.

En premier lieu, nous appliquerons cette méthode dans le cadre de la classification de patients atteints de troubles cognitifs légers et de patients atteints d'Alzheimer. Puis, dans un second temps, pour l'étude de l'effet d'un traitement dans le cadre d'un essai clinique de sclérose en plaques avec IRM multimodale.

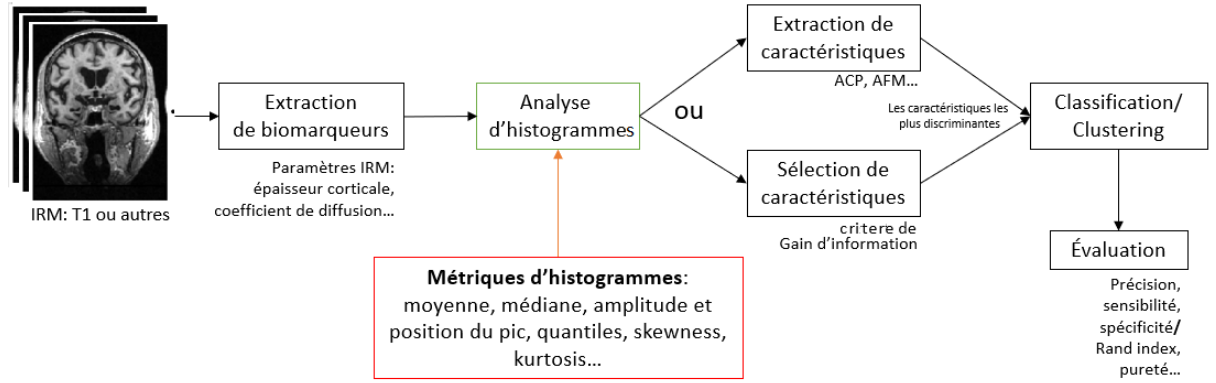


FIGURE 3.1 – Schéma résumé de l'approche basée sur les métriques d'histogrammes

3.1.1 Production d'histogrammes

Les différentes étapes de production d'histogrammes à partir d'images sont présentées ci-après :

1. Prétraitement des images : recalage, segmentation (Exclusion des tissus non cérébraux et le LCS).
2. Mesure des cartes paramétriques : mesure du paramètre IRM¹ dans chaque voxel. Les valeurs du paramètre IRM remplacent l'intensité des voxels correspondants.
3. Extraction des histogrammes à partir des cartes paramétriques, en choisissant le nombre de bins adapté.
4. Lissage des histogrammes afin d'atténuer l'apparition des spikes².
5. Normalisation des histogrammes, ce qui permet de rendre les histogrammes indépendants des variations de la taille du cerveau des sujets de l'étude.

3.1.2 Définition des métriques d'histogrammes

Ces métriques résument l'histogramme en quelques paramètres décrits ci-dessous :

— Moyenne empirique

Soit $h^p = (h_1 \dots h_N)$ est l'histogramme p ou N est le nombre de bins, la moyenne du paramètre IRM étudié est :

$$\bar{x} = \frac{\sum_{j=1}^N x_j h_j^p}{\sum_{j=1}^N h_j^p} \quad (3.1)$$

avec x_j la valeur du paramètre au centre du j -ème bin.

Il est important de noter que la moyenne peut être calculé sans générer d'histogrammes, car il s'agit de la valeur moyenne de tous les voxels de la carte paramétrique représentés dans l'histogramme.

1. Ce paramètre pourrait être : ADC, FA, MTR, T1...

2. discontinuités qui apparaissent sur l'histogramme

— Amplitude et position du pic

h_i^p est l'amplitude³ du pic de l'histogramme p si et seulement si :

$$h_i^p \geq h_j^p, \quad j \neq i, \quad j = 1, \dots, N \quad (3.2)$$

La position du pic est aussi appelé "le mode". Elle décrit la valeur la plus représentée du paramètre IRM. i est la position du pic de l'histogramme p si et seulement si h_i^p est l'amplitude du pic.

L'histogramme couvre l'intervalle x_{min} à x_{max} et le i -ème bin couvre l'intervalle x_i^- à x_i^+ et a pour centre x_i^c :

$$x_i^c = x_{min} + (i - 1)\Delta \quad (3.3)$$

avec Δ la taille du bin :

$$\Delta = \frac{(x_{max} - x_{min})}{N - 1} \quad (3.4)$$

— Les quantiles

Les quantiles sont des moyens pratiques de caractériser la forme d'un histogramme. Le n -ème quantile d'une distribution est la valeur x où $n\%$ des données (ou zone de l'histogramme) se trouvent à gauche du bin. x_{25} est sensible à la structure dans la partie gauche de l'histogramme (où les lésions ont tendance à se situer⁴ sur un histogramme MTR).

k est le numéro de bin du n -ème quantile de l'histogramme p si et seulement si k est la plus grande valeur qui satisfait le critère suivant :

$$\sum_{i=1}^k h_i^p \leq (n\%) \sum_{j=1}^N h_j^p \quad (3.5)$$

Quand $n=50\%$ on obtient **la médiane**, quand $n=0\%$ on obtient le minimum, quand $n=100\%$ on obtient le maximum.

— Skewness

Skewness ou coefficient d'asymétrie (Pearson, 1895) est défini par la formule suivante, où μ_2 et μ_3 sont respectivement le deuxième et troisième moment central :

$$\gamma_1 = \mu_3 / \mu_2^{3/2} \quad (3.6)$$

En règle générale, une asymétrie négative indique que la moyenne des valeurs de données est inférieure à la médiane et que la distribution des données est asymétrique à gauche (ce qui correspond au cas présenté sur la figure ??). Une asymétrie positive indiquerait que la moyenne des valeurs de données est supérieure à la médiane et que la distribution des données est asymétrique à droite. Dans le cas d'une gaussienne, la distribution est symétrique donc $\gamma_1 = 0$.

3. ou la hauteur

4. Les paramètres T1 et DM ont des histogrammes où la direction de la pathologie est inversée, le LCS et le tissu qui a partiellement perdu sa structure ont des valeurs anormalement élevées de T1 ou de DM, de sorte que ces voxels sont situés à droite de l'histogramme.

— Kurtosis

Kurtosis ou coefficient d'aplatissement (Pearson, 1905) est défini par la formule suivante, où μ_2 et μ_4 sont respectivement le deuxième et quatrième moment central :

$$\gamma_1 = \mu_4 / \mu_2^2 - 3 \quad (3.7)$$

Le kurtosis est négatif pour une distribution moins "aplatie" qu'une distribution normale et positif sur une distribution plus aplatie qu'une distribution normale.

3.1.3 Classification basée sur les métriques d'histogrammes

3.1.3.1 Réduction de dimension

Après avoir extrait les différentes métriques d'histogrammes, la dimensionnalité des données se trouve considérablement augmentée. En effet, le nombre de variables, aussi appelées "caractéristiques"⁵, se multiplie ce qui va introduire des difficultés à plusieurs niveaux : complexité, temps de calcul, plus grand risque d'avoir des données bruitées et "le fléau de la dimension"⁶. Pour surmonter ces problèmes, il est nécessaire de trouver un moyen de réduire le nombre de caractéristiques considérées. Deux techniques sont souvent utilisées (tableau 3.1) :

1. Extraction de caractéristiques⁷ :

Les techniques d'extraction de caractéristiques ont pour but de réduire la dimension de l'espace d'état de l'échantillon initial afin d'en simplifier l'analyse et d'en supprimer les redondances.

De nombreuses méthodes statistiques, essentiellement linéaires, ont été proposées. Les plus connues sont l'analyse en composantes principales (ACP), l'analyse discriminante linéaire (ADL) (Fisher, 1936), le positionnement multi-dimensionnel (MDS⁸) (Messick and Abelson, 1956)... Dans cette section, nous nous intéresserons plus particulièrement à l'ACP dont le principe se retrouve dans la plupart des autres techniques. L'ACP fait partie de la famille des méthodes factorielles. Elle transforme les variables quantitatives corrélées en un nombre plus petit de variables quantitatives non corrélées, appelées composantes principales, par projection des données sur un sous-espace vectoriel de l'espace initial.

Soit $X = (x_1, \dots, x_N)$ un échantillon de variables aléatoires iid prenant leurs valeurs dans \mathbb{R}^n . Le premier axe propre de l'ACP est le sous-espace affine de dimension 1 de \mathbb{R}^n pour lequel la variance des données projetées est maximale. On peut sans perte de généralité supposer l'échantillon centré, le problème se formule alors mathématiquement comme un programme

5. "Features" en anglais.

6. ou "the curse of dimensionality" (Bellman, 1957). Ce phénomène est observé lorsque des variables sont ajoutées sans augmenter également le nombre d'échantillons d'apprentissage, ce qui conduit à un sur-apprentissage.

7. Plus connue sous le nom anglais "feature extraction" dans la communauté du "machine learning"

8. "Multi Dimensional Scaling"

d'optimisation convexe :

$$\begin{cases} \min_{v \in \mathbb{R}^n} \sum_{i=1}^N \langle v, x_i \rangle^2 \\ \langle v, v \rangle = 1 \end{cases} \quad (3.8)$$

En utilisant les multiplicateurs de Karish Kuhn Tucker, on obtient la forme lagrangienne :

$$\max_{\lambda \in \mathbb{R}^+} \min_{v \in \mathbb{R}^n} \sum_{i=1}^N \langle v, x_i \rangle^2 + \lambda (\langle v, v \rangle - 1) \quad (3.9)$$

qui admet pour solution :

$$\begin{cases} \left(\sum_{i=1}^N x_i x_i^t \right) v = kv, k \in \mathbb{R} \\ \|v\| = 1 \end{cases} \quad (3.10)$$

Le premier axe propre est donc une droite passant par moyenne empirique de l'échantillon et de vecteur directeur un vecteur propre unitaire de la matrice :

$$\tilde{X} = \sum_{i=1}^N x_i x_i^t$$

Le second axe propre s'obtient par le même procédé après avoir projeté orthogonalement sur le sous-espace supplémentaire du premier axe propre, et ainsi de suite par récurrence pour les axes propres suivants. On remarque que l'ACP peut s'obtenir en une seule opération en réalisant une décomposition en vecteurs propres de \tilde{X} , qui sont automatiquement orthogonaux, et en remarquant que chaque valeur propre donne la variance en projection de l'échantillon.

En pratique, on ne garde dans l'ACP que les axes propres dont la variance en projection est la plus élevée, en arrêtant la décomposition lorsqu'une part suffisante de la variance initiale a été expliquée (typiquement 0.95 ou 0.99). Pour plus de détails, voir l'ouvrage de [Jolliffe \(1986\)](#).

2. Sélection de caractéristiques⁹ :

Il s'agit de sélectionner, parmi les caractéristiques existantes, un sous-ensemble constitué de celles qui sont les plus informatives selon un critère donné. Le processus de sélection supprime les caractéristiques non pertinentes ou redondantes. Contrairement aux méthodes d'extraction de caractéristiques, les techniques de sélection de caractéristiques ne transforment pas la représentation originale des données. L'un des objectifs de ces méthodes est d'éviter le sur-apprentissage¹⁰. Les méthodes de sélection de caractéristiques se divisent en trois catégories :

- (i) "Filter" qui extrait des caractéristiques des données sans aucun apprentissage ([John et al., 1994](#)).
- (ii) "Wrapper" qui utilise des techniques d'apprentissage pour évaluer les caractéristiques utiles ([Kohavi and John, 1997](#)).

9. Plus connue sous "feature selection" en machine learning

10. "overfitting" en anglais

(iii) Les méthodes intégrées "embedded" combinent les avantages des deux précédentes méthodes. Elles incorporent la sélection de caractéristiques lors du processus d'apprentissage.

Dans la dernière classe, on trouve des algorithmes basés sur des approches dites parcimonieuses qui réalisent des optimisations par rapport à des critères L^1 tels LASSO (least absolute shrinkage and selection operator).

Méthode	Avantages	Inconvénients
Extraction de caractéristiques	<ul style="list-style-type: none"> - Pouvoir discriminant supérieur - Contrôle le sur-apprentissage quand c'est non-supervisé 	<ul style="list-style-type: none"> - Perte d'interprétabilité - Transformation peut-être coûteuse - Difficile de relier les caractéristiques de l'espace d'origine aux nouvelles caractéristiques
Sélection de caractéristiques	<ul style="list-style-type: none"> - Préserve les caractéristiques : Facilite l'interprétation - Meilleures performances d'apprentissage 	<ul style="list-style-type: none"> - Couche supplémentaire de complexité dans la modélisation - Temps additionnel pour l'apprentissage - Risque plus élevé de sur-apprentissage

Tableau 3.1 – Comparaison des méthodes d'extraction de caractéristiques et de sélection de caractéristiques

3.1.3.2 Les algorithmes de classification

La classification a pour objectif d'identifier les classes auxquelles appartiennent des objets à partir de traits descriptifs. Les algorithmes de classification sont utilisés de façon récurrente dans le domaine biomédical. Parmi ces algorithmes on distingue deux grandes familles d'algorithmes de classification : la classification supervisée et la classification non-supervisée ¹¹.

- **La classification supervisée** : elle consiste à construire une règle de classement à partir d'un ensemble de données étiquetées (telles que des images) pour prédire la population d'appartenance de nouvelles observations. On retrouve les algorithmes : Machine à vecteurs de support (SVM ¹²), k plus proches voisins (KNN ¹³), réseaux de neurones (ANN ¹⁴), deep learning...
- **La classification non-supervisée/Clustering** : elle consiste à trouver une partition des données selon des critères de partitionnement (telles que les mesures de distances). On retrouve des algorithmes du type k-means, k-medoids, classification hiérarchique...

Certains de ces algorithmes seront décrits plus loin dans le manuscrit.

3.2 Application 1 : classification des populations de la maladie d'Alzheimer

En clinique, des tests neuropsychologiques sont utilisés pour établir un diagnostic probable de la maladie d'Alzheimer. Cependant les modifications anatomiques cérébrales associées à la pathologie apparaissent plusieurs années avant l'apparition des premiers troubles cognitifs. Dans cette perspective, les

11. On retrouve plus récemment la classification semi-supervisée (Zhu, 2006) et l'apprentissage par renforcement (Kaelbling et al., 1996).

12. "Support vector machine"

13. "k-nearest neighbors"

14. "Artificial neural network"

biomarqueurs anatomiques capables de quantifier les modifications de morphologie cérébrale pourraient jouer un rôle important dans le diagnostic précoce de la maladie. Dans notre étude, l'épaisseur et la courbure corticale ont été utilisées en tant que biomarqueur anatomique afin de quantifier les modifications morphologiques du ruban cortical causées par la pathologie.

3.2.1 Matériels et méthodes

3.2.1.1 Construction de la population d'étude

3.2.1.1.1 La base de données ADNI

La population de l'étude provient de la phase 1 (ADNI1) de la base de données ADNI, Alzheimer's Disease Neuroimaging Initiative ([Mueller et al., 2005](#)) librement mise à la disposition de la communauté scientifique. Il s'agit d'un projet initié en 2004 par le National Institute on Aging (NIA), le National Institute of Biomedical Imaging and Bioengineering (NIBIB) et la Food and Drug Administration (FDA), dont le principal investigateur est le Pr. Michael Weiner (Docteur Honoris Causa 2019 de l'Université Paul Sabatier). Cette base de données permet de disposer d'un large échantillon d'individus et d'images IRM de bonne qualité provenant de différents centres et hôpitaux. Le but est de développer des biomarqueurs cliniques, génétiques, biochimiques et d'imagerie pour le diagnostic précoce et le suivi de la maladie d'Alzheimer. Pour des informations à jour, voir [le site ADNI](#).

3.2.1.1.2 Le choix des données

Pour s'assurer de la cohérence des groupes de patients que nous avons constitués et de la qualité des données traitées, nous n'avons pas retenu l'ensemble de la population ADNI1, qui rassemble 800 sujets (200 HC, 200 AD, 400 MCI) Dans ce paragraphe, nous explicitons nos critères de sélection des données.

Nous avons basé notre étude sur une population constituée de sujets sains (HC), de sujets diagnostiqués Alzheimer (AD) et de sujets atteints de déficiences cognitives légères (MCI¹⁵). Nous avons choisi de répartir les MCI en deux sous-groupes, selon qu'ils convertissent en AD (MCI progressifs, notés pMCI) durant un suivi de 3ans ou ne convertissent pas, sujets dits MCI stables et notés sMCI.

En plus de la contrainte concernant le suivi, nous avons ajouté quatre contraintes afin d'améliorer la fiabilité et la robustesse de notre étude :

- **Pertinence du diagnostic** : pour cela nous nous sommes basés sur les informations fournies par ADNI. ADNI affecte un niveau de confiance au diagnostic (HC, MCI ou AD) selon une échelle allant de 1 à 4, soit du peu fiable au très fiable ; nous avons donc décidé de ne conserver que les images associées à un niveau de confiance de 3 ou 4 ;
- **Qualité des images** : les images doivent permettre l'estimation des biomarqueurs, c'est-à-dire qu'elles doivent passer avec succès les étapes de

15. "Mild cognitive impairment" en anglais

prétraitement, réorientation, rééchantillonnage, segmentation... (cf. section sur le prétraitement des images);

- **Cohérence des épaisseurs corticales** : suppression des données conduisant à une estimation aberrante de l'épaisseur corticale moyenne; à partir de la littérature nous avons choisi comme seuil 4.5mm;
- **Appariement en âge** : l'âge a une influence sur la morphologie du cortex car l'épaisseur corticale décroît avec l'âge même durant le vieillissement normal (Thambisetty et al., 2010); cette étape est donc importante pour pouvoir réaliser des comparaisons intergroupes en supprimant le facteur de confusion dû à l'âge, ce qui évitera de confondre l'effet de la MA et celui de l'âge; l'appariement en âge est effectué en moyenne (via un test d'ANOVA) et en variance (via le test de Levene); si ces tests donnent des résultats significatifs, l'appariement est nécessaire et nous utiliserons alors une procédure de choix aléatoire à l'instar de celle utilisée par Vanquin (2015) dans son travail de thèse.

3.2.1.1.3 Population finale

Nous avons appliqué les critères définis précédemment sur les données de la base ADNI1. La population résultante est ainsi constituée d'un total de 279 sujets, les détails sont sur le tableau 3.2.

- Pertinence du diagnostic : en appliquant ce critère sur toutes les images de la base, nous éliminons environ la moitié des sujets de la base ADNI1.
- Qualité des images : nous avons trouvé une trentaine d'images avec des artefacts. La majorité de ces artefacts étaient dû à la morphologie du sujet (excès de couche lipidique). Par précaution, nous avons supprimé le sujet en entier et non pas juste l'image de mauvaise qualité.
- Cohérence des épaisseurs corticales : plus d'une cinquantaine de sujets ont été exclus dans chaque groupe à cause des valeurs aberrantes de l'épaisseur corticale.
- Appariement en âge : les groupes résultants sont appariés en âge.

	HC	AD	sMCI	pMCI	valeur-p
Nombre de sujets N=279	71	77	45	86	-
Âge [intervalle]	76.1±5.6 [62.9-87]	77.5±5.4 [64.3-87.8]	77.3±5.5 [61-86.4]	76.0±5.7 [55.2-88.3]	0.24
Genre Femme/Homme	38/33	45/32	18/27	38/48	0.15
MMSE	29.2±1.1	22.7±2.1	27.0±1.9	26.8±1.9	<0.01

Tableau 3.2 – Information démographique de la population de notre étude

± représentent moyennes ± écart-type. La valeur p est basée sur le test ANOVA dans le cas de l'âge et MMSE et sur le test khi-deux χ^2 pour Genre, avec $\alpha < 0.05$ (Abrv : HC sujets sains; AD, patients atteints d'Alzheimer; sMCI, MCI stables; pMCI, MCI progressifs; MCI, patients atteints de troubles cognitifs légers.)

3.2.1.2 Acquisitions IRM

Toutes les images utilisées sont des Images par Résonance Magnétique, pondérées en T1, avec un champ magnétique de 1.5 Tesla, acquises selon la

séquence MP-RAGE. Dans une image pondérée en T1, le cortex apparaît hyperintense et le liquide cébrospinal (LCS) hypointense. Nous avons donc des images bien contrastées.

La séquence 3D MP-RAGE (3-Dimensional Magnetization-Prepared Rapid Acquisition Gradient Echo) a été utilisée avec les paramètres présentés dans le tableau 3.3. Les images MP-RAGE sont considérées comme les meilleures du point de vue de la qualité. Cette séquence a été choisie par ADNI pour ses bonnes propriétés concernant le rapport signal sur bruit (et contraste sur bruit) de l'image ainsi que pour sa bonne résolution spatiale qui est souvent proche du mm^3 , le tout dans un temps d'acquisition acceptable pour la pratique clinique (environ 7 à 9 minutes). Ainsi, nous obtenons des images de qualité qui facilitent la segmentation des tissus et l'estimation de biomarqueurs structuraux.

B_0	Bobine	N_x	N_y	N_z	FoV (mm)	TR (ms)	TI (ms)	Flip (deg)	Δz	Time (min :s)
1.5 T	BC	192	192	184–208	240 x 240	3000	1000	8	1.2	9 :36–9 :38
1.5 T	MA	192	192	160–170	240 x 240	2300–2400	1000	8	1.2	7 :11–7 :24

Tableau 3.3 – Les paramètres de l'acquisition MP-RAGE

Abrv : MA= multicoil phased-array head coil ; BC= Birdcage or volume head coil ; N_x , N_y , N_z = taille de la matrice avec N_z nombre de coupe ; FoV= champ de vision, TR : temps de répétition ; TI=Temps d'inversion ; Δz = épaisseur de la coupe

3.2.1.3 Prétraitement des images

La chaîne de prétraitement des images structurales comprend plusieurs étapes (résumées sur la figure 3.2) :

1. **Redimensionner**¹⁶ **les voxels** : il est important de commencer par une mise en forme isométrique afin d'avoir les mêmes dimensions de voxels dans chaque direction ($1\text{mm} \times 1\text{mm} \times 1\text{mm}$). Pour cela, un algorithme d'interpolation trilineaire a été utilisé. Il s'agit d'une extension de l'interpolation linéaire aux espaces de dimension 3 (i.e. images 3D). Toutes les images ont été normalisées sur un template de dimension $256 \times 256 \times 180$.
2. **Réorientation dans le plan AC/PC** : tous les volumes sont réorientés dans le plan AC/PC (Han and Park, 2003). AC (commissure antérieure) et PC (commissure postérieure) sont deux repères anatomiques dans le cerveau. Un cerveau qui est aligné selon le plan AC/PC, signifie que l'AC et le PC sont dans le même plan axial. Ce plan est attrayant pour des applications cliniques individuelles, car il conserve la taille originale du cerveau, tout en fournissant une orientation commune pour chaque cerveau.
3. **Segmentation** à l'aide de SPM8 version 6313 via Matlab version R2016b. SPM est mis gratuitement à la disposition de la communauté de neuro-imagerie, afin de promouvoir la collaboration et un schéma d'analyse commun à tous les laboratoires. Le logiciel représente la mise en œuvre des concepts théoriques de la cartographie statistique paramétrique¹⁷ dans un package Matlab complet.
La méthode de segmentation en différentes classes de tissus se fait à l'aide

16. "Reslice" en anglais

17. Plus connu sous "statistical parametric mapping"

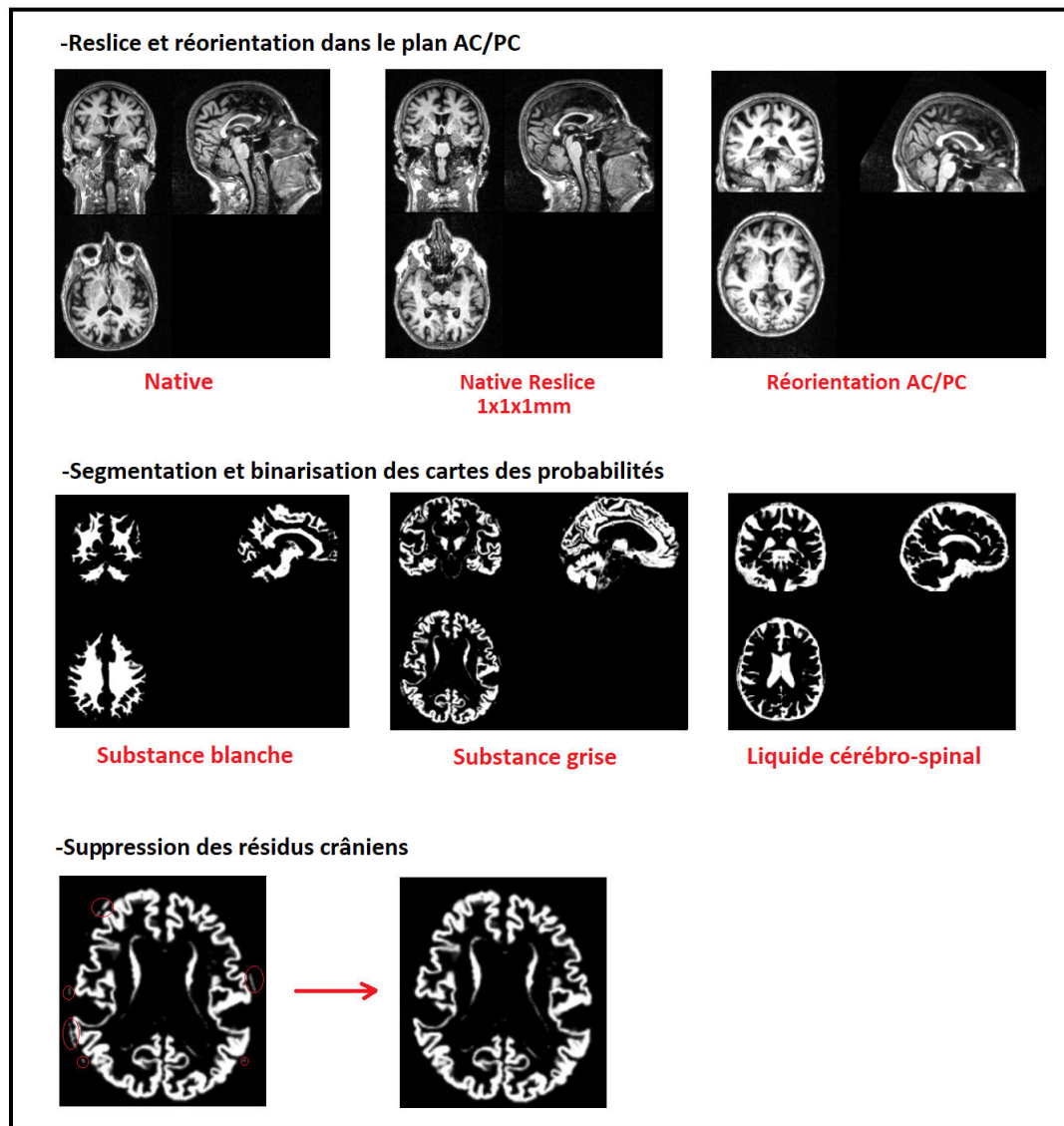


FIGURE 3.2 – Schéma représentant les différentes étapes de prétraitement des volumes IRM

d'un modèle modifié de mélange gaussien via l'algorithme itératif de type Expectation-Maximisation (EM) ¹⁸. En connaissant la probabilité spatiale ¹⁹ de chaque voxel d'appartenir à la substance grise, substance blanche ou liquide cérébro-spinal, il est possible d'obtenir une classification plus robuste. De plus, une étape de correction de la non-uniformité d'intensité est également incluse, ce qui rend le procédé plus adapté aux images affectées par des variations d'intensité. Pour davantage de détails, voir [Ashburner \(2000\)](#).

4. **Suppression des résidus crâniens** : après la segmentation, des résidus crâniens sont susceptibles d'apparaître et risquent de fausser l'estimation des biomarqueurs tels que l'épaisseur corticale et la courbure. Dans sa thèse, [Vanquin \(2015\)](#) a proposé une méthode pour détecter et supprimer ces résidus. Cette méthode se base sur la connexité des voxels

18. L'étape E est le calcul des probabilités d'appartenance et l'étape M, l'estimation des paramètres des distributions et la correction de non-uniformité.

19. Il s'agit de la binarisation des cartes de probabilités.

(26-connextité). Elle consiste à détecter les amas de voxels correspondant à la SG et à éliminer les autres amas avec l'hypothèse que la SG correspond à l'amas ayant le plus grand nombre de voxels. Pour cela, un voisinage V^{20} est défini. Les voxels de l'image ayant le même voisinage V signifie que ces voxels sont connectés et par transition nous obtenons un amas de voxels. Par la suite, l'amas ayant le plus grand nombre de voxels (correspondant au ruban cortical) est préservé, les autres amas sont supprimés. Toutefois, dans le cas où les fragments de crâne sont en contact avec le ruban cortical, il est impossible de dissocier ces résidus parasites de la SG ²¹.

3.2.1.4 Estimation des biomarqueurs

Nous avons choisi des biomarqueurs anatomiques capables de quantifier les modifications de la morphologie cérébrale. Dans notre étude, l'épaisseur et la courbure corticale ont été utilisées pour la SG. Cette section présente les outils utilisés pour estimer ces biomarqueurs.

3.2.1.4.1 Épaisseur corticale

L'épaisseur corticale (EC) est estimée en résolvant l'équation de Laplace. Cette approche a été initialement développée par Jones et al. (2000). L'estimation de l'EC a été effectuée à l'aide de la toolbox *Corthizon* développée sous Matlab[®] au sein de l'unité INSERM ToNIC par Nicolas Chauveau et basée sur le travail de thèse de Querbes (2009).

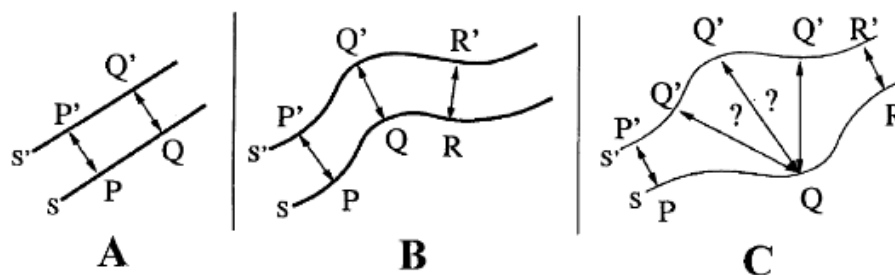


FIGURE 3.3 – Trois géométries possibles pour le ruban de substance grise dans le cerveau. (A) montre l'épaisseur entre les deux surfaces au sens de la projection orthogonale. (B) montre l'exemple d'un endroit plus courbé du cerveau et (C) montre un exemple de géométrie pour lequel la projection orthogonale est applicable pour P - P' et R - R' mais pas pour Q (d'après Jones et al., 2000)

— Résolution de l'équation de Laplace

Soient les deux surfaces, l'interface SB/SG S et l'interface SG/LCS S' (figure 3.3). L'épaisseur est alors la distance du vecteur reliant P à P' . Ceci induit 3 propriétés désirables pour une définition mathématique robuste :

- Association bijective : quelle que soit la surface, chacun de ses points doit être associé à un et un seul point de l'autre surface ; on notera (P, P') cette paire de points,

20. Un cube de 3 pixels de côté et qui balaye l'image afin de déterminer les amas de voxels.

21. Querbes (2009) a proposé dans sa thèse d'utiliser la morphologie mathématique : une succession d'érosions et de dilations afin de supprimer ces zones parasites. Cette approche présente aussi la limite de ne pas pouvoir supprimer les résidus trop volumineux et augmenter le nombre d'érosions et de dilations risque de supprimer des voxels appartenant à la SG.

- Réciprocité : quelle que soit la paire de points (P, P'), l'estimation de l'épaisseur ne doit dépendre que de ces points et non de l'ordre dans lequel ils sont parcourus,

- Distance minimale : les points (P, P') sont les plus proches parmi l'ensemble des possibilités de paires, selon un certain critère de distance.

Une approche basée sur les fonctions analytiques permet de satisfaire ces contraintes. Elle se traduit par une équation aux dérivées partielles du second ordre du type de celle utilisée pour un potentiel électrostatique ψ situé entre S et S'. En imposant un potentiel constant ψ_0 sur S et ψ_1 sur S' ($\psi_1 \neq \psi_0$) et en résolvant l'équation de Laplace dans le volume isolé par S' et S, nous pouvons calculer le potentiel en chaque voxel de la substance grise, puis à partir de ces potentiels un ensemble de sous-couches (les surfaces équipotentielles) permettant une transition douce entre S et S'.

$$\begin{aligned}\nabla^2\psi &= \frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} + \frac{\partial^2\psi}{\partial z^2} = 0 \\ \psi &= \psi_0 \text{ sur } S \\ \psi &= \psi_1 \text{ sur } S'\end{aligned}\tag{3.11}$$

Cette équation décrit un ensemble de surfaces effectuant une transition lisse entre S et S'. La longueur des lignes de champ (les courbes orthogonales aux surfaces équipotentielles) est alors assimilée à l'épaisseur corticale en chaque point du volume corticale.

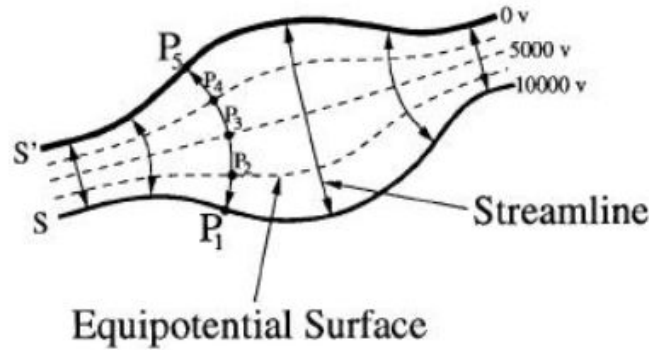


FIGURE 3.4 – Surfaces équipotentielles entre S et S'.

Les valeurs de potentiels de 0V et 10000V sont des conditions aux frontières pour résoudre l'équation. Le chemin connectant P₁ à P₅ montre un exemple de résultat de l'astuce présentée dans le paragraphe précédent... Les lignes fléchées correspondent aux rayons corticaux (d'après Jones et al., 2000)

En effet, en procédant de couche en couche, nous construisons une série de points permettant de relier un point P de S à un point P' de S'. Le principe de réciprocité est respecté, puisque le rayon cortical (la courbe P₁P₂P₃P₄P₅) est le même si nous partons de S (du point P₁) ou de S' (du point P₂). L'EC, que nous associons ensuite à chaque point P_i (e.g. P_{1...5} sur la figure 3.4), est alors définie comme la somme des distances entre sous-couches du ruban cortical.

— La carte d'épaisseur corticale

Après avoir estimé l'épaisseur corticale, il est possible de générer une carte d'épaisseur corticale, dans l'espace natif ou dans l'espace MNI. Il s'agit d'une carte paramétrique dont chaque voxel du cortex représente l'épaisseur corticale qui lui est associée (figure 3.5).

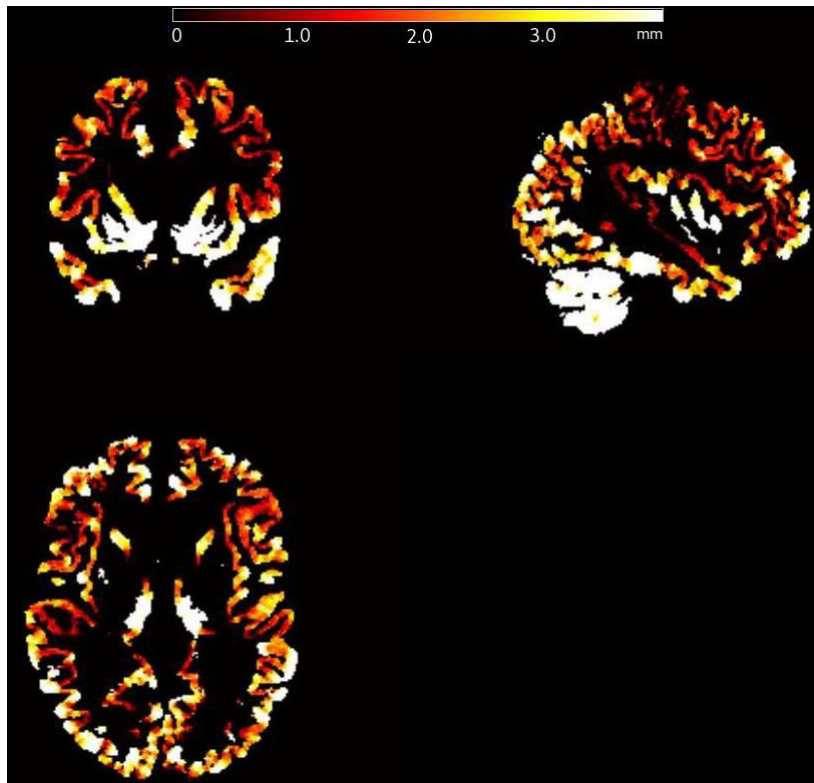


FIGURE 3.5 – Coupes de cartes EC d'un patient atteint de la maladie d'Alzheimer. On observe clairement la présence du cervelet, particulièrement sur la coupe sagittale (droite-haut). Ce dernier sera supprimé en utilisant le template de Brodmann.

3.2.1.4.2 Courbure corticale

La morphologie des surfaces corticales est altérée en présence de la maladie d'Alzheimer (Cai et al., 2017). La morphologie des sillons et des gyri²² (figures 3.6 et 3.7) est étudiée à l'aide de biomarqueurs morphologiques, telle que la courbure corticale.

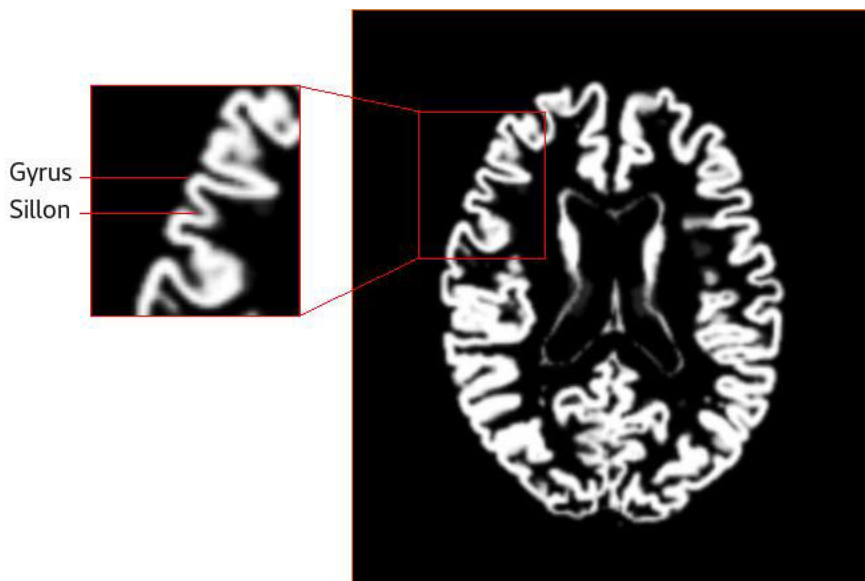


FIGURE 3.6 – Anatomie des surfaces corticales. Les gyri sont un ensemble de replis sinueux du cortex cérébral, entourés par des sillons qui marquent la surface du cerveau.

22. aussi appelés "circonvolutions"

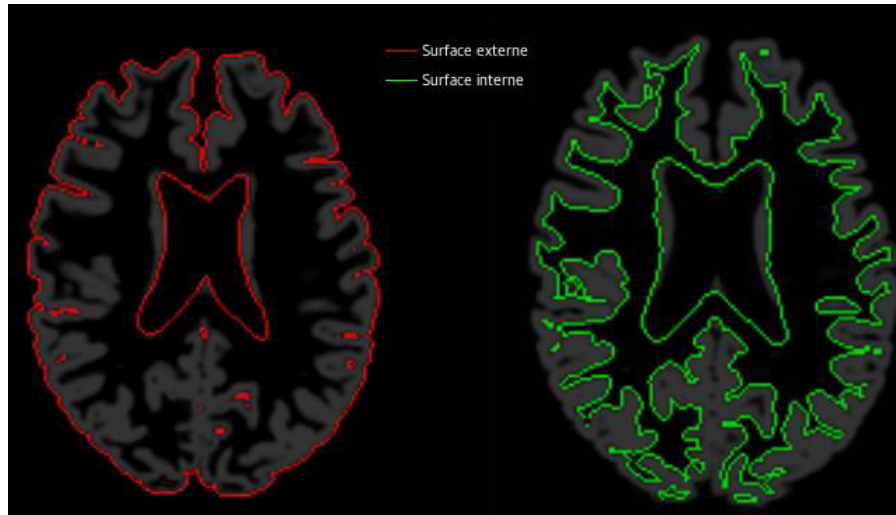


FIGURE 3.7 – ruban cortical (en gris) et les surfaces corticales internes (en vert) et externes (en rouge). Les surfaces externes et internes correspondent respectivement aux interfaces SG/LCS et SG/SB.

— Estimation de la courbure corticale

L'estimation de la courbure corticale a été développée au sein de l'unité INSERM U1214 lors de la thèse de [Vanquin \(2015\)](#) en se basant sur la méthode de [Monga and Benayoun \(1995\)](#) et adaptée par Nicolas Chauveau pour être compatible à la toolbox Corthizon. L'approche utilisée permet d'estimer la courbure des sillons et celle des gyri des surfaces internes et externes du ruban cortical (obtenant ainsi 4 types de courbures).

La méthode d'estimation de la courbure est présentée dans l'annexe III et pour d'avantages de détails voir la thèse de [Vanquin \(2015\)](#).

Le signe de la courbure est utilisé afin de pouvoir distinguer les sillons et les gyri. Ainsi, les gyri sont définis positivement et les sillons négativement. Comme nous analysons séparément les sillons et les gyri, nous ne prenons que les amplitudes des courbures (i.e. les valeurs absolues).

— Carte des courbures corticales

La courbure du cortex entier est estimée pour chaque voxel des surfaces corticales internes et externes, au niveau des gyri et des sillons, obtenant ainsi une carte des courbures corticales présentée sur la figure 3.8.

3.2.1.5 L'analyse d'histogrammes

À partir des cartes paramétriques d'épaisseur corticale et de courbure corticale, nous avons extrait cinq histogrammes :

- Épaisseur corticale (EC)
- Courbure corticale interne au niveau des gyri (GyrInt)
- Courbure corticale interne au niveau des sillons (SilInt)
- Courbure corticale externe au niveau des gyri (GyrExt)
- Courbure corticale externe au niveau des sillons (SilExt)

Il a été nécessaire au préalable de retirer les voxels correspondant au cervelet en utilisant le template de Brodmann, obtenant ainsi uniquement des images

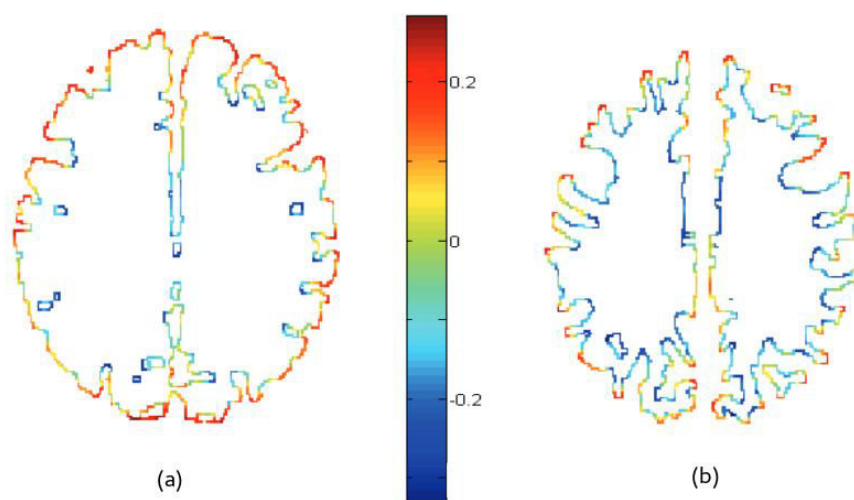


FIGURE 3.8 – Carte des courbures corticales. (a) Carte des courbures de la surface corticale externe; (b) Carte des courbures de la surface corticale interne. Les sillons et les gyri sont respectivement caractérisés par une courbure négative et positive (Vanquin, 2015).

paramétriques du cortex cérébral. De plus, afin d’atténuer les discontinuités qui apparaissent sur les histogrammes, un filtre médian 3D d’ordre 8 (26-neighbors) a été appliqué. Pour le choix du nombre de bins des histogrammes, étant donné qu’il n’existe pas de règle adaptée à toute situation, nous avons préféré une approche empirique à une approche statistique. Nous avons fait varier le nombre de bins entre 50 et 250. Nous avons ainsi choisi le nombre de bins égal à 101. Sachant que le volume cérébral varie considérablement d’un sujet à l’autre, nous avons normalisé les histogrammes, en choisissant une normalisation “totale”, i.e. diviser toutes les valeurs de l’histogramme par le nombre total de voxels et par la taille du bin (formule 1.9), ainsi les histogrammes sont indépendants de la taille des bins.

Les 8 métriques décrites dans la section 3.1.2 ont été extraites de chaque histogramme : Moyenne (Moy), médiane (Med), amplitude (PH) et position (PL) du pic, 10ème quantile (Q10), 25ème quantile (Q25), 75ème quantile (Q75) et le 90ème quantile (Q90). Ce sont les variables utilisées dans la classification de la population ADNI.

3.2.1.6 Classification de la population

3.2.1.6.1 Sélection des caractéristiques

Le but de cette classification est de classer les sujets dans les 4 groupes : HC, AD, sMCI et pMCI. Les variables utilisées sont les 8 métriques des histogrammes des 5 biomarqueurs (EC, curv_int sillons, curv_int gyri, curv_ext sillons, curv_ext gyri). Nous avons donc un total de 40 variables. Nous cherchons à réduire cette dimensionnalité en utilisons une approche “filtre” qui permet d’extraire les variables les plus informatives : le gain d’information (GI). Notre choix s’est porté sur une méthode filtre car elle ne nécessite aucun apprentissage, elle est ainsi moins coûteuse en calcul et le temps d’exécution est plus rapide.

Le GI a pour but de sélectionner l’attribut qui a le gain le plus élevé. Il mesure la diminution de l’entropie, ou la quantité de connaissance gagnée si une variable A

est présente ou absente. Elle est définie par :

$$GI(A) = H(S) - \sum \frac{S_i}{S} H(S_i), \quad (3.12)$$

avec $H(S)$ l'entropie de la classe donnée, et $H(S_i)$ l'entropie de la i -ème classe générée par la classe S sachant la variable A .

Dans l'apprentissage automatique, le GI est utilisé pour aider à classer les variables. Une variable avec un gain d'information élevé est classée plus haut que les autres, car elle a un pouvoir de classification plus élevé. Ainsi, dans notre cas, nous choisirons les 10 variables avec les GI les plus élevés.

3.2.1.6.2 Classification par SVM

Les machines à vecteurs de support (SVM²³) représentent la méthode de classification supervisée la plus utilisée en neuroimagerie, que ce soit pour le diagnostic, le pronostic ou la prédiction de maladies diverses, telles que la maladie d'Alzheimer, la sclérose en plaques, la maladie de Parkinson... Voir la revue de [Orrù et al. \(2012\)](#) qui présente des études en neuroimagerie ayant utilisé les SVM et pour des revues plus récentes sur l'utilisation des algorithmes d'apprentissage en imagerie médicale voir [Pellegrini et al. \(2018\)](#) et [Mateos-Pérez et al. \(2018\)](#).

L'approche supervisée vise à développer un algorithme permettant la classification d'observations en classes distinctes. Cela implique d'optimiser les paramètres de l'algorithme à l'aide de données d'apprentissage, puis d'évaluer ses performances à l'aide de données test. Dans cette section nous présenterons le principe général. Pour les détails mathématiques, voir [Cortes and Vapnik \(1995\)](#).

— Principe général des SVM

L'algorithme SVM construit un ensemble d'hyperplans dans un espace de dimension élevée ou infinie (figure 3.9-a). L'objectif est de déterminer l'hyperplan "optimal" (figure 3.9-b) i.e. celui pour lequel la distance minimale aux ensembles d'apprentissage (appelée marge) est maximale ([Boser et al., 1992](#)) , d'où son second nom de "Séparateur à Vaste Marge" et cela en résolvant un problème d'optimisation quadratique par la méthode des multiplicateurs de Lagrange ([Scholkopf and Smola, 2001](#)).

Il arrive souvent que les ensembles à discriminer ne soient pas séparables linéairement dans cet espace (voir la figure 3.9-c). Pour cette raison, les noyaux peuvent être utilisés pour apprendre les limites de décision non linéaires. Dans ce cas, un noyau non-linéaire est utilisé pour cartographier implicitement les données de l'espace d'entrée (là où aucune séparation linéaire des données n'est possible) à un espace de dimension plus élevée, où les données peuvent être séparées par un hyperplan linéaire (voir la figure 3.9-d). En théorie, les noyaux non linéaires permettent au classifieur de résoudre des problèmes de classification plus difficiles que les noyaux linéaires.

23. "Support Vector Machine" en anglais.

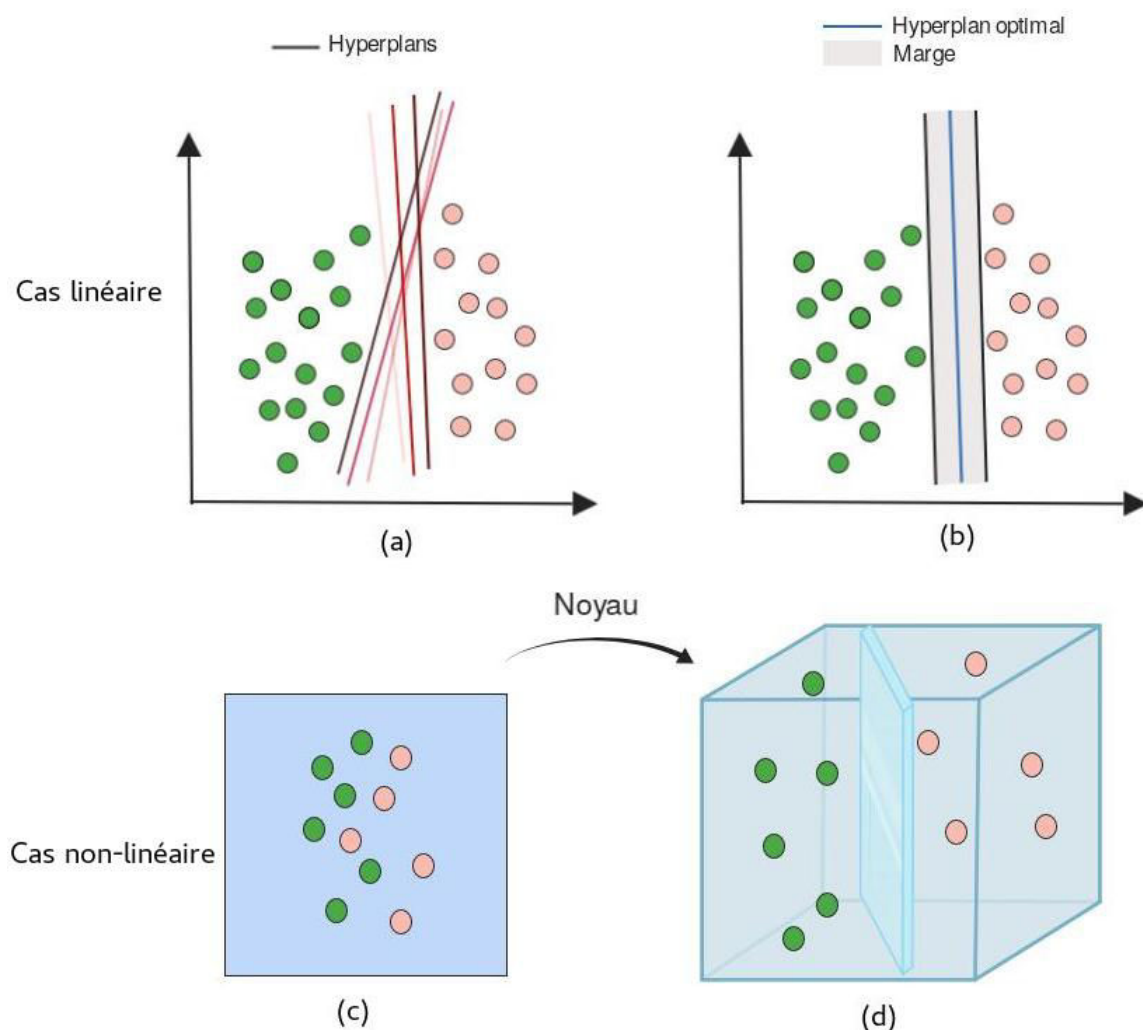


FIGURE 3.9 – Schéma des SVM : (a) et (b) représentent le cas linéaire, (c) et (d) représentent le cas non-linéaire. (a) il existe une infinité d'hyperplans permettant de séparer les ensembles. (b) L'hyperplan optimal est celui pour lequel la marge est maximale. (c) Données non séparables linéairement. (d) Plongement dans un espace de dimension plus élevée où les données peuvent être séparées par un hyperplan.

— Cross-validation

La cross-validation (i.e. validation croisée) est principalement utilisée dans l'apprentissage automatique pour estimer les compétences d'un modèle d'apprentissage automatique sur des données invisibles i.e. utiliser un échantillon limité afin d'estimer comment le modèle pourrait se généraliser lorsqu'il est utilisé pour faire des prédictions sur des données test. Il en résulte généralement une estimation moins biaisée ou moins optimiste sur la capacité du modèle que la simple répartition train/test, et diminuer ainsi le risque que le modèle sur-apprenne.

Nous avons choisi une 10-fold cross-validation (10CV). La procédure générale est la suivante :

- Diviser le jeu de données en 10 folds.
- Stratification : Il est important lors de la division des données en 10 folds que chaque fold ait la même proportion d'observations.
- Pour chaque fold :
 - Prendre le fold comme un ensemble test.

- Prendre les groupes restants comme un ensemble de données d'entraînement.
 - Ajuster le modèle sur l'ensemble d'apprentissage et l'évaluer sur l'ensemble test.
 - Conserver le score d'évaluation.
- Résumer les compétences du modèle à l'aide des scores d'évaluation du modèle.

La cross-validation a été répétée 10 fois. A chaque répétition, les folds sont divisés de manière différente. Après chaque répétition, la métrique d'évaluation du modèle est calculée (e.g. précision). Les scores de toutes les répétitions sont finalement moyennés pour obtenir le score final d'évaluation du modèle. Cela donne un score d'évaluation du modèle plus "robuste" que d'effectuer une validation croisée une seule fois.

— Choix du noyau et ses paramètres

Notre choix s'est porté sur le noyau gaussien RBF²⁴ qui est un noyau qui se présente sous la forme d'une fonction de base radiale (plus précisément, une fonction gaussienne). Les paramètres de ce noyau sont : "sigma" et "C". Le paramètre sigma définit l'influence d'un seul exemple d'entraînement. Le paramètre C²⁵ contrôle la complexité du classifieur dans la mesure où il détermine le coût d'une mauvaise classification.

Nous avons testé notre classificateur avec différentes valeurs de C et sigma en utilisant une grille de recherche²⁶. ROC (Receiver Operating Characteristic) a été utilisé pour sélectionner le modèle optimal en utilisant la plus grande valeur. Nous avons utilisé le package e1071 du logiciel R.

— Évaluation de la classification

Des mesures ont été utilisées afin d'évaluer la classification : Précision (PR)²⁷, valeur prédictive positive (VPP), valeur prédictive négative (VPN), F-score et AUC (Area Under the Curve). Toutes ces mesures sont déduites à partir de la matrice de confusion (tableau 3.4).

–Précision : la proportion des prédictions correctes effectuées par le modèle.

$$PR = \frac{VP + VN}{VP + VN + FP + FN}$$

–Sensibilité : la probabilité de détection (e.g. la proportion de patients atteints de la MA et qui sont identifiés comme tels).

$$SE = \frac{VP}{VP + FP}$$

24. "Radial basis function"

25. Paramètre "coût".

26. sigma = [0, 0.01, 0.02, 0.025, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1, 0.25, 0.5, 0.75, 0.9] et C = [0, 0.01, 0.05, 0.1, 0.25, 0.5, 0.75, 1, 1.5, 2.5].

27. "Accuracy" en anglais. Attention en anglais "precision" correspond à la "valeur prédictive positive" et non à l'accuracy.

–Spécificité : mesure la proportion de vrais négatifs correctement identifiés (e.g. le pourcentage de sujets sains qui sont correctement identifiés comme ne présentant pas l'état pathologique).

$$SP = \frac{VN}{VN + FP}$$

–Valeur prédictive positive : la probabilité que le patient, dont le test est positif, soit effectivement malade.

$$VPP = \frac{2VP}{2VP + FP}$$

–Valeur prédictive négative : la probabilité que le patient, dont le test est négatif, ne soit pas malade.

$$VPN = \frac{2VN}{2FN + VN}$$

–F₁-score : la moyenne harmonique de la sensibilité et de la valeur prédictive positive.

$$F_1 = \frac{2VP}{2VP + FP + FN}$$

–AUC : Aire sous la courbe ROC. Une courbe ROC est un graphique représentant les performances d'un modèle de classification. Elle trace le taux de vrais positifs en fonction du taux de faux positifs. Plus l'AUC est élevée, "meilleur" est le modèle.

		Classe réelle	
		+	-
Classe Prédite	+	Vrais positifs (VP)	Faux positifs (FP)
	-	Faux négatifs (FN)	Vrais négatifs (VN)

Tableau 3.4 – Matrice de confusion. Elle mesure la qualité d'un système de classification.

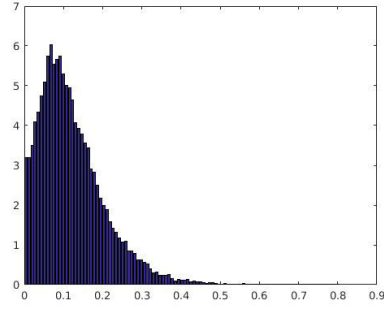
3.2.2 Résultats

3.2.2.1 Métriques d'histogrammes et analyses statistiques

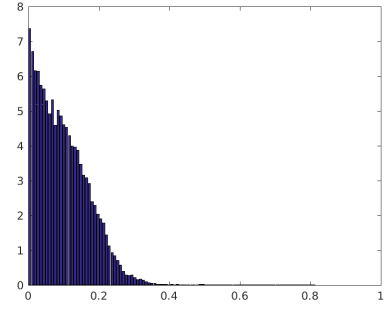
Tous les histogrammes extraits des images paramétriques d'épaisseur corticale et de courbure corticale (figure 3.10) sont normalisés et filtrés. Huit métriques d'histogrammes sont utilisées pour résumer les histogrammes des 5 biomarqueurs : EC, GyrExt, GyrInt, SilExt, SilInt. Les différences de ces métriques entre groupes ont été mesurées à l'aide d'une analyse MANOVA²⁸ de dimension 40 (8 métriques × 5 histogrammes) à un seul facteur (le diagnostic : HC, AD, sMCI et pMCI).

Nous avons alors observé une interaction significative entre les mesures et les groupes : $F(3,275)=1.95$; $p<0.001$ (Détails sur le tableau 3.5).

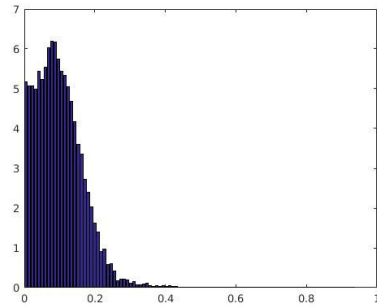
28. Analyse de variance multivariée [Hand and Taylor \(1987\)](#). MANOVA est une généralisation de l'analyse de la variance (ANOVA), qui est univariée i.e. qui ne porte que sur une seule variable dépendante.



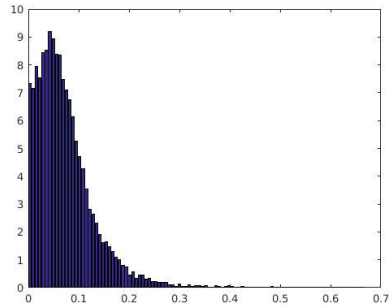
(a) Courbure interne au niveau des gyri



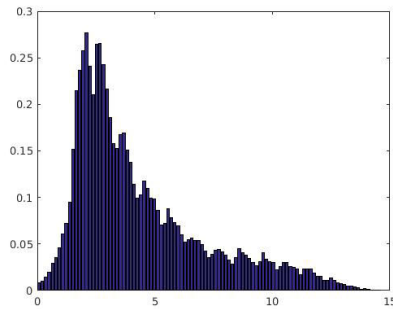
(b) Courbure interne au niveau des sillons



(c) Courbure externe au niveau des gyri



(d) Courbure externe au niveau des sillons



(e) Épaisseur corticale

FIGURE 3.10 – Histogramme extraits des images paramétriques d'épaisseur corticale (e) et de courbure corticale (a,b,c,d). Tous les histogrammes sont normalisés car le volume cérébral varie d'un sujet à l'autre et filtrés pour atténuer les discontinuités, avec un nombre de bins=101.

3.2.2.2 Classification de la population ADNI

Les SVM effectuent une classification binaire. Ainsi, nous procéderons à 6 classifications séparées : les groupes AD/HC, HC/sMCI, HC/pMCI, AD/sMCI, AD/pMCI et sMCI/pMCI.

3.2.2.2.1 Les variables discriminantes

D'après la procédure décrite dans la section 3.2.1.6.1, nous avons déterminé pour les 6 groupes de la classification, les variables optimales, i.e. les variables qui discriminent le mieux les différents groupes. Passant ainsi de 40 variables initiales à 10 variables optimales. Ces variables sont présentées sur le tableau de l'annexe III. Les caractéristiques les plus discriminantes selon le GI sont principalement celles de l'épaisseur corticale et de la courbure corticale externe.

Variables	HC	AD	sMCI	pMCI	valeur p
moy_EC	2.41 ± 0.29	2.02 ± 0.23	2.26 ± 0.21	2.22 ± 0.28	<0.001
med_EC	2.02 ± 0.26	1.69 ± 0.19	1.89 ± 0.18	1.86 ± 0.24	<0.001
EC_Q10	0.17 ± 0.03	0.12 ± 0.03	0.15 ± 0.03	0.15 ± 0.03	<0.001
EC_Q25	0.69 ± 0.09	0.56 ± 0.08	0.65 ± 0.07	0.63 ± 0.09	<0.001
EC_Q75	4.48 ± 0.56	3.76 ± 0.43	4.16 ± 0.37	4.10 ± 0.52	<0.001
EC_Q90	7.87 ± 1.00	6.59 ± 0.80	7.29 ± 0.73	7.17 ± 1.00	<0.001
EC_PL	1.59 ± 0.12	1.47 ± 0.18	1.56 ± 0.09	1.56 ± 0.11	<0.001
EC_PH	0.05 ± 0.01	0.05 ± 0.01	0.05 ± 0.00	0.05 ± 0.01	<0.001
moy_GyrExt	0.10 ± 0.01	0.10 ± 0.01	0.10 ± 0.00	0.10 ± 0.01	<0.001
med_GyrExt	0.09 ± 0.01	0.09 ± 0.01	0.09 ± 0.01	0.09 ± 0.01	<0.001
Q10_GyrExt	0.02 ± 0.00	0.02 ± 0.00	0.02 ± 0.00	0.02 ± 0.00	<0.001
Q25_GyrExt	0.04 ± 0.00	0.05 ± 0.00	0.04 ± 0.00	0.04 ± 0.00	<0.001
Q75_GyrExt	0.14 ± 0.01	0.15 ± 0.01	0.14 ± 0.01	0.14 ± 0.01	<0.001
Q90_GyrExt	0.18 ± 0.01	0.19 ± 0.01	0.18 ± 0.01	0.19 ± 0.01	<0.001
GyrExt_PL	0.01±0.02	0.01±0.03	0.01±0.03	0.01±0.02	n.s
GyrExt_PH	0.03±0.00	0.03±0.00	0.03±0.01	0.03±0.00	n.s
moy_SilExt	0.10±0.00	0.10±0.01	0.10±0.00	0.10±0.00	<0.001
med_SilExt	0.08±0.00	0.09±0.01	0.08±0.00	0.08±0.00	<0.001
Q10_SilExt	0.20±0.01	0.21±0.01	0.20±0.01	0.20±0.01	<0.001
Q25_SilExt	0.14±0.01	0.14±0.01	0.14±0.01	0.14±0.01	<0.001
Q75_SilExt	0.05±0.00	0.05±0.00	0.04±0.00	0.04±0.00	<0.001
Q90_SilExt	0.02±0.00	0.02±0.00	0.02±0.00	0.02±0.00	<0.001
SilExt_PL	0.06±0.01	0.06±0.01	0.06±0.01	0.06±0.01	n.s
SilExt_PH	0.04±0.00	0.04±0.00	0.04±0.00	0.04±0.00	0.001
moy_GyrInt	0.10±0.00	0.11±0.01	0.10±0.00	0.11±0.00	n.s
med_GyrInt	0.09±0.00	0.09±0.00	0.09±0.00	0.09±0.00	n.s
Q10_GyrInt	0.02±0.00	0.02±0.00	0.02±0.00	0.02±0.00	n.s
Q25_GyrInt	0.05±0.00	0.05±0.00	0.05±0.00	0.05±0.00	n.s
Q75_GyrInt	0.14±0.01	0.14±0.01	0.14±0.01	0.15±0.01	n.s
Q90_GyrInt	0.21±0.01	0.21±0.01	0.21±0.01	0.21±0.01	0.011
GyrInt_PL	0.06±0.01	0.06±0.01	0.06±0.01	0.06±0.01	n.s
GyrInt_PH	0.04±0.00	0.04±0.00	0.04±0.00	0.04±0.00	n.s
moy_SilInt	0.09±0.00	0.09±0.00	0.09±0.00	0.09±0.01	0.036
med_SilInt	0.08±0.00	0.08±0.00	0.08±0.00	0.08±0.00	0.013
Q10_SilInt	0.19±0.01	0.19±0.01	0.19±0.01	0.19±0.01	0.032
Q25_SilInt	0.14±0.01	0.14±0.01	0.14±0.01	0.14±0.01	0.011
Q75_SilInt	0.04±0.00	0.04±0.00	0.04±0.00	0.04±0.00	0.045
Q90_SilInt	0.01±0.00	0.01±0.00	0.01±0.00	0.01±0.00	n.s
SilInt_PL	0.003±0.00	0.003±0.00	0.003±0.00	0.003±0.00	n.s
SilInt_PH	0.04±0.01	0.04±0.01	0.04±0.01	0.03±0.01	0.004

Tableau 3.5 – Résultats du test MANOVA sur les métriques des 5 histogrammes : EC, GyrExt, SilExt, GyrInt, SilInt.
± représentent moyennes ± écart-type.

Le sous-ensemble de variables optimales obtenu a été intégré dans l'algorithme SVM.

3.2.2.2.2 La classification SVM

— Les paramètres SVM

Les résultats des réglages des paramètres du noyau RBF pour les 6 classifications SVM sont présentés sur le tableau 3.6.

	AD/HC	HC/sMCI	HC/pMCI	AD/sMCI	AD/pMCI	sMCI/pMCI
σ	0.09	0.9	0.25	0.05	0.05	0.5
C	1	0.25	0.5	0.25	0.5	1

Tableau 3.6 – Les paramètres du noyau RBF pour chaque groupe de classification SVM

— Les performances de la classification

Les performances sont pour la plupart très satisfaisantes en effet la classification SVM discrimine les sujets normaux et les patients AD avec une précision de 93,1% (AUC=0.96) mais apparaît moins effective pour la classification des patients atteints de troubles cognitifs légers stables sMCI et progressifs pMCI, avec une précision de 65.2% (AUC=0.63). Nous rappelons que les capacités de prédiction dans la conversion vers la maladie sont évalués sur un horizon temporel de conversion de 36 mois. Les performances de classification pour chaque groupe sont synthétisées dans le tableau 3.7.

	AD/HC	HC/sMCI	HC/pMCI	AD/sMCI	AD/pMCI	sMCI/pMCI
PR(%)	93.1	65.2	80.7	75	75	65.3
SE(%)	93.3	85.7	78.6	86.7	73.3	88.2
SP(%)	92.8	33.3	82.3	55.6	76.5	22.2
VPP(%)	93.3	66.6	78.6	76.5	73.3	68.2
VPN(%)	92.9	60	82.4	71.4	76.5	50
F-score	0.93	0.75	0.79	0.81	0.73	0.77
AUC	0.96	0.75	0.84	0.80	0.74	0.63

Tableau 3.7 – Résultats de la classification SVM de la population ADNI avec des variables issues des histogrammes des épaisseurs et courbures corticales.

Abrv : PR, précision ; SE, sensibilité ; SP, spécificité ; VPP, valeur prédictive positive ; VPN, valeur prédictive négative ; AUC, aire sous la courbe ROC

3.2.3 Discussion

Dans cette étude, nous avons voulu mettre en avant l'utilisation de métriques dérivées des histogrammes des biomarqueurs IRM. Pour cela, nous avons choisi comme application la classification des populations de la maladie d'Alzheimer²⁹ à l'aide de biomarqueurs du ruban cortical. En effet, de nombreuses études basées sur l'IRM structurale ont démontré que l'estimation de l'atrophie dans des régions cérébrales telles que le cortex reflète la MA et est prédictive de la progression du trouble cognitif léger vers la MA (Bakkour et al., 2009; Querbes, 2009; Busovaca et al., 2016; Racine et al., 2018). De plus, nombre de ces études atteignent une précision de classification élevée, autour de 75%-90% pour la discrimination des sujets normaux et patients atteint de MA et autour de

29. Par population Alzheimer on entend les patients atteints de la MA mais aussi les patients ayant des troubles cognitifs légers (stables ou progressives).

60%-70% pour la prédiction de la conversion des patients atteints de troubles cognitifs légers vers la MA. Les performances de classification varient du fait des différents choix de biomarqueurs, d'algorithmes de classification, types d'analyses d'image (ROIs, au niveau du voxel...) et le nombre de sujets de l'étude.

Nous avons comparé notre approche qui consiste à procéder par une analyse d'histogrammes en prélevant des métriques (e.g. moyenne, quantiles...) des histogrammes d'épaisseur corticale et de courbures corticales afin de discriminer la population Alzheimer, à des études qui ont utilisé d'autres types d'analyse d'image. En vue du nombre d'études sur la classification des populations de la maladie d'Alzheimer utilisant l'apprentissage automatique ([Pellegrini et al., 2018](#)), nous avons restreint notre comparaison à des études utilisant des biomarqueurs issus de l'IRM pondérée en T1 et ayant utilisé un classifieur SVM. Pour évaluer et comparer les performances de notre méthode avec les approches classiques tels que l'analyse par régions d'intérêt ou voxel à voxel, six classifications ont été réalisées : AD/HC, sMCI/pMCI, HC/pMCI, HC/sMCI, AD/pMCI et AD/sMCI. Pour chaque classification, les variables³⁰ les plus discriminantes ont été sélectionnées à l'aide de la mesure du gain d'information et les valeurs optimales des paramètres du noyau SVM ont été déterminées à l'aide d'une grille de recherche et de 10CV sur la base d'apprentissage. De cette manière, nous avons obtenu des estimations non biaisées des performances pour chaque classification. Les tableaux [3.8](#), [3.9](#), [3.10](#), [3.11](#), [3.12](#) et [3.13](#) présentent les résultats des comparaisons.

Pour la discrimination des sujets sains et des patients atteints de MA, nous avons atteint une précision de 93.1% (AUC=0.96) avec notre approche basée sur l'analyse d'histogrammes d'images quantitatives d'épaisseur et de courbure corticale ce qui surpasse les résultats des études qui utilisent des approches type ROI ou voxel à voxel (tableau [3.8](#)). [Park et al. \(2012\)](#) ont entraîné des SVM avec comme variables l'épaisseur corticale et la profondeur des sillons afin de différencier les patients AD et les sujets contrôles. Ils ont testé les performances de la classification en combinant les biomarqueurs avec une approche ROI et ont obtenu une précision de 85%. A titre de comparaison, [Cai et al. \(2017\)](#) ont combiné la largeur des sillons et trois autres mesures morphologiques courantes (épaisseur corticale, volume cortical et volume sous-cortical), également avec une analyse par ROI, et ont obtenu une précision de 90.9%. [Vanquin \(2015\)](#), dans sa thèse, a mesuré l'épaisseur corticale et les courbures corticales au niveaux des sillons et des gyri dans différents régions du cortex et les a intégré, séparément, dans un classifieur SVM. Il a obtenu un maximum de 0.89 AUC, correspondant à l'épaisseur corticale (et un minimum de 0.54 AUC pour la courbure interne des gyri).

En termes de capacité prédictive (sur un horizon de 36 mois), i.e. la discrimination des MCI stables et les MCI progressives, en comparant avec d'autres études (tableau [3.9](#)), la précision reste relativement basse que ce soit avec notre approche ou les analyses par ROI ou voxel à voxel. Cela peut s'expliquer par le fait que le sous-groupe sMCI est considéré comme un groupe

30. Elles représentent dans notre cas les différentes métriques d'histogrammes des 5 biomarqueurs : Épaisseur corticale, courbure corticale interne au niveau des gyri, courbure corticale interne au niveau des sillons, courbure corticale externe au niveau des gyri et courbure corticale externe au niveau des sillons.

3.2. APPLICATION 1 : CLASSIFICATION DES POPULATIONS DE LA MALADIE D'ALZHEIMER

Étude	Type	AD vs HC					
		PR (%)	SE (%)	SP (%)	VPP (%)	VPN (%)	AUC
Cuingnet et al. (2011)	Atlas	-	79	90	87	84	-
Casanova et al. (2013)	Voxel	87.1	84.3	88.9	-	-	-
Min et al. (2014)	Atlas	87.1	81.3	91.5	-	-	-
Liu et al. (2015)	ROI	92.5	92.9	88.3	-	-	0.95
Vanquin (2015)	ROI	-	77	84	-	-	0.89
Cai et al. (2017)	ROI	90.9	88.9	88.4	-	-	0.83
Samper-González et al. (2018)	Voxel	87	-	-	-	-	-
Samper-González et al. (2018)	ROI	84	-	-	-	-	-
Ruiz et al. (2018)	ROI + Analyse d'histogrammes	81.3	76.6	85.2	-	-	0.89
Notre approche	Analyse d'histogrammes	93.1	93.3	92.8	93.3	92.9	0.96

Tableau 3.8 – Comparaison des performances de classification AD/HC de notre approche avec d'autres approches.

"instable" étant donné qu'il contient à la fois des MCI stabilisés mais aussi des MCI qui vont convertir dès la sortie du suivi (donc possiblement des pMCI). Nous obtenons une précision de 65.3% (AUC=0.63), alors que d'autres études (Min et al., 2014; Samper-González et al., 2018; Vanquin, 2015) atteignent, respectivement, une précision de 71.1% , 70% et une AUC de 0.72, utilisant une approche basée sur un atlas ou sur les voxels. Cette différence peut-être expliquée de différentes manières. Tout d'abord, l'horizon de conversion des patients atteints de troubles cognitifs légers vers la MA diffère entre études, en effet les études de Vanquin (2015); Min et al. (2014) ont un horizon de 2 ans alors que celui de notre étude est de 3 ans, ce qui complique la prédiction. Ensuite, la différence peut aussi provenir du fait que Vanquin (2015) ait utilisé une méthode de cross-validation pour le paramétrage de ces SVM contrairement à nous qui avons utilisé une grille de recherche.

L'avantage de notre approche basée sur l'analyse d'histogrammes est sa simplicité, sa rapidité et le fait d'éviter tout biais ou a priori quant aux zones du cerveau susceptibles d'être affectées par la MA, en se gardant de placer des ROIs sur les images. Notre méthode fournit une mesure globale et est utile particulièrement dans le cas de différences bien marquées, c'est pour cela qu'elle surpasse tous les autres types d'analyse d'images pour la discrimination des groupes AD/HC (tableau 3.8), HC/pMCI (tableau 3.10), HC/sMCI (tableau 3.11), AD/pMCI (tableau 3.12) et AD/sMCI (tableau 3.13). Toutefois, il semble que notre approche atteigne ses limites dans des cas plus subtils tels que la différenciation entre les patients ayant converti vers la MA (pMCI) et ceux restés stables durant le suivi (sMCI) (tableau 3.9). Ce résultat n'est pas surprenant car bien que l'histogramme fournisse des informations sur l'hétérogénéité du tissu d'intérêt, l'une de ses limites est l'absence de spécificité spatiale et si les effets de la maladie ne sont localisés que sur une partie du cerveau, la sensibilité aux changements sera réduite. De plus, l'approche *analyse d'histogrammes* consiste à résumer toute une image en quelques simples indices numériques, ce qui pourrait finalement effacer d'importantes informations. Il serait ainsi plus intéressant d'utiliser ce type d'approche dans le cas de maladies diffuses telles que la Sclérose en plaques.

Étude	Type	sMCI vs pMCI					
		PR (%)	SE (%)	SP (%)	VPP (%)	VPN (%)	AUC
Cuingnet et al. (2011)	ROI	-	24	82	43	66	-
Ye et al. (2011)	Atlas	56.1	94.1	40.8	-	-	-
Casanova et al. (2013)	voxel	61.5	45.8	75.5	-	-	-
Adaszewski et al. (2013)	Density-map	62	-	-	-	-	-
Min et al. (2014)	Atlas	71.1	73.9	68.2	-	-	-
Vanquin (2015)	ROI	-	70	66	-	-	0.72
Samper-González et al. (2018)	Voxel	66	-	-	-	-	-
Samper-González et al. (2018)	ROI	70	-	-	-	-	-
Notre approche	Analyse d'histogrammes	65.3	88.2	22.2	68.2	50	0.63

Tableau 3.9 – Comparaison des performances de classification sMCI/pMCI de notre approche avec d'autres approches.

Étude	Type	HC vs pMCI					
		PR (%)	SE (%)	SP (%)	VPP (%)	VPN (%)	AUC
Cuingnet et al. (2011)	ROI	-	57	93	78	82	-
Wolz et al. (2011)	ROI	77	89	62	-	-	-
Muñoz-Ruiz et al. (2012)	Voxel	78	72	82	-	-	-
Casanova et al. (2013)	Voxel	72.3	66.3	78.9	-	-	-
Samper-González et al. (2018)	Voxel	74	-	-	-	-	-
Samper-González et al. (2018)	ROI	78	-	-	-	-	-
Notre approche	Analyse d'histogrammes	80.7	78.6	82.3	78.6	82.4	0.84

Tableau 3.10 – Comparaison des performances de classification HC/pMCI de notre approche avec d'autres approches.

Étude	Type	HC vs sMCI					
		PR (%)	SE (%)	SP (%)	VPP (%)	VPN (%)	AUC
Muñoz-Ruiz et al. (2012)	Voxel	60	56	68	-	-	-
Casanova et al. (2013)	Voxel	61.4	58.2	65.5	-	-	-
Notre approche	Analyse d'histogrammes	65.2	85.7	33.3	66.6	60	0.96

Tableau 3.11 – Comparaison des performances de classification HC/sMCI de notre approche avec d'autres approches.

Étude	Type	AD vs pMCI					
		PR (%)	SE (%)	SP (%)	VPP (%)	VPN (%)	AUC
Muñoz-Ruiz et al. (2012)	Voxel	51	54	43	-	-	-
Liu et al. (2015)	ROI	-	56	61	-	-	-
Notre approche	Analyse d'histogrammes	75	73.3	76.5	73.3	76.5	0.74

Tableau 3.12 – Comparaison des performances de classification AD/pMCI de notre approche avec d'autres approches.

Étude	Type	AD vs sMCI					
		PR (%)	SE (%)	SP (%)	VPP (%)	VPN (%)	AUC
Muñoz-Ruiz et al. (2012)	Voxel	65	62	66	-	-	-
Liu et al. (2015)	ROI	-	77	73	-	-	-
Notre approche	Analyse d'histogrammes	75	86.7	55.6	76.5	71.4	0.80

Tableau 3.13 – Comparaison des performances de classification AD/sMCI de notre approche avec d'autres approches.

3.3 Application 2 : essai clinique de Sclérose en plaques

L'IRM cérébrale joue un rôle primordial dans le diagnostic, la réponse au traitement et la prédiction de la progression de la sclérose en plaques. La sclérose

en plaques est une maladie inflammatoire, démyélinisante et neurodégénérative caractérisée par la présence de multiples lésions dans le système nerveux central et cliniquement par des poussées et une aggravation du handicap.

Au cours des deux dernières décennies, des progrès significatifs ont été accomplis dans le traitement de la sclérose en plaques avec l'introduction de traitements efficaces pour la forme rémittente récurrente de la maladie ([Miller and Rhoades, 2012](#)). Mais à ce jour, les résultats des essais cliniques sur la forme progressive de la maladie (SEP-SP et SEP-PP) ont été généralement décevants. En effet, les nouveaux traitements semblent plus efficaces pour prévenir les poussées, les lésions ou pour atténuer la progression du handicap à court terme qu'à long terme. [De Angelis et al. \(2018\)](#) présentent une revue des traitements récents utilisés dans la SEP progressive.

La compréhension des mécanismes diffus sous-jacents de la SEP est complexe et nous pensons qu'une approche multimodale est nécessaire pour générer des mesures combinées reflétant les poids respectifs de ces mécanismes. Aujourd'hui, nous avons accès à un large éventail de biomarqueurs et il a été admis qu'un biomarqueur unique ne peut, à lui seul, refléter tous les mécanismes de la maladie. Par conséquent, de nombreuses études se sont intéressées à la combinaison de biomarqueurs pour améliorer la précision du diagnostic. Pour ce faire, des méthodes linéaires ([Su and Liu, 1993](#); [Pepe and Thompson, 2000](#); [Liu et al., 2011](#); [Yin and Tian, 2014](#); [Kang et al., 2016](#)) ou des méthodes non linéaires ([Kouskoumvekaki et al., 2008](#); [Huang and Fong, 2014](#); [Xu et al., 2015](#)) ont été utilisées pour combiner plusieurs biomarqueurs dans diverses maladies neurologiques.

L'IRM conventionnelle fournit une information cruciale sur les mécanismes de la SEP, telle que la quantification de l'atrophie cérébrale en mesurant le volume des tissus cérébraux et l'épaisseur corticale ([Steenwijk et al., 2016](#)). Cependant, les séquences classiques d'IRM ne sont pas en mesure de détecter ou de quantifier les caractéristiques hétérogènes de la maladie (e.g. la démyélinisation, la remyélinisation et la perte axonale). Les séquences IRM non conventionnelles fournissent des biomarqueurs capables de surmonter, au moins partiellement, ces limites. En effet, de nombreuses études suggèrent que le ratio de transfert d'aimantation (MTR) est principalement un biomarqueur de démyélinisation ([Schmierer et al., 2004](#)), et que la fraction d'anisotropie (FA) et le coefficient de diffusion apparent (ADC) de l'imagerie de diffusion sont représentatifs de la destruction tissulaire et de la perte axonale ([Cercignani et al., 1999](#); [Rovaris et al., 2005](#)).

La méthode de lecture d'un essai thérapeutique doit être aussi précise que possible pour déduire l'effet d'un traitement. C'est pourquoi nous proposons d'aller plus loin que les études traditionnelles qui proposent une analyse individuelle des biomarqueurs basée sur une analyse statistique entre le groupe traité et le groupe placebo. L'idée est d'effectuer une analyse d'histogrammes et d'extraire un certain nombre de métriques de ces histogrammes. En effet, certaines études ont étudié les avantages de l'analyse par histogramme dans les maladies neurologiques, d'après [Tofts et al. \(2004a\)](#), les histogrammes fournissent des informations utiles sur les mécanismes sous-jacents de la maladie et permettent la quantification du changement pathologique diffus. L'utilisation de cette approche est particulièrement intéressante dans le cas de la SEP en

raison de la présence simultanée de lésions dans le SNC mais aussi une altération diffuse dans les tissus cérébraux. Les mesures IRM tel que le MTR ([Dehmeshki et al., 2001a,b](#)) et les biomarqueurs de diffusion (i.e. ADC et FA) sont sensibles à ces subtiles anomalies diffuses et nous utilisons donc l'analyse par histogramme de ces biomarqueurs pour prendre en compte les changements diffus dans la SEP. Après l'analyse d'histogrammes, nous utilisons une analyse factorielle appelée *analyse factorielle multiple* (AFM) pour extraire des combinaisons linéaires des biomarqueurs (i.e. les composantes principales). Enfin, nous avons utilisé un algorithme de classification hiérarchique basé sur ces composantes principales (HCPC³¹).

3.3.1 Matériels

3.3.1.1 La population d'étude

3.3.1.1.1 La base de donnée : essai clinique MS-SPI

Notre laboratoire a participé à l'essai thérapeutique MS-SPI conduit par les laboratoires MEdDay sur le MD1003. Il s'agit d'un essai multicentrique randomisé de phase III, contrôlé par placebo³², visant à étudier l'effet de la biotine à forte dose (MD1003) sur la SEP progressive. Il regroupe au total 154 patients recrutés dans 16 centres avec 103 patients randomisés dans le bras MD1003 et 51 dans le bras placebo. Le MD1003 est un médicament expérimental susceptible de promouvoir l'activité de plusieurs enzymes, notamment l'acétylCoA carboxylase, impliquées dans la production d'énergie et la synthèse de la myéline ([Peyro Saint Paul et al., 2016](#)). Son mode d'action peut affecter les cibles liées à la SEP progressive. L'étude pilote indique une amélioration clinique (i.e. EDSS et T25FW) chez 91% des participants dans les deux à huit mois suivant le début du traitement par MD1003. Pour plus de détails sur les critères d'inclusion, d'exclusion et sur le MD1003, voir [Sedel et al. \(2015\)](#); [Toubah et al. \(2016b\)](#) ou encore la thèse de [Dutilleul \(2015\)](#).

Les données de notre étude ont été obtenues à partir d'une étude IRM ancillaire de l'essai clinique MS-SPI qui regroupe 75 participants. Au cours d'une phase contrôlée par placebo de 12 mois, les patients ont été randomisés pour recevoir trois fois par jour (TID) le MD1003 (100 mg de biotine) ou un placebo par voie orale, suivi par le MD1003 pour tous les patients à partir du 12ème mois. Le schéma de l'essai clinique est présenté sur la figure 3.11.

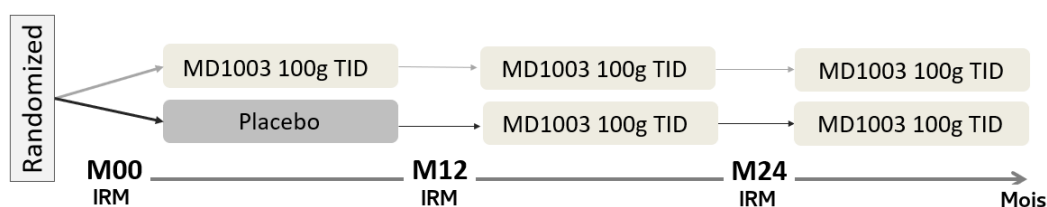


FIGURE 3.11 – Plan d'expérience de l'étude IRM ancillaire de l'essai MS-SPI.

31. "Hierarchical Clustering on Principal Components"

32. Le double aveugle a été levé dans le courant de la thèse, nous avons ainsi pu faire nous même les analyses statistiques.

3.3.1.1.2 La population analysée

Parmi les 75 participants de l'étude IRM ancillaire de l'essai MS-SPI, nos analyses ont été menées sur 40 patients. Parmi ces patients, 29 ont reçu le traitement MD1003 et 11 patients ont reçu un placebo durant les 12 premiers mois. Lors du suivi longitudinal de 2 ans, certains patients ne se sont plus présentés aux examens annuels d'IRM, et la méthode d'analyse que nous avons choisie, décrite plus bas, n'opère pas sur des données manquantes. De plus, nous avons sélectionné les sujets de sorte qu'ils soient appariés en âge et en sexe (voir les détails démographiques sur le tableau 3.14).

	MD1003	Placebo	valeur-p
Nombre de sujet N=40	29	11	–
Âge	51.4±9.0	50.5±7.5	0.9
Genre Femme/Homme	13/16	7/4	0.3
Forme de SEP (SEP-PP/SEP-SP)	14/15	6/5	0.8
Durée de la maladie	15±6.7	19.6±7.7	0.2
EDSS	5.8±1.0	6.3±0.5	0.06

Tableau 3.14 – Information démographique de la population de notre étude
 ± représentent moyennes ± écart-type. Toutes les valeurs p sont basées sur les tests U de Mann-Whitney, à l'exception du Genre et les formes de SEP, qui sont basées sur le test khi-deux χ^2 . Le MD1003 est le traitement évalué dans l'essai clinique. Les valeurs de la durée de la maladie et l'EDSS sont celles de la baseline M00 (Abrv : SEP-PP, SEP primaire progressive; SEP-SP, SEP secondairement progressive; EDSS, Expanded Disability Status Scale)

3.3.1.2 Acquisitions IRM

L'essai MS-SPI est un essai multicentrique, tous les paramètres des acquisitions IRM ont été homogénéisés entre les six centres participant : Clermont-Ferrand, Marseille, Paris, Reims, Rennes et le notre à Toulouse. Toutefois, les images du centre de Clermont-Ferrand n'ont pas été exploitées car le centre a utilisé une résolution spatiale différente que celle des autres centres pour la séquence 3D-T1.

Cette sous-étude IRM réalisée à 3 Tesla regroupe les séquences suivantes :

- 3D-T1
- T1 Spin Echo
- T1 Spin Echo avec injection de Gadolinium (0.2mL/kg avec 5min de délai)
- T2-2D
- DP-2D
- FLAIR-3D
- Diffusion
- Spectroscopie par résonance magnétique
- Relaxometrie
- Transfert d'aimantation

Toutefois, les analyses de notre étude ont été effectuées sur les séquences T1-3D, imagerie de diffusion et de transfert d'aimantation. Les paramètres d'acquisition de ces différentes séquences IRM sont résumés dans le tableau 3.15.

		SIEMENS	PHILIPS
T1-3D	Nombre de coupes	176	182
	Résolution(mm ³)	1.0×1.0×1.0	1.0×1.0×1.0
	TE(ms)	2.26	4.6
	TR(ms)	1900	9.8
	Matrice (mm)	256×256×176	256×256×182
Diffusion	Nombre de coupes	60	60
	Résolution(mm ³)	2.0×2.0×2.0	2.0×2.0×2.0
	TE(ms)	99	64
	TR(ms)	11000	10000
	Matrice (mm)	256×256×120	256×256×120
Transfert d'aimantation	Nombre de coupes	44	120
	Résolution(mm ³)	1.0×1.0×3.0	1.0×1.0×1.0
	TE(ms)	4.5	5
	TR(ms)	750	62
	Matrice (mm)	192×256×132	256×256×120

Tableau 3.15 – Paramètres d'acquisition des différentes séquences utilisées dans notre étude

3.3.1.3 Analyse par IRM multimodale

L'analyse des données des différentes séquences a été répartie entre les différents centres :

- Marseille : Transfert d'aimantation
- Rennes : Imagerie du tenseur de diffusion
- Toulouse : Analyse structurelle (telle que la volumétrie). Ainsi nous sommes chargés, à Toulouse, de l'analyse des acquisitions 3D-T1.

De ces trois séquences, quatre biomarqueurs ont été extraits : Épaisseur corticale (EC), coefficient de diffusion apparent (ADC), Fraction d'anisotropie (FA) et le ratio de transfert d'aimantation (MTR), comme cela a été résumé dans le tableau 3.16 et dont les cartes paramétriques sont représentées sur la figure 3.12.

	Centre d'analyse	Biomarqueurs et cartes paramétriques	Région	Approche / Logiciel
T1-3D	Toulouse	Épaisseur corticale (EC)	Cortex entier	<i>Matlab Toolbox Corthizon</i>
Diffusion	Rennes	- Coefficient de diffusion apparent (ADC) - Fraction d'anisotropie (FA)	Cerveau entier*	Formules 1.5 et 1.6 <i>MedINRIA</i>
Transfert d'aimantation	Marseille	Ratio de transfert d'aimantation (MTR)	Cerveau entier*	formule 1.7 <i>SPM8</i>

Tableau 3.16 – Les différentes séquences IRM utilisées dans notre étude et les biomarqueurs correspondants

* Les voxels correspondant au LCS ont été exclus de nos analyses

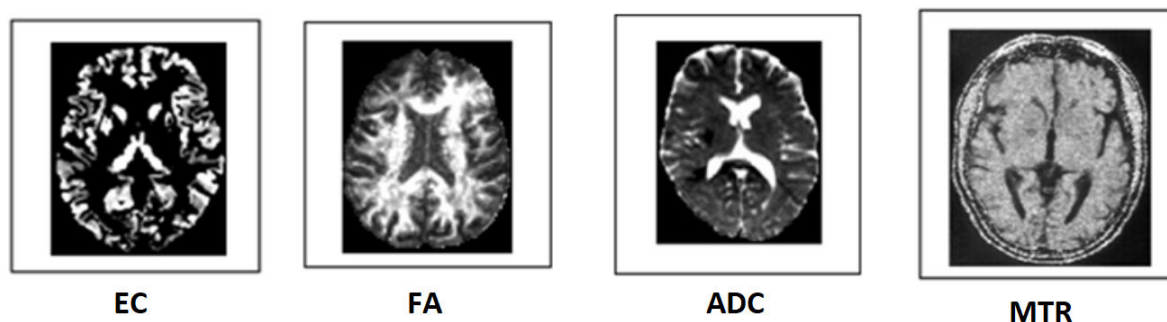


FIGURE 3.12 – Les différents biomarqueurs IRM utilisés dans cette étude.

3.3.1.3.1 Imagerie structurale

Nous n'avons pas effectué d'analyses volumétriques car [Dutilleul \(2015\)](#), dans sa thèse, a déjà exploré ce marqueur dans le cadre de l'essai MS-SPI. De plus nous avons besoin d'un biomarqueur structural qui pourrait être estimé dans chaque voxel de l'image, ce qui nous permettrait d'en extraire l'histogramme et d'y appliquer l'approche décrite dans la section [3.1](#).

La mesure IRM que nous avons choisi d'extraire de l'IRM pondérée en T1 est l'épaisseur corticale (EC). Tout comme dans la précédente application elle a été estimée à l'aide de la toolbox Matlab CorThiZon. L'EC a été mesurée sur tout le cortex à l'aide de l'équation de Laplace ([Querbes et al., 2009](#)) comme cela a été décrit dans la section [3.2.1.4.1](#).

3.3.1.3.2 Tenseur de diffusion

Le prétraitement des données DTI comprend trois étapes : Correction des distorsions causées par les courants de Foucault, l'extraction du crâne, et l'estimation du tenseur de diffusion. Ce prétraitement est réalisé par un opérateur du centre de Rennes utilisant le logiciel MedINRIA ([Toussaint et al., 2007](#)). MedINRIA est un logiciel de traitement et de visualisation d'images médicales regroupant différents modules. Le module "DTI Track" fournit tous les outils nécessaires pour une analyse en profondeur des données DTI et le suivi des fibres. Mais aussi l'estimation du champ de tenseur de diffusion, le calcul des cartes FA/ADC, lissage du tenseur et l'extraction des fibres.

L'estimation du tenseur se fait à l'aide d'une famille de mesures riemanniennes appelée log-euclidean. Ces mesures possèdent d'excellentes propriétés théoriques et donnent des résultats similaires dans la pratique, mais avec des calculs beaucoup plus simples et plus rapides. Cette nouvelle approche est basée sur une nouvelle structure d'espace vectoriel pour tenseurs. Dans ce cadre, les calculs riemanniens peuvent être convertis en calculs euclidiens une fois que les tenseurs ont été transformés en logarithmes matriciels. Pour plus de détails voir la thèse de [Fillard \(2008\)](#).

Après diagonalisation du tenseur de diffusion, le coefficient de diffusion apparent (ADC) et de la fraction d'anisotropie (FA) ont été mesurés dans chaque voxel de l'image (formule [1.5](#) et [1.6](#)) et ont été cartographiés. Les voxels du LCS n'ont pas été pris en compte. Le détail des calculs de ces paramètres de diffusion se trouve dans la partie *État de l'art* section [1.1.1.2.1](#).

3.3.1.3.3 Imagerie de transfert d'aimantation

Les cartes du ratio de transfert d'aimantation (MTR) ont été calculées à l'aide d'une méthode semi-automatique voxel à voxel, selon l'équation ([1.7](#)). Les cartes MTR ont été recalées sur les images correspondantes pondérées en T2 de chaque sujet. Les cartes MTR ont été normalisées dans l'espace MNI à l'aide du template anatomique T1 fourni dans le logiciel SPM8. Des pixels avec un pourcentage de tissu (SG + SB) supérieur à 90% ont été utilisés pour masquer la carte MTR normalisée. Ces analyses ont été effectuées par un opérateur au centre de Marseille. Pour plus de détails sur l'imagerie de transfert d'aimantation voir la section [1.1.1.2.2](#).

3.3.2 Méthodes

Après avoir extrait les différentes métriques des histogrammes des biomarqueurs, nous procédons à une étape d'extraction de caractéristique afin de réduire la dimensionnalité en utilisant une analyse factorielle multiple (AFM). Notre objectif étant de conclure plus efficacement sur l'apport d'un traitement dans un essai clinique, il était important de combiner les biomarqueurs de la manière la plus précise et efficace. Dans notre étude, nous avons commencé par utiliser l'approche classique consistant à évaluer le traitement sur la base d'une analyse individuelle et indépendante de chaque biomarqueur. Ensuite, nous avons appliqué notre approche, combinant d'abord les métriques d'histogrammes des biomarqueurs utilisant l'AFM, puis un clustering hiérarchique basé sur les composantes principales obtenus à partir de l'AFM pour classer les patients du bras MD1003 et ceux du bras placebo.

3.3.2.1 Analyse d'histogrammes

À partir des cartes paramétriques nous avons extrait quatre histogrammes pour chaque patient de l'étude : EC, ADC, FA et MTR. Pour la création de ces histogrammes, nous avons procédé de la même manière que la précédente application (section 3.2.1.5), nous avons appliqué un filtre médian3D pour atténuer les spikes des histogrammes et choisi un nombre de bins égale à 101. Puis nous avons normalisé les histogrammes et enfin nous avons extrait 11 métriques des histogrammes des 4 biomarqueurs : Q0% (min), Q10%, Q20%, Q30%, Q40%, Q50%(médiane), Q60%, Q70%, Q80%, Q90% and Q100% (max). Ainsi, pour chaque patient, nous avons un total de 44 variables à chaque point temporel M00, M12 et M24.

3.3.2.2 Analyses individuelles indépendantes

L'analyse individuelle, qui correspond à l'approche classique utilisée dans les essais thérapeutiques, consiste à effectuer des analyses statistiques sur les moyennes des biomarqueurs sur tout le cerveau³³, séparément.

- Une comparaison des moyennes des biomarqueurs entre le groupe MD1003 et le groupe Placebo est effectuée à l'aide du test U de Mann-Whitney, une alternative non paramétrique au test t à deux échantillons non appariés. Il est utilisé lorsque les données ne sont pas normalement distribuées.
- Afin d'évaluer l'effet du traitement, les variations des différents biomarqueurs mesurés chez les sujets placebo et chez les sujets traités ont également été comparées, par des tests de Mann-Whitney. La variation des biomarqueurs a été mesurée selon la formule (3.13).

$$Var(\%) = \frac{biomarqueur(M12) - biomarqueur(M00)}{biomarqueur(M00)} \times 100 \quad (3.13)$$

avec M12 la valeur du biomarqueur à 12mois. Il s'agit ainsi de la variation sur un an. La mesure de la variation sur 2 ans se fait de la même façon en remplaçant M12 par M24.

33. Ou sur des tissus spécifiques (e.g. SBAN, SGAN...), mais dans notre cas les mesures des biomarqueurs ont été effectuées sur tout le cerveau (en supprimant les voxels correspondant au LCS).

3.3.2.3 Analyse Factorielle Multiple

L'analyse factorielle multiple (AFM) (Escofier and Pagès, 1990) est une extension pondérée de l'analyse en composantes principales (ACP) qui permet d'analyser des données multivariées ou les individus sont décrits par plusieurs ensembles de variables structurés en groupes/blocs. Elle est aussi appelée ACP multiblocs (ou multitables). L'AFM permet d'analyser simultanément les tableaux de variables et d'obtenir des résultats, notamment graphiques, permettant d'étudier la relation entre les observations, les variables et les tableaux. Elle est utilisée lorsque plusieurs ensembles de variables ont été mesurés sur le même ensemble d'observations. La MFA se déroule en deux étapes : premièrement, elle effectue une ACP pour chaque bloc de données (M00, M12 et M24) et "normalise" chaque bloc en divisant tous leurs éléments par la première valeur propre de l'ACP. Puis, tous les blocs de données normalisés sont regroupés dans un tableau de données qui a son tour est analysé via une ACP. Pour plus de détails, voir Abdi et al. (2013).

Dans notre cas, les ensembles de données sont mesurés sur les mêmes observations (i.e. les 40 patients atteints de SEP progressive de l'essai) à différentes occasions (i.e. trois points temporels M00, M12 et M24). Le premier ensemble de données correspond aux données collectées à l'examen M00, le second aux données collectées 12 mois après (M12) et le troisième aux données collectées 24 mois après (M24). Le but de l'analyse est donc d'évaluer l'évolution de la position des observations (tableau 3.17). La MFA a été réalisée à l'aide du package FactoMineR du logiciel R.

Les différents constituants de la MFA sont :

- Les blocs de l'AFM : les 3 points du suivi temporel représentent les 3 blocs (M00, M12 et M24).
- Les variables actives : 11 quantiles pour chaque biomarqueur. Ainsi, un total de 44 variables pour chaque bloc.
- Variables supplémentaires/illustratives : il s'agit d'une variable qualitative qui informe si le patient fait parti du bras traité MD1003 ou du bras non traité Placebo. Une variable supplémentaire n'a aucune influence sur la création des composantes principales mais reste néanmoins importante pour l'interprétation une fois projetée sur les plans factoriels.

Sujets	Drug	M00				M12				M24			
		ADC	FA	MTR	EC	ADC	FA	MTR	EC	ADC	FA	MTR	EC
Sujet1		Q0...Q100	Q0...Q100	Q0...Q100	Q0...Q100	Q0...Q100	Q0...Q100	Q0...Q100	Q0...Q100	Q0...Q100	Q0...Q100	Q0...Q100	Q0...Q100
.	
.	
Sujet40	

Tableau 3.17 – Tableau représentatif des données de notre étude sur l'essai MS-SPI.
 "M00", "M12" et "M24" représentent les 3 blocs actifs de l'AFM, "Drug" est une variable supplémentaire qui nous informe si le patient est du bras placebo ou du bras MD1003 et "Q0...Q100" représente les 11 quantiles extraits des histogrammes des biomarqueurs (Q0%, Q10%, Q20%, Q30%, Q40%, Q50%, Q60%, Q70%, Q80%, Q90%, Q100%).

3.3.2.4 Clustering hiérarchique

L'approche HCPC (clustering hiérarchique sur composantes principales, d'après Josse 2010) nous permet de combiner les méthodes standards utilisées dans l'analyse de données multivariées : les méthodes des composantes

principales (e.g. ACP et AFM) et la classification ascendante hiérarchique.

La MFA a réduit la dimensionnalité des données à un nombre restreint de variables (i.e. composantes principales) dérivées des variables d'entrée, contenant les plus importantes informations des données. La MFA peut être considérée comme une étape de prétraitement et de débruitage pouvant conduire à une mise en cluster plus stable. En effet, elle permet de prendre en compte la structure des groupes et de rendre le clustering plus robuste en supprimant les dernières dimensions.

La classification ascendante hiérarchique (CAH) est une technique statistique visant à partitionner une population en différentes classes ou sous-groupes. La classification est ascendante car elle part des observations individuelles (tous les individus sont dans une unique classe) ; elle est hiérarchique car elle produit des classes ou groupes de plus en plus vastes, incluant eux mêmes des sous-groupes. Le principe de la CAH est de rassembler des individus selon un critère de ressemblance défini au préalable qui s'exprimera sous la forme d'une matrice de distances, exprimant la distance existant entre les individus pris deux à deux. La CAH va ensuite rassembler les individus de manière itérative afin de produire un dendrogramme (arbre de classification). De nombreuses méthodes d'agrégation existe pour construire le dendrogramme (saut minimum, distance maximum, moyenne, Ward...). Nous privilégierons la méthode de Ward. Elle cherche à minimiser l'inertie intra-classe et à maximiser l'inertie interclasse afin d'obtenir des classes les plus homogènes possibles. L'étape du découpage de l'arbre en sous-classes est importante, car c'est ainsi que nous obtenons les différents regroupements et le nombre de classe. De notre cas, nous avons choisi d'obtenir deux classes : une classe regrouperait les patients SEP faisant parti du bras *MD1003* et l'autre classe regrouperait ceux du bras *Placebo*. La classification HCPC a été réalisée à l'aide du package FactoMineR du logiciel R.

Afin d'évaluer le clustering obtenu à l'aide de HCPC, nous avons utilisé 3 mesures d'évaluation externes : Pureté, rand index et l'entropie. Les mesures d'évaluation externes permettent de comparer le regroupement proposé par le clustering aux étiquettes réelles des objets.

— Pureté

Afin de calculer la pureté ([Manning et al., 2008](#)), chaque cluster est affecté à la classe la plus fréquente dans le cluster. Puis, la précision de cette affectation est mesurée en comptant le nombre d'objets correctement assignés et en divisant par le nombre total d'objets. C'est le pourcentage du nombre total d'objets qui ont été classés correctement.

$$\text{Pureté} = \frac{1}{N} \sum_{i=1}^k \max_j |c_i \cap t_j| \quad (3.14)$$

où N est le nombre d'objets étiquetés, k est le nombre de clusters, c_i est l'ensemble des objets de la classe i , et t_j l'ensemble des objets appartenant au cluster j et $|c_i \cap t_j|$ sont les objets à la fois dans le cluster i et de la classe j .

— Rand Index :

L'indice de Rand ([Rand, 1971](#)) consiste à vérifier si les couples d'objets de

la même classe, d'après les connaissances disponibles, ont été placés dans un même cluster.

$$RI = \frac{VP + VN}{VP + FP + FN + VN} \quad (3.15)$$

Vrai positif (VP) si les deux objets sont de la même classe et sont placés dans le même cluster; vrai négatif (VN) quand les deux objets sont de classes différentes et sont placés dans deux clusters différents; faux positif (FP) correspond à deux objets de classes différentes placés dans le même cluster; faux négatif (FN) correspond à deux objets de la même classe dans deux clusters différents.

— Entropie :

L'entropie est une mesure d'incertitude des clusters et des classes :

$$H = - \sum_k \frac{|c_k|}{N} \log \frac{|c_k|}{N}. \quad (3.16)$$

3.3.3 Résultats

3.3.3.1 Analyses statistiques

— Analyses à M00, M12 et M24 de chacune des mesures

D'après le U-test de Mann-Whitney, aucune différence significative n'a été retrouvée entre les patients du bras Placebo et ceux du bras MD1003 à la baseline M00, à M12 ou à M24. Les résultats sont présentés sur la figure 3.13.

— Effet du traitement sur chacune des mesures

Lorsque nous comparons l'évolution de l'épaisseur corticale des patients de l'essai, nous observons très légèrement plus d'atrophie, en un an et en deux ans, dans le groupe Placebo que dans le groupe traité. En ce qui concerne le transfert d'aimantation, le MTR diminue pour le groupe Placebo et augmente pour le groupe traité, même observation pour la fraction d'anisotropie. Ces différences sont plus ou moins légères mais aucune n'est cependant significative entre les deux groupes. Toutefois, nous observons une tendance ($p=0.05$) entre les deux groupes dans la variation de l'ADC en un an qui pourrait sous entendre une légère remyélinisation, mais cette tendance est effacée au bout de 2 ans, ce qui pourrait être expliqué du fait que le groupe "Placebo" à M24 correspond aux patients qui ont reçu le traitement durant un an et le groupe "MD1003" sont les patients qui ont reçu le MD1003 durant deux ans.

Les résultats sont résumés dans les tableaux 3.18 et 3.19.

Variation % M00→M12	MD1003	Placebo	valeur p
EC	0.40±3.06	0.35±3.76	n.s
ADC	-0.19±1.35	0.97±1.90	0.05
FA	2.21±5.86	-0.10±3.10	n.s
MTR	0.29±2.76	-0.08± 3.15	n.s

Tableau 3.18 – Résultats de l'évolution des biomarqueurs (%), en un an (M00→M12).

Variation % M00→M24	MD1003	Placebo	valeur p
EC	-1.2±2.4	-0.98±2.9	n.s
ADC	1.9±1.1	2.5±1.4	n.s
FA	2.30±1.7	-1.4±0.9	n.s
MTR	-0.33±2.55	-0.26±3.18	n.s

Tableau 3.19 – Résultats de l'évolution des biomarqueurs (%), en deux ans (M00→M24).

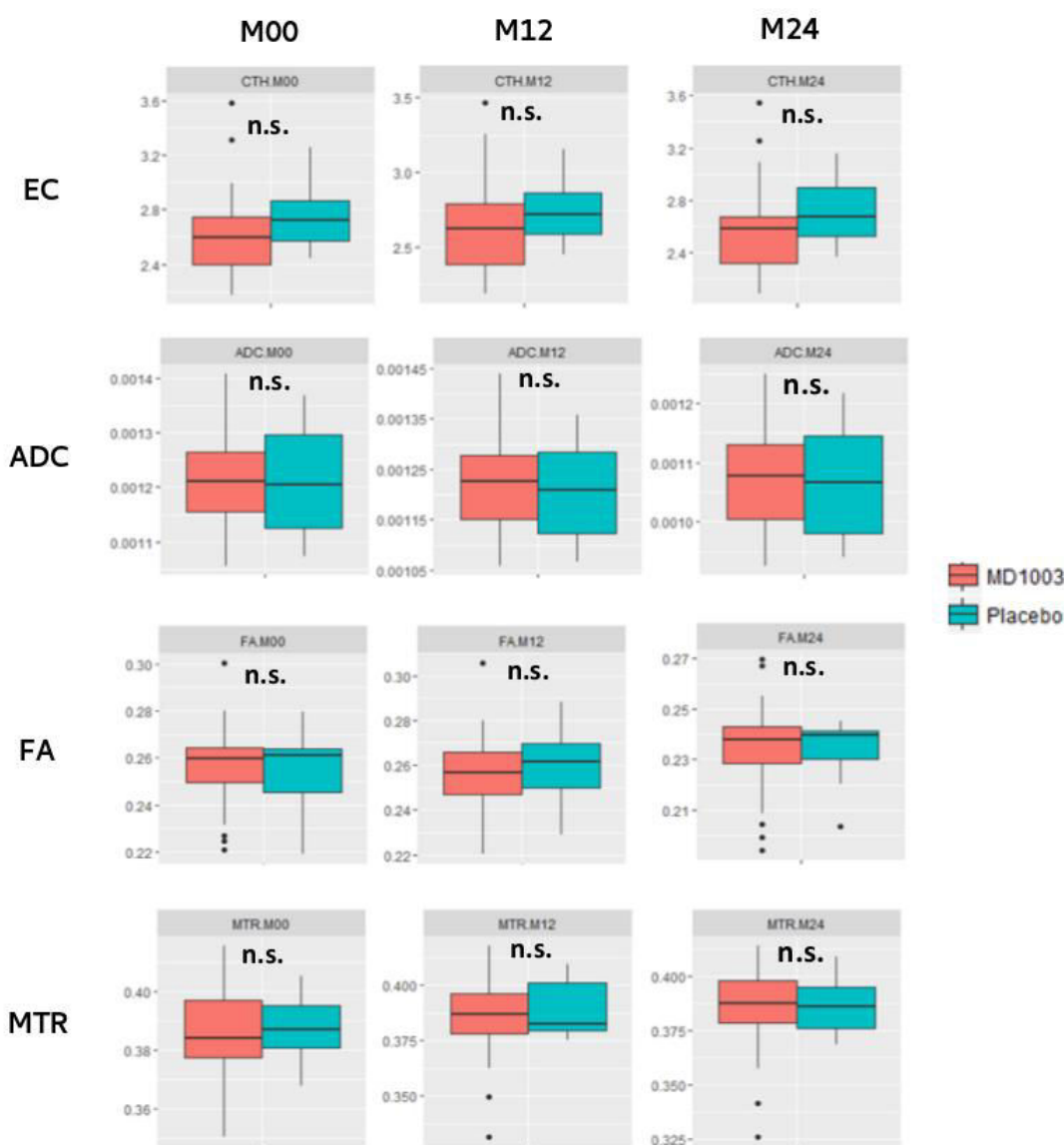


FIGURE 3.13 – Boxplot à M00, M12 et M24 de l'EC (mm), ADC ($\times 10^{-3} \text{mm}^2/\text{s}$), FA et MTR. Les deux groupes MD1003 et Placebo ont été comparés à l'aide du test de Mann-Whitney ($\alpha = 0.05$).

3.3.3.2 L'analyse factorielle multiple

Les groupes de l'AFM sont représentés par les 3 dates M00, M12 et M24 ; les variables de l'AFM par les quantiles de chaque biomarqueur. Nous avons ajouté un groupe supplémentaire appelé "Drug" qui décrit si un patient appartient au groupe MD1003 ou au groupe Placebo.

Nous choisissons le nombre d'axes principaux de façon à conserver le maximum d'inertie avec le minimum de facteur. Pour cela, nous utilisons deux critères empiriques : la règle de Kaiser ([Kaiser, 1960](#)) qui suggère de ne conserver que les axes associés aux valeurs propres supérieures à 1³⁴ et la règle du "coude" qui suggère de sélectionner les axes avant le décrochement qu'on peut observer sur l'éboulis des valeurs propres. Ainsi, comme il est démontré sur

34. Une valeur propre > 1 indique que la composante principale concernée représente plus de variance par rapport à une seule variable d'origine.

la figure 3.14, nous avons conservé les 3 premières composantes principales (PC) expliquant 71% de la variabilité totale portée par les données.

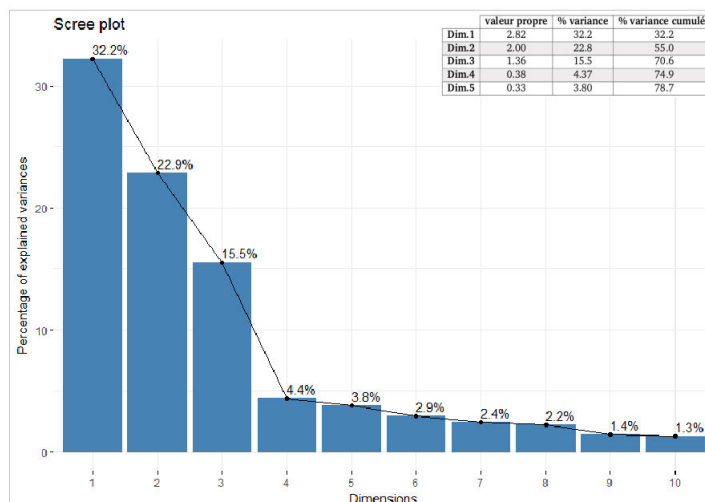


FIGURE 3.14 – Le pourcentage de variance expliquée pour chaque dimension. Nous observons un décrochement au 4ème axe, puis décroissance régulière donc d’après le critère du coude, ce sont les 3 premières dimensions qui auraient un éventuel intérêt. De plus, il apparaît que les valeurs propres des 3 premières dimensions sont supérieures à 1, respectant ainsi le critère de Kaiser.

L’AFM tout comme l’ACP, fournissent plusieurs mesures et graphiques qui permettent d’interpréter les résultats. Nous nous intéressons particulièrement aux représentations des individus, des groupes et au coefficient RV :

— La représentation des individus

La représentation des individus sur les axes principaux (figure 3.15), ne montre pas deux groupes distincts. En effet, les ellipses de confiance associées aux deux groupes MD1003 et Placebo se chevauchent, ce qui signifie que les deux groupes sont relativement similaires.

— La représentation des groupes

Sur la figure 3.16, nous observons que les coordonnées des 3 groupes actifs (M00, M12 et M24) sur les 1ère, 2ème et 3ème dimensions sont quasiment identiques, elles contribuent de la même manière à la construction des composantes principales. À la différence du groupe supplémentaire ”Drug”, qui nous le rappelons, définit si les patients sont du bras traité ou du bras placebo, ses coordonnées sont très proches de l’origine des axes indiquant une contribution faible, voire nulle, à l’AFM.

— Coefficient RV

Le coefficient RV (Robert and Escoufier, 1976) permet d’évaluer la corrélation entre deux matrices de variables quantitatives ou deux tableaux issus d’analyses multivariées. Il peut être interprété comme une généralisation du coefficient de corrélation entre deux ensembles de variables.

Le tableau 3.20 montre que les groupes actifs (M00, M12 et M24) sont fortement corrélés les uns aux autres ($RV > 0.75$) et à la MFA ($RV > 0.9$). Contrairement au groupe supplémentaire ”Drug” qui n’est pas corrélé aux groupes actifs et à l’AFM ($RV < 0.05$), ce qui signifie que les informations contenues dans les groupes actifs M00, M12 et M24 sont très similaires et d’autre part, le coefficient RV souligne à nouveau la faible contribution du groupe supplémentaire ”Drug”.

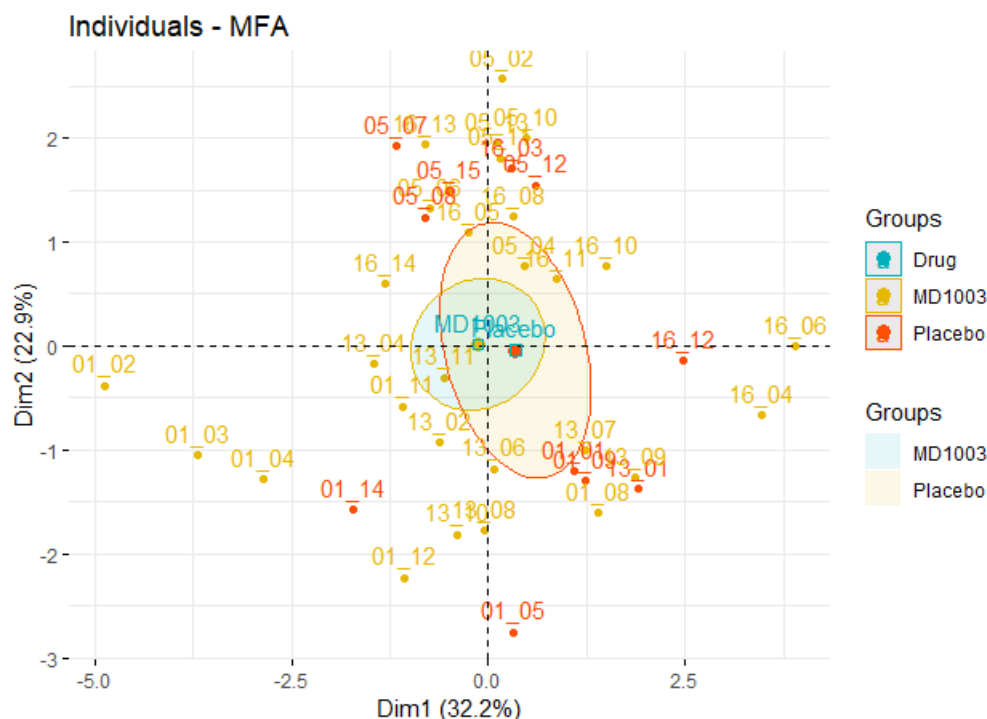


FIGURE 3.15 – Représentation des individus. Les individus (patients) sont présentés par des points sur le nuage de points projeté sur les deux premiers axes principaux de MFA. Chaque individu est coloré suivant la variable qualitative supplémentaire "Drug" (placebo ou MD1003). Les ellipses sont des ellipses de confiance de 95% autour des points moyens du groupe 'Drug'.

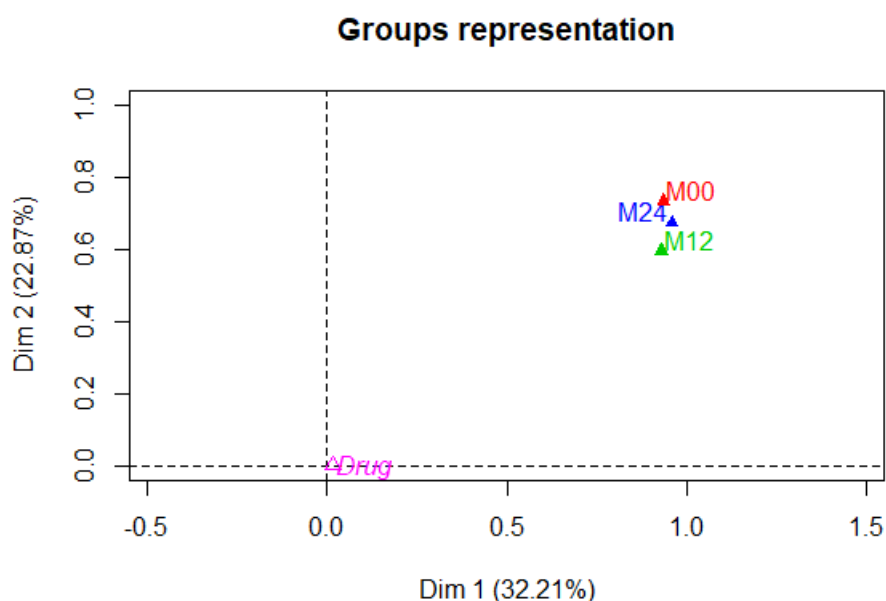


FIGURE 3.16 – Représentation des groupes MFA. M00, M12 et M24 sont les groupes MFA. "Drug" est un groupe supplémentaire.

3.3.3.3 Évaluation du clustering hiérarchique

Le clustering hiérarchique a été réalisé en utilisant le critère de Ward sur les 3 premières composantes principales de l'AFM. Un nombre de 2 clusters a été fixé pour correspondre au nombre de groupes de patients (MD1003 et Placebo). La distance euclidienne est la métrique choisie. Le résultat HCPC est représenté sur

3.3. APPLICATION 2 : ESSAI CLINIQUE DE SCLÉROSE EN PLAQUES

	Drug	M00	M12	M24	MFA
Drug	1.0	0.03	0.04	0.04	0.04
M00	0.03	1.0	0.79	0.84	0.94
M12	0.04	0.79	1.0	0.82	0.92
M24	0.04	0.84	0.82	1.0	0.94
MFA	0.04	0.94	0.92	0.94	1.0

Tableau 3.20 – Coefficients RV entre les groupes (actifs et supplémentaire) et l'AFM

les dendrogrammes de la figure 3.17.

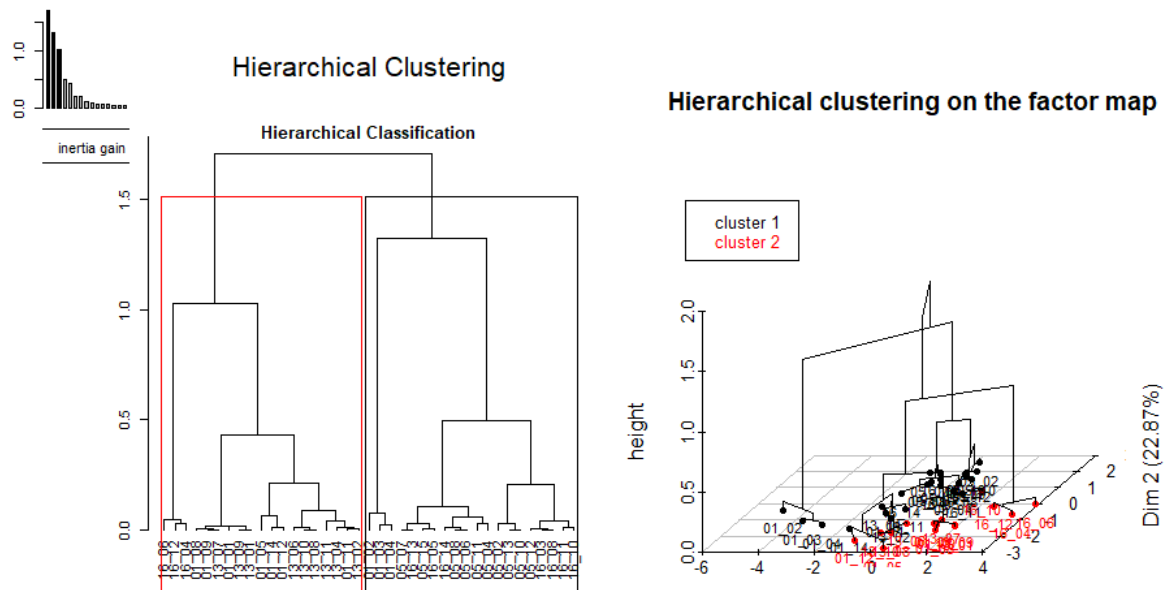


FIGURE 3.17 – Dendrogramme 2D (gauche) et 3D projeté sur les deux premiers axes principaux (droite) de la classification par HCPC ayant pour variables les 3 premières composantes principales de l'AFM. 2 clusters sont représentés (en noir et rouge).

Afin d'évaluer le clustering obtenu à l'aide de HCPC, nous avons utilisé 3 mesures : Pureté, Rand index et l'entropie. Les résultats sont présentés sur la figure 3.18. Les deux groupes, MD1003 et Placebo, ne sont pas séparés correctement (Pureté=62%). Le clustering HCPC a validé une observation faite lors de l'étude de l'AFM : le groupe MD1003 est similaire au groupe Placebo.

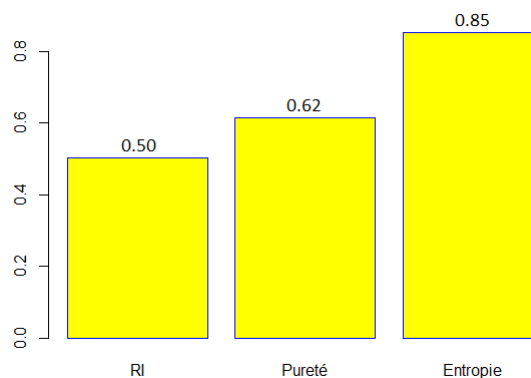


FIGURE 3.18 – Évaluation du clustering hiérarchique (Placebo VS MD1003), basé sur les composantes principales obtenues à partir de l'AFM.

3.3.4 Discussion et limites

Les résultats de l'analyse factorielle multiple basée sur les métriques des histogrammes des biomarqueurs, permettent d'une part d'étudier la relation entre les patients recevant le traitement et ceux recevant le placebo et d'autre part d'évaluer l'évolution dans le temps. En effet, comme présenté dans les résultats, le graphique des individus AFM montre que les ellipses de confiance des groupes MD1003 et placebo se chevauchent, ce qui atteste de la similitude de ces groupes. Cette conclusion rejoint la conclusion établie par les analyses individuelles des biomarqueurs. De plus, le clustering hiérarchique sur les composantes principales n'a pas permis de séparer les patients MD1003 des patients Placebo. Ainsi, nous pouvons conclure que, sur la base des biomarqueurs IRM, à savoir EC, FA, ADC et MTR, le traitement testé ne semble pas avoir d'effet sur la SEP progressive. En ce qui concerne la progression dans le temps, la représentation des groupes AFM (M00, M12 et M24) et les coefficients RV montrent que les groupes actifs sont fortement corrélés les uns aux autres ($RV > 0.75$), ce qui signifie qu'il n'y a pas d'évolution qu'il s'agisse du groupe MD1003 ou du groupe Placebo.

Deux précédentes études d'imagerie (Tourbah et al., 2016a; Dutilleul, 2015), se sont également intéressées à l'apport du traitement MD1003 à la SEP progressive. Ces études montrent que le MD1003 est associé à une diminution du volume cérébral du cerveau entier et au niveau de la substance grise. Les effets du MD1003 sur les mesures volumétriques peuvent être dus à un phénomène de pseudo-atrophie. Dans notre étude, aucune différence significative entre les patients du groupe MD1003 et les patients du groupe Placebo n'a été observée en ce qui concerne l'épaisseur corticale, ce qui pourrait s'expliquer par le fait que les patients inclus dans cet essai clinique présentent une forme progressive de SEP et sont déjà à un stade avancé de la maladie et il aurait donc été surprenant d'observer une atrophie cérébrale après un an où deux ans de suivi. En ce qui concerne l'imagerie de transfert d'aimantation, nous n'avons pas observé de changements significatifs entre les deux groupes dans leurs mesures MTR. Nous nous attendions à une augmentation significative des mesures MTR dans le groupe MD1003 puisque le traitement a pour rôle de stimuler la production d'énergie dans les axones démyélinisés, améliorer la synthèse de la myéline et donc conduire à la remyélinisation. Toutefois, nous avons observé une petite tendance statistique dans la mesure ADC au bout d'un an de traitement. En effet, l'ADC diminue dans le groupe MD1003 alors qu'elle augmente dans le groupe Placebo. Cela pourrait indiquer une possible remyélinisation.

Le fait qu'il n'y ait pas de différences significatives entre les patients du bras placebo et ceux du bras MD1003 et qu'il n'y ait pas d'évolution dans le temps, ne signifie pas nécessairement que le traitement n'a aucun effet sur la SEP progressive en général. En effet, notre étude se limite à quatre biomarqueurs IRM (EC, MTR, ADC et FA), et pour mener une étude plus complète et prendre en compte l'aspect clinique de la maladie, il serait intéressant de combiner des biomarqueurs IRM et des marqueurs cliniques tel que l'EDSS. En outre, bien que l'analyse des biomarqueurs d'histogrammes du cerveau entier soit sensible aux changements subtils et à la détection des anomalies hétérogènes, nous pourrions affiner encore plus l'étude en nous concentrant non seulement sur les histogrammes du cerveau entier, mais également sur les histogrammes de la substance grise et de la substance blanche d'apparence normale. En effet,

Iannucci et al. (1999, 2000) suggèrent que l'analyse d'histogrammes du MTR au niveau de la substance blanche d'apparence normale détecte une anomalie alors que l'analyse d'histogrammes MTR du cerveau entier n'en détecte aucune.

Le but de cette étude était de présenter une méthode permettant de combiner des biomarqueurs IRM basés sur l'analyse histogramme de biomarqueurs au lieu d'une mesure unique de la tendance centrale (e.g. moyenne). Mais, nous sommes bien conscient que ce traitement ne semble pas avoir la capacité de séparer le groupe traité du groupe placebo en fonction des quatre biomarqueurs IRM (EC, ADC, FA, MTR). Il serait intéressant, dans une future étude, de travailler sur un ensemble de données comprenant des groupes clairement distincts (e.g. sujets sains VS sujets SEP) pour quantifier la supériorité de notre méthode et pour ce faire, nous comparerions les performances de classification dans les deux cas : avant de combiner les biomarqueurs et après avoir appliqué notre approche de biomarqueurs combinés décrite plus haut.

Chapitre 4

Distance entre histogrammes

Sommaire

4.1 Description de l'approche	103
4.1.1 Contexte	103
4.1.2 Notion de distance	104
4.2 Application : classification de la population de Sclérose en plaques	107
4.2.1 Matériels et méthodes	108
4.2.2 Résultats	111
4.2.3 Discussion	112

4.1 Description de l'approche

4.1.1 Contexte

Dans le chapitre précédent, nous avons proposé de travailler avec les histogrammes de paramètres IRM. Pour ce faire, nous avons utilisé une approche qui consiste à extraire des résumés d'histogrammes tels que moyenne et quantiles. Toutefois, il s'agit tout de même d'un résumé, ce qui signifie que des parties de l'information contenues dans les paramètres IRM sont effacées. Dans ce chapitre, nous proposons une méthode qui remédie à cela. En observant les histogrammes des différents biomarqueurs, nous avons émis l'hypothèse que, puisque ces distributions n'étaient pas gaussiennes, l'analyse de la moyenne, médiane et écart type n'était pas suffisamment sensibles aux changements subtils dus aux variations structurelles des tissus. Ainsi, les paramètres de notre étude ne sont plus des résumés d'histogrammes mais les histogrammes eux-mêmes. Il s'agira de préserver ces histogrammes dans leur globalité et de mesurer des distances¹ entre eux. Étant donné que différentes mesures de (dis)similarité sont plus ou moins sensibles à la forme, à l'étendue, à l'offset, etc, de l'histogramme, nous avons donc étudié différentes familles de distances et recherché des distances capables de refléter les différences entre les sujets de notre population.

1. Ici, le terme "distance" est utilisé en général. Il existe cependant d'autres notions de distance statistique, plus spécialisées, pour lesquelles les termes *divergence* ou *dissimilarité* sont utilisés.

4.1.2 Notion de distance

Après avoir extrait les histogrammes des images paramétriques pour chaque sujet, des mesures de dissimilarité sont effectuées entre les histogrammes.

Une distance est définie comme la longueur qui sépare deux objets. Plus formellement, en mathématiques, on appelle distance sur un ensemble E toute application d définie sur le produit $E^2 = E \times E$ et à valeurs dans l'ensemble \mathbb{R}^+ des réels positifs,

$$d : E \times E \rightarrow \mathbb{R}^+$$

vérifiant les axiomes suivants :

$$\begin{aligned} &\mathbf{1- Symétrie : } \forall (a, b) \in E^2, d(a, b) = d(b, a) \\ &\mathbf{2- Séparation : } \forall (a, b) \in E^2, d(a, b) = 0 \Leftrightarrow a = b \\ &\mathbf{3- Inégalité triangulaire : } \forall (a, b, c) \in E^3, d(a, c) \leq d(a, b) + d(b, c) \end{aligned} \quad (4.1)$$

Les mesures satisfaisant ces propriétés sont appelées distances et si elles ne respectent pas tous les axiomes, elles sont parfois appelées *divergences*.

Il existe deux approches pour mesurer la dissimilarité entre distributions : une approche vectorielle et une approche probabiliste. Étant donné que chaque bin d'histogramme est supposé être indépendant des autres bins, un histogramme ou une distribution peuvent être considérés comme des vecteurs, i.e. un point de l'espace euclidien ou un système de coordonnées cartésien. Par conséquent, de nombreuses distances géométriques peuvent être appliquées pour comparer ces distributions. Il existe de nombreux ouvrages sur les versions discrètes de diverses divergences dans la théorie des probabilités et de l'information (Toussaint, 1974; Kailath, 1967). Le calcul de la distance entre deux distributions peut être considéré comme identique au calcul de la probabilité de Bayes (Duda et al., 2000). Cela équivaut à mesurer le chevauchement entre deux distributions comme étant leur distance. L'approche probabiliste est basée sur le fait qu'un histogramme d'une mesure fournit la base d'une estimation empirique de la distribution.

Il existe plus de 40 mesures de dissimilarité qui peuvent être regroupées dans 7 familles aux propriétés communes (Cha, 2007). Dans cette section nous allons présenter ces familles et quelques-unes des mesures de dissimilarité (les formules de toutes les mesures sont présentées dans l'annexe III).

1. Famille "Minkowski L_p "

La famille Minkowski regroupe les mesures de distance les plus simples : *Euclidienne L_2* (Duda et al., 2000) qui est la formule la plus familière de distance entre deux points, généralisée au cas des histogrammes ; *Manhattan L_1* (Duda et al., 2000) qui additionne simplement les différences absolues entre les bins de l'histogramme ; *Chebyshev L_{\inf}* (Webb and Copsey, 2011) qui est simplement la valeur maximale de la différence entre les histogrammes pour tous les bins. Ces mesures sont les plus simples, mais elles ne permettent pas de distinguer des histogrammes de formes similaires qui sont simplement décalés les uns des autres. Les formules associées à ces métriques sont les suivantes :

— Euclidienne L_2

$$d_{euc} = \sqrt{\sum_{i=1}^d |P_i - Q_i|^2}$$

— Manhattan L_1

$$d_{man} = \sum_{i=1}^d |P_i - Q_i|$$

— Chebyshev L_{inf}

$$d_{cheb} = \max_i |P_i - Q_i|$$

avec d le nombre de bins de l'histogramme, P et Q deux distributions où $P_i = H_i(X)/n$, avec X un ensemble de n éléments dont les valeurs sont discrètes et finies. Un histogramme $H(X)$ d'un ensemble X représente la fréquence de chaque valeur. La valeur de fréquence du i -ème bin est notée $H_i(X)$.

Les mesures *Canberra* et *Lorentzian* sont des variations de L_1 . La mesure de *Canberra* ([Webb and Copsey, 2011](#)) normalise les différences absolues avec la somme des valeurs des deux bins et est connu pour être sensible aux petits changements proches de zéro, et la mesure de *Lorentzian* ([Deza and Deza, 2014](#)) est essentiellement le log de L_1 , bien que l'unité soit ajoutée pour assurer la non-négativité et éviter le log de zéro.

— Canberra

$$d_{can} = \sum_{i=1}^d \frac{P_i - Q_i}{P_i + Q_i}$$

— Lorentzian

$$d_{lor} = \sum_{i=1}^d \ln(1 + |P_i - Q_i|)$$

2. Famille "Intersection"

La mesure de *non-intersection* ([Duda et al., 2000](#)) est basée sur la valeur minimale entre chaque bin des histogrammes. Puisque nous nous intéressons ici aux différences entre les histogrammes et que nous parlons d'histogrammes normalisés, cette mesure est égale à zéro si les histogrammes sont identiques.

— Non-intersection

$$d_{NI} = 1 - \sum_{i=1}^d \min(P_i, Q_i) = \frac{1}{2} \sum_{i=1}^d |P_i - Q_i|$$

3. Famille "Inner product"

Les mesures de cette famille traitent les deux histogrammes comme des vecteurs et calculent le produit interne normalisé par un facteur $(P \cdot Q)$. La mesure *Cosine* ([Webb and Copsey, 2011](#)) est le produit interne normalisé par la racine carrée de la somme des carrés de chaque élément de l'histogramme. Cette famille contient également les célèbres mesures de *Jaccard* ([Jaccard, 1901](#)) et *Dice* ([Dice, 1945](#)), qui contiennent également le produit interne mais ont un facteur de normalisation différent.

— Cosine

$$d_{cos} = \frac{\sum_{i=1}^d P_i Q_i}{\sqrt{\sum_{i=1}^d P_i^2} \sqrt{\sum_{i=1}^d Q_i^2}}$$

— Jaccard

$$d_{jac} = \frac{\sum_{i=1}^d (P_i - Q_i)^2}{\sum_{i=1}^d P_i^2 + \sum_{i=1}^d Q_i^2 - \sum_{i=1}^d P_i Q_i}$$

— Dice

$$d_{dic} = \frac{\sum_{i=1}^d (P_i - Q_i)^2}{\sum_{i=1}^d P_i^2 + \sum_{i=1}^d Q_i^2}$$

4. Famille "Fidelity"

Cette famille contient des mesures qui utilisent la somme des moyennes géométriques des histogrammes modifiées en utilisant la racine carrée. Les mesures *Hellinger* et *Squared-Chord* (Deza and Deza, 2014), sont les versions les plus communément utilisées. La distance de *Bhattacharyya* (Bhattacharyya, 1946) s'est avérée être liée à la probabilité bayésienne d'erreur de classification minimale, sa forme étant liée à la distance de *Matusita* (Matusita, 1951).

— Hellinger

$$d_H = 2 \sqrt{1 - \sum_{i=1}^d \sqrt{P_i Q_i}}$$

— Squared-Chord

$$d_{sqc} = \sum_{i=1}^d (\sqrt{P_i} - \sqrt{Q_i})^2$$

— Bhattacharyya

$$d_B = -\ln \sum_{i=1}^d \sqrt{P_i Q_i}$$

5. Famille "L2 quadratique χ^2 "

La mesure *quadratique* χ^2 est utilisée (Deza and Deza, 2014) car son facteur de normalisation est symétrique dans les deux histogrammes (par opposition à la mesure de *Pearson* et de *Neyman* χ^2 qui normalisent les différences au carré du numérateur par la valeur bin de l'un ou l'autre des histogrammes). Cette mesure est essentiellement la distance euclidienne normalisée entre deux vecteurs.

— Euclidienne quadratique

$$d_{sqe} = \sum_{i=1}^d (P_i - Q_i)^2$$

— Pearson χ^2

$$d_P = \sum_{i=1}^d \frac{(P_i - Q_i)^2}{Q_i}$$

— Neyman χ^2

$$d_P = \sum_{i=1}^d \frac{(P_i - Q_i)^2}{P_i}$$

6. Famille "Shannon entropy"

Ces mesures prennent la forme de l'entropie de Shannon avec divers choix de facteurs de normalisation. Appliqué à deux histogrammes, elles mesurent l'entropie croisée minimale de deux distributions de probabilité. Les mesures de cette famille ne sont pas de vraies distances puisqu'elles ne sont pas symétriques. La *divergence de Kullback-leibler* (Kullback and Leibler, 1951) permet de calculer exactement combien d'information est perdue lorsque nous "approximons" une distribution avec une autre. Il y a aussi les mesures de *Jensen-Shannon* et *Jeffreys* (Jeffreys Harold, 1946) qui est la forme symétrique de la divergence KL.

— Divergence Kullback-Leibler

$$d_{KL} = \sum_{i=1}^d P_i \ln \frac{P_i}{Q_i}$$

— Jeffreys

$$d_j = \sum_{i=1}^d (P_i - Q_i) \ln \frac{P_i}{Q_i}$$

— Jensen-Shannon

$$d_{js} = \frac{1}{2} \left[\sum_{i=1}^d \left(P_i \ln \frac{2P_i}{P_i + Q_i} \right) + \sum_{i=1}^d \left(Q_i \ln \frac{2Q_i}{P_i + Q_i} \right) \right]$$

7. Famille "Combinaison"

Cette famille présente des mesures qui sont des combinaisons d'autres mesures. *Kumar-Johnson* combine la mesure symétrique χ^2 et la divergence moyenne arithmétique et géométrique. La mesure *Average* combine les mesures de Manhattan et de Chebyshev.

— Kumar-Johnson

$$d_{KJ} = \sum_{i=1}^d \left(\frac{(P_i^2 - Q_i^2)^2}{2(P_i Q_i)^{3/2}} \right)$$

— Average(L_1, L_∞)

$$d_{avg} = \frac{\sum_{i=1}^d |P_i - Q_i| + \max_i |P_i - Q_i|}{2}$$

4.2 Application : classification de la population de Sclérose en plaques

La sclérose en plaques (SEP) est une maladie neurodégénérative démyélinisante. En raison de l'aspect diffus de la maladie, plusieurs études ont

choisi l'analyse d'histogrammes afin de quantifier les changements pathologiques diffus de la maladie. Un inconvénient commun de ces études est que l'ensemble des informations inclus dans l'histogramme n'est pas utilisé ; seules des mesures arbitraires sont choisies pour décrire l'histogramme ; incluant la moyenne, la médiane, les quantiles... Dans notre étude, nous proposons une deuxième alternative à l'utilisation des histogrammes qui consiste à inclure toute les informations de l'histogramme, et pas seulement quelques descripteurs locaux, dans un classifieur utilisant les k plus proches voisins (kNN), dans le but d'améliorer les performances de classification de la population SEP. La figure 4.1 résume notre approche.

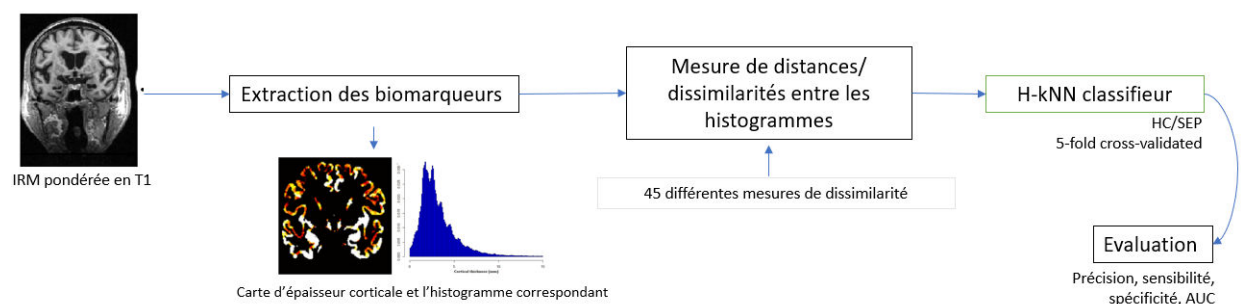


FIGURE 4.1 – Schéma résumé de l'approche basée sur les distances entre histogrammes.

4.2.1 Matériels et méthodes

4.2.1.1 Population de l'étude

Notre population est constituée de 111 sujets. Une partie de cette population provient de la base ADNI (cf. section 3.2.1.1.1) et l'autre de l'essai clinique MS-SPI (cf. section 3.3.1.1.1). Nous avons extrait les 71 sujets contrôles d'ADNI (HC) et les 40 patients atteints de SEP progressive de MS-SPI. Les informations démographiques sont renseignées sur le tableau 4.1. MS-SPI étant une étude longitudinale, seules les volumes IRM de l'examen initial ont été utilisées (i.e. avant de recevoir le traitement). Les groupes sont appariés en sexe mais pas en âge. En effet, si on sait que la maladie d'Alzheimer touche des sujets âgés, la sclérose en plaques peut toucher des sujets très jeunes. De plus, au vue du nombre restreint de sujets de notre étude, nous ne pouvions pas nous permettre d'en supprimer plus.

	HC	SEP	valeur p
Effectif	71	40	-
Âge [intervalle]	76.1 ± 5.6 [63 – 87]	51.3 ± 8.5 [35 – 69]	<0.01
Sexe (F/H)	38 / 33	20 / 18	0.9

Tableau 4.1 – Information démographique de la population de notre étude
 ± représentent moyennes ± écart-type. Les valeurs p sont basées sur le test Anova pour Âge et le test khi-deux χ^2 pour Sexe (Abrv : HC, sujets sains ; SEP, sclérose en plaques)

4.2.1.2 Acquisitions IRM et analyses d'images

Les paramètres des acquisitions des volumes IRM pondérées en T1, sont identiques à ceux décrits dans les parties 3.2.1.2 pour les sujets HC et 3.3.1.2 pour les patients SEP. Les volumes 3D IRM pondérés en T1 ont été principalement acquis avec la séquence MPRAGE. Toutefois, MS-SPI étant un essai multicentrique, d'autres paramètres d'acquisition sont spécifiques aux centres concernés. La chaîne de prétraitement a été présentée dans la section 3.2.1.3.

4.2.1.3 Histogramme d'épaisseur corticale

L'atrophie de la substance grise, évaluée par IRM, est un marqueur crucial de la neurodégénérescence et a donc été utilisé dans plusieurs études sur la SEP (Filippi et al., 2014; Steenwijk et al., 2016; Eshaghi et al., 2017). L'épaisseur corticale (EC) a été mesurée à partir d'une IRM pondérée en T1 à l'aide de la Toolbox Matlab CorThiZon, qui comprend les étapes de normalisation, de segmentation et de mesure d'épaisseur corticale. Les EC ont été calculées sur tout le ruban cortical à l'aide d'un algorithme basé sur l'équation de Laplace (Querbes et al., 2009). Le détail de la méthode d'estimation de l'EC a été présentée dans la partie 3.2.1.4.1.

En mesurant l'épaisseur corticale sur le ruban cortical entier, une carte 3D d'épaisseur corticale a été obtenue, qui a ensuite été recalée dans l'espace MNI, à partir de laquelle un histogramme d'EC a été extrait suivant la procédure décrite dans la partie 3.1.1. Pour rappel : nous filtrons la carte EC à l'aide d'un filtre médian, nous choisissons de façon empirique le nombre de bins et nous normalisons les histogrammes.

4.2.1.4 Classification KNN basée sur les histogrammes

Cette section présentera la méthode de classification des k plus proches voisins basée sur les histogrammes qu'on notera " H - kNN " dans la suite du manuscrit.

4.2.1.4.1 Principe général des kNN

L'algorithme des k plus proches voisins est un algorithme d'apprentissage non paramétrique est dit "paresseux". "Paresseux" car à l'inverse de beaucoup d'autres méthodes d'apprentissage automatique (tels que la régression logistique, les réseaux de neurones, etc.), aucun apprentissage n'est réellement réalisé, i.e. il n'y a pas de phase de détermination de paramètres d'une fonction par le biais d'une optimisation mathématique. L'algorithme classe les objets en fonction des données d'apprentissage les plus proches dans l'espace des variables. Dans le processus de classification, l'objet non étiqueté à classer est simplement affecté à l'étiquette de ses k voisins les plus proches, avec k le nombre de voisins. La figure 4.2 illustre l'algorithme.

4.2.1.4.2 L'algorithme des k plus proches voisins basé sur les histogrammes

C'est la méthode de classification la plus utilisée en raison de sa facilité et de son efficacité pratique : elle n'a pas besoin d'appliquer de modèle et il a été prouvé que ses performances étaient supérieures pour la classification de plusieurs types de données. Cependant, les performances supérieures de classification des

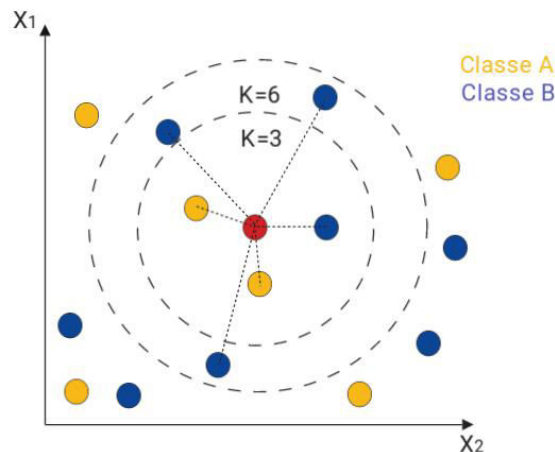


FIGURE 4.2 – Illustration de l’algorithme classique des k plus proches voisins kNN. Le but est de classer l’objet non-classé (rouge) dans la classe A (jaune) ou B (bleue). L’algorithme kNN attribue à un point de la base test l’étiquette la plus fréquente de ses k plus proches voisins dans l’ensemble d’apprentissage. Dans cet exemple, pour $k=3$, l’objet est classé dans la classe A (2 jaunes vs 1 bleu). Pour $k=6$, il est assigné à la classe B (2 jaunes vs 4 bleus).

kNN dépendent grandement de la métrique utilisée pour calculer les distances entre les points de données. La distance la plus communément utilisée dans le cas classique est la distance euclidienne, mais comme on l’a vu dans la section 4.1.2, il existe d’autres métriques. Ces mesures peuvent être effectuées entre différents types d’objets tels que des vecteurs ou encore des distributions de probabilités. Ainsi, l’algorithme des k plus proches voisins basé sur les histogrammes est une extension de l’algorithme classique (cf. illustration 4.3). La principale différence entre kNN basé sur les histogrammes (H-kNN) et l’algorithme classique kNN est que les distances/dissimilarités sont mesurées entre les histogrammes d’épaisseurs corticales et non entre des points uniques telles que la moyenne de l’EC.

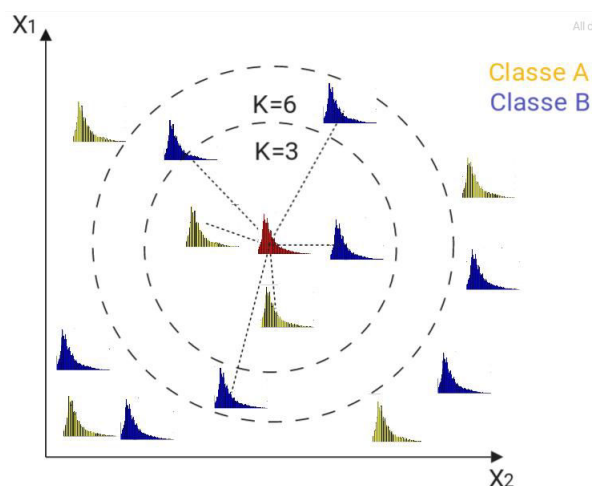


FIGURE 4.3 – Illustration de l’algorithme des k plus proches voisins basé sur les histogrammes H-kNN. L’algorithme H-kNN est une extension de l’algorithme kNN appliqué aux histogrammes. Dans notre application, ces histogrammes représentent les histogrammes d’épaisseur corticale de chaque sujet.

4.2.1.4.3 Classification et validation

L’algorithme H-kNN a été utilisé pour classer les patients atteints de SEP et les sujets contrôles HC. L’implémentation a été faite à la main avec le logiciel R

version 3.5.1 (2018).

— Choix du nombre de voisins

En ce qui concerne le nombre k de voisins, nous avons utilisé une grille de recherche $k=(1,..,21)$ et s'agissant d'une classification binaire, nous avons pris soin de choisir des valeurs impaires de k . La grille de recherche consiste à entraîner notre modèle plusieurs fois sur une gamme de paramètres que nous spécifions. De cette manière, nous pouvons tester notre modèle avec chaque valeur du paramètre k et déterminer les valeurs optimales pour obtenir les meilleurs résultats.

— Choix des mesures de dissimilarité

Il existe de nombreuses métriques qui permettent de mesurer la distance entre deux histogrammes. Il y a deux catégories de mesures de dissimilarité : celles qui considèrent les histogrammes en tant que vecteurs et celles qui les considèrent en tant que distributions de probabilités. Ces mesures ont été réparties par [Cha \(2007\)](#) dans 7 familles présentées dans la section 4.1.2. Les formules des 45 mesures sont présentées dans l'annexe III.

— Évaluation de la classification

Les deux groupes HC et SEP ont été classés en utilisant une cross-validation à 5-fold (5CV) stratifiée. Il s'agit de la même approche que 10-fold cross-validation (section 3.2.1.6.2) à la différence que le jeu de données a été divisé en 5 folds au lieu de 10. Quatre mesures ont permis d'évaluer la classification : précision, sensibilité, spécificité et AUC, toutes définies dans la partie 3.2.1.6.2.

4.2.2 Résultats

La sélection de la mesure de distance appropriée est l'un des défis rencontrés par les chercheurs lorsqu'ils utilisent un algorithme basé sur des mesures de distances. La grande variété des mesures de dissimilarité peut être source de confusion pour choisir une mesure appropriée. Dans nos travaux, 45 mesures de dissimilarité ont été utilisées lors de la classification H-kNN. Nous avons fait varier le nombre de voisins k et avons évalué la classification pour chacune de ces mesures. Les résultats sont présentés dans le tableau 4.2. La famille de mesures habituellement utilisée est la famille Minkowski car elle réunit les mesures les plus simples. Nous avons pris soin d'utiliser d'autres familles qui contrairement à la famille Minkowski sont sensibles aux petits décalages entre les histogrammes. Par exemple, si deux histogrammes sont identiques mais sont légèrement décalés, les mesures de Minkowski ont tendance à surestimer la distance entre les histogrammes.

Dans l'ensemble, les résultats indiquent que "average" de la famille *combinaisons* et "divergence de Kullback-Leiber" de la famille de *Shannon entropy* sont les mesures les plus précises pour la classification HC/SEP avec une précision de 83% (90% de sensibilité et 70% de spécificité) pour 3 plus proches voisins. En moyenne, il s'agit de la famille de l'entropie de Shannon qui permet de mieux classer les patients SEP et les sujets contrôles. Par contre, la famille "inner product" ne semble pas adaptée à notre jeu de données.

4.2. APPLICATION : CLASSIFICATION DE LA POPULATION DE SCLÉROSE EN PLAQUES

Dissimilarités	Familles	K	Précision	Sensibilité	Spécificité	AUC
Euclidienne L_2	Minkowski_Lp	5	0.80	0.87	0.68	0.77
Manhattan L_1	Minkowski_Lp	3	0.80	0.87	0.68	0.79
Chebyshev L_∞	Minkowski_Lp	3	0.80	0.87	0.68	0.77
Sorensen	Minkowski_Lp	3	0.81	0.87	0.70	0.79
Soergel	Minkowski_Lp	3	0.81	0.87	0.70	0.79
Lorentzian	Minkowski_Lp	3	0.81	0.87	0.70	0.79
Kulczynski	Minkowski_Lp	3	0.81	0.87	0.70	0.79
Gower	Minkowski_Lp	3	0.81	0.87	0.70	0.79
Canberra	Minkowski_Lp	5	0.69	0.83	0.45	0.64
Intersection	Intersection	19	0.43	0.62	0.10	0.36
Non-intersection	Intersection	3	0.81	0.87	0.70	0.79
Tanimoto	Intersection	3	0.81	0.87	0.70	0.79
Motyka	Intersection	3	0.81	0.87	0.70	0.79
Czekanowski	Intersection	3	0.81	0.87	0.70	0.79
Wavehedges	Intersection	5	0.70	0.86	0.42	0.64
Kulczynski_s	Intersection	19	0.43	0.62	0.10	0.36
Jaccard	Inner Product	5	0.78	0.86	0.65	0.78
Dice	Inner Product	5	0.82	0.92	0.65	0.78
Inner product	Inner Product	7	0.63	0.99	0.00	0.49
Cosine	Inner Product	13	0.44	0.62	0.12	0.37
Hassebrook	Inner Product	15	0.44	0.63	0.10	0.37
Ruzicka	Inner Product	19	0.43	0.62	0.10	0.36
Moyenne harmonique	Inner Product	21	0.42	0.59	0.12	0.36
Squared chord	Fidelity	3	0.80	0.87	0.68	0.77
Matusita	Fidelity	3	0.80	0.87	0.68	0.77
Hellinger	Fidelity	3	0.80	0.87	0.68	0.77
Bhattacharyya	Fidelity	3	0.80	0.87	0.68	0.77
Fidelity	Fidelity	21	0.41	0.58	0.12	0.35
Euclidienne quadratique	L2 quadratique	5	0.81	0.90	0.65	0.78
Chi quadratique	L2 quadratique	3	0.80	0.87	0.68	0.77
Probabilistic symmetric	L2 quadratique	3	0.80	0.87	0.68	0.77
Pearson	L2 quadratique	3	0.80	0.87	0.68	0.77
Neyman	L2 quadratique	3	0.80	0.87	0.68	0.77
Additive symmetric	L2 quadratique	3	0.80	0.87	0.68	0.77
Divergence	L2 quadratique	3	0.72	0.83	0.52	0.68
Clark	L2 quadratique	3	0.72	0.83	0.52	0.68
Kullback-Leibler	Shannon entropy	3	0.83	0.90	0.70	0.80
Jensen-shannon	Shannon entropy	3	0.81	0.87	0.70	0.79
Topsoe	Shannon entropy	3	0.81	0.87	0.70	0.79
Jensen difference	Shannon entropy	3	0.81	0.87	0.70	0.79
Jeffreys	Shannon entropy	3	0.81	0.87	0.70	0.79
K-Divergence	Shannon entropy	3	0.81	0.87	0.70	0.79
Average	Combinaisons	3	0.83	0.90	0.70	0.80
Taneja	Combinaisons	3	0.80	0.87	0.68	0.77
Kumar-Johnson	Combinaisons	3	0.80	0.87	0.68	0.77

Tableau 4.2 – Résultat de la classification des kNN basé sur les histogrammes. K représente le nombre optimal de voisins trouvé à l'aide d'une grille de recherche pour chaque mesure.

4.2.3 Discussion

Afin de classer les sujets contrôles et les patients atteints de sclérose en plaques, nous avons étendu l'algorithme des k plus proches voisins aux histogrammes en mesurant les distances entre les histogrammes. Nous avons obtenu un maximum de 83% de précision à l'aide de la divergence de

Kullback-Leibler. La distance euclidienne a atteint quand à elle une précision de 80%. Cela s'explique par le fait que les mesures de la famille Minkowski sont basées sur une distance géométrique entre deux objets. Dans notre cas, les histogrammes d'EC ne peuvent pas être distingués en utilisant des caractéristiques géométriques où les distributions liées à ces objets ne sont pas considérées. D'autres applications telles que l'analyse des données incertaines, l'analyse de données symboliques et la récupération de l'information ont montré la supériorité de la divergence de Kullback-Leibler pour classer des objets². Pour surmonter la limite des mesures géométriques et capturer les différences de distribution entre les objets, Priyadharshini et al. (2014) proposent d'utiliser la divergence de Kullback-Leibler comme mesure de dissimilarité et de l'intégrer à l'algorithme k-médoides pour regrouper des données incertaines. De même, dans le domaine de la récupération de l'information, pour la catégorisation automatique de textes, Bigi (2003) a proposé Kullback-Leibler comme mesure de divergence entre les distributions et a conclu que la DKL permet d'obtenir des améliorations substantielles par rapport à la méthode classique.

Dans la littérature, il existe un nombre très limité d'études sur la modélisation mathématique et la classification de patients atteints de sclérose en plaques. Notre approche d'analyse des histogrammes EC permet d'améliorer la précision de la classification HC/SEP par rapport, par exemple à des analyses types régions d'intérêts. Richiardi et al. (2012) ont classé avec succès les patients atteints de sclérose en plaques et les sujets contrôles avec une sensibilité et une spécificité de 80%, identifiant la zone droite du temporal moyen comme la région la plus discriminante. Toujours à l'aide d'une approche par ROI, Wottschel (2017) a proposé de classer les patients SEP et les patients avec un syndrome cliniquement isolé ayant converti au bout d'un an et a obtenu à l'aide d'un classifieur SVM une précision de 74.1%. Karaca et al. (2017) ont proposé de classer la population SEP en utilisant une combinaison de marqueurs IRM (i.e. les lésions, leurs tailles, leurs rayons...) et de marqueurs cliniques (i.e. l'EDSS), afin d'obtenir, à l'aide de l'algorithme classique kNN, une précision de 73.7%.

Dans ce chapitre, nous avons mis en avant une approche alternative à l'analyse des histogrammes qui permet de caractériser les modifications diffuses subtiles dans la SEP en utilisant des histogrammes dérivés des cartes d'épaisseur corticales. L'histogramme, particulièrement dans le cadre de l'étude de maladies dites diffuses, est un outil objectif et sensible permettant d'obtenir de nouvelles informations sur les mécanismes sous-jacents de la pathologie étudiée (Tofts et al., 2004a). Diverses résumés ont déjà été utilisés pour décrire l'histogramme. Ces derniers sont essentiellement des descripteurs locaux de l'histogramme (i.e. une partie seulement de l'histogramme). Étant donné que ces résumés sont locaux et choisis de manière arbitraire, il est peu probable qu'ils soient optimaux.

Cette étude présente majoritairement deux limites, un nombre relativement bas de sujets et une application clinique assez simple et peu étudiée (i.e. classification de patients SEP et sujets contrôles à l'aide d'un unique biomarqueur IRM structurel). Toutefois, cette méthode alternative d'analyse des histogrammes présente deux avantages majeurs par rapport à la méthode existante. Premièrement, elle prend en compte l'histogramme entier et pas seulement quelques marqueurs arbitraires. Deuxièmement, au lieu d'effectuer un test t pour

2. Plus particulièrement de la classification non-supervisée.

comparer des groupes de patients, les patients sont classés en fonction de leurs histogrammes individuels. En utilisant cette approche, les patients atteints de SEP peuvent être classés de façon plus efficace.

Chapitre 5

Distribution de probabilité et géométrie de l'information

Sommaire

5.1	État de l'art sur la géométrie de l'information	116
5.1.1	Contexte : Comparer deux distributions	116
5.1.2	Géométrie différentielle	121
5.2	Application 1 : classification de la population de la maladie d'Alzheimer	127
5.2.1	Conception mathématique	127
5.2.2	Application aux données de la maladie d'Alzheimer	134
5.3	Application 2 : caractérisation des retards dans le transport aérien	141
5.3.1	Contexte	141
5.3.2	Conception mathématique	142
5.3.3	Application : clustering des retards aéroportuaires	148

La géométrie de l'information est un domaine récent des mathématiques qui utilise la géométrie différentielle pour expliquer les propriétés des estimateurs statistiques (Rao, 1945; Chentsov, 1982; Amari and Nagaoka, 2000). Ces dernières années, de nombreuses applications pratiques ont exploré cette approche, telles que l'imagerie tenseur de diffusion (Fletcher and Joshi, 2004), le traitement d'images et de vidéos (Tuzel et al., 2008), l'interface cerveau-ordinateur (Barachant and Congedo, 2014) et le traitement d'images radar (Barbaresco, 2008). L'information de Fisher (Amari and Nagaoka, 2000), qui est associée de manière naturelle à des familles de densités de probabilités, permet de munir ces dernières d'une structure de variété riemannienne dont les propriétés intrinsèques, telle la courbure, donne des informations importantes sur le comportement des estimateurs. Il est également possible d'exploiter la distance géodésique entre densités de probabilité pour en déduire des approximations optimaux, soit dans le cas où le problème possède des variables cachées, soit lorsque l'on souhaite réduire le nombre de paramètres.

Toutes ces notions seront expliquées dans une première partie *état de l'art*. Suivie de deux applications, une dans le domaine médical et l'autre dans le domaine aéronautique.

5.1 État de l'art sur la géométrie de l'information

Cette première partie présentera l'état de l'art de la géométrie de l'information et les outils indispensables à la compréhension des deux applications de ce chapitre.

5.1.1 Contexte : Comparer deux distributions

Les principaux concepts permettant de mesurer une similarité entre deux distributions de probabilité vont être présentés ici de façon essentiellement intuitive. Les outils mathématiques, en particulier ceux liés à la géométrie différentielle seront présentés plus bas.

5.1.1.1 La divergence Kullback-Leibler

La divergence de Kullback-Leibler ([Kullback and Leibler, 1951](#)) ou entropie relative est une mesure qui a été développée dans le contexte de la théorie de l'information pour mesurer la dissimilarité entre deux distributions de probabilité. Elle est largement utilisée pour l'optimisation dans le domaine de l'apprentissage automatique lorsque des distributions sont impliquées.

La divergence de Kullback-Leibler, notée $D_{KL}(p||q)$, de deux distributions de probabilité $p(\cdot)$ et $q(\cdot)$, est définie (dans le cas continu) :

$$D_{KL}(p||q) = \int_{-\infty}^{+\infty} p(x) \log \frac{p(x)}{q(x)} dx \quad (5.1)$$

Propriétés de la divergence de Kullback-Leibler :

$$D_{KL}(p||q) \geq 0$$

$$p = q \text{ presque partout ssi } D_{KL}(p||q) = 0$$

Plus la valeur D_{KL} est petite et plus les distributions sont similaires

La divergence de Kullback-Leibler a de nombreuses applications. [Clim et al. \(2018\)](#) présentent l'utilisation de la divergence de Kullback-Leibler dans des applications de santé, à l'aide d'algorithmes d'apprentissage automatique.

- Avantage(s) : Le principal avantage de cette mesure et qu'elle est très simple et rapide à utiliser.
- Inconvénient(s) :
 - Il ne s'agit pas d'une distance. En effet, elle n'est pas symétrique et ne satisfait pas l'inégalité triangulaire. De plus, le fait que la divergence KL n'ait pas de limite supérieure bien définie est un inconvénient majeur.
 - Nous ne pouvons pas l'utiliser pour comparer p et q quand l'une est discrète et que l'autre est continue.

5.1.1.2 Métrique de Wasserstein

La métrique de Wasserstein ([Wasserstein, 1969](#)) découle de la théorie dite du transport optimal où une distribution de probabilité est "transportée" vers une autre distribution. Le coût minimum de transport définit une distance entre les

deux distributions, appelée distance de Wasserstein. La distance de Wasserstein se définit de façon générale entre deux lois de probabilité P et Q par :

$$\mathcal{W}(P, Q) = \inf\{[E(X_1 - X_2)^2]^{1/2} : \mathcal{L}(X_1) = P, \mathcal{L}(X_2) = Q\},$$

où $\mathcal{L}(X_1)$ signifie : la loi de X_1 , et E : l'espérance mathématique.

Dans le cas de lois unidimensionnelles, on peut l'exprimer de façon simple à l'aide des fonctions de répartition F, G de respectivement P et Q :

$$\mathcal{W}(P, Q) = \int_0^1 |F^{-1}(t) - G^{-1}(t)|^p dt, p \geq 1 \quad (5.2)$$

La distance de Wasserstein a été largement étudiée et appliquée à l'analyse de forme, de par sa capacité de comparaison intrinsèque des similitudes entre les formes (Schmitzer and Schnörr, 2013; Su et al., 2015). Mais, elle est aussi de plus en plus utilisée dans le domaine des statistiques et de l'apprentissage automatique (cf. la revue de Kolouri et al. (2017)).

— Avantages :

- Il s'agit d'une véritable distance.
- Cette distance prend en compte la géométrie sous-jacente de l'espace.
- Lorsque nous mesurons la moyenne de plusieurs objets, tels que des distributions ou des images, nous voudrions nous assurer de récupérer un objet similaire. La figure 5.1 montre quatre distributions gaussiennes avec différentes moyennes et variances. La figure 5.1-(f) montre le barycentre de Wasserstein et le résultat est bien une gaussienne contrairement à la densité correspondant à la moyenne euclidienne (figure 5.1-(e)) qui n'est plus de la famille de distribution gaussienne.

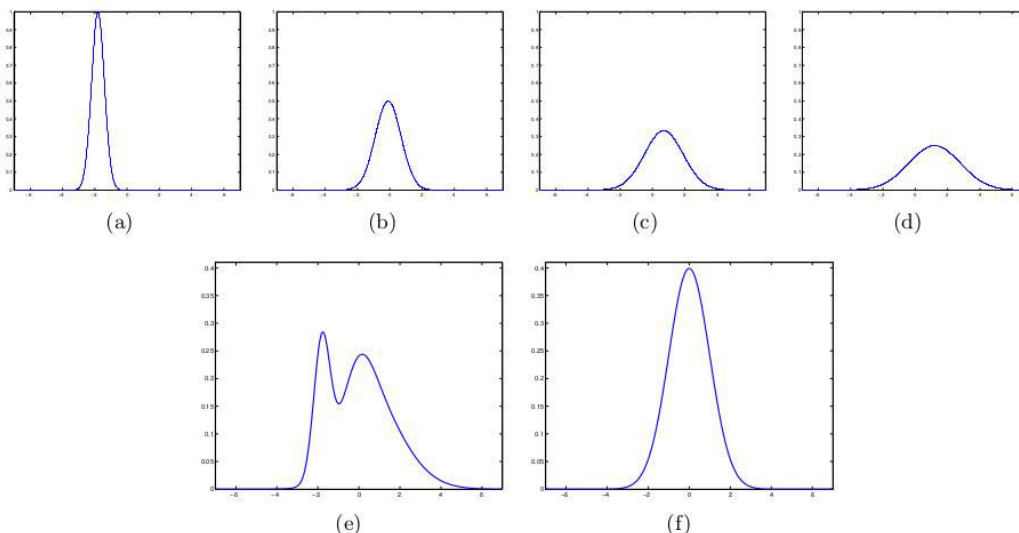


FIGURE 5.1 – Moyenne de gaussiennes : Comparaison de la distance euclidienne et la distance de Wasserstein. (a,b,c,d) Graphes de densités gaussiennes $f_1 \dots f_4$ avec différentes moyennes et variances. (e) la moyenne euclidienne de $f_1 \dots f_4$. Il ne s'agit plus d'une gaussienne. (f) Barycentre de Wasserstein, la famille gaussienne est conservée (d'après Bigot et al. (2013)).

- Supposons que nous voulions créer un chemin de distributions P_t (une géodésique) qui interpole entre deux distributions P_0 et P_1 . Nous

voudrions que les distributions P_t préservent la structure de base des distributions. La figure 5.2 montre un exemple. La ligne du haut montre le chemin entre P_0 et P_1 (qui sont des distributions gaussiennes) en utilisant la distance de Wasserstein, celle du bas utilise la distance euclidienne. On remarquera que dans ce dernier cas, l'interpolation ne donne pas des distributions gaussiennes, sauf pour les points extrêmes.

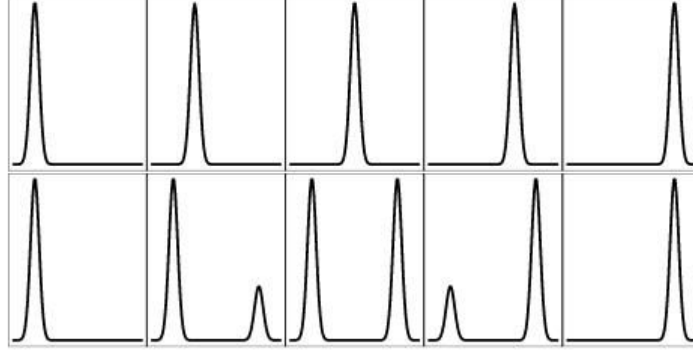


FIGURE 5.2 – Chemin de distributions : Comparaison de la distance euclidienne et la distance de Wasserstein. Rangée supérieure : chemin géodésique de P_0 à P_1 . Rangée du bas : chemin euclidien de P_0 à P_1 (d'après Wasserman (2018)).

— Inconvénients :

- Les calculs sont complexes et coûteux. La résolution numérique du problème de transport optimal peut se faire en le reformulant comme un problème linéaire, mais cette approche ne permet pas de traiter des instances de grande dimension. C'est la raison principale pour laquelle elle a été peu utilisée dans le traitement d'images jusqu'à récemment.
- Elle ne préserve pas toujours la famille de distribution, sauf s'il s'agit de la famille gaussienne (que les exemples figure 5.1 et figure 5.2 représentent)

5.1.1.3 Métrique d'information de Fisher

L'information de Fisher est à la base de la géométrie de l'information. Elle a été introduite par R.A. Fisher et quantifie l'information contenue dans une distribution relativement à un paramètre.

L'information de Fisher est définie pour un modèle statistique paramétrique $\{p_\theta | \theta \in \Theta\}$ par :

$$I(\theta) = E_\theta[\partial_\theta \ell_\theta(X) \cdot \partial_\theta \ell_\theta(X)^t], \quad \ell_\theta = \log p_\theta.$$

En estimation paramétrique, l'information de Fisher donne une limite à la précision de l'estimation donnée par un estimateur non biaisé T de la fonction θ d'un échantillon de taille n (borne de Cramer-Rao).

$$\text{Var}_\theta(T) \geq (nI(\theta))^{-1}$$

L'information de Fisher est la courbure (dérivée seconde) de la divergence de Kullback-Leibler $K(p, q) = E_p[\log(p/q)]$

$$\partial_\theta K(\theta^*, \theta)|_{\theta=\theta^*} = 0, \quad \partial_{\theta_i} \partial_{\theta_j} K(\theta^*, \theta)|_{\theta=\theta^*} = I(\theta^*)_{i,j}$$

L'information de Fisher présente deux caractéristiques fondamentales qui la rendent naturelle pour construire une distance entre densités de probabilités :

- Elle est invariante si l'on remplace la variable aléatoire X de loi p_θ intervenant dans sa définition par $F(X)$ avec F statistique suffisante. Le théorème de Chensov et ses nombreuses généralisations montre qu'elle est la seule à le vérifier sous certaines hypothèses.
- Elle est compatible avec les changements de paramètres par des difféomorphismes :

$$I(\eta) = J^t(\eta)I(\theta(\eta))J(\eta)$$

où J est la matrice jacobienne du changement de variable $\eta \mapsto \theta(\eta)$

Cette dernière propriété permet de la considérer comme une métrique riemannienne sur une variété dont un système de coordonnées locales est donné par les paramètres θ de la famille de densités p_θ .

L'intérêt de disposer d'une telle métrique va maintenant être brièvement expliqué, avant d'y revenir de façon plus détaillée dans une section dédiée aux concepts de base de la géométrie différentielle.

Une **variété topologique** M est définie comme un espace topologique localement "plat". Plus précisément chaque point d'une variété de dimension n a un voisinage homéomorphe¹ à l'espace euclidien de dimension n . Une courbe est une variété de dimension 1 (e.g. lignes, cercles), et une surface est une variété de dimension 2 (e.g. plan, sphère, tore). Il est possible de décrire une variété en utilisant une collection de "cartes"² réunies en un atlas mathématique qui recouvre toute la variété indiquant comment passer d'une carte à l'autre. Un bon exemple d'une variété est le globe terrestre puisqu'il peut être présenté par une collection de cartes géographiques (figure 5.3).

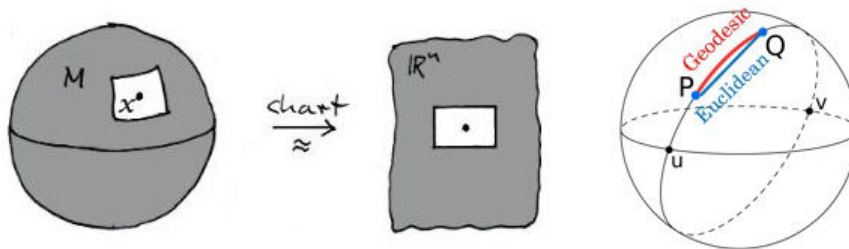


FIGURE 5.3 – L'exemple de la sphère.

À l'aide de la **métrique Riemannienne**, nous pouvons calculer des angles, des longueurs, des distances. La métrique Riemannienne est un produit scalaire défini localement sur chaque espace tangent.

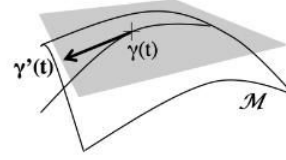
$$\langle v, w \rangle_x = g_x(v, w), \quad x \in M, \quad v, w \in T_x M$$

- Longueur d'un chemin lisse $\gamma : [0, 1] \rightarrow M$ chemin en M

1. Application bijective continue, d'un espace topologique dans un autre. Autrement dit : deux espaces topologiques sont "le même" vu différemment.

2. Ou "chart" en anglais.

$$L(\gamma) = \int_0^1 \sqrt{g(\gamma'(t), \gamma'(t))} dt.$$



- Les chemins de longueur minimale (localement) sont appelés **géodésiques**. La distance géodésique entre deux points est :

$$d(x, y) = \inf_{\gamma : \gamma(0)=x, \gamma(1)=y} L(\gamma).$$

→ Avantage(s) : Distance réelle. Il s'agit d'une métrique naturelle lorsque sont considérées des distributions de probabilité.

→ Inconvénient(s) : Complexe et coûteuse.

- Quelques exemples :
 - Tout espace euclidien est une variété (courbure nulle).
 - La sphère est une variété de courbure positive.
 - Les géodésiques de la sphère sont les arcs de grands cercles (figure 5.3 droite). Les géodésiques du plan sont des droites.

• Exemple illustratif :

L'utilisation de la géométrie de l'information, en particulier le calcul des géodésiques, permet de comparer des distributions de probabilité. La mesure de la distance entre les distributions normales univariées est la démonstration parfaite (Costa et al., 2014) (figure 5.4). Une distance entre deux points $A = (\mu_1, \sigma_1)$ et $B = (\mu_2, \sigma_2)$ dans le demi-plan (figure 5.4 droite) doit refléter la dissimilarité entre les distributions de probabilité associées (figure 5.4 gauche). Nous pouvons observer que les distributions C et D ont un écart-type plus élevé. En effet, elles sont plus dispersées et se chevauchent plus que les distributions A et B (figure 5.4 gauche). Par conséquent, la dissimilarité entre A et B est supérieure à la dissimilarité entre C et D. Mais, dans le demi-plan $(\mu; \sigma)$, en utilisant la géométrie euclidienne (i.e. distance euclidienne), la distance entre A et B est la même que C et D (figure 5.4 droite). Ce qui démontre la limite de la géométrie euclidienne et de la distance euclidienne. La géométrie de l'information propose une distance appropriée à partir de la matrice d'information de Fisher : Géodésique.

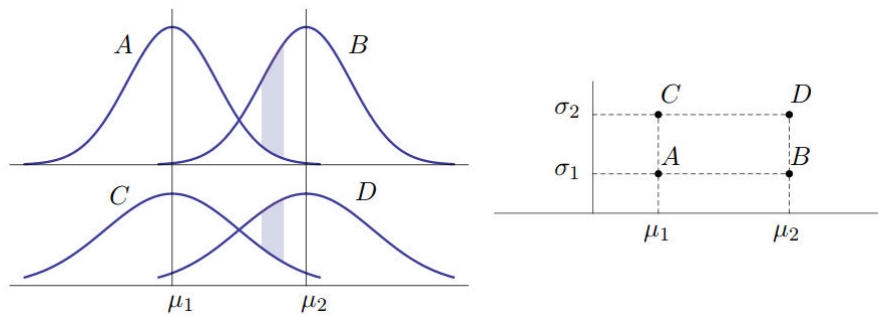


FIGURE 5.4 – Les distributions normales univariées (à gauche) et leurs représentations dans le demi-plan $(\mu; \sigma)$ (à droite), d'après Costa et al. (2014)

5.1.2 Géométrie différentielle

5.1.2.1 Variétés

Une variété différentielle est intuitivement un objet qui se comporte localement comme un espace vectoriel, bien que ses propriétés globales puisse s'en écarter notablement. Les exemples les plus courants de variétés différentielles sont donnés par des courbes ou des surfaces dans un espace vectoriel, le plus souvent \mathbb{R}^2 ou \mathbb{R}^3 . Le cas le plus simple que l'on peut considérer est celui des surfaces plongées de dimension n , que l'on représente comme l'image par une application ϕ d'un ouvert U de \mathbb{R}^n dans \mathbb{R}^p , $p > n$, ϕ étant un difféomorphisme sur $\phi(U)$. Le cas $n = 1$ donne les courbes plongées, $n = 2$ les surfaces plongées au sens usuel. La figure (5.5) donne un exemple de surface plongée en dimension 2. On remarque que le choix d'un repère cartésien sur U donne un système de coordonnées locales sur $\phi(U)$. Si x_0 est un point de U , il existe pour tout $k = 1 \dots n$ un réel positif ϵ tel que pour tout $t \in]-\epsilon, \epsilon[$, $x_0 + te_k \in U$, avec e_k le k -ième vecteur de la base canonique. La courbe $t \in]-\epsilon, \epsilon[\mapsto \phi(x_0 + te_k)$ est dite courbe coordonnée sur $\phi(U)$. Les surfaces plongées

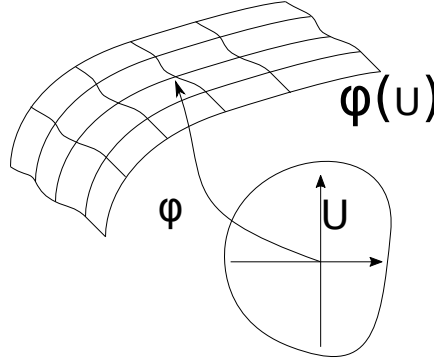


FIGURE 5.5 – Surface plongée de dimension 2

ne permettent pas de décrire tous les objets géométriques courants : c'est le cas de la sphère \mathbb{S}^2 , du tore de révolution \mathbb{T}^2 . De plus, il est nécessaire de disposer du plongement ϕ , qui n'est pas toujours facile à obtenir sous une forme explicite. Dans le cadre général des variétés différentielles abstraites, on procédera en sens inverse, à partir de la notion de carte.

Définition 1. Soit E un ensemble. Une carte sur E est un couple (U, ϕ) où U est une partie de E et ϕ est une bijection de U sur un ouvert $\phi(U)$ d'un espace vectoriel.

Définition 2. Deux cartes $(U, \phi), (V, \psi)$ sur un ensemble E avec $U \cap V \neq \emptyset$ sont dites compatibles de classe C^k , $k \in \mathbb{N}$ si l'application :

$$\psi \circ \phi^{-1} : \phi(U \cap V) \rightarrow \psi(U \cap V)$$

est un difféomorphisme de classe C^k .

Définition 3. Soit E un ensemble. Un atlas de classe C^k sur E est la donnée d'une famille $(U_i, \phi_i)_{i \in I}$ de cartes deux à deux compatibles de classe C^k telle que $E = \cup_{i \in I} U_i$.

En général la réunion de deux atlas sur un même ensemble E n'est pas un atlas, car des cartes appartenant à des atlas différents ne vérifient a priori aucune relation de compatibilité. Dans le cas où cela est vrai, on dit que les deux atlas

sont équivalents. C'est une relation d'équivalence sur l'ensemble des atlas sur E . La réunion de tous les atlas d'une classe d'équivalence est encore un atlas, qui est donc maximal : on en déduit la définition suivante pour une variété abstraite.

Definition 4. Une variété différentielle de classe C^k est un couple (M, \mathcal{A}) où \mathcal{A} est un atlas maximal de classe C^k sur M .

En pratique, on se donnera un atlas non nécessairement maximal, celui-ci ne servant que pour identifier de façon unique la structure de variété mise sur l'ensemble M . Une variété différentielle de classe C^∞ sera dite lisse.

Une variété (M, \mathcal{A}) est un espace topologique lorsqu'on l'on munit M de la topologie initiale pour les cartes. On imposera une condition technique, la paracompacité, qui permet d'assurer l'existence de partitions de l'unité.

Un exemple très classique, mais important, de variété lisse est donné par la sphère \mathbb{S}^n , dont on peut facilement exhiber un atlas à deux cartes, en utilisant des projections stéréographiques à partir de points antipodaux.

Proposition 1. Soit $\mathbb{S}^n = \{(x_0, \dots, x_n) \mid \sum_{i=0}^n x_i^2 = 1\}$ la sphère de dimension n de \mathbb{R}^{n+1} . Soient les points $x_N = (1, 0, \dots, 0)$, $x_S = (-1, 0, \dots, 0)$. On définit les deux projections stéréographiques $\phi_N: \mathbb{S}^n - \{x_N\} \rightarrow \mathbb{R}^n$, $\phi_S: \mathbb{S}^n - \{x_S\} \rightarrow \mathbb{R}^n$ par :

$$\begin{aligned}\phi_N(x_0, x_1, \dots, x_n) &= \left(\frac{x_1}{1 - x_0}, \dots, \frac{x_n}{1 - x_0} \right) \\ \phi_S(x_0, x_1, \dots, x_n) &= \left(\frac{x_1}{1 + x_0}, \dots, \frac{x_n}{1 + x_0} \right)\end{aligned}$$

Les deux cartes $(\mathbb{S}^n - \{x_N\}, \phi_N)$, $(\mathbb{S}^n - \{x_S\}, \phi_S)$ forment un atlas lisse sur \mathbb{S}^n .

Démonstration. On vérifie par le calcul que l'application $\phi_S \circ \phi_N^{-1} : \mathbb{R}^n - \{0\} \rightarrow \mathbb{R}^n - \{0\}$ s'écrit :

$$\phi_S \circ \phi_N^{-1}(u_1, \dots, u_n) = \frac{1}{\sum_{i=1}^n u_i^2} (u_1, \dots, u_n)$$

ce qui montre le caractère C^∞ de l'atlas. □

Tout ouvert U de \mathbb{R}^n définit une variété dont l'unique carte est $(U, \text{Id}|_U)$. C'est un cas couramment rencontré en géométrie de l'information. Dans toute la suite, on supposera que toutes les variétés sont lisses.

5.1.2.2 Applications différentiables

Definition 5. Soient (M, \mathcal{A}) , (N, \mathcal{B}) deux variétés lisses. Une application $f: M \rightarrow N$ est dite différentiable au point $x \in M$ si pour toute carte $(U, \phi) \in \mathcal{A}$ dont le domaine contient x et toute carte $(V, \psi) \in \mathcal{B}$ dont le domaine contient $f(x)$, l'application $\psi \circ f \circ \phi^{-1}$ est différentiable au point $\phi(x)$.

On notera que dans la définition précédent, l'application $\psi \circ f \circ \phi^{-1}$, que l'on appelle aussi application f lue dans les cartes (U, ϕ) , (V, ψ) , est définie entre deux ouverts de \mathbb{R}^n , ce qui permet de parler de différentiabilité en un point.

Une application sera dite différentiable si elle est différentiable au sens précédent en tout point de M . On définira de même les applications différentiables de classe C^k .

Deux cas particuliers très importants se rencontrent en pratique : Une courbe $\gamma:]a, b[\rightarrow M$ de classe C^k est une application de classe C^k entre la variété $(]a, b[, Id]_{a,b})$ et M . Une observable $\phi: M \rightarrow \mathbb{R}$ de classe C^k est une application de classe C^k entre M et la variété (\mathbb{R}, Id) .

5.1.2.3 Vecteurs tangents

Les courbes et les observables précédemment définies sont des objets extrêmement importants en géométrie. On remarquera en particulier que si γ est une courbe de classe C^k et ϕ une observable de classe C^k , alors $\phi \circ \gamma$ est une application de classe C^k d'un intervalle de \mathbb{R} dans \mathbb{R} , dont on peut calculer la dérivée en un point.

Definition 6. Soit (M, \mathcal{A}) une variété lisse et soit $x \in M$. Deux courbes lisses γ_1, γ_2 sur M passant par x sont dites tangentes en ce point si il existe une carte (U, ϕ) dont le domaine contient x telle que les courbes $\phi \circ \gamma_1, \phi \circ \gamma_2$ sont tangentes en $\phi(x)$.

On notera que si la propriété de tangence en x est vraie pour une carte, alors elle sera vraie pour toute carte contenant x , en raison de la condition de compatibilité. La définition ne perd aucune généralité en se limitant aux applications lisses. Enfin, la tangence en x est une relation d'équivalence, conséquence immédiate de cette propriété dans un espace vectoriel réel.

Definition 7. Soit (M, \mathcal{A}) une variété lisse et soit $x \in M$. Un vecteur tangent à M en x est une classe d'équivalence de courbes tangentes en x .

La définition précédente peut recevoir une interprétation très concrète : dans une même classe d'équivalence C de courbes tangentes en x , et pour une carte (U, ϕ) dont le domaine contient x , toutes les applications $\phi \circ \gamma, \gamma \in C$ ont même vecteur tangent $v \in \mathbb{R}^n$ en $\phi(x)$. Réciproquement, pour un vecteur $v \in \mathbb{R}^n$ quelconque, l'application $t \mapsto \phi^{-1}(\phi(x) + tv)$ définit un représentant d'une classe d'équivalence de courbes lisses passant par x . Si la carte (U, ϕ) est fixée, on remarque que l'association entre v et le vecteur tangent de représentant $t \mapsto \phi^{-1}(\phi(x) + tv)$ est bijective. On peut ainsi transporter la structure d'espace vectoriel de \mathbb{R}^n sur l'ensemble des vecteurs tangents en x .

Definition 8. Soit (M, \mathcal{A}) une variété lisse et soit $x \in M$. L'espace tangent à M en x , noté $T_x M$, est l'espace vectoriel de dimension n des vecteurs tangents en x .

Il est très important de remarquer les espaces tangents $T_x M, T_y M$ pour $x \neq y$ sont des objets différents, car ils correspondent à des ensembles de courbes distincts. En tant qu'espaces vectoriels, ils sont en revanche isomorphes. La réunion $\cup_{x \in M} (x, T_x M)$, appelé fibré tangent de M et noté TM , peut être muni d'une structure de variété en utilisant des cartes construites à partir de celles de l'atlas de la variété de base.

Proposition 2. Soit TM le fibré tangent d'une variété lisse (M, \mathcal{A}) . A toute carte $(U, \phi) \in \mathcal{A}$, on associe le couple $(\tilde{U}, \tilde{\phi})$ où $\tilde{U} = \cup_{x \in U} (x, T_x M)$ et $\tilde{\phi}: (x, [\gamma]) \mapsto (\phi(x), (\phi \circ \gamma)'(x))$ avec $[\gamma]$ la classe d'équivalence des courbes lisses passant par x et de représentant γ . L'ensemble des couples $(\tilde{U}, \tilde{\phi})$ pour (U, ϕ) parcourant \mathcal{A} est un atlas lisse sur TM .

TM ayant une structure de variété lisse, on peut définir par récurrence TTM, T^3M, \dots . La dimension de $T^k M$ est kn si la dimension de (M, \mathcal{A}) est n .

Il existe une application lisse canonique de TM sur M , notée π , qui à un couple $(x, T_x M)$ associe le point x . Elle est appelée projection du fibré tangent TM .

Définition 9. Une section (lisse) $s: M \rightarrow TM$ est une application lisse telle que $\pi \circ s = Id_M$

Une section est également appelée champ de vecteur. L'ensemble des sections de TM est noté $\Gamma(TM)$. Une section X de $\Gamma(TM)$ associe à un point $x \in M$ un couple $(x, [\gamma])$ où $[\gamma]$ est un vecteur tangent au point x . Dans une carte $(\tilde{U}, \tilde{\phi})$ de TM dont le domaine contient x , il sera représenté sous la forme $(\phi(x), v)$, avec $v \in \mathbb{R}^n$. Localement, un champ de vecteur sur une variété peut être vu comme un champ de vecteurs sur un ouvert de \mathbb{R}^n . Dans la suite, le point de vue local sera toujours implicitement adopté.

Définition 10. Soit $X \in \Gamma(TM)$ un champ de vecteurs. A toute observable lisse $f: M \rightarrow \mathbb{R}$, on peut associer l'observable notée $X(f)$ définie en tout point $x \in M$ par :

$$X(f)(x) = \frac{d}{dt} f \circ \phi^{-1}(\phi(x) + tv)$$

avec (ϕ, v) la représentation de $X(x)$ dans la carte $(\tilde{U}, \tilde{\phi})$.

Proposition 3. Soit $X \in \Gamma(TM)$. L'application $f \in C^\infty(M, \mathbb{R}) \mapsto X(f)$ est une dérivation sur l'algèbre $C^\infty(M, \mathbb{R})$:

$$\begin{aligned} \forall \lambda \in \mathbb{R}, f, g \in C^\infty(M, \mathbb{R}), X(\lambda f + g) &= \lambda X(f) + X(g) \\ \forall f, g \in C^\infty(M, \mathbb{R}), X(fg) &= X(f)g + fX(g) \end{aligned}$$

Pour des variétés de dimension finie telle que considérées dans ce travail, il existe une correspondance bijective entre les dérivations sur $C^\infty(M, \mathbb{R})$ et les champs de vecteurs sur M . Selon le contexte, il est plus simple d'utiliser l'une ou l'autre des représentations.

Au voisinage d'un point $x \in M$, il est toujours possible de définir localement des champs de vecteurs privilégiés que l'on appelle champs coordonnées et que l'on note $\partial_i, i = 1 \dots n$. Dans une carte (U, ϕ) , on les définit par leur action sur les observables :

$$\partial_i(f)(y) = \frac{d}{dt} f \circ \phi^{-1}(\phi(x) + te_i) = \frac{\partial}{\partial y_i} f \circ \phi^{-1}(y)|_{y=x}$$

Il est important de remarquer que cette écriture n'a de sens que dans une carte locale. Tout champ de vecteur s'y écrit alors sous la forme :

$$X(x) = \sum_{i=1}^n X_i(x) \partial_i$$

avec $X_i, i = 1 \dots n$ des observables lisses définies sur le domaine de la carte locale. Pour simplifier les notations, on adoptera la convention d'Einstein de sommation des indices répétés et on écrira plus simplement :

$$X = X^i \partial_i$$

Definition 11. Soient $X, Y \in \Gamma(TM)$. On définit le champ de vecteur lisse $[X, Y]$:

$$\forall f \in C^\infty(M, \mathbb{R}), [X, Y](f) = X(Y(f)) - Y(X(f))$$

$[X, Y]$ est appelé crochet de Lie de X et Y

Le crochet de Lie traduit la non-commutativité de deux champs de vecteurs. Le lemme de Schwarz montre directement que $[\partial_i, \partial_j] = 0$.

Proposition 4. Le crochet de Lie est un opérateur \mathbb{R} -bilinéaire qui vérifie :

$$[X, X] = 0$$

$$[X, Y] = -[Y, X]$$

$$[X, [Y, Z]] + [Z, [X, Y]] + [Y, [Z, X]] = 0 \text{ (identité de Jacobi).}$$

5.1.2.4 Connexions

Contrairement aux observables pour lesquelles il est facile d'obtenir la dérivée par action d'un champ de vecteur, la dérivée d'un champ de vecteur par rapport à une autre est plus difficile à définir, la raison principale étant que l'ensemble des valeurs est maintenant une variété, en l'occurrence TM . Il existe plusieurs façons équivalentes de résoudre ce problème, mais la plus simple à énoncer est celle qui consiste à utiliser le concept de dérivation, cette fois sur le module des champs de vecteurs.

Definition 12. Une connexion ∇ est une application \mathbb{R} -bilinéaire sur le module des champs de vecteurs qui vérifie :

$$\nabla_{fX+Y}Z = f\nabla_XZ + \nabla_YZ, f \in C^\infty(M, \mathbb{R}), X, Y, Z \in \Gamma(TM)$$

$$\nabla_X(fY) = X(f)Y + f\nabla_XY, f \in C^\infty(M, \mathbb{R}), X, Y \in \Gamma(TM)$$

On peut écrire en coordonnées locale une connexion en utilisant les champs ∂_i et les deux relations précédentes. Soient $X = X^i\partial_i, Y = Y^j\partial_j$. On a :

$$\nabla_XY = X^i\nabla_{\partial_i}Y^j\partial_j = X^i(\partial_i(Y^j)\partial_j + Y^j\nabla_{\partial_i}\partial_j)$$

En posant :

$$\nabla_{\partial_i}\partial_j = \Gamma_{ij}^k\partial_k$$

il vient, après un changement d'indices :

$$\nabla_XY = X^i\partial_i(Y^k)\partial_k + Y^jX^i\Gamma_{ij}^k\partial_k$$

On remarque avec cette écriture qu'une connexion est entièrement déterminée par les observables Γ_{ij}^k , que l'on appelle symboles de Christoffel de ∇ .

Definition 13. La torsion d'une connexion ∇ est définie comme l'application \mathbb{R} -bilinéaire :

$$T(X, Y) = \nabla_XY - \nabla_YX - [X, Y], X, Y \in \Gamma(TM)$$

Une connexion est dite sans torsion si $T = 0$.

Definition 14. La courbure d'une connexion ∇ est définie comme l'application \mathbb{R} -trilinéaire :

$$R(X, Y)Z = \nabla_X\nabla_YZ - \nabla_Y\nabla_XZ - \nabla_{[X, Y]}Z, X, Y, Z \in \Gamma(TM)$$

On peut montrer que la courbure est en réalité trilinéaire sur le module des champs de vecteurs. Une connexion dont la courbure est identiquement nulle est dite plate.

5.1.2.5 Équations des géodésiques

Soit ∇ une connexion. Une courbe lisse γ sur M est une géodésique de ∇ si $\nabla_{\gamma'}\gamma' = 0$. En utilisant des coordonnées locales, cette équation est équivalente au système différentiel suivant :

$$\frac{d^2}{dt^2}\gamma^k(t) + \Gamma_{ij}^k \frac{d}{dt}\gamma^i(t) \frac{d}{dt}\gamma^j(t) = 0 \quad (5.3)$$

Par le théorème de Cauchy, il existera une unique solution locale vérifiant les conditions initiales $\gamma(0) = x, \gamma'(0) = v$.

Les géodésiques sont souvent associées à une connexion particulière, et correspondent dans ce cas à des courbes de longueur minimum. Pour cela, il est nécessaire de disposer d'une notion de distance locale.

Definition 15. Soit M une variété lisse. Une métrique riemannienne sur M est la donnée en tout point $x \in M$ d'un produit scalaire noté g_x dans $T_x M$ qui est tel que l'application :

$$x \mapsto g_x(X(x), Y(x)), X, Y \in \Gamma(TM)$$

soit une observable lisse.

On note souvent $g(x; X(x), Y(x))$ au lieu de $g_x(X(x), Y(x))$. De même, l'observable de la définition (15) est souvent abrégée en $g(X, Y)$. En coordonnées locales, pour deux champs $X = X^i \partial_i, Y = Y^j \partial_j$ on peut écrire :

$$g(x; X, Y) = g_{ij}(x) X^i Y^j$$

avec g_{ij} des observables lisses. Une variété munie d'une métrique riemannienne est appelée variété riemannienne.

Definition 16. On dit qu'une connexion ∇ est métrique si elle vérifie :

$$X(g(Y, Z)) = g(\nabla_X Y, Z) + g(Y, \nabla_X Z)$$

Le caractère métrique est fortement contraignant, comme le montre la proposition suivante :

Proposition 5. Il existe une unique connexion métrique sans torsion que l'on appelle connexion de Levi-Civita.

Démonstration. La connexion de Levi-Civita est donnée par la formule de Koszul, conséquence directe des hypothèses :

$$\begin{aligned} 2g(\nabla_X Y, Z) = & X(g(Y, Z)) + Y(g(X, Z)) - Z(g(X, Y)) \\ & + g([X, Y], Z) - g([X, Z], Y) - g([Y, Z], X) \end{aligned}$$

□

Les symboles de christoffel de la connexion de Levi-Civita peuvent s'obtenir facilement à l'aide de la formule de Koszul, dans laquelle les termes liés aux commutateurs d'annulent :

$$g(\nabla_{\partial_i} \partial_j, \partial_k) = \frac{1}{2} (\partial_i g_{jk} + \partial_j g_{ik} - \partial_k g_{ij})$$

Ce qui donne finalement :

$$g_{mk} \Gamma_{ij}^m = \frac{1}{2} (\partial_i g_{jk} + \partial_j g_{ik} - \partial_k g_{ij})$$

Definition 17. Soit (M, g) une variété riemannienne et soit un chemin lisse $\gamma: [0, 1] \rightarrow M$. La longueur de γ , notée $l(\gamma)$ est définie comme :

$$l(\gamma) = \int_0^1 g(\gamma'(t), \gamma'(t))^{1/2} dt$$

Definition 18. Soit (M, g) une variété riemannienne et soient $x, y \in M$. Un chemin de longueur minimum entre x et y est appelé géodésique minimisante.

Proposition 6. Une géodésique minimisante est une géodésique pour la connexion de Levi-Civita.

5.2 Application 1 : classification de la population de la maladie d'Alzheimer

La métrique d'information de Fisher fournit des densités de probabilité paramétrées avec une structure de variété riemannienne, donnant la géométrie dite d'information. La géométrie de l'information de la variété gamma associée à la famille des distributions gamma a été bien étudiée. Cependant, seuls quelques résultats sont connus pour la famille gamma généralisée, qui ajoute un paramètre de forme supplémentaire. Cette section donne de nouveaux résultats sur la variété gamma généralisée et introduit également une application en imagerie médicale qui est la classification de la population atteinte de la maladie d'Alzheimer.

Dans le domaine médical, un nombre croissant de techniques d'analyse d'images quantitatives ont été mises au point, notamment l'analyse par histogramme, largement utilisée pour quantifier les changements pathologiques diffus de certaines maladies neurologiques. Cette méthode présente plusieurs inconvénients. En effet, toutes les informations incluses dans l'histogramme ne sont pas utilisées et l'histogramme est une estimation trop simpliste d'une distribution de probabilité. Ainsi, dans cette étude, nous présentons comment l'utilisation de la géométrie de l'information et de la variété gamma généralisée a amélioré les performances de la classification de la population Alzheimer.

5.2.1 Conception mathématique

La distribution gamma généralisée a été introduite par Stacy (1962) et peut être considérée comme un cas particulier de la distribution Amoroso (Amoroso, 1925) dans laquelle le paramètre de décalage est supprimé (Crooks, 2010). Outre la distribution gamma, elle généralise également la distribution de Weibull et est couramment utilisée dans les modèles de survie. Le présent travail a pour objectif d'examiner certaines propriétés de la géométrie de l'information de la famille gamma généralisée, en particulier lorsqu'elles se limitent à la sous-variété gamma. Tout d'abord, l'information de Fisher en tant que métrique riemannienne et les résultats obtenus dans le cas de la variété gamma seront brièvement présentés. Ensuite, le cas de la variété gamma généralisée sera détaillé, en utilisant une approche basée sur des groupes de difféomorphismes. Puis, la courbure extrinsèque de la sous-variété gamma sera calculée. Dans les sections

suivantes, deux exemples d'application dans le domaine de l'imagerie médicale et de l'aéronautique seront présentés.

5.2.1.1 Géométrie de l'information et la variété gamma

La géométrie de l'information traite des familles paramétrées de distributions dont les paramètres sont interprétés comme des coordonnées locales sur une variété munie d'une structure riemannienne par la métrique de Fisher. Dans la suite, Θ sera une variété lisse et $(p_\theta), \theta \in \Theta$ une famille de fonctions de densité de probabilité définies sur un espace d'événements commun Ω et dépendant des paramètres θ .

Définition 1. La métrique d'information de Fisher sur Θ est définie au point $\theta \in \Theta$ par le tenseur d'ordre 2 symétrique :

$$g = g_{ij} d\theta^i \otimes d\theta^j$$

où :

$$g_{ij} = E_{p_\theta} [\partial_{\theta_i} l \partial_{\theta_j} l], \quad l(\theta) = \log p_\theta$$

Lorsque le support des fonctions de densité p_θ ne dépend pas de θ , la métrique d'informations peut être réécrite comme suit :

$$g_{ij} = -E_{p_\theta} [\partial_{\theta_i} \partial_{\theta_j} l] \quad (5.4)$$

Il en résulte une métrique riemannienne sur Θ .

Lorsque l'espace événement sous-jacent Ω est également une variété lisse, la métrique de Fisher possède une propriété d'invariance, qui correspond à la conservation d'informations par des statistiques suffisantes :

Proposition 1. Soit $\tilde{\Omega}$ une variété lisse et $\Phi: \Omega \rightarrow \tilde{\Omega}$ un difféomorphisme lisse. Soit \tilde{g} la métrique d'informations de Fisher associée à la famille d'images $\Phi_* p_\theta$ définie dans l'espace d'événements $\tilde{\Omega}$. Alors $\tilde{g} = g$.

La métrique de Fisher a une expression très simple lorsque la famille paramétrée p_θ est de type exponentiel naturel. Dans un tel cas, en supposant que, par souci de simplicité, Θ et Ω soient des sous-ensembles ouverts d'espaces vectoriels réels de dimensions finies, la fonction de densité p_θ peut s'écrire ainsi :

$$p_\theta(x) = \exp(\langle \theta, F(x) \rangle - \phi(\theta) + g(x)) \quad (5.5)$$

La fonction ϕ dans l'équation (5.5) est appelée la fonction potentiel de la densité et une application immédiate de la définition (1) donne l'expression de la métrique d'information de Fisher :

$$g_{ij}(\theta) = \frac{\partial^2 \phi}{\partial \theta_i \partial \theta_j}(\theta) \quad (5.6)$$

Une variété avec une telle métrique riemannienne est appelée une structure Hessienne (Shima, 2013). De nombreux outils importants de la géométrie riemannienne, comme la connexion Levi-Civita, sont grandement simplifiés dans ce cadre. Dans la suite, toutes les dérivées partielles ∂_{θ_i} seront abrégées par ∂_i .

Proposition 2. Pour une famille de densité paramétrée $p_\theta, \theta \in \Theta$ appartenant à la classe exponentielle avec la fonction potentielle Ψ , les symboles de Christoffel du premier type de la structure Hessienne associée sont donnés par [Duistermaat \(2001\)](#) :

$$\Gamma_{ijk} = \frac{1}{2} \partial_i \partial_j \partial_k \phi$$

La distribution gamma peut être écrite comme une famille exponentielle à deux paramètres (α, λ) , définis sur l'espace $\mathbb{R}^{+*} \times \mathbb{R}^{+*}$ par :

Définition 2. La distribution gamma est la loi de probabilité sur \mathbb{R}^{+*} avec une densité relative à la mesure de Lebesgue donnée par :

$$p(x; \alpha, \lambda) = \frac{1}{\Gamma(\lambda) \alpha^\lambda} x^{\lambda-1} e^{-\frac{x}{\alpha}}, x > 0 \quad (5.7)$$

avec les paramètres $\alpha > 0, \lambda > 0$.

La proposition suivante vient directement de la définition :

Proposition 3. La distribution gamma définit une famille exponentielle avec des paramètres naturels $\lambda, \eta = \alpha^{-1}$ et admet pour fonction potentiel $\phi(\eta, \lambda) = \log(\Gamma(\lambda)) - \lambda \log(\eta)$.

En utilisant (5.6), la métrique de Fisher est obtenue par un calcul simple :

$$g(\eta, \lambda) = \begin{pmatrix} \frac{\lambda}{\eta^2} & -\frac{1}{\eta} \\ -\frac{1}{\eta} & \psi'(\lambda) \end{pmatrix} \quad (5.8)$$

où ψ est la fonction digamma.

Il est parfois pratique de modifier le paramétrage afin d'avoir une forme diagonale pour la métrique. La proposition suivante est d'utilisation courante et permet de calculer une métrique pullback en coordonnées locales :

Proposition 4. Soit \mathcal{M} une variété lisse et (\mathcal{N}, g) une variété riemannienne lisse. Pour un difféomorphisme lisse $f: \mathcal{M} \rightarrow \mathcal{N}$, la matrice de la métrique pullback f^*g s'exprime en coordonnées locales au point $m \in \mathcal{M}$ par :

$$J_f^t(m) G(f(m)) J_f(m) \quad (5.9)$$

avec $J_f(m)$ la matrice jacobienne de f en m et $G(n)$ la matrice de la métrique g en $n \in \mathcal{N}$.

Par le changement de paramétrage : $f: (\mu, \beta) \mapsto (\eta = \beta/\mu, \lambda = \beta)$ on obtient :

$$J_f(\mu, \beta) = \begin{pmatrix} -\frac{\mu}{\beta^2} & \frac{1}{\beta} \\ 0 & 1 \end{pmatrix}$$

L'utilisation de la proposition 4 donne alors pour la matrice de la métrique pullback :

$$G(\mu, \beta) = \begin{pmatrix} \frac{\beta}{\mu^2} & 0 \\ 0 & \psi(\beta)' - \frac{1}{\beta} \end{pmatrix}.$$

La géométrie de l'information de la distribution gamma est étudiée en détail dans ([Arwini et al., 2008](#)), avec des calculs explicites des symboles de Christoffel et de l'équation géodésique.

5.2.1.2 La géométrie de la variété gamma généralisée

Bien que la distribution gamma soit bien adaptée pour étudier des cas qui dévie de loi exponentielle (Arwini et al., 2008), elle n'est pas assez générale dans de nombreuses applications. En particulier, la distribution de Weibull, qui généralise également la distribution exponentielle, n'est pas une distribution gamma. Une famille plus générale a donc été introduite, en ajoutant un terme de puissance.

Définition 3. La distribution gamma généralisée est la mesure de probabilité sur \mathbb{R}^+ avec une densité respective à la mesure de Lebesgue donnée par :

$$p(x; \alpha, \lambda, \beta) = \frac{\beta x^{\beta\lambda-1}}{\alpha^{\beta\lambda}\Gamma(\lambda)} e^{-\left(\frac{x}{\alpha}\right)^\beta}, \quad x > 0 \quad (5.10)$$

Où $\alpha > 0, \lambda > 0, \beta > 0$.

En raison de l'exposant β , la distribution gamma généralisée ne définit pas une famille exponentielle naturelle. Cependant, en laissant β fixé, $\Phi_\beta: x \mapsto x^\beta$ est un difféomorphisme de \mathbb{R}^+ sur lui-même, et la densité image de $p(\alpha, \lambda, \beta)$ sous Φ_β est une densité gamma avec les paramètres (α^β, λ) . Pour tout $\kappa > 0$, la sous-variété $\beta = \kappa$ de la variété gamma généralisée est difféomorphe à la variété gamma. En utilisant l'invariance de la métrique de Fisher par difféomorphismes, la métrique induite sur la sous-variété ci-dessus peut être obtenue.

Proposition 5. Soit $\kappa > 0$ un nombre réel fixe. La métrique de Fisher induite G_κ sur la sous-variété $(\alpha, \lambda, \kappa)$ de la variété gamma généralisée est donnée en coordonnées locales par :

$$G_\kappa(\alpha, \lambda) = \begin{pmatrix} \frac{\lambda\kappa^2}{\alpha^2} & -\frac{\kappa}{\alpha} \\ -\frac{\kappa}{\alpha} & \psi'(\lambda) \end{pmatrix}.$$

Proposition 6. En coordonnées locales, la métrique d'information de Fisher de la variété gamma généralisée est donnée par :

$$G(\alpha, \lambda, \beta) = \begin{pmatrix} \frac{\beta^2\lambda}{\alpha^2} & -\frac{\beta}{\alpha} & \frac{-\lambda\psi(\lambda)-1}{\alpha} \\ -\frac{\beta}{\alpha} & \psi'(\lambda) & -\frac{\psi(\lambda)}{\beta} \\ \frac{-\lambda\psi(\lambda)-1}{\alpha} & -\frac{\psi(\lambda)}{\beta} & \frac{\lambda\psi(\lambda)^2+2\psi(\lambda)+\lambda\psi'(\lambda)+1}{\beta^2} \end{pmatrix} \quad (5.11)$$

La définition 3 habituelle de la distribution gamma généralisée découle de la distribution gamma par un simple changement de variable, rendant ainsi certains calculs moins naturels. En commençant par le difféomorphisme Φ_β ci-dessus et en l'appliquant à une distribution gamma, on obtient une forme équivalente, mais plus intuitive. De plus, il est généralement préférable d'exprimer la densité gamma en tant que distribution exponentielle naturelle :

$$p(x; \eta, \lambda) = \frac{\eta^\lambda x^{\lambda-1} e^{-\eta x}}{\Gamma(\lambda)}, \quad x > 0,$$

où $\lambda > 0, \eta > 0$ sont des paramètres naturels de la distribution.

Définition 4. La distribution gamma généralisée sur \mathbb{R}^+ est la mesure de probabilité avec la densité :

$$p(x; \eta, \lambda, \beta) = \frac{\beta \eta^\lambda x^{\beta\lambda-1} e^{-\eta x^\beta}}{\Gamma(\lambda)}, \quad x > 0,$$

avec $\eta > 0, \lambda > 0$ et $\beta > 0$.

En raison de la propriété d'invariance par difféomorphisme de la métrique d'information de Fisher, la métrique induite sur les sous-variétés $\beta = \text{cte}$ est indépendante de β et est exactement celle de la variété gamma, donnée ici par :

$$g(\eta, \lambda) = \begin{pmatrix} \frac{\lambda}{\eta^2} & -\frac{1}{\eta} \\ -\frac{1}{\eta} & \psi'(\lambda) \end{pmatrix}. \quad (5.12)$$

La famille des difféomorphismes Φ_β vérifie la propriété de semi-groupe $\Phi_{\beta_1} \circ \Phi_{\beta_2} = \Phi_{\beta_1 \beta_2}$. Il s'avère que tous les calculs peuvent être effectués dans le cas général d'un groupe de Lie, comme indiqué ci-dessous.

Soit $p_\theta, \theta \in \Theta$ une famille paramétrée définie sur un ouvert U of \mathbb{R}^n et, soit W un groupe de Lie agissant U par des difféomorphismes préservant l'orientation. Pour tout w dans W et θ dans Θ , la densité image $\tilde{p}_{w,\theta}$ par le difféomorphisme $x \in U \mapsto \xi(w, x) = w.x$ est donnée par :

$$\forall x \in U, \quad \tilde{p}_{w,\theta}(x) = p_\theta(\xi(w, x)) |\det \partial_2 \xi(w, x)|.$$

Les difféomorphismes préservant l'orientation la valeur absolue peut être supprimée dans l'expression ci-dessus. En notant $\tilde{l}(x, \theta, w)$ la log-vraisemblance de $\tilde{p}_{w,\theta}(x)$ et $l(x, \theta)$ celle de $p_\theta(x)$, un calcul immédiat donne :

$$\forall x \in U, \quad \tilde{l}(x, \theta, w) = l(\xi(w, x), \theta) + \log \det \partial_2 \xi(w, x).$$

Dans cette section la notation ∂_i indique une dérivée partielle par rapport à la i -ième variable. Les dérivées d'ordre supérieur sont notées de même $\partial_{i\dots i, j\dots j, \dots}$ en répétant l'indice k fois pour indiquer une dérivée partielle d'ordre k .

Proposition 7. Pour tout $x \in U, w \in W$:

$$\partial_1 \xi(w, x) = \partial_1 \xi(e, \xi(w, x)) T_w R_{w^{-1}},$$

où e est l'élément neutre de W et R_w est l'application translation à droite $h \in W \mapsto R_w.h = h.w$.

Démonstration. comme ξ dérive d'une action de groupe :

$$\xi(h, \xi(w, x)) = \xi(h.w, x).$$

En prenant la dérivée par rapport à h en l'identité :

$$\partial_1 \xi(e, \xi(w, x)) = \partial_1 \xi(w, x) T_e R_w.$$

Comme $T_e R_w T_w R_{w^{-1}} = Id$ par le théorème des application composées, on en déduit le résultat. \square

Cette proposition permet de calculer facilement l'information de Fisher

Proposition 8. L'élément $g_{w,\theta}$ de la matrice de l'information de Fisehr de $\tilde{p}_{w,\theta}$ est donné par :

$$- \int_U \partial_{12} l(x, \theta) \partial_1 \xi(e, x) p_\theta(x) dx T_w R_{w^{-1}}.$$

Démonstration. Comme :

$$\tilde{l}(x, \theta, w) = l(\xi(w, x), \theta) + \log \det \partial_2 \xi(w, x),$$

Il vient :

$$\partial_2 \tilde{l}(x, \theta, w) = \partial_2 l(\xi(w, x), \theta),$$

et donc :

$$\partial_{23} \tilde{l}(x, \theta, w) = \partial_{12} l(\xi(w, x), \theta) \partial_1 \xi(w, x).$$

En utilisant la Proposition 7 :

$$\partial_{23} \tilde{l}(x, \theta, w) = \partial_{12} l_\theta(\xi(w, x), \theta) \partial_1 \xi(e, \xi(w, x)) T_w R_w^{-1}.$$

En prenant l'espérance par rapport à $\tilde{p}_{w, \theta}$:

$$E[\partial_{23}] = \int_U \partial_{12} l(\xi(w, x), \theta) \partial_1 \xi(e, \xi(w, x)) \tilde{p}_{w, \theta}(x) dx T_w R_w^{-1}$$

et le résultat s'obtient finalement par le changement de variable $y = \xi(w, x)$. \square

Le cas des éléments $g_{w, w}$ est un peu plus compliqué, en raison d'un terme supplémentaire non nul dans $\tilde{l}(x, \theta, w)$. En prenant la dérivée première par rapport à w il vient :

$$\forall x \in U, \quad \partial_3 \tilde{l}(x, \theta, w) = \partial_1 l(\xi(w, x), \theta) \partial_1 \xi(w, x) + \text{tr} (\partial_{12} \xi(w, x) \partial_2 \xi(w, x)^{-1}),$$

où tr représente la trace de l'application linéaire par rapport aux composantes en x . Le second terme en partie droite peut être encore simplifié en utilisant la proposition suivante, qui est une conséquence directe de (7).

Proposition 9. Pour tout $\theta \in \Theta, w \in W, x \in U$:

$$\partial_{12} \xi(e, \xi(w, x)) \partial_2 \xi(w, x) = \partial_{12} \xi(w, x) T_e R_w.$$

En l'appliquant à la dérivée de la log-vraisemblance et en utilisant à nouveau la Proposition 7, on obtient :

$$\forall x \in U, \quad \partial_3 \tilde{l}(x, \theta, w) = (\partial_1 l(\xi(w, x), \theta) \partial_1 \xi(e, \xi(w, x)) + \text{tr} (\partial_{12} \xi(e, \xi(w, x)))) T_w R_w^{-1}.$$

Proposition 10. L'élément $g_{w, w}$ de la matrice de l'information de Fisher de $\tilde{p}_{w, \theta}$ est donnée par :

$$T_w R_w^{-1} \int_U h_{w, \theta}(x)^T h_{w, \theta}(x) p_\theta(x) dx T_w R_w^{-1},$$

avec :

$$h_{w, \theta}(x) = \partial_1 l(x, \theta) \partial_1 \xi(e, x) + \text{tr} (\partial_{12} \xi(e, x)).$$

Démonstration. A partir de la définition :

$$g_{w, w} = E[(\partial_3 \tilde{l})^T (\partial_3 \tilde{l})],$$

le résultat s'obtient par le changement de variable $y = \xi(w, x)$ dans l'espérance. \square

Un important corollaire des propositions 8 et 10 est que l'information de Fisher est une métrique invariante à droite sur le groupe de Lie \mathcal{W} .

Les propositions 8 et 10 permettent le calcul des coefficients $g_{\eta\beta}, g_{\lambda\beta}, g_{\beta\beta}$.

Proposition 11. *La matrice de l'information de Fisher en coordonnée naturelles s'exprime par :*

$$\begin{aligned} g_{\eta\eta} &= \frac{\lambda}{\eta^2}, \\ g_{\eta\lambda} &= -\frac{1}{\eta}, \\ g_{\lambda\lambda} &= \psi'(\lambda), \\ g_{\eta\beta} &= \frac{\lambda}{\eta\beta} (\psi(\lambda+1) - \log \eta), \\ g_{\lambda\beta} &= \frac{1}{\beta} (\log \eta - \psi(\lambda)), \\ g_{\beta\beta} &= \frac{1}{\beta^2} [1 + \lambda \log^2 \eta - 2\lambda\psi(\lambda+1) \log \eta + \lambda\psi^2(\lambda+1) + \lambda\psi'(\lambda+1)]. \end{aligned}$$

En utilisant l'expression suivante pour le calcul des symboles de Christoffel de première espèce :

$$\Gamma_{kij} = \frac{1}{2} (\partial_i g_{jk} + \partial_j g_{ik} - \partial_k g_{ij}),$$

on peut les obtenir explicitement :

$$\begin{aligned} \Gamma_{111} &= -\frac{\lambda}{\eta^3}, & \Gamma_{211} &= \frac{1}{2\eta^2}, & \Gamma_{311} &= \frac{\lambda(-1+\log \eta - \psi(\lambda+1))}{\eta^2\beta}, \\ \Gamma_{121} &= \Gamma_{211} = \Gamma_{112}, & \Gamma_{221} &= \Gamma_{212} = 0, & \Gamma_{321} &= \Gamma_{312} = \frac{1-\log \eta + \psi(\lambda+1) + \lambda\psi'(\lambda+1)}{2\eta\beta}, \\ \Gamma_{122} &= 0, & \Gamma_{222} &= \frac{1}{2}\psi''(\lambda), & \Gamma_{322} &= -\frac{\psi'(\lambda)}{\beta}, \\ & & \Gamma_{131} &= \Gamma_{113} = 0, & & \\ & & \Gamma_{231} &= \Gamma_{213} = \frac{1+\log \eta - \psi(\lambda+1) - \lambda\psi'(\lambda+1)}{2\eta\beta}, & & \\ & & \Gamma_{331} &= \Gamma_{313} = \frac{\lambda(\log \eta - \psi(\lambda+1))}{\eta\beta^2}, & & \\ & & \Gamma_{132} &= \Gamma_{123} = \frac{-1-\log \eta + \psi(\lambda+1) + \lambda\psi'(\lambda+1)}{2\eta\beta}, & & \\ & & \Gamma_{232} &= \Gamma_{223} = 0, & & \\ \Gamma_{332} &= \Gamma_{323} = \frac{\psi'(\lambda+1)(1-2\lambda \log \eta) - 2\psi(\lambda+1)(\log \eta - \lambda\psi'(\lambda+1)) + \log^2 \eta + \psi(\lambda+1)^2 + \lambda\psi''(\lambda+1)}{2\beta^2}, & & \\ & & \Gamma_{133} &= 0, & & \\ \Gamma_{233} &= -\frac{-2\psi(\lambda+1)(\log \eta - \lambda\psi'(\lambda+1)) + \psi'(\lambda+1)(1-2\lambda \log \eta) + \log \eta(\log \eta + 2) + \psi(\lambda+1)^2 - 2\psi(\lambda) + \lambda\psi''(\lambda+1)}{2\beta^2}, & & \\ & & \Gamma_{333} &= -\frac{\lambda \log^2 \eta + \lambda(-2 \log \eta \psi(\lambda+1) + \psi(\lambda+1)^2 + \psi'(\lambda+1)) + 1}{\beta^3}. \end{aligned} \tag{5.13}$$

5.2.1.3 La sous-variété gamma

Les sous-variétés $\beta = \text{cte}$ de la variété gamma généralisée sont toutes isométriques à la variété gamma. Cette section est consacrée à l'étude de leurs propriétés à l'aide des équations de Gauss-Codazzi. Dans cette section, la variété

gamma généralisée sera notée M tandis que $N_\kappa, \kappa > 0$ correspondra à la sous-variété plongée³ $\beta = \kappa$.

Proposition 12. *Le fibré normal N_κ est engendré au point (η, λ) de la sous-variété gamma par le vecteur :*

$$n(\eta, \lambda) = (-\eta(\lambda\psi'(\lambda)(\psi(\lambda+1) - \log(\eta)) + \log(\eta) - \psi(\lambda)), -1, \kappa(\lambda\psi'(\lambda) - 1))$$

Proposition 13. *Le terme $a_{12}^2 - a_{11}a_{22}$ est strictement positif.*

Proposition 14. *La courbure sectionnelle de la variété gamma généralisée dans (e_1, e_2) satisfait :*

$$\overline{K}(e_1, e_2) \xrightarrow{\lambda \rightarrow 0^+} \frac{12 - \pi^2}{2(\pi^2 - 6)}.$$

On conjecture que la courbure sectionnelle de la variété gamma généralisée dans les directions $\partial_\eta, \partial_\lambda$ est strictement positive, bornée supérieurement par 1/2 comme cela semble être le cas numériquement.

5.2.2 Application aux données de la maladie d'Alzheimer

L'imagerie par résonance magnétique (IRM) cherche à identifier, à localiser et à mesurer différentes parties de l'anatomie du système nerveux central. Elle s'est révélée être un incontournable marqueur des maladies neurodégénératives telles que la maladie d'Alzheimer, principale cause de démence (Vemuri and Jack, 2010; Cuingnet et al., 2011; Lama et al., 2017). En effet, l'atrophie cérébrale mesurée par IRM structurale a été proposée comme marqueur pour le diagnostic précoce de la maladie d'Alzheimer (Pini et al., 2016; Busovaca et al., 2016).

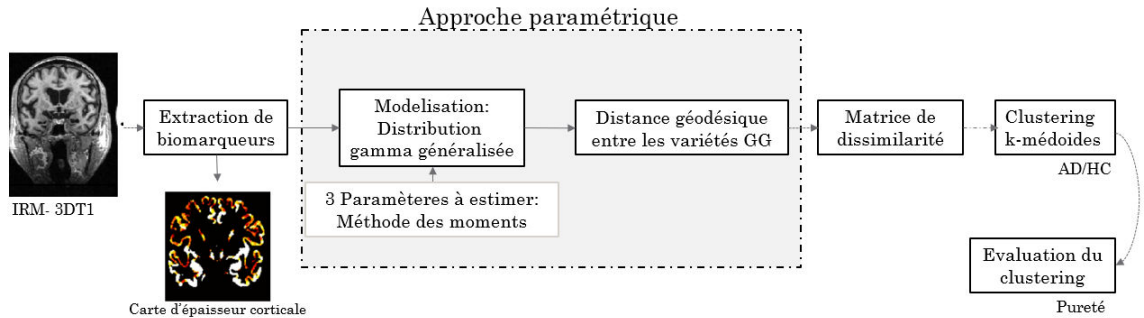


FIGURE 5.6 – Approche basée sur la géométrie de l'information dans le cadre de la classification de la population Alzheimer

Beaucoup de ces études se sont limitées à l'utilisation des mesures de tendance centrale telles que la moyenne ou la médiane. Les plus récentes ont utilisé l'analyse par histogramme (Ruiz et al., 2018; Giulietti et al., 2018) pour représenter un biomarqueur plutôt que d'utiliser la distribution du biomarqueur dans tout le cerveau ou dans des tissus spécifiques. Dans cette section, nous présentons l'une des applications possibles de la géométrie de l'information sur plusieurs distributions de probabilités et démontrons l'utilisation de ces distributions dans le contexte de la classification de la population Alzheimer. Le schéma 5.6 résume cette application.

3. "Embedded" en anglais.

5.2.2.1 Matériels et méthodes

Il s'agira de comparer 5 scénarios de classification de la population Alzheimer. Les trois premiers se basent sur l'approche classique étudiée dans la littérature, le quatrième sur les résumés d'histogrammes et le dernier présentera l'approche paramétrique basée sur la géométrie de l'information :

1. Classification à partir de l'épaisseur corticale moyenne.
2. Classification à partir de l'épaisseur corticale moyenne et 22 régions d'intérêt du cortex.
3. Classification à partir de l'épaisseur corticale moyenne et 22 régions d'intérêt du cortex en utilisant une méthode de sélection de caractéristiques afin d'extraire les zones les plus discriminantes.
4. Classification à partir des métriques des histogrammes d'épaisseur corticale.
5. Classification à partir de l'approche, décrite dans cette section, basée sur les distributions de probabilité et les outils de la géométrie de l'information.

5.2.2.1.1 La population d'étude

Les sujets de notre étude proviennent de la population d'étude basée sur les métriques d'histogrammes de la section 3.2.1.1 : Même base de données (ADNI1) et même processus de sélection des sujets (présenté dans la section 3.2.1.1.2). Toutefois cette étude réunie uniquement des patients atteints de la maladie d'Alzheimer (AD) et des sujets sains (HC) et non les patients MCI (cf. tableau 5.1).

	HC ($n = 71$)	AD ($n = 72$)	Valeur- p
Âge	76.1 ± 5.6	77.4 ± 5.5	0.17
Sexe (F/H)	38/33	41/31	0.20
MMSE	29 ± 0.9	23.2 ± 2.1	<0.001

Tableau 5.1 – Caractéristiques démographiques et cliniques de la population étudiée. Les valeurs \pm représentent des moyennes \pm écart type. Toutes les valeurs p sont basées sur le test ANOVA, à l'exception de Sexe, qui repose sur le test chi-2 ($\alpha < 0.05$). Abréviations : HC, sujets sains ; AD, patients atteints de la maladie d'Alzheimer ; MMSE, mini mental state examination.

5.2.2.1.2 Acquisitions IRM et prétraitement des images

Les paramètres des acquisitions des volumes IRM pondérées en T1, sont identiques à ceux décrits dans les parties 3.2.1.2. Les volumes 3D IRM pondérés en T1 ont été principalement acquis avec la séquence MPRAGE. La chaîne de prétraitement a été présentée dans la section 3.2.1.3.

5.2.2.1.3 Distribution d'épaisseur corticale

L'implication de la substance grise dans la maladie d'Alzheimer est étudiée indirectement par des mesures de l'atrophie de la substance grise. L'épaisseur corticale est un marqueur de référence dans la quantification de l'atrophie corticale. Nous avons ainsi choisi ce dernier dans le cadre de notre étude.

L'épaisseur corticale (EC) a été mesurée à partir d'une IRM pondérée en T1 à l'aide de la Toolbox Matlab CorThiZon, qui comprend les étapes de normalisation, de segmentation et de mesure d'épaisseur corticale. Les EC ont été calculés sur tout le ruban cortical à l'aide d'un algorithme basé sur l'équation de Laplace (Querbes et al., 2009).

Le détail de la méthode d'estimation de l'EC a été présentée dans la partie 3.2.1.4.1.

— **Estimation paramétrique**

Sur chaque vecteur d'épaisseurs corticales et pour chaque sujet, une densité gamma généralisée sous la forme (5.10) est estimée à l'aide de la méthode des moments décrite dans Stacy and Mihran (1965), donnant les estimations de trois paramètres gamma généralisés (α, λ, β) qui sont ensuite convertis en paramètres naturels. À la fin de la phase de traitement des données, nous avons un jeu de données de 143 densités gamma généralisées estimées représentant les distributions d'épaisseur corticale.

— **Validation statistique**

Pour chaque sujet, une fois que les paramètres (α, λ, β) des gamma généralisées représentant la distribution de probabilité d'épaisseur corticale ont été estimés, nous avons effectué un test d'homogénéité aussi appelé "goodness of fit". Nous avons choisi le test de Cramér-von Mises, le cas à deux échantillons (Anderson, 1962) qui est une approche non paramétrique permettant de tester si une variable continue X suit une loi de distribution fixée. Ce test étudie l'écart entre les deux fonctions de répartition que nous souhaitons comparer en se basant sur la somme des différences. L'hypothèse nulle H_0 est "les deux fonctions de répartition sont équivalentes", cette hypothèse est rejetée si $p \leq 0.05$.

5.2.2.1.4 Classification de la population Alzheimer

Le clustering, également appelé classification non supervisée, a fait l'objet d'études approfondies pendant des années dans de nombreux domaines, tels que l'exploration de données, la reconnaissance de formes, la segmentation d'images et la bioinformatique. Cette technique est principalement utilisée pour segmenter ou classer une base de données ou extraire des connaissances pour tenter d'identifier des sous-ensembles de données difficiles à distinguer. L'objectif est de regrouper les ensembles de données de manière à maximiser la similarité intra-cluster tout en minimisant la similarité inter-cluster. Il existe principalement trois catégories de classification dans la littérature : la classification par partition, la classification hiérarchique et la classification basée sur la densité.

La classification non supervisée a été choisie à contrario de la classification supervisée, car cette dernière consiste à prendre en compte le label des sujets (i.e. HC et AD), qui correspond au diagnostic des sujets, or dans la maladie d'Alzheimer le diagnostic se fait, entre autre, à l'aide des critères du NINCDS-ADRDA qui sont des tests cliniques et non des examens IRM. Ainsi, le clustering, contrairement à la classification supervisée, ne prendra en compte que les données d'imagerie.

L'approche k-médoïdes, comme tous les algorithmes de clustering, tente d'organiser les données en k clusters. Pour ce faire, la méthode comprend deux phases : la phase de construction ("building") et la phase de permutation

("swapping"). La phase de construction consiste à sélectionner au hasard les représentants k initiaux (i.e. les médoïdes). Les objets non sélectionnés sont affectés au représentant le plus similaire. Ensuite, lors de la phase de permutation, les médoïdes sont remplacés de manière itérative par des objets non représentatifs (voir l'algorithme 1).

Algorithm 1 Algorithme des k-médoïdes

1. **Initialisation** : Sélectionner aléatoirement k échantillons comme objets représentatifs initiaux (i.e. médoïdes)
 2. **Repeat**
 - i. Calculer une mesure de dissimilarité entre chaque médoïde m et les objets restants
 - ii. Affecter l'objet non représentatif o_i au médoïde le plus proche m
 - iii. Calculer le coût total δS de permuter le médoïde m avec o_i ; le coût total est défini comme étant la somme des carrés des résidus SSE du clustering résultant
 - iv. Si $\delta S < 0$, permuter m avec o_i pour former le nouvel ensemble de médoïdes.
 3. **Until**
Le critère de convergence est satisfait (i.e. aucun changement dans les médoïdes ou dans le coût total δS)
-

1. Clustering basé sur l'épaisseur corticale moyenne

L'épaisseur corticale a été mesurée sur le cortex entier et c'est cette mesure qui a été utilisée pour classer les sujets HC et les patients AD.

2. Clustering basé sur l'épaisseur corticale moyenne et sur 22 régions d'intérêt

Nous nous sommes aussi intéressés à l'épaisseur corticale dans une parcellisation du cerveau constituée par 22 regroupements d'aires de Brodmann⁴, décrits dans l'annexe III ou l'épaisseur corticale moyenne a été mesurée dans chacune de ces zones à l'aide de la toolbox Matlab CorThiZon (figure 5.7). Ainsi, l'algorithme des k-médoïdes classera AD et HC à partir de 23 variables : l'épaisseur corticale moyenne du cortex entier et les épaisseurs corticales dans 22 ROI.

3. Clustering basé sur l'épaisseur corticale moyenne et sur 22 régions d'intérêt, avec sélection de caractéristiques

Comme il a été décrit précédemment, il s'agit de classer les sujets HC et AD à partir de 23 variables (i.e. EC moyenne du cortex entier + 22 ROI). Toutefois, nous avons utilisé la méthode de sélection de caractéristiques qui permet de sélectionner les régions les plus discriminantes. Nous choisissons la mesure du gain d'information décrite dans la section 3.2.1.6.1 et la formule (3.12). Nous avons choisi les 10 variables avec les GI les plus élevées, présentées dans le

4. Les aires de Brodmann sont des délimitations du cortex du cerveau humain (Brodmann, 1909).

tableau 5.9.

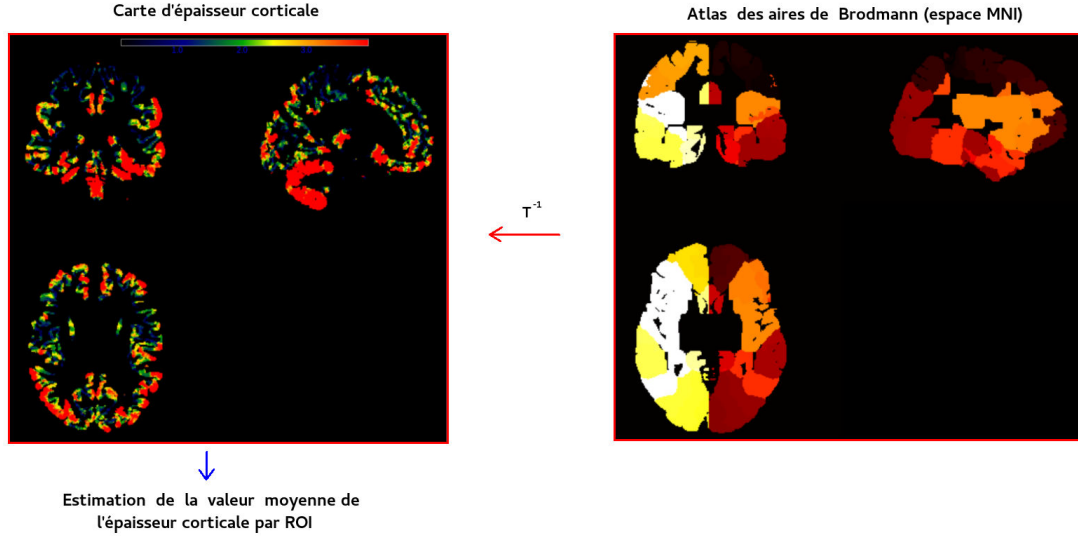


FIGURE 5.7 – Estimation de l'épaisseur corticale moyenne par ROI. L'IRM T1 native du sujet est normalisée sur le template T1 MNI obtenant la matrice de transformation affine T ; la transformation inverse T^{-1} est ensuite appliquée à l'atlas des régions de Brodmann afin de la recaler celui-ci sur la carte d'épaisseur corticale située dans l'espace natif du sujet. Une valeur moyenne de l'épaisseur corticale est ainsi calculée pour chaque ROI.

4. Clustering basé sur des métriques d'histogrammes

Cette approche a été décrite dans la section 3.1. En résumé, elle consiste à résumer les histogrammes d'images quantitatives à l'aide de descripteurs locaux tels que la moyenne, les quantiles, skewness etc... Ces métriques d'histogrammes sont utilisées comme variables dans un algorithme de classification (supervisé ou non) afin de discriminer des populations. Ici, il s'agira de produire des histogrammes d'épaisseur corticale (comme expliqué dans la section 3.1.1) puis d'en extraire des résumés : moyenne, médiane, 10ème quantile, 25ème quantile, 75ème quantile, 90ème quantile, amplitude et position du pic, skewness et kurtosis (pour plus de détails voir section 3.1.2).

5. Clustering basé sur les distributions de probabilité

L'algorithme populaire k -means s'applique aux échantillons vectoriels de \mathbb{R}^p et trouve une partition de l'ensemble de données original D en k sous-ensembles D_1, \dots, D_k qui minimise la variance intra-classe totale :

$$\sum_{i=1}^k \#D_i \text{Var } D_i \quad (5.15)$$

où $\#D_i$ indique le nombre d'éléments dans D_i . Il peut être formulé comme un problème de quantification vectorielle, c'est-à-dire trouver une séquence optimale c_1, \dots, c_k de vecteurs à partir de \mathbb{R}^p qui minimise :

$$\sum_{i=1}^k \sum_{x \in D_i} \|x - c_i\|^2 \quad (5.16)$$

où D_i est le sous-ensemble de points de l'ensemble de données situé dans la cellule de voronoï du centre c_i . Il s'est avéré que c'était NP-difficile⁵ (Garey et al., 1982), même avec seulement deux classes. Les algorithmes existants ne rechercheront donc que des solutions localement optimales. Un choix courant est l'algorithme de Lloyd (1982), qui est une procédure de minimisation locale basée sur un gradient. Il peut être étendu au cas riemannien (Brigant and Puechmorel, 2019), mais nécessite à chaque itération le calcul des géodésiques entre un échantillon et les centres de cellules. Dans notre étude, les expériences ont été conduites en utilisant un algorithme de partitionnement légèrement différent, les k -mediods (Kaufman and Rousseeuw, 1987). Comparé à l'algorithme de Lloyd, il est généralement considéré comme plus lent, mais nécessite uniquement une mesure de dissimilarité entre des paires de points de l'ensemble de données au lieu d'une distance réelle. Il est plus adapté aux petits échantillons, car les itérations sont effectuées sans qu'il soit nécessaire de recalculer les distances. Dans le contexte du clustering des variétés riemanniennes, il s'agit d'un avantage remarquable, car les calculs géodésiques sont coûteux.

La matrice de dissimilarité utilisée dans l'algorithme k -médoides a été calculée à l'aide des mesures de similarité suivantes :

- Distance géodésique sur la variété gamma généralisée (DGG1).
- Distance géodésique approximative sur la variété gamma généralisée (DGG2).
- Distance géodésique sur la variété gamma (DG).
- Valeur absolue entre moyennes empiriques (DM).
- Divergence de Kullback-Leibler pour les distributions gamma généralisées (KL).

Les distances (DGG1) et (DG) sont calculées entre deux points x, y de la variété respective en résolvant l'équation géodésique :

$$\frac{d^2}{dt^2}\gamma^k(t) + \Gamma_{ij}^k \frac{d}{dt}\gamma^i(t) \frac{d}{dt}\gamma^j(t) = 0 \quad (5.17)$$

$$\gamma(0) = x \quad (5.18)$$

$$\gamma(1) = y \quad (5.19)$$

Une méthode appelée "shooting method" a été sélectionnée pour résoudre ce problème aux limites. Il y a convergence dans tous les cas pour (DG) mais ne parvient pas à converger sur dix paires pour (DGG1). La distance géodésique approximative (DGG2) a été définie pour contourner ce problème. Elle est basée sur l'observation que la variété gamma est isométriquement plongée dans la variété gamma généralisée lorsque β est constant. Il y a donc une submersion riemannienne π définie dans les coordonnées par $(\eta, \lambda, \beta) \mapsto (\eta, \lambda)$. On peut alors obtenir une distance approximative en considérant séparément les parties verticales et horizontales. Soit $p(\eta_2, \lambda_2, \beta_2), p(\eta_1, \lambda_1, \beta_1)$ deux densités gamma généralisées. L'énergie E_1 du

5. "NP" signifie "non déterministe polynomial". C'est un problème vers lequel on peut ramener tout problème de la classe NP par une réduction polynomiale.

chemin vertical $t \in [0, 1] \mapsto \gamma_\beta(t) = (\eta_1, \lambda_1, (1-t)\beta_1 + t\beta_2)$ est calculé à l'aide de la formule :

$$E_1 = \int_0^1 g\left(\gamma_\beta(t); \frac{d}{dt}\gamma_\beta(t), \frac{d}{dt}\gamma_\beta(t)\right) dt = (\beta_2 - \beta_1)^2 \int_0^1 g_{\beta\beta}(\gamma_\beta(t)) dt$$

Ensuite, l'énergie E_2 de la géodésique rejoignant $p(\eta_1, \lambda_1, \beta_2)$ et $p(\eta_2, \lambda_2, \beta_2)$ est calculée sur la sous-variété gamma comme la borne inférieure des intégrales :

$$\int_0^1 g\left(\gamma(t); \frac{d}{dt}\gamma(t), \frac{d}{dt}\gamma(t)\right) dt$$

où γ est un chemin lisse reliant les deux points précédents avec β constante égale à β_2 . La distance approximative globale est alors supposée être $\sqrt{E_1 + E_2}$. On remarque qu'il ne s'agit que d'une mesure de similarité, telle que la divergence de Kullback-Liebler, car elle n'est pas symétrique.

5.2.2.2 Résultats

5.2.2.2.1 Test d'homogénéité

Le test de Cramér-von Mises à deux échantillons a été utilisé pour tester, pour chaque sujet, l'adéquation du modèle de gamma généralisée à la distribution empirique de l'épaisseur corticale. 88% des distributions correspondent bien à des gamma généralisées ($p > 0.05$).

5.2.2.2.2 Évaluation de la classification

Dans notre cas, le but était d'évaluer avec quelle précision notre approche classerait les patients atteints de la maladie d'Alzheimer et les sujets sains. Ainsi, nous avons choisi $k = 2$ comme nombre de classes dans l'algorithme k-médoides, un cluster représente les patients atteints de MA et l'autre les sujets HC. Ces clusters sont comparés aux vraies étiquettes des sujets à l'aide de la mesure de pureté qui est une mesure d'évaluation externe. Les mesures d'évaluation externes permettent de comparer le regroupement proposé par le clustering aux étiquettes réelles des objets. La mesure pureté a été détaillée dans (3.16). Les résultats pour les méthodes classiques sont présentés dans le tableau 5.2 et les résultats du clustering basé sur les distributions sont représentés dans le tableau 5.3. Notez que comme l'algorithme k-médoides a une initialisation aléatoire, les valeurs indiquées dans le tableau 5.3 ont été calculées comme étant la moyenne de la pureté sur 100 exécutions.

Pour le clustering basé sur les distributions, nous avons obtenu **Pureté=0.84**, ce qui signifie que les deux classes de l'algorithme k-médoides basé sur les distributions correspondent à 84% des étiquettes réelles suivant les distances géodésiques. La variété gamma généralisée discrimine mieux les deux groupes de sujet comparée à la variété gamma. En utilisant les méthodes classiques, basées sur la valeur moyenne de l'épaisseur corticale dans le cortex entier ou dans des régions d'intérêt, la pureté est moindre. On obtient un maximum de 80% pour la classification basée sur les ROI, après avoir utilisée une méthode de sélection des régions les plus discriminantes. Quant à la méthode basée sur des résumés d'histogrammes une précision de 81% est obtenue.

	Moyenne	Analyse par ROI		Analyse d'histogrammes
	EC_moy	EC_moy + 22 ROI	EC_moy + 22 ROI (feature selection)	Métriques d'histogrammes
Pureté (%)	79.1	77.7	80.4	81

Tableau 5.2 – Résultat du clustering k-médoides AD/HC suivant 4 scénarios classiques.

	Clustering basé sur les distributions				
	DGG1	DGG2	DG	DM	KL
Pureté (%)	84	84	78	80	78

Tableau 5.3 – Résultat du clustering k-medoids basé sur les distributions suivant différents modèles et mesures de dissimilarité

5.2.2.3 Discussion

Dans cette étude nous avons comparé cinq scénarios afin de classer la population Alzheimer. Nous avons comparé une classification basée sur les distributions de probabilité d'un biomarqueur IRM à d'autres approches déjà utilisées précédemment, basées sur la mesure de la valeur moyenne d'un biomarqueur dans un tissu entier ou dans des régions d'intérêt ou plus récemment basées sur des métriques d'histogrammes. Les résultats mettent en avant le pouvoir discriminant de la distribution de probabilité en comparaison aux analyses par régions d'intérêt ou d'histogrammes.

Notre étude comprend une limite qui est celle de la simplicité de la problématique clinique (i.e. classification des patients atteints de MA et les sujets sains). Toutefois, ce fut un choix. Nous avons besoin d'une application simple avec une problématique déjà résolue afin de mettre en avant notre méthodologie. Cette étude présente ainsi une preuve de concept à l'utilisation des distributions et la géométrie non-euclidienne.

L'algorithme k-médoides basé sur la distribution permet d'obtenir une classification précise de la population atteinte de la maladie d'Alzheimer (Pureté=84%). En effet, la géométrie de l'information offre des outils appropriés permettant d'utiliser correctement la distribution de probabilité, ce qui augmente considérablement les performances de la classification de la maladie par rapport aux approches classiques (Cuingnet et al., 2011).

5.3 Application 2 : caractérisation des retards dans le transport aérien

5.3.1 Contexte

Le retard est l'un des facteurs les plus influents dans l'expérience de voyage, quel que soit le moyen de transport utilisé. Dans le système de transport aérien (ATS), les retards sont souvent associés à des expériences très négatives, en particulier lorsqu'un vol en correspondance est manqué en raison d'une arrivée tardive. La manière de faire face aux retards est basée sur la gestion du flux de trafic aérien et le suivi des vols en fonction des opérations planifiées, puis révisé

de manière continue (Niarchakou and Cech, 2018). Pour ce faire, le gestionnaire de réseau doit avoir la meilleure prédiction possible de la situation lors de l'opération. L'un des principaux problèmes de cette prévision consiste à anticiper la manière dont un retard donné peut se propager dans l'ensemble du réseau.

Dans le système de transport aérien, les aéroports sont reliés par des compagnies exploitant des vols de différentes manières. L'une consiste à utiliser un système de navette, avec un grand nombre de vols quotidiens réguliers. Dans un tel cas, un retard survenant sera lissé en utilisant l'effet tampon des vols restants en rotation. D'autre part, les entreprises opérant une seule rotation ne pourront atténuer le retard en utilisant la durée de rotation. C'est la même chose pour les vols de transit. La manière dont un retard pourrait être compensé dépendra de nombreux facteurs difficiles à anticiper (Xu et al., 2008). Même la motivation de la compagnie aérienne à prendre une décision dépend de divers éléments tels que le coût, l'efficacité en termes de temps, la disponibilité des avions, la durée de rotation et le coût dans un aéroport donné, la disponibilité de l'équipage...

Cette étude a été réalisée en collaboration avec Stéphane Puechmorel (chercheur à l'École Nationale de l'Aviation Civile) et des stagiaires ISESA2018 de l'École Nationale de l'Aviation civile, Alexandre Bevigiani, Elie Bonvalot et Laurent Zeggwagh.

5.3.2 Conception mathématique

5.3.2.1 Une vision probabiliste des retards

Les retards ont de nombreuses origines possibles et sont la plupart du temps le résultat d'une chaîne d'événements. Une caractérisation approfondie nécessitera beaucoup de travail, si cela est réalisable, et peut ne pas être très informative en raison de sa complexité. Une vue globale des retards est beaucoup plus pratique et fournira un indicateur de performance essentiel. Une approche grossière consisterait à estimer la moyenne et l'écart-type de la distribution des retards, même s'il peut être intéressant de disposer d'une vue globale de la performance, cette dernière ne caractérise pas le degré d'interdépendance. Une approche paramétrique basée sur un modèle statistique est plus adaptée à la caractérisation des retards. Dans ce cadre, les échantillons sont supposés provenir d'une loi de probabilité sous-jacente appartenant à une famille de distribution donnée en fonction d'un nombre fini de paramètres. On peut ensuite obtenir un modèle des retards en trouvant l'ensemble optimal de paramètres par rapport aux observations. Le principal problème consiste à identifier une famille de distributions qui représentera au mieux la vraie loi de probabilité sous-jacente, tout en ayant un nombre minimal de paramètres à estimer. Dans la suite, les retards sont supposés être positifs, même si des retards négatifs sont présents dans les données opérationnelles. Ces valeurs ont été écartés dans la présente étude.

5.3.2.2 La distribution gamma

D'un point de vue probabiliste, les retards peuvent être modélisés comme les réalisations d'une variable aléatoire X prenant en charge les nombres réels

positifs. Lorsque les retards sont indépendants, X a la propriété dite "perte de mémoire".

Définition 1. Une variable aléatoire X sur \mathbb{R}^+ est sans mémoire si pour tout couple $(t, s) \in \mathbb{R}^{+2}$:

$$P(X > t + s | X > t) = P(X > s)$$

Dans le contexte des retards, la propriété de perte de mémoire est liée à l'indépendance, ou au pur hasard. La proposition suivante est bien connue (Ross, 2009) et montre que les variables aléatoires sans mémoire sont entièrement caractérisées.

Proposition 1. Si X est sans mémoire alors elle suit une loi exponentielle :

$$\forall t \in \mathbb{R}^+, P(X < t) = 1 - \exp(-\lambda t)$$

Les distributions exponentielles sont donc de bons candidats pour la caractérisation des retards non corrélés. Dans un tel cas, seul le paramètre taux λ doit être estimé. Cela peut être fait en utilisant une estimation par maximum de vraisemblance, donnant un estimateur simple $\hat{\lambda}$ basé sur la moyenne empirique :

$$\hat{\lambda}^{-1} = n^{-1} \sum_{i=1}^n t_i \quad (5.20)$$

avec $(t_i)_{i=1 \dots n}$ les échantillons.

La densité exponentielle est non nulle à $t = 0$ et décroît rapidement lorsque t tend vers l'infini. C'est un modèle largement utilisé pour les temps de service dans la théorie des files d'attente (Gross et al., 2011), et il est bien adapté pour décrire une source de retard élémentaire dans une chaîne d'événements.

Dans le contexte du système de transport aérien, il est peu probable que l'on observe des retards indépendants, mais plutôt la somme de ces retards, du fait que les vols traversent une succession d'états, chacun d'entre eux étant sujet à un certain retard. Un modèle simple pour décrire une telle situation est donc une somme de distributions exponentielles indépendantes. Lorsque les taux sont distincts par paires, la proposition suivante prouve que la distribution résultante est un mélange de distributions exponentielles.

Proposition 2. Soit $(X_i)_{i=1}^n$ variables aléatoires indépendantes distribuées de manière exponentielle avec des taux respectifs $\lambda_i, i = 1 \dots n$. Si $\lambda_i \neq \lambda_j$ pour $i \neq j$, alors la densité de la variable aléatoire $S = \sum_{i=1}^n X_i$ est donnée par :

$$p_S(t) = \sum_{i=1}^n \frac{\prod_{j=1}^n \lambda_j}{\prod_{j \neq i} (\lambda_j - \lambda_i)} e^{-\lambda_i t}$$

Lorsque deux taux sont égaux, ce qui précède n'est plus valable. Au lieu de cela, on a la proposition suivante :

Proposition 3. soient $(X_i)_{i=1}^n$ des variables aléatoires indépendantes distribuées de manière exponentielle avec le même taux λ . La densité de la variable aléatoire $S = \sum_{i=1}^n X_i$ est donné par :

$$P_S(t) = \frac{\lambda^n t^{n-1}}{(n-1)!} e^{-\lambda t}$$

Cette distribution est connue sous le nom de distribution d'Erlang. C'est un cas particulier de la distribution gamma :

Définition 2. La distribution gamma avec les paramètres $\alpha > 0$ et $\beta > 0$ est définie comme suit :

$$\Gamma_{\alpha,\beta}(t) = \frac{\beta^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\beta t}$$

Le paramètre α est appelé la forme de la distribution et le paramètre β est appelé le taux. La distribution Erlang est un cas particulier de la distribution gamma lorsque le paramètre de forme est un entier. Une distribution exponentielle est une distribution gamma de forme 1.

Proposition 4. La fonction caractéristique d'une variable aléatoire X avec distribution $\Gamma_{\alpha,\beta}$ is :

$$E[e^{i\xi X}] = \left(1 - \frac{i\xi}{\beta}\right)^{-\alpha}$$

On remarque que la proposition 3 découle directement de la proposition 4.

En réunissant les propositions 2 et 3, il apparaît que la forme la plus générale de distribution de retards résultant d'une somme de retards répartis de manière exponentielle indépendante est un mélange de distributions gamma.

5.3.2.3 Un modèle de mélange pour les retards

Sur la base de la discussion ci-dessus, un modèle statistique approprié pour les retards consistera à supposer une distribution paramétrique de la forme :

$$p(t; (\alpha_i)_{i=1}^q, (\beta_i)_{i=1}^q, (\mu_i)_{i=1}^q) = \sum_{i=1}^q \mu_i \Gamma_{\alpha_i, \beta_i}(t) \quad (5.21)$$

avec :

$$\sum_{i=1}^q \mu_i = 1$$

Proposition 5. Les estimateurs de maximum de vraisemblance $\hat{\alpha}, \hat{\beta}$ pour les paramètres de forme et taux sur les échantillons $(t_i)_{i=1}^N$ sont des solutions du système d'équations :

$$\begin{cases} \log(\hat{\alpha}) - \psi(\hat{\alpha}) = \log \bar{t}_i - \overline{\log t_i} \\ \hat{\beta} = \hat{\alpha} / \bar{t}_i \end{cases} \quad (5.22)$$

où \bar{x}_i désigne la moyenne empirique :

$$\bar{x}_i = \frac{1}{N} \sum_{i=1}^N x_i$$

et ψ est la fonction digamma.

Le système (5.22) est un système d'équations non linéaire impliquant la fonction spéciale ψ qui n'est pas simple à résoudre. Un algorithme itératif, basé sur l'itération de Newton, est proposé dans (Minka, 2002) et est assez efficace en pratique. On remarque que, bien que la distribution exponentielle soit un cas particulier d'une distribution gamma, l'estimateur ci-dessus ne peut pas être utilisé directement car $\log(t_i)$ tendra vers $-\infty$ si l'un des t_i est nul. Il est donc conseillé d'utiliser explicitement une loi exponentielle dans le modèle de mélange (5.21) en laissant $\alpha_1 = 1$.

Pour estimer les paramètres dans (5.21), une approche standard consiste à utiliser l'algorithme EM (Dempster et al., 1977). Les paramètres à déterminer correspondent aux paramètres α_i et β_i de chaque loi qui constitue le mélange ainsi que leur poids respectifs. L'algorithme le plus fréquemment utilisé pour la détermination d'une densité de mélange est l'algorithme d'Espérance-Maximisation (EM). Celui-ci permet de trouver de manière itérative les paramètres permettant de maximiser la vraisemblance à partir d'un nombre d'observations, en ne connaissant pas toutes les informations relatives à ces observations, par exemple, l'influence ou l'appartenance d'une composante sur une observation donnée.

L'algorithme EM procède en deux étapes :

- L'étape "Espérance" (étape E) : le calcul, à chaque itération, des influences des composantes du mélange qui sont des données nécessaires pour la maximisation de la vraisemblance. En supposant que l'algorithme est à l'itération k , les probabilités conditionnelles $P(Z = i|t_j, \alpha, \beta)$ sont estimées comme suit :

$$a_{ij}^{(k)} = P(Z = i|t_j, \alpha^{(k)}, \beta^{(k)}) = \frac{\mu_i^{(k)} \Gamma_{\alpha_i^{(k)}, \beta_i^{(k)}}(t)}{\sum_{j=1}^q \mu_j^{(k)} \Gamma_{\alpha_j^{(k)}, \beta_j^{(k)}}(t)} \quad (5.23)$$

et permettent de fournir une estimation de $\mu_i^{(k+1)}$:

$$\mu_i^{(k+1)} = \frac{1}{N} \sum_{j=1}^N a_{ij}^{(k)} \quad (5.24)$$

- L'étape "Maximisation" (étape M) : correspond à la maximisation de la vraisemblance rendue possible à la suite du calcul des influences réalisé lors de l'étape E. Les paramètres respectifs de chaque loi ainsi que les poids sont mis à jour. Les valeurs de $\alpha_i^{(k+1)}, \beta_i^{(k+1)}$ maximisent la vraisemblance conditionnelle :

$$l(\alpha, \beta) = \sum_{i=1}^q \sum_{j=1}^N a_{ij}^{(k)} \quad (5.25)$$

Dans le cas de la distribution Gamma, l'étape de maximisation implique la résolution d'une équation non linéaire dans α , qui est une version pondérée de la première ligne du système (5.20). Étant donné que cela augmente considérablement la quantité de calcul nécessaire, il est conseillé d'utiliser une approche différente et de remplacer la maximisation exacte dans α . C'est l'idée principale de l'algorithme présenté dans (Almhana et al., 2006a), qui remplace le

M par :

$$\beta_i^{k+1} = \alpha_i^{(k)} \frac{\sum_{j=1}^N a_{ij}^{(k)}}{\sum_{j=1}^N t_j a_{ij}^{(k)}} \quad (5.26)$$

$$\alpha_i^{(k+1)} = \alpha_i^{(k)} + k^{-1} H_i^{(k)} \quad (5.27)$$

$$H_i^{(k)} = N^{-1} \sum_{j=1}^N \left(\log t_j + \log \beta_i^{(k)} - \psi(\alpha_i^{(k)}) \right) a_{ij}^{(k)} \quad (5.28)$$

On remarque que le facteur k^{-1} assure la convergence de l'algorithme mais peut être remplacé par toute série divergente de terme positif s_k .

Les paramètres de mélange sont mis à jour selon la méthode présentée dans (Almhana et al., 2006b). À chaque itération, les valeurs sont recalculées et mises à jour :

$$\mu_i^{k+1} = \frac{\sum_{j=1}^N t_j a_{ij}^{(k)}}{\sum_{j=1}^N a_{ij}^{(k)}} \quad (5.29)$$

$$\alpha_i^{(k+1)} = \alpha_i^{(k)} + k^{-1} G_i^{(k)} \quad (5.30)$$

$$G_i^{(k)} = N^{-1} \sum_{j=1}^N \frac{\log t_j - \log \mu_i^{(k)} + \log(\alpha_i^{(k)} - \psi(\alpha_i^{(k)}))}{\psi'(\alpha_i^{(k)}) - \frac{1}{\alpha_i^{(k)}}} a_{ij}^{(k)} \quad (5.31)$$

5.3.2.4 Distance géodésique entre les modèles

5.3.2.4.1 Contexte

Pour un nombre déterminé de composants, le modèle de mélange est entièrement caractérisé par les proportions du mélange et les paramètres des distributions gamma. Pour pouvoir comparer deux de ces mélanges, il faut introduire une distance sur l'ensemble des mélanges. C'est un problème notoirement difficile, abordé par plusieurs auteurs. Il est possible d'utiliser une approche géométrique, en utilisant l'information de Fisher comme métrique riemannienne sur une multitude de modèles statistiques, mais c'est presque impossible du point de vue du calcul. Une référence récente (Chen et al., 2017) utilise une distance de Wasserstein pour un mélange de gaussiennes, les composants du mélange étant supposés provenir d'une loi de probabilité discrète, les distributions gaussiennes étant dirigées sur un sous-variété de l'espace des densités. Cette vue originale permettant de trouver une distance entre les mélanges peut être étendue au cas des mélanges gamma utilisant la géométrie de l'information.

5.3.2.4.2 Calcul de la distance géodésique

Une géodésique correspond au chemin le plus court entre deux points d'un espace. Afin de caractériser la différence entre les types de retard récoltés, par exemple selon différents critères tels que la plateforme aéroportuaire ou la compagnie aérienne, nous calculerons la distance géodésique séparant deux lois.

Afin de caractériser la différence entre les types de retard récoltés, nous calculerons la distance géodésique séparant deux lois.

1. Mise en équation d'une géodésique entre deux points

Une loi gamma se caractérisant par les deux paramètres α et β , nous nous plaçons dans une variété riemannienne de dimension 2, soit une variété différentielle munie d'une métrique riemannienne.

Soit $\gamma(t) = \begin{pmatrix} \gamma^\beta(t) \\ \gamma^\alpha(t) \end{pmatrix}$ une courbe reliant deux lois gamma. La géodésique correspond à la courbe qui minimise la longueur

$$L(\gamma) = \int_0^1 \sqrt{g(\gamma'(t), \gamma'(t))} dt.$$

La métrique utilisée ici pour calculer les distances sera l'information de Fisher. Celle-ci peut se mettre sous la forme suivante :

$$g_{i,j} = -E \left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} \log(f(X; \vec{\theta})) \right]$$

Le vecteur $\vec{\theta}$ est de dimension 2 et ses coordonnées correspondent au couple (α, β) . L'information de Fisher est ainsi une matrice de dimension 2×2 . Pour le paramétrage suivant de la loi gamma : $f(x, \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} e^{-\beta x} x^{\alpha-1}$, la métrique à un point (α, β) donné de la variété s'écrit :

$$[g_{i,j}] = \begin{bmatrix} \frac{\alpha}{\beta^2} & \frac{-1}{\beta} \\ \frac{-1}{\beta} & \Psi(\alpha) \end{bmatrix}$$

$\Psi(\alpha)$ est la fonction digamma : $\Psi(\alpha) = \frac{d}{d\alpha} \log(\Gamma(\alpha))$. Le système d'équations des géodésiques permet de calculer l'arc paramétré qui définit le plus court chemin entre deux lois gamma. Ces équations peuvent être déterminées à partir des symboles de Christoffel (proposition 2). Ces derniers peuvent directement être calculés à partir de la métrique.

En utilisant ces symboles, le problème se ramène à l'équation suivante :

$$\gamma''^k(t) + \Gamma_{i,j}^k \gamma'^i(t) \gamma'^j(t) = 0$$

où $\Gamma_{i,j}^k$ représente un symbole de Christoffel. Dans notre cas, pour une loi $\Gamma(\alpha, \beta)$ on aura donc :

$$\gamma''^\alpha(t) + \Gamma_{\alpha,\alpha}^\alpha \cdot \gamma'^\alpha(t)^2 + 2\Gamma_{\alpha,\beta}^\alpha \cdot \gamma'^\alpha(t) \cdot \gamma'^\beta(t) + \Gamma_{\beta,\beta}^\alpha \gamma'^\beta(t)^2 \quad (5.32)$$

$$\gamma''^\beta(t) + \Gamma_{\alpha,\alpha}^\beta \cdot \gamma'^\alpha(t)^2 + 2\Gamma_{\alpha,\beta}^\beta \cdot \gamma'^\alpha(t) \cdot \gamma'^\beta(t) + \Gamma_{\beta,\beta}^\beta \gamma'^\beta(t)^2 \quad (5.33)$$

2. Calcul de la distance géodésique entre deux lois gamma mélangées

La distance géodésique pour les lois gamma mélangées est la distance de Wasserstein calculée à partir de lois discrètes ayant pour probabilité respectives leur poids associés. En considérant deux mélanges de taille n et m , la fonction de coût à optimiser est une fonction de $n \times m$ variables, qui s'écrit sous la forme suivante :

$$C(p_{11}, \dots, p_{nm}) = \sum_{i=1}^n \sum_{j=1}^m p_{ij} d_{ij} \quad (5.34)$$

Les variables sont les p_{ij} les d_{ij} , elles correspondent à la distance géodésique entre la loi i du mélange d'origine et la loi j du mélange destination qui sont désormais calculables.

L'optimisation de cette fonction se fait sous les contraintes suivantes :

$$\sum_{i=1}^n = w_j \quad \text{pour tout } j$$

et

$$\sum_{j=1}^m = z_i \quad \text{pour tout } i$$

Dans ces équations les w_j et z_i représentent respectivement les poids du mélange d'origine ainsi que les poids du mélange destination. Nous devons rajouter à cela les contraintes de positivité des variables de la fonction de coût :

$$p_{ij} > 0 \quad \text{pour tout } i \text{ et } j$$

Nous nous retrouvons donc avec un problème d'optimisation d'une fonction linéaire sous contraintes linéaires. Nous avons choisi la méthode du Simplexe qui permet d'optimiser une fonction linéaire à partir de contraintes linéaires. La bibliothèque Apache Common Math sur Java nous fournit cette méthode.

5.3.3 Application : clustering des retards aéroportuaires

Dans la partie précédente, les différents concepts mathématiques permettant de modéliser les retards ont été présentés. Dans cette partie, nous nous sommes intéressés à la récupération, au stockage et au clustering des données de retards.

5.3.3.1 Fichiers DDR2 Eurocontrol

Les fichiers "Demand Data Repository" (DDR) fournis par Eurocontrol⁶ ont pour objectif de fournir une vision précise du trafic aérien passé, actuel et futur en Europe. Ils sont destinés aux principaux prestataires de services de la navigation aérienne et aux projets de recherche sur le management du trafic aérien. Ces fichiers contiennent l'intégralité des vols commerciaux traversant l'espace aérien européen. Les échantillons de vols décrivent entre autres les caractéristiques temporelles du vol, son aéroport de départ et d'arrivée, l'opérateur du vol, le type d'aéronef utilisé ou encore le nombre de passagers. Historiquement déclinés en deux versions, seuls les fichiers DDR2 contenant des informations supplémentaires par rapport aux fichiers DDR1 seront utilisés. Les données de retard que nous allons récolter sont mises à disposition par Eurocontrol par le biais de fichiers DDR2. Ces fichiers contiennent un ensemble d'informations relatives aux vols telles que le plan de vol, l'opérateur, les temps d'arrivée et de départ.

Nous avons à notre disposition 7 fichiers DDR2 reflétant le trafic européen de la première semaine de septembre 2018. Au total, 170 209 vols sont récoltés et stockés en base de données. Environ 1000 données de retard sont nécessaires pour

6. <https://www.eurocontrol.int/ddr>

pouvoir réaliser une étude statistique. Quarante terrains européens et trente huit compagnies aérienne possèdent un trafic de plus de 1000 vols sur cette période et seront exploitables. En effet, un nombre important de données est nécessaire pour avoir des résultats statistiques valides.

5.3.3.2 Extraction de données

On souhaite extraire des fichiers DDR2 un ensemble d'informations nécessaires au projet concernant les vols.

Pour chacun des vols, les informations que l'on extrait sont les suivantes :

1. **Aéroport de départ** : identifiant ICAO⁷ de l'aéroport d'où le vol décolle.
2. **Aéroport d'arrivée** : identifiant ICAO de l'aéroport d'où le vol atterrit.
3. **Identification de l'avion** : combinaison de l'identifiant ICAO de la compagnie aérienne opérant le vol et d'un numéro de vol
4. **Identification de l'opérateur de l'avion** : identifiant ICAO de la compagnie aérienne opérant le vol
5. **Temps réel de départ de l'avion de l'aire de stationnement**⁸ : date et heure à laquelle l'avion quitte sa position de parking
6. **Temps initial de départ de l'avion de l'aire de stationnement** : date et heure à laquelle l'avion devait initialement quitter sa position de parking
7. **"Tactical System Flight Identification"** : identifiant du vol correspondant au "Enhanced Tactical Flow Management System" d'Eurocontrol

Les deux champs "Off-Block Time" (5 et 6) permettent de calculer le retard d'un vol. En effet, la différence de ces deux champs fournit le retard au départ d'un vol. Les autres champs choisis seront utilisés comme critères de filtrage afin de faire une sélection parmi la liste de tous les vols. On pourra ainsi choisir la liste des vols pour lesquels on estime une loi statistique décrivant le retard au départ.

5.3.3.3 Clustering : Regroupement des terrains

Une fois les données filtrées, les paramètres des lois gamma estimés à l'aide de la méthode EM (obtenant ainsi pour chaque composante gamma 3 paramètres : le poids, α et β) comme expliqué dans la section 5.3.2.3, une des possibilités d'exploitation des résultats est de classer les lois statistiques en groupe ayant des caractéristiques proches.

Par exemple, après avoir estimé la loi statistique décrivant le retard au départ à partir d'une série d'aéroports, si l'on arrive à regrouper les lois semblables en groupe, on obtiendra alors des ensembles d'aéroports ayant des caractéristiques sur les retards proches. Cela est valable pour une multitude d'autres critères. On pourra ainsi comparer les retards selon les compagnies aériennes et les ranger par type de retards similaires. On pourra alors vérifier si le type de compagnie aérienne (bas coûts, compagnie régulière etc...) a une influence sur la caractérisation du retard.

7. Codes à trois lettres, attribués par l'Organisation de l'aviation civile internationale aux compagnies aériennes du monde entier.

8. Le temps de départ de l'avion de l'aire de stationnement est aussi appelé "Off-Block Time".

On va donc chercher à classer les différentes lois statistiques par groupe selon leurs similitudes en prenant en compte comme critère de similitude la distance géodésique entre ces lois.

Pour cela on va appliquer l'algorithme des "k-médoïdes" (voir algorithme 1) et plus précisément sa réalisation par la méthode PAM (Partition Around Medoids). Cet algorithme, à partir d'une valeur "k" fixée, regroupe un ensemble d'éléments en k groupes distincts autour d'un élément central appelé médoïde de manière à minimiser l'erreur quadratique moyenne (qui est la somme de la distance entre les différents points du cluster et le point central).

A partir d'une matrice contenant les distances entre les différentes lois gamma et le nombre de cluster souhaité, il permet de spécifier l'appartenance de chaque loi gamma à un cluster.

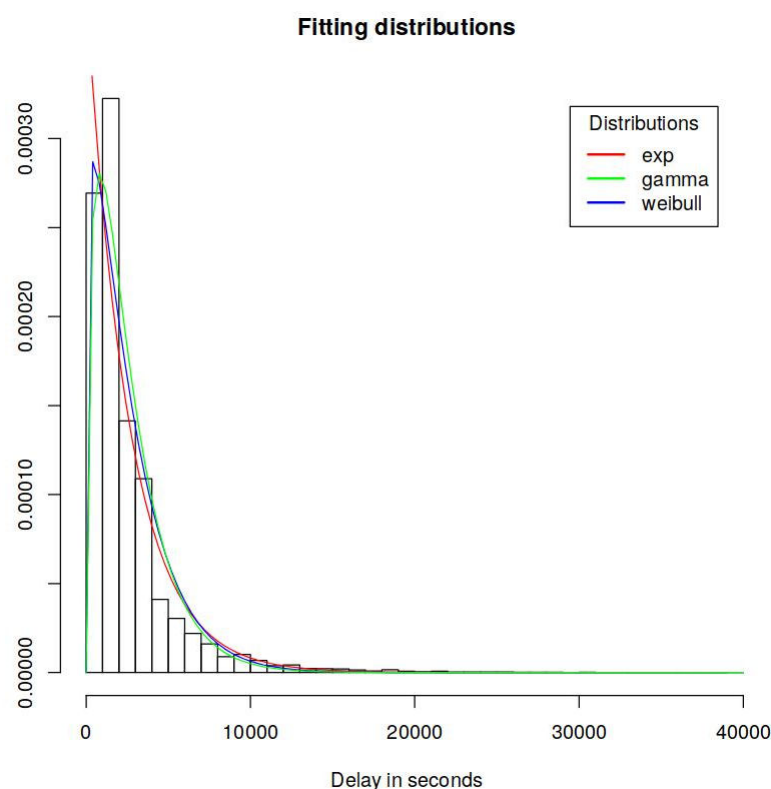


FIGURE 5.8 – Distribution des retards de départ d'un aéroport selon trois modèles : Gamma, Weibull et exponentiel

5.3.3.4 Résultats

Nous avons 40 aéroports pour la région Europe modélisables par une loi gamma (figure 5.8). Dans un premier temps, nous avons fait le choix d'utiliser une loi gamma simple pour leur caractérisation et un regroupement en trois groupes. L'algorithme k-médoïdes étant stochastique, il peut donner des résultats différents. Dans le cas présent nous convergions vers deux résultats différents après une série de huit tests. Les résultats sont présentés sur les tableaux 5.4 et 5.5. Les centres des clusters sont représentés en gras.

On constate une certaine stabilité dans le regroupement des aéroports. En effet seuls quatre aéroports changent de cluster. Les aéroports changeant de cluster d'un test à l'autre sont représentés en rouge.

Cluster 1	Cluster 2	Cluster 3
EHAM (Amsterdam-Schiphol)	EBBR (Bruxelles-National)	EGSS (Londres- Stansted)
LEPA (Palma de Majorque)	LTBA (Istanbul Atatürk)	LPPT (Lisbonne)
LLBG (Tel Aviv)	LFPO (Paris-Orly)	EGKK (Londres-Gatwick)
LEBL (Barcelone-El Prat)	LSGG (Genève)	LEMG (Malaga-Costa del Sol)
LFPG (Paris-CDG)	LIRF (Rome Fiumicino)	EFHK (Helsinki-Vantaa)
EDDF (Francfort)	EDDM (Munich)	ESSA (Stockholm-Arlanda)
LKPR (Prague-Václav-Havel)	LTFJ (Istanbul Sabiha Gökçen)	LEMD (Madrid-Barajas)
EGLL (Londres-Heathrow)	EKCH (Copenhague)	EIDW (Dublin)
EDDL (Düsseldorf)	LIMC (Milan Malpensa)	EGCC (Manchester)
EDDH (Hambourg)	LFMN (Nice-Côte d'Azur)	LROP (Bucharest-Otopeni)
EDDS (Stuttgart)	LSZH (Zurich)	LGAV (Athènes)
LTAI (Antalya)	LOWW (Vienne-Schwechat)	EGGW (Londres-Luton)
		EDDK (Cologne-Bonn)
		EPWA (Varsovie-Chopin)
		ENGM (Oslo-Gardermoen)
		EDDT (Berlin-Tegel)

Tableau 5.4 – Premier résultat du clustering k-médoides des aéroports

Cluster 1	Cluster 2	Cluster 3
EHAM (Amsterdam-Schiphol)	LGAV (Athènes)	EGSS (Londres- Stansted)
LEPA (Palma de Majorque)	LTBA (Istanbul Atatürk)	LPPT (Lisbonne)
LLBG (Tel Aviv)	LFPO (Paris-Orly)	EPWA (Varsovie-Chopin)
LEBL (Barcelone-El Prat)	LSGG (Genève)	LEMG (Malaga-Costa del Sol)
LFPG (Paris-CDG)	LIRF (Rome Fiumicino)	EFHK (Helsinki-Vantaa)
EDDF (Francfort)	EDDM (Munich)	ESSA (Stockholm-Arlanda)
LKPR (Prague-Václav-Havel)	LTFJ (Istanbul Sabiha Gökçen)	ENGM (Oslo-Gardermoen)
EGLL (Londres-Heathrow)	EKCH (Copenhague)	EIDW (Dublin)
EDDL (Düsseldorf)	LIMC (Milan Malpensa)	EGGW (Londres-Luton)
EDDH (Hambourg)	LFMN (Nice-Côte d'Azur)	LROP (Bucharest-Otopeni)
EDDS (Stuttgart)	EDDK (Cologne-Bonn)	EDDT (Berlin-Tegel)
LTAI (Antalya)	LOWW (Vienne-Schwechat)	EGCC (Manchester)
	EBBR (Bruxelles-National)	
	LEMD (Madrid-Barajas)	
	EGKK (Londres-Gatwick)	
	LSZH (Zurich)	

Tableau 5.5 – Second résultat du clustering k-médoides des aéroports

5.3.3.5 Discussion

Ce travail a porté sur l'étude des retards dans le transport aérien que nous avons modélisé à l'aide d'outils récents exploitant la géométrie de l'information. Le but de cette étude préliminaire était avant tout de fournir les outils permettant à un futur utilisateur de réaliser des études plus poussées.

Plusieurs autres études se sont intéressées à la modélisation des retards d'avion par des distributions de probabilité, mais très peu ont utilisé les outils de la géométrie de l'information. Parmi ces derniers ; [Arwini and Dodson \(2008\)](#) utilisent une approche basée sur la géométrie de l'information afin de calculer la distance géodésique séparant deux loi gamma dans une variété associée à ces lois. D'autres part, nous avons proposé de modéliser les retards à l'aide de mélanges de loi gamma sachant l'existence d'un ensemble de causes de retard pouvant se réaliser ou non : le passage à des lois gamma mélangées permet de couvrir ce dernier cas. De plus, il s'agit de la première étude qui porte sur le

calcul de distances géodésiques entre lois gamma mélangées appliquées aux retards dans le transport aérien.

Cette étude est toujours en cours, tous les outils et conceptions mathématiques ont été développés et implémentés. Toutefois, l'interprétation des résultats nécessite une expertise particulière dans le domaine du transport aérien qui se fera prochainement.

Troisième partie

Discussion générale

Dans cette partie nous rappelons les principaux résultats apportés par ce travail de thèse, en présentant les limites et les perspectives associées.

En étudiant la littérature, que cela soit en imagerie médicale ou en aéronautique, nous avons constaté que les statistiques les plus communément utilisées sont des mesures agrégées : la moyenne, la médiane et l'écart-type. Par exemple, afin d'étudier l'effet d'un traitement sur des patients atteints de sclérose en plaques, c'est la valeur moyenne du biomarqueur IRM (e.g. coefficient de diffusion, coefficient de transfert d'aimantation...) qui est mesurée dans le cerveau entier ou dans des tissus d'intérêt. Puis, ces valeurs de tendance centrale sont utilisées dans les analyses statistiques entre un groupe traité et un groupe placebo. Le risque avec cet estimateur moyen est que des informations qui peuvent être primordiales se trouvent résumées, voire effacées.

Dans ce projet de thèse, nous avons proposé trois différentes approches permettant de prendre en compte non pas uniquement la moyenne d'un marqueur mais sa distribution dans sa globalité. Les deux premières méthodes correspondent à des approches non-paramétriques et la troisième à une approche paramétrique :

— Les métriques d'histogrammes

L'analyse d'histogrammes est un moyen alternatif aux analyses par régions d'intérêt et voxel-à-voxel classiques. Les histogrammes représentant les valeurs des paramètres IRM mesurés dans tout le cerveau sont de plus en plus utilisés pour caractériser des maladies diffuses qui affectent de grandes parties du cerveau (Tofts et al., 2004a). Cette analyse d'histogrammes consiste à extraire des *métriques* d'histogrammes, qui sont des résumés d'histogrammes, tels que la moyenne, la médiane, l'amplitude du pic et plus rarement les quantiles.

Une première application correspond à la classification de la population Alzheimer à l'aide de biomarqueurs du ruban cortical. Ces biomarqueurs sont représentés par des histogrammes dont on extrait des métriques. Ces métriques sont intégrées dans un classifieur SVM afin de discriminer différents groupes de sujets de la population Alzheimer. L'avantage de notre approche basée sur l'analyse d'histogrammes est sa simplicité, sa rapidité et le fait d'éviter tout biais ou a priori quant aux régions du cerveau susceptibles d'être affectées par la maladie d'Alzheimer. Elle surpasse tous les autres types d'analyse d'images pour la discrimination de tous les groupes, sauf dans le cas de la différenciation entre les patients ayant converti vers la MA (pMCI) et ceux restés stables durant le suivi (sMCI). Ce résultat révèle les limites de l'analyse d'histogrammes. En effet, l'absence de spécificité spatiale et le fait de résumer toute une image paramétrique en quelques simples indices numériques, efface d'importantes informations. Il serait ainsi plus intéressant d'utiliser ce type d'approche dans le cadre de maladies diffuses telles que la sclérose en plaques. Dans le cas des pathologies localisées, une approche hybride pourrait être développée fusionnant analyse par régions d'intérêt et analyse d'histogrammes, comme l'ont proposée très récemment Ruiz et al. (2018).

La seconde application porte sur une étude IRM ancillaire d'un essai thérapeutique de sclérose en plaques. Les images proviennent de différentes modalités, dont l'IRM structurelle, l'IRM de diffusion et l'IRM de transfert d'aimantation. En effet, la compréhension des mécanismes diffus sous-jacents de

la SEP est complexe et une approche multimodale est nécessaire pour quantifier les caractéristiques hétérogènes de la maladie. Les résultats concernant la thérapeutique de l'essai clinique ne sont pas concluants mais dans ce travail de thèse, il ne s'agit pas de juger de l'efficacité du traitement mais de proposer une méthodologie permettant de combiner des biomarqueurs IRM basés sur l'analyse d'histogrammes de biomarqueurs au lieu de l'approche classique, qui effectue des analyses statistiques individuelles indépendantes basées sur la valeur moyenne de chaque biomarqueur. Il sera tout de même intéressant, dans une future étude, de travailler sur un ensemble de données comprenant des groupes clairement distincts pour quantifier la supériorité de notre méthode et pour ce faire, nous comparerions les performances de classification dans les deux cas : avant de combiner les biomarqueurs et après avoir appliqué notre approche de biomarqueurs combinés.

— Distance entre histogrammes

Cette seconde approche non-paramétrique est une alternative à l'usage des histogrammes d'images quantitatives. En effet, par rapport à l'approche classique d'analyse d'histogrammes, cette seconde méthode ne se limite pas à l'utilisation de mesures arbitraires telles que la moyenne ou les quantiles d'histogrammes. En effet, elle permet de conserver les histogrammes de biomarqueurs dans leur globalité et de mesurer des distances entre eux, obtenant ainsi une matrice de dissimilarité qui sera par la suite intégrée dans un classifieur afin de discriminer des groupes d'une population donnée. Cette méthode a été appliquée pour la classification de la population de sclérose en plaques, en utilisant quarante mesures de dissimilarité et l'algorithme des k-plus proches voisins qui a été étendue aux données histogrammes. Cette approche a permis d'améliorer les résultats de classification de la population SEP par rapport à des analyses type régions d'intérêt (Richiardi et al., 2012; Wottschel, 2017). Les résultats ont aussi révélés l'importance du choix de la mesure de dissimilarité ; en effet la précision de classification varie considérablement d'une mesure à l'autre et c'est la divergence de Kullback-Leibler qui permet de mieux classer les populations. De plus, au lieu d'effectuer des tests statistiques pour comparer des groupes de patients, dans notre étude les patients sont classés en fonction de leurs histogrammes individuels. Nous pensons qu'il s'agit de la première application d'une telle technique de classification aux données type histogrammes de biomarqueurs.

— Distribution de probabilité et géométrie de l'information

Dans la plupart des études récentes portant sur l'analyse d'histogrammes (Busovaca et al., 2016; Cercignani et al., 2018), les auteurs ont utilisé les histogrammes comme des approximations des fonctions de densité de probabilités. Seules quelques-unes de leurs caractéristiques, telles que la moyenne, les quantiles, la position et l'amplitude du pic, le skewness et le kurtosis ont été utilisées. Dans notre approche précédente, dans le contexte de la sclérose en plaques, l'intégralité des informations de l'histogramme a été intégrée dans un classifieur k-plus proches voisins, avec des performances de classification supérieures à celles des études précédentes (Rebbah et al., 2019a). Cependant, la qualité de l'estimation de l'histogramme dépend du choix du nombre de classes,

du type de filtrage et de normalisation. L'histogramme peut alors être une estimation trop grossière de la densité de probabilité et fournir de mauvaises estimations des caractéristiques du biomarqueur en question. Pour surmonter cet inconvénient et utiliser l'ensemble des informations qu'offre le biomarqueur, les distributions de probabilité sous-jacentes doivent elles-mêmes être utilisées comme biomarqueur. Le cadre général de la géométrie de l'information, dans lequel les distributions de probabilité sont considérées comme des points sur une variété, est particulièrement pertinent pour atteindre cet objectif. Nous avons utilisés les outils de la géométrie de l'information dans deux applications distinctes :

La première application est une preuve de concept démontrant la supériorité de l'utilisation de distributions de probabilité. Pour ce faire, nous avons choisi une simple problématique clinique qui est la classification des patients atteints de la maladie d'Alzheimer et des sujets contrôles. L'approche basée sur la géométrie de l'information a consisté à modéliser les distributions d'épaisseur corticale par des lois gamma généralisées, puis de mesurer les distances géodésiques entre ces lois gamma généralisées qui sont intégrées dans l'algorithme de clustering k-médoides. Chaque étape représente une nouveauté dans le domaine. En effet, dans le cadre de la géométrie de l'information très peu d'études se sont intéressées à la loi gamma généralisée qui dispose d'un paramètre de forme supplémentaire par rapport à la distribution gamma classique. Aussi, dans les études qui cherchent à classer ou à prédire l'évolution d'une pathologie, la distance utilisée dans les algorithmes de classification est habituellement la distance euclidienne, or dans notre étude nous avons choisi la distance géodésique qui est une métrique naturelle lorsque sont considérées des distributions de probabilité. Enfin, l'algorithme k-médoides de clustering a été étendue avec succès en utilisant la distance géodésique. En plus de l'originalité de la méthode, les résultats sont très satisfaisants. En effet, l'approche basée sur la géométrie de l'information et le modèle gamma généralisée surpasse les approches d'analyse d'histogrammes classique et par régions d'intérêt. Ce travail a fait l'objet d'une publication dans le journal *Mathematics* ([Rebbah et al., 2019b](#)). Dans un travail futur, il serait intéressant d'étudier des problématiques cliniques plus complexes et d'étendre les outils de la géométrie non-euclidienne aux études longitudinales et multimodales.

La seconde application a porté sur la caractérisation des retards dans le transport aérien, où nous avons proposé une approche paramétrique basée sur le modèle de lois gamma mélangées. Dans cette étude, toutes les conceptions mathématiques portant sur la géométrie de l'information ont été développées et implémentées : de l'estimation des paramètres du modèle des retards à la mesure des distances géodésiques. Des clusters convergents sont obtenus mais l'appréciation de la pertinence des résultats obtenus sur l'application des retards aéroportuaires nécessite l'interprétation d'un professionnel du trafic aérien.

Bibliographie

- Abdi, H., Williams, L. J., and Valentin, D. (2013). Multiple factor analysis : principal component analysis for multitable and multiblock data sets. *Wiley Interdisciplinary Reviews : Computational Statistics*, 5(2) :149–179.
- Acosta-Cabronero, J., Alley, S., Williams, G. B., Pengas, G., and Nestor, P. J. (2012). Diffusion tensor metrics as biomarkers in Alzheimer’s disease. *PloS One*, 7(11) :e49072.
- Adaszewski, S., Dukart, J., Kherif, F., Frackowiak, R., and Draganski, B. (2013). How early can we predict Alzheimer’s disease using computational anatomy? *Neurobiology of Aging*, 34(12) :2815–2826.
- Agosta, F., Rovaris, M., Pagani, E., Sormani, M. P., Comi, G., and Filippi, M. (2006). Magnetization transfer MRI metrics predict the accumulation of disability 8 years later in patients with multiple sclerosis. *Brain : A Journal of Neurology*, 129(Pt 10) :2620–2627.
- Almhana, J., Liu, Z., Choulakian, V., and McGorman, R. (2006a). A recursive algorithm for gamma mixture models. In *2006 IEEE International Conference on Communications*, volume 1, pages 197–202.
- Almhana, J., Liu, Z., Choulakian, V., and McGorman, R. (2006b). A Recursive Algorithm for Gamma Mixture Models. In *2006 IEEE International Conference on Communications*, volume 1, pages 197–202.
- Alzheimer, A. (1906). Über einen eigenartigen schweren Erkrankungsprozeß der Hirnrinde. *Neurol Central*, 25(1134).
- Al-Radaideh, A., Mougin, O. E., Lim, S.-Y., Chou, I.-J., Constantinescu, C. S., and Gowland, P. (2015). Histogram analysis of quantitative T1 and MT maps from ultrahigh field MRI in clinically isolated syndrome and relapsing–remitting multiple sclerosis. *NMR in Biomedicine*, 28(11) :1374–1382.
- Amari, S.-i. and Nagaoka, H. (2000). *Methods of Information Geometry*, volume 191. Translations of Mathematical Monographs.
- Amlie, I. K. and Fjell, A. M. (2014). Diffusion tensor imaging of white matter degeneration in Alzheimer’s disease and mild cognitive impairment. *Neuroscience*, 276 :206–215.
- Amoroso, L. (1925). Ricerche intorno alla curva dei redditi. *Annali di Matematica Pura ed Applicata*, 2(1) :123–159.
- Anderson, T. W. (1962). On the Distribution of the Two-Sample Cramer-von Mises Criterion. *The Annals of Mathematical Statistics*, 33(3) :1148–1159.
- Arwini, K., Dodson, C., Doig, A., Sampson, W., Scharcanski, J., and Felipussi, S. (2008). *Information Geometry : Near Randomness and Near Independence*. Information Geometry : Near Randomness and Near Independence. Springer.
- Arwini, K. and Dodson, C. T. J. (2008). *Information Geometry : Near Randomness and Near Independence*. Lecture Notes in Mathematics. Springer-Verlag, Berlin Heidelberg.

- Ashburner, J. (2000). *Computational Neuroanatomy, Chapter5 : Image Segmentation*. PhD thesis, University College London.
- Ashburner, J. and Friston, K. J. (2000). Voxel-based morphometry—the methods. *NeuroImage*, 11(6 Pt 1) :805–821.
- Astrakas, L. G. and Argyropoulou, M. I. (2010). Shifting from region of interest (ROI) to voxel-based analysis in human brain mapping. *Pediatric Radiology*, 40(12) :1857–1867.
- Azevedo, C. J. and Pelletier, D. (2016). Whole-brain atrophy : ready for implementation into clinical decision-making in multiple sclerosis? *Current Opinion in Neurology*, 29(3) :237–242.
- Bakkour, A., Morris, J. C., and Dickerson, B. C. (2009). The cortical signature of prodromal AD : regional thinning predicts mild AD dementia. *Neurology*, 72(12) :1048–1055.
- Bakkour, A., Morris, J. C., Wolk, D. A., and Dickerson, B. C. (2013). The effects of aging and Alzheimer’s disease on cerebral cortical anatomy : specificity and differential relationships with cognition. *NeuroImage*, 76 :332–344.
- Ballard, C., Gauthier, S., Corbett, A., Brayne, C., Aarsland, D., and Jones, E. (2011). Alzheimer’s disease. *Lancet (London, England)*, 377(9770) :1019–1031.
- Bammer, R., Augustin, M., Strasser-Fuchs, S., Seifert, T., Kapeller, P., Stollberger, R., Ebner, F., Hartung, H. P., and Fazekas, F. (2000). Magnetic resonance diffusion tensor imaging for characterizing diffuse and focal white matter abnormalities in multiple sclerosis. *Magnetic Resonance in Medicine*, 44(4) :583–591.
- Barachant, A. and Congedo, M. (2014). A Plug&Play P300 BCI Using Information Geometry. *arXiv :1409.0107 [cs, stat]*. arXiv : 1409.0107.
- Barbaresco, F. (2008). Innovative tools for radar signal processing Based on Cartan’s geometry of SPD matrices and Information Geometry. In *2008 IEEE Radar Conference*, pages 1–6.
- Barkhof, F. (2002). The clinico-radiological paradox in multiple sclerosis revisited. *Current Opinion in Neurology*, 15(3) :239–245.
- Bellman, R. E. (1957). *Dynamic Programming*. Princeton University Press.
- Bester, M., Jensen, J. H., Babb, J. S., Tabesh, A., Miles, L., Herbert, J., Grossman, R. I., and Inglese, M. (2015). Non-Gaussian diffusion MRI of gray matter is associated with cognitive impairment in multiple sclerosis. *Multiple Sclerosis (Houndmills, Basingstoke, England)*, 21(7) :935–944.
- Bhattacharyya, A. (1946). On a Measure of Divergence between Two Multinomial Populations. *Sankhyā : The Indian Journal of Statistics (1933-1960)*, 7(4) :401–406.
- Bigi, B. (2003). Using kullback-leibler distance for text categorization. In *Proceedings of the 25th European Conference on IR Research, ECIR’03*, pages 305–319, Berlin, Heidelberg. Springer-Verlag.
- Bigot, J., Gouet, R., Klein, T., and López, A. (2013). Geodesic PCA in the Wasserstein space. *arXiv :1307.7721 [math, stat]*. arXiv : 1307.7721.
- Bobinski, M., de Leon, M. J., Wegiel, J., Desanti, S., Convit, A., Saint Louis, L. A., Rusinek, H., and Wisniewski, H. M. (2000). The histological validation of post mortem magnetic resonance imaging-determined hippocampal volume in Alzheimer’s disease. *Neuroscience*, 95(3) :721–725.
- Boser, B. E., Guyon, I. M., and Vapnik, V. N. (1992). A Training Algorithm for Optimal Margin Classifiers. In *Proceedings of the Fifth Annual Workshop on Computational Learning Theory, COLT ’92*, pages 144–152, New York, NY, USA. ACM. event-place : Pittsburgh, Pennsylvania, USA.

- Bozzali, M., Cercignani, M., Sormani, M. P., Comi, G., and Filippi, M. (2002). Quantification of Brain Gray Matter Damage in Different MS Phenotypes by Use of Diffusion Tensor MR Imaging. *American Journal of Neuroradiology*, 23(6) :985–988.
- Bozzali, M., Franceschi, M., Falini, A., Pontesilli, S., Cercignani, M., Magnani, G., Scotti, G., Comi, G., and Filippi, M. (2001). Quantification of tissue damage in AD using diffusion tensor and magnetization transfer MRI. *Neurology*, 57(6) :1135–1137.
- Braak, H. and Braak, E. (1991). Neuropathological staging of Alzheimer-related changes. *Acta Neuropathologica*, 82(4) :239–259.
- Brayne, C. and Miller, B. (2017). Dementia and aging populations—A global priority for contextualized research and health policy. *PLoS Medicine*, 14(3).
- Brigant, A. L. and Puechmorel, S. (2019). Quantization and clustering on riemannian manifolds with an application to air traffic analysis. *Journal of Multivariate Analysis*, 173 :685 – 703.
- Brodmann, K. (1909). *Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues*. Leipzig : Barth.
- Buchem, M. A. v., McGowan, J. C., Kolson, D. L., Polansky, M., and Grossman, R. I. (1996). Quantitative volumetric magnetization transfer analysis in multiple sclerosis : Estimation of macroscopic and microscopic disease burden. *Magnetic Resonance in Medicine*, 36(4) :632–636.
- Busovaca, E., Zimmerman, M. E., Meier, I. B., Griffith, E. Y., Grieve, S. M., Korgaonkar, M. S., Williams, L. M., and Brickman, A. M. (2016). Is the Alzheimer’s disease cortical thickness signature a biological marker for memory? *Brain imaging and behavior*, 10(2) :517–523.
- Cai, K., Xu, H., Guan, H., Zhu, W., Jiang, J., Cui, Y., Zhang, J., Liu, T., and Wen, W. (2017). Identification of Early-Stage Alzheimer’s Disease Using Sulcal Morphology and Other Common Neuroimaging Indices. *PLoS ONE*, 12(1).
- Casanova, R., Hsu, F.-C., Sink, K. M., Rapp, S. R., Williamson, J. D., Resnick, S. M., Espeland, M. A., and Initiative, f. t. A. D. N. (2013). Alzheimer’s Disease Risk Assessment Using Large-Scale Machine Learning Methods. *PLOS ONE*, 8(11) :e77949.
- Cassol, E., Ranjeva, J.-P., Ibarrola, D., Mékies, C., Manelfe, C., Clanet, M., and Berry, I. (2004). Diffusion tensor imaging in multiple sclerosis : a tool for monitoring changes in normal-appearing white matter. *Multiple Sclerosis (Houndmills, Basingstoke, England)*, 10(2) :188–196.
- Castriota-Scanderbeg, A., Sabatini, U., Fasano, F., Floris, R., Fraracci, L., Mario, D. M., Nocentini, U., and Caltagirone, C. (2002). Diffusion of water in large demyelinating lesions : a follow-up study. *Neuroradiology*, 44(9) :764–767.
- Ceccarelli, A., Rocca, M. A., Falini, A., Tortorella, P., Pagani, E., Rodegher, M., Comi, G., Scotti, G., and Filippi, M. (2007). Normal-appearing white and grey matter damage in MS. A volumetric and diffusion tensor MRI study at 3.0 Tesla. *Journal of Neurology*, 254(4) :513–518.
- Cercignani, M., Bozzali, M., Iannucci, G., Comi, G., and Filippi, M. (2001a). Magnetisation transfer ratio and mean diffusivity of normal appearing white and grey matter from patients with multiple sclerosis. *Journal of Neurology, Neurosurgery, and Psychiatry*, 70(3) :311–317.
- Cercignani, M., Dowell, N. G., and Tofts, P. S. (2018). *Quantitative MRI of the Brain : Principles of Physical Measurement*. Series in medical physics and biomedical engineering. CRC Press, second edition edition.
- Cercignani, M., Iannucci, G., and Filippi, M. (1999). Diffusion-weighted imaging in multiple sclerosis. *Italian Journal of Neurological Sciences*, 20(5 Suppl) :S246–249.
- Cercignani, M., Iannucci, G., Rocca, M. A., Comi, G., Horsfield, M. A., and Filippi, M. (2000). Pathologic damage in MS assessed by diffusion-weighted and magnetization transfer MRI. *Neurology*, 54(5) :1139–1144.

- Cercignani, M., Inglese, M., Pagani, E., Comi, G., and Filippi, M. (2001b). Mean Diffusivity and Fractional Anisotropy Histograms of Patients with Multiple Sclerosis. *American Journal of Neuroradiology*, 22(5) :952–958.
- Cha, S.-H. (2007). Comprehensive Survey on Distance/Similarity Measures Between Probability Density Functions. *International Journal Of Mathematical Models And Methods In Applied Sciences*, 1(4) :300–307.
- Chao, L. L., Schuff, N., Kramer, J. H., Du, A. T., Capizzano, A. A., O'Neill, J., Wolkowitz, O. M., Jagust, W. J., Chui, H. C., Miller, B. L., Yaffe, K., and Weiner, M. W. (2005). Reduced medial temporal lobe N-acetylaspartate in cognitively impaired but nondemented patients. *Neurology*, 64(2) :282–289.
- Chard, D. T., Griffin, C. M., McLean, M. A., Kapeller, P., Kapoor, R., Thompson, A. J., and Miller, D. H. (2002). Brain metabolite changes in cortical grey and normal-appearing white matter in clinically early relapsing-remitting multiple sclerosis. *Brain : A Journal of Neurology*, 125(Pt 10) :2342–2352.
- Chard, D. T., Griffin, C. M., Rashid, W., Davies, G. R., Altmann, D. R., Kapoor, R., Barker, G. J., Thompson, A. J., and Miller, D. H. (2004). Progressive grey matter atrophy in clinically early relapsing-remitting multiple sclerosis. *Multiple Sclerosis (Houndmills, Basingstoke, England)*, 10(4) :387–391.
- Chen, J. T., Collins, D. L., Atkins, H. L., Freedman, M. S., Arnold, D. L., and Canadian MS/BMT Study Group (2008). Magnetization transfer ratio evolution with demyelination and remyelination in multiple sclerosis lesions. *Annals of Neurology*, 63(2) :254–262.
- Chen, Y., Georgiou, T. T., and Tannenbaum, A. (2017). Optimal transport for Gaussian mixture models. *arXiv :1710.07876 [cs, math]*. arXiv : 1710.07876.
- Chentsov, N. N. (1982). *Statistical Decision Rules and Optimal Inference*. American Mathematical Society. Google-Books-ID : kTJJtQAACAAJ.
- Ciccarelli, O., Werring, D. J., Wheeler-Kingshott, C. A., Barker, G. J., Parker, G. J., Thompson, A. J., and Miller, D. H. (2001). Investigation of MS normal-appearing brain using diffusion tensor MRI with clinical correlations. *Neurology*, 56(7) :926–933.
- Clim, A., Zota, R. D., and TinicĂ, G. (2018). The Kullback-Leibler Divergence Used in Machine Learning Algorithms for Health Care Applications and Hypertension Prediction : A Literature Review. *Procedia Computer Science*, 141 :448–453.
- Coclitu, C., Constantinescu, C. S., and Tanasescu, R. (2016). The future of multiple sclerosis treatments. *Expert Review of Neurotherapeutics*, 16(12) :1341–1356.
- Correale, J., Gaitán, M. I., Ysraelit, M. C., and Fiol, M. P. (2017). Progressive multiple sclerosis : from pathogenic mechanisms to treatment. *Brain*, 140(3) :527–546.
- Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine learning* 20.3, pages 273–297.
- Costa, S. I. R., Santos, S. A., and Strapasson, J. E. (2014). Fisher information distance : a geometrical reading. *arXiv :1210.2354 [math-ph, stat]*. arXiv : 1210.2354.
- Crooks, G. E. (2010). The amoroso distribution. *arXiv preprint arXiv :1005.3274*.
- Cuingnet, R., Gerardin, E., Tessieras, J., Auzias, G., Lehericy, S., Habert, M.-O., Chupin, M., Benali, H., Colliot, O., and Alzheimer's Disease Neuroimaging Initiative (2011). Automatic classification of patients with Alzheimer's disease from structural MRI : a comparison of ten methods using the ADNI database. *NeuroImage*, 56(2) :766–781.
- De Angelis, F., Plantone, D., and Chataway, J. (2018). Pharmacotherapy in Secondary Progressive Multiple Sclerosis : An Overview. *CNS drugs*, 32(6) :499–526.

- De Stefano, N., Iannucci, G., Sormani, M. P., Guidi, L., Bartolozzi, M. L., Comi, G., Federico, A., and Filippi, M. (2002). MR correlates of cerebral atrophy in patients with multiple sclerosis. *Journal of Neurology*, 249(8) :1072–1077.
- De Stefano, N., Matthews, P. M., Fu, L., Narayanan, S., Stanley, J., Francis, G. S., Antel, J. P., and Arnold, D. L. (1998). Axonal damage correlates with disability in patients with relapsing-remitting multiple sclerosis. Results of a longitudinal magnetic resonance spectroscopy study. *Brain : A Journal of Neurology*, 121 (Pt 8) :1469–1477.
- De Stefano, N., Stromillo, M. L., Giorgio, A., Bartolozzi, M. L., Battaglini, M., Baldini, M., Portaccio, E., Amato, M. P., and Sormani, M. P. (2016). Establishing pathological cut-offs of brain atrophy rates in multiple sclerosis. *Journal of Neurology, Neurosurgery, and Psychiatry*, 87(1) :93–99.
- Dehmeshki, J., Ruto, A. C., Arridge, S., Silver, N. C., Miller, D. H., and Tofts, P. S. (2001a). Analysis of MTR histograms in multiple sclerosis using principal components and multiple discriminant analysis. *Magnetic Resonance in Medicine*, 46(3) :600–609.
- Dehmeshki, J., Silver, N. C., Leary, S. M., Tofts, P. S., Thompson, A. J., and Miller, D. H. (2001b). Magnetisation transfer ratio histogram analysis of primary progressive and other multiple sclerosis subgroups. *Journal of the Neurological Sciences*, 185(1) :11–17.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1) :1–38.
- Deza, M. M. and Deza, E. (2014). *Encyclopedia of Distances*. Springer-Verlag, Berlin Heidelberg, 3 edition.
- Dice, L. R. (1945). Measures of the Amount of Ecologic Association Between Species. *Ecology*, 26(3) :297–302.
- Dickerson, B. C., Bakkour, A., Salat, D. H., Feczko, E., Pacheco, J., Greve, D. N., Grodstein, F., Wright, C. I., Blacker, D., Rosas, H. D., Sperling, R. A., Atri, A., Growdon, J. H., Hyman, B. T., Morris, J. C., Fischl, B., and Buckner, R. L. (2009a). The cortical signature of Alzheimer's disease : regionally specific cortical thinning relates to symptom severity in very mild to mild AD dementia and is detectable in asymptomatic amyloid-positive individuals. *Cerebral Cortex (New York, N.Y. : 1991)*, 19(3) :497–510.
- Dickerson, B. C., Feczko, E., Augustinack, J. C., Pacheco, J., Morris, J. C., Fischl, B., and Buckner, R. L. (2009b). Differential effects of aging and Alzheimer's disease on medial temporal lobe cortical thickness and surface area. *Neurobiology of Aging*, 30(3) :432–440.
- Dickerson, B. C., Stoub, T. R., Shah, R. C., Sperling, R. A., Killiany, R. J., Albert, M. S., Hyman, B. T., Blacker, D., and Detolledo-Morrell, L. (2011). Alzheimer-signature MRI biomarker predicts AD dementia in cognitively normal adults. *Neurology*, 76(16) :1395–1402.
- Dickerson, B. C. and Wolk, D. (2013). Biomarker-based prediction of progression in MCI : Comparison of AD-signature and hippocampal volume with spinal fluid amyloid- and tau. *Frontiers in Aging Neuroscience*, 5.
- Dousset, V., Grossman, R. I., Ramer, K. N., Schnall, M. D., Young, L. H., Gonzalez-Scarano, F., Lavi, E., and Cohen, J. A. (1992). Experimental allergic encephalomyelitis and multiple sclerosis : lesion characterization with magnetization transfer imaging. *Radiology*, 182(2) :483–491.
- Droogan, A. G., Clark, C. A., Werring, D. J., Barker, G. J., McDonald, W. I., and Miller, D. H. (1999). Comparison of multiple sclerosis clinical subgroups using navigated spin echo diffusion-weighted imaging. *Magnetic Resonance Imaging*, 17(5) :653–661.
- Duda, R. O., Hart, P. E., and Stork, D. G. (2000). *Pattern Classification (2Nd Edition)*. Wiley-Interscience, New York, NY, USA.

- Duistermaat, J. (2001). On hessian riemannian structures. *Asian journal of mathematics*, 5 :79–91.
- Dutilleul, C. (2015). *Evaluation de traitement de la sclérose en plaques par analyses morphométriques de données d'imagerie par résonance magnétique*. Toulouse 3.
- Es, A. C. G. M. v., Grond, J. v. d., Dam, V. H. t., Craen, A. J. M. d., Blauw, G. J., Westendorp, R. G. J., Admiraal-Behloul, F., Buchem, M. A. v., and Group, f. t. P. S. (2010). Associations between Total Cerebral Blood Flow and Age Related Changes of the Brain. *PLOS ONE*, 5(3) :e9825.
- Escofier, B. and Pagès, J. (1990). *Analyses factorielles simples et multiples : objectifs, méthodes et interprétation*. Dunod.
- Eshaghi, A., Marinescu, R. V., Young, A. L., Firth, N. C., Prados, F., Cardoso, M. J., Tur, C., Angelis, F. D., Cawley, N., Brownlee, W., Stefano, N. D., Stromillo, M. L., Battaglini, M., Ruggieri, S., Gasperini, C., Filippi, M., Rocca, M. A., Rovira, A., Sastre-Garriga, J., Geurts, J., Vrenken, H., Wottschel, V., Leurs, C. E., Uitdehaag, B., Pirpamer, L., Enzinger, C., Ourselin, S., Wheeler-Kingshott, C. A. G., Chard, D., Thompson, A. J., Barkhof, F., Alexander, D. C., and Ciccarelli, O. (2017). Progression of regional grey matter atrophy in multiple sclerosis. *bioRxiv*, page 190116.
- Evangelou, N., Konz, D., Esiri, M. M., Smith, S., Palace, J., and Matthews, P. M. (2000). Regional axonal loss in the corpus callosum correlates with cerebral white matter lesion volume and distribution in multiple sclerosis. *Brain : A Journal of Neurology*, 123 (Pt 9) :1845–1849.
- Ferini-Strambi, L., Bozzali, M., Cercignani, M., Oldani, A., Zucconi, M., and Filippi, M. (2000). Magnetization transfer and diffusion-weighted imaging in nocturnal frontal lobe epilepsy. *Neurology*, 54(12) :2331–2333.
- Fernando, K. T. M., Tozer, D. J., Miskiel, K. A., Gordon, R. M., Swanton, J. K., Dalton, C. M., Barker, G. J., Plant, G. T., Thompson, A. J., and Miller, D. H. (2005). Magnetization transfer histograms in clinically isolated syndromes suggestive of multiple sclerosis. *Brain*, 128(12) :2911–2925.
- Filippi, M. (2001). Linking structural, metabolic and functional changes in multiple sclerosis. *European Journal of Neurology*, 8(4) :291–297.
- Filippi, M., editor (2015). *Oxford Textbook of Neuroimaging*. Oxford Textbooks in Clinical Neurology. Oxford University Press, Oxford, New York.
- Filippi, M., Cercignani, M., Inglese, M., Horsfield, M. A., and Comi, G. (2001). Diffusion tensor magnetic resonance imaging in multiple sclerosis. *Neurology*, 56(3) :304–311.
- Filippi, M., De Stefano, N., Dousset, V., and McGowan, J. (2005). *MR Imaging in White Matter Diseases of the Brain and Spinal Cord*. Springer Berlin Heidelberg, medical radiology - diagnostic imaging and radiation oncology edition.
- Filippi, M., Dousset, V., McFarland, H. F., Miller, D. H., and Grossman, R. I. (2002). Role of magnetic resonance imaging in the diagnosis and monitoring of multiple sclerosis : consensus report of the White Matter Study Group. *Journal of magnetic resonance imaging : JMRI*, 15(5) :499–504.
- Filippi, M., Iannucci, G., Cercignani, M., Assunta Rocca, M., Pratesi, A., and Comi, G. (2000a). A quantitative study of water diffusion in multiple sclerosis lesions and normal-appearing white matter using echo-planar imaging. *Archives of Neurology*, 57(7) :1017–1021.
- Filippi, M., Iannucci, G., Tortorella, C., Minicucci, L., Horsfield, M. A., Colombo, B., Sormani, M. P., and Comi, G. (1999). Comparison of MS clinical phenotypes using conventional and magnetization transfer MRI. *Neurology*, 52(3) :588–594.
- Filippi, M., Preziosa, P., and Rocca, M. A. (2014). Magnetic resonance outcome measures in multiple sclerosis trials : time to rethink? *Current Opinion in Neurology*, 27(3) :290–299.
- Filippi, M., Preziosa, P., and Rocca, M. A. (2017). Microstructural MR Imaging Techniques in Multiple Sclerosis. *Neuroimaging Clinics of North America*, 27(2) :313–333.

- Filippi, M., Tortorella, C., Rovaris, M., Bozzali, M., Possa, F., Sormani, M., Iannucci, G., and Comi, G. (2000b). Changes in the normal appearing brain tissue and cognitive impairment in multiple sclerosis. *Journal of Neurology, Neurosurgery, and Psychiatry*, 68(2) :157–161.
- Fillard, P. (2008). Riemannian processing of tensors for diffusion MRI and computational anatomy of the brain.
- Fisher, R. A. (1936). The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics*, 7(2) :179–188.
- Fjell, A. M., Walhovd, K. B., Fennema-Notestine, C., McEvoy, L. K., Hagler, D. J., Holland, D., Brewer, J. B., and Dale, A. M. (2009). One year brain atrophy evident in healthy aging. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 29(48) :15223–15231.
- Fletcher, P. T. and Joshi, S. (2004). Principal Geodesic Analysis on Symmetric Spaces : Statistics of Diffusion Tensors. In Sonka, M., Kakadiaris, I. A., and Kybic, J., editors, *Computer Vision and Mathematical Methods in Medical and Biomedical Image Analysis*, Lecture Notes in Computer Science, pages 87–98. Springer Berlin Heidelberg.
- Garey, M., Johnson, D., and Witsenhausen, H. (1982). The complexity of the generalized lloyd - max problem (corresp.). *IEEE Transactions on Information Theory*, 28(2) :255–256.
- Ge, Y., Law, M., Johnson, G., Herbert, J., Babb, J. S., Mannon, L. J., and Grossman, R. I. (2004). Preferential occult injury of corpus callosum in multiple sclerosis measured by diffusion tensor imaging. *Journal of Magnetic Resonance Imaging*, 20(1) :1–7.
- Giulietti, G., Torso, M., Serra, L., Spanò, B., Marra, C., Caltagirone, C., Cercignani, M., Bozzali, M., and Alzheimer’s Disease Neuroimaging Initiative (ADNI) (2018). Whole brain white matter histogram analysis of diffusion tensor imaging data detects microstructural damage in mild cognitive impairment and alzheimer’s disease patients. *Journal of magnetic resonance imaging : JMRI*.
- Goodyear, B. G., Zayed, N. M. H., Cortese, F., Trufyn, J., and Costello, F. (2015). Skewness of Fractional Anisotropy Detects Decreased White Matter Integrity Resulting From Acute Optic Neuritis. *Investigative Ophthalmology & Visual Science*, 56(12) :7597–7603.
- Graff-Radford, J. and Kantarci, K. (2013). Magnetic resonance spectroscopy in Alzheimer’s disease. *Neuropsychiatric Disease and Treatment*, 9 :687–696.
- Grande-Barreto, J. and del Pilar Gomez-Gil, M. (2017). Partial Volume Segmentation inMagnetic Resonance Imaging (MRI). Technical Report No. CCC-17-007, National Institute for Astrophysics Optics and Electronics, Tonantzintla, Puebla.
- Grignon, Y., Duyckaerts, C., Benneceb, M., and Hauw, J. J. (1998). Cytoarchitectonic alterations in the supramarginal gyrus of late onset Alzheimer’s disease. *Acta Neuropathologica*, 95(4) :395–406.
- Gross, D., Shortle, J., Thompson, J., and Harris, C. (2011). *Fundamentals of Queueing Theory*. Wiley Series in Probability and Statistics. Wiley.
- Han, Y. and Park, H. W. (2003). Automatic brain MR image registration based on Talairach reference system. In *Proceedings 2003 International Conference on Image Processing (Cat. No.03CH37429)*.
- Hand, D. J. and Taylor, C. C. (1987). *Multivariate analysis of variance and repeated measures : A practical approach for behavioural scientists*. Multivariate analysis of variance and repeated measures : A practical approach for behavioural scientists. Chapman & Hall/CRC, Boca Raton, FL.

- Hanyu, H., Asano, T., Iwamoto, T., Takasaki, M., Shindo, H., and Abe, K. (2000). Magnetization transfer measurements of the hippocampus in patients with Alzheimer's disease, vascular dementia, and other types of dementia. *AJNR. American journal of neuroradiology*, 21(7) :1235–1242.
- Hiehle, J. F., Lenkinski, R. E., Grossman, R. I., Dousset, V., Ramer, K. N., Schnall, M. D., Cohen, J. A., and Gonzalez-Scarano, F. (1994). Correlation of spectroscopy and magnetization transfer imaging in the evaluation of demyelinating lesions and normal appearing white matter in multiple sclerosis. *Magnetic Resonance in Medicine*, 32(3) :285–293.
- Hoa, D. (2007). *L'Irm Pas Pas (Edition Noir & Blanc)*. Lulu.com.
- Hodges, J. R. (2006). Alzheimer's centennial legacy : origins, landmarks and the current status of knowledge concerning cognitive aspects. *Brain : A Journal of Neurology*, 129(Pt 11) :2811–2822.
- Huang, Y. and Fong, Y. (2014). Identifying optimal biomarker combinations for treatment selection via a robust kernel method. *Biometrics*, 70(4) :891–901.
- Iannucci, G., Minicucci, L., Rodegher, M., Sormani, M. P., Comi, G., and Filippi, M. (1999). Correlations between clinical and MRI involvement in multiple sclerosis : assessment using T(1), T(2) and MT histograms. *Journal of the Neurological Sciences*, 171(2) :121–129.
- Iannucci, G., Tortorella, C., Rovaris, M., Sormani, M. P., Comi, G., and Filippi, M. (2000). Prognostic Value of MR and Magnetization Transfer Imaging Findings in Patients with Clinically Isolated Syndromes Suggestive of Multiple Sclerosis at Presentation. *American Journal of Neuroradiology*, 21(6) :1034–1038.
- Inglese, M., Rovaris, M., Bianchi, S., La Mantia, L., Mancardi, G. L., Ghezzi, A., Montagna, P., Salvi, F., and Filippi, M. (2001). Magnetic resonance imaging, magnetisation transfer imaging, and diffusion weighted imaging correlates of optic nerve, brain, and cervical cord damage in Leber's hereditary optic neuropathy. *Journal of Neurology, Neurosurgery, and Psychiatry*, 70(4) :444–449.
- Jaccard, P. (1901). Distribution de la flore alpine dans le bassin des Dranses et dans quelques régions voisines. *Bulletin de la Société Vaudoise des Sciences Naturelles*, 37 :241 – 272.
- Jeffreys Harold (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 186(1007) :453–461.
- John, G. H., Kohavi, R., and Pfleger, K. (1994). Irrelevant Features and the Subset Selection Problem. In *ICML*.
- Jolliffe, I. T. (1986). *Principal Component Analysis*. Springer Series in Statistics. Springer-Verlag, New York.
- Jones, S. E., Buchbinder, B. R., and Aharon, I. (2000). Three-dimensional mapping of cortical thickness using Laplace's equation. *Human Brain Mapping*, 11(1) :12–32.
- Josse, J. (2010). Principal component methods - hierarchical clustering - partitional clustering : why would we need to choose for visualizing data ?
- Juottonen, K., Laakso, M. P., Partanen, K., and Soininen, H. (1999). Comparative MR analysis of the entorhinal cortex and hippocampus in diagnosing Alzheimer disease. *AJNR. American journal of neuroradiology*, 20(1) :139–144.
- Kaelbling, L. P., Littman, M. L., and Moore, A. W. (1996). Reinforcement Learning : A Survey. *Journal of Artificial Intelligence Research*, 4 :237–285.
- Kailath, T. (1967). The Divergence and Bhattacharyya Distance Measures in Signal Selection. *IEEE Transactions on Communication Technology*, 15(1) :52–60.

- Kaiser, H. F. (1960). The Application of Electronic Computers to Factor Analysis. *Educational and Psychological Measurement*, 20(1) :141–151.
- Kalkers, N. F., Hintzen, R. Q., Waesberghe, J. H. T. M. v., Lazeron, R. H. C., Schijndel, R. A. v., Adèr, H. J., Polman, C. H., and Barkhof, F. (2001). Magnetization transfer histogram parameters reflect all dimensions of MS pathology, including atrophy. *Journal of the Neurological Sciences*, 184(2) :155–162.
- Kang, L., Liu, A., and Tian, L. (2016). Linear combination methods to improve diagnostic/prognostic accuracy on future observations. *Statistical methods in medical research*, 25(4) :1359–1380.
- Kang, Y., Choi, S. H., Kim, Y.-J., Kim, K. G., Sohn, C.-H., Kim, J.-H., Yun, T. J., and Chang, K.-H. (2011). Gliomas : Histogram analysis of apparent diffusion coefficient maps with standard- or high-b-value diffusion-weighted MR imaging–correlation with tumor grade. *Radiology*, 261(3) :882–890.
- Kantarci, K., Weigand, S. D., Petersen, R. C., Boeve, B. F., Knopman, D. S., Gunter, J., Reyes, D., Shiung, M., O'Brien, P. C., Smith, G. E., Ivnik, R. J., Tangalos, E. G., and Jack, C. R. (2007). Longitudinal 1h MRS changes in mild cognitive impairment and Alzheimer's disease. *Neurobiology of aging*, 28(9) :1330–1339.
- Kantarci, K., Xu, Y., Shiung, M. M., O'Brien, P. C., Cha, R. H., Smith, G. E., Ivnik, R. J., Boeve, B. F., Edland, S. D., Kokmen, E., Tangalos, E. G., Petersen, R. C., and Jack, C. R. (2002). COMPARATIVE DIAGNOSTIC UTILITY OF DIFFERENT MR MODALITIES IN MILD COGNITIVE IMPAIRMENT AND ALZHEIMER'S DISEASE. *Dementia and geriatric cognitive disorders*, 14(4) :198–207.
- Karaca, Y., Zhang, Y.-D., Cattani, C., and Ayan, U. (2017). The Differential Diagnosis of Multiple Sclerosis Using Convex Combination of Infinite Kernels. *CNS & neurological disorders drug targets*, 16(1) :36–43.
- Kaufman, L. and Rousseeuw, P. (1987). Clustering by means of Medoids. *Statistical Data Analysis Based on the L1-Norm and Related Methods*, edited by Y. Dodge, North-Holland, pages 405–416.
- Kaunzner, U. W. and Gauthier, S. A. (2017). MRI in the assessment and monitoring of multiple sclerosis : an update on best practice. *Therapeutic Advances in Neurological Disorders*, 10(6) :247–261.
- Khaleeli, Z., Altmann, D. R., Cercignani, M., Ciccarelli, O., Miller, D. H., and Thompson, A. J. (2008). Magnetization transfer ratio in gray matter : a potential surrogate marker for progression in early primary progressive multiple sclerosis. *Archives of Neurology*, 65(11) :1454–1459.
- Killiany, R. J., Hyman, B. T., Gomez-Isla, T., Moss, M. B., Kikinis, R., Jolesz, F., Tanzi, R., Jones, K., and Albert, M. S. (2002). MRI measures of entorhinal cortex vs hippocampus in preclinical AD. *Neurology*, 58(8) :1188–1196.
- Klunk, W. E., Panchalingam, K., Moossy, J., McClure, R. J., and Pettegrew, J. W. (1992). N-acetyl-L-aspartate and other amino acid metabolites in Alzheimer's disease brain : a preliminary proton nuclear magnetic resonance study. *Neurology*, 42(8) :1578–1585.
- Kohavi, R. and John, G. H. (1997). Wrappers for feature subset selection. *Artif. Intell.*
- Kolouri, S., Park, S. R., Thorpe, M., Slepcev, D., and Rohde, G. K. (2017). Optimal Mass Transport : Signal processing and machine-learning applications. *IEEE Signal Processing Magazine*, 34(4) :43–59.
- Kouskoumvekaki, I., Yang, Z., Jónsdóttir, S. O., Olsson, L., and Panagiotou, G. (2008). Identification of biomarkers for genotyping Aspergilli using non-linear methods for clustering and classification. *BMC bioinformatics*, 9 :59.

- Kullback, S. and Leibler, R. A. (1951). On Information and Sufficiency. *The Annals of Mathematical Statistics*, 22(1) :79–86.
- Kurtzke, J. F. (1983). Rating neurologic impairment in multiple sclerosis : an expanded disability status scale (EDSS). *Neurology*, 33(11) :1444–1452.
- Lama, R. K., Gwak, J., Park, J.-S., and Lee, S.-W. (2017). Diagnosis of Alzheimer’s Disease Based on Structural MRI Images Using a Regularized Extreme Learning Machine and PCA Features. *Journal of Healthcare Engineering*, 2017.
- Lan, S., Clarke, J.-P., and Barnhart, C. (2006). Planning for Robust Airline Operations : Optimizing Aircraft Routings and Flight Departure Times to Minimize Passenger Disruptions. *Transportation Science*, 40(1) :15–28.
- Law, M., Young, R., Babb, J., Pollack, E., and Johnson, G. (2007). Histogram analysis versus region of interest analysis of dynamic susceptibility contrast perfusion MR imaging data in the grading of cerebral gliomas. *AJNR. American journal of neuroradiology*, 28(4) :761–766.
- Le Bihan, D. and Breton, E. (1985). Imagerie de diffusion in-vivo par résonance magnétique nucléaire. *Comptes-Rendus de l’Académie des Sciences*, 93(5) :27–34.
- Lerch, J. P. and Evans, A. C. (2005). Cortical thickness analysis examined through power analysis and a population simulation. *NeuroImage*, 24(1) :163–173.
- Liu, C., Liu, A., and Halabi, S. (2011). A min-max combination of biomarkers to improve diagnostic accuracy. *Statistics in Medicine*, 30(16) :2005–2014.
- Liu, M., Zhang, D., and Shen, D. (2015). View-centralized multi-atlas classification for Alzheimer’s disease diagnosis. *Human Brain Mapping*, 36(5) :1847–1865.
- Lloyd, S. (1982). Least squares quantization in pcm. *IEEE Transactions on Information Theory*, 28(2) :129–137.
- Magy, L. (2009). La sclérose en plaques. *Actualités Pharmaceutiques Hospitalières*, 5(19) :14–19.
- Manning, C. D., Raghavan, P., and Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press, New York, NY, USA.
- Martínez-Bisbal, M. C., Arana, E., Martí-Bonmatí, L., Mollá, E., and Celda, B. (2004). Cognitive impairment : classification by 1h magnetic resonance spectroscopy. *European Journal of Neurology*, 11(3) :187–193.
- Mascalchi, M., Ginestroni, A., Bessi, V., Toschi, N., Padiglioni, S., Ciulli, S., Tessa, C., Giannelli, M., Bracco, L., and Diciotti, S. (2013). Regional analysis of the magnetization transfer ratio of the brain in mild Alzheimer disease and amnesic mild cognitive impairment. *AJNR. American journal of neuroradiology*, 34(11) :2098–2104.
- Mateos-Pérez, J. M., Dadar, M., Lacalle-Aurioles, M., Iturria-Medina, Y., Zeighami, Y., and Evans, A. C. (2018). Structural neuroimaging as clinical predictor : A review of machine learning applications. *NeuroImage : Clinical*, 20 :506–522.
- Matusita, K. (1951). On the theory of statistical decision functions. *Annals of the Institute of Statistical Mathematics*, 3(1) :17–35.
- McDonald, W. I., Compston, A., Edan, G., Goodkin, D., Hartung, H. P., Lublin, F. D., McFarland, H. F., Paty, D. W., Polman, C. H., Reingold, S. C., Sandberg-Wollheim, M., Sibley, W., Thompson, A., van den Noort, S., Weinshenker, B. Y., and Wolinsky, J. S. (2001). Recommended diagnostic criteria for multiple sclerosis : guidelines from the International Panel on the diagnosis of multiple sclerosis. *Annals of Neurology*, 50(1) :121–127.

- McFarland, H. F., Barkhof, F., Antel, J., and Miller, D. H. (2002). The role of MRI as a surrogate outcome measure in multiple sclerosis. *Multiple Sclerosis (Houndmills, Basingstoke, England)*, 8(1) :40–51.
- McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., and Stadlan, E. M. (1984). Clinical diagnosis of Alzheimer's disease : report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*, 34(7) :939–944.
- McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R., Kawas, C. H., Klunk, W. E., Koroshetz, W. J., Manly, J. J., Mayeux, R., Mohs, R. C., Morris, J. C., Rossor, M. N., Scheltens, P., Carrillo, M. C., Thies, B., Weintraub, S., and Phelps, C. H. (2011). The diagnosis of dementia due to Alzheimer's disease : recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia : The Journal of the Alzheimer's Association*, 7(3) :263–269.
- Messick, S. and Abelson, R. (1956). The additive constant problem in multidimensional scaling. *Psychometrika*, 21(1) :1–15.
- Metastasio, A., Rinaldi, P., Tarducci, R., Mariani, E., Feliziani, F. T., Cherubini, A., Pelliccioli, G. P., Gobbi, G., Senin, U., and Mecocci, P. (2006). Conversion of MCI to dementia : Role of proton magnetic resonance spectroscopy. *Neurobiology of Aging*, 27(7) :926–932.
- Miller, A. E. and Rhoades, R. W. (2012). Treatment of relapsing-remitting multiple sclerosis : current approaches and unmet needs. *Current Opinion in Neurology*, 25 Suppl :S4–10.
- Min, R., Wu, G., Cheng, J., Wang, Q., and Shen, D. (2014). Multi-atlas based representations for Alzheimer's disease diagnosis. *Human Brain Mapping*, 35(10) :5052–5070.
- Minka, T. P. (2002). *Estimating a Gamma distribution*.
- Molyneux, P. D., Barker, G. J., Barkhof, F., Beckmann, K., Dahlke, F., Filippi, M., Ghazi, M., Hahn, D., MacManus, D., Polman, C., Pozzilli, C., Kappos, L., Thompson, A. J., Wagner, K., Yousry, T., Miller, D. H., and European Study Group on Interferon Beta-1b in Secondary Progressive MS (2001). Clinical-MRI correlations in a European trial of interferon beta-1b in secondary progressive MS. *Neurology*, 57(12) :2191–2197.
- Monga, O. and Benayoun, S. (1995). Using partial derivatives of 3d images to extract typical surface features. *Computer Vision and Image Understanding*, 61(2) :171–189.
- Morris, M. C. (2009). The role of nutrition in Alzheimer's disease : epidemiological evidence. *European Journal of Neurology*, 16 Suppl 1 :1–7.
- Mueller, E. and Chatterji, G. (2002). Analysis of Aircraft Arrival and Departure Delay Characteristics. In *AIAA's Aircraft Technology, Integration, and Operations (ATIO) 2002 Technical Forum*. American Institute of Aeronautics and Astronautics.
- Mueller, S. G., Schuff, N., and Weiner, M. W. (2006). Evaluation of treatment effects in Alzheimer's and other neurodegenerative diseases by MRI and MRS. *NMR in biomedicine*, 19(6) :655–668.
- Mueller, S. G., Weiner, M. W., Thal, L. J., Petersen, R. C., Jack, C., Jagust, W., Trojanowski, J. Q., Toga, A. W., and Beckett, L. (2005). The Alzheimer's Disease Neuroimaging Initiative. *Neuroimaging clinics of North America*, 15(4) :869–xii.
- Muñoz-Ruiz, M., Hartikainen, P., Koikkalainen, J., Wolz, R., Julkunen, V., Niskanen, E., Herukka, S.-K., Kivipelto, M., Vanninen, R., Rueckert, D., Liu, Y., Lötjönen, J., and Soininen, H. (2012). Structural MRI in Frontotemporal Dementia : Comparisons between Hippocampal Volumetry, Tensor-Based Morphometry and Voxel-Based Morphometry. *PloS one*, 7 :e52531.
- Niarchakou, S. and Cech, M. (2018). *ATFCM OPERATIONS MANUAL*, volume 22.1. Network Manager - EUROCONTROL.

- Nir, T. M., Jahanshad, N., Villalon-Reina, J. E., Toga, A. W., Jack, C. R., Weiner, M. W., and Thompson, P. M. (2013). Effectiveness of regional DTI measures in distinguishing Alzheimer's disease, MCI, and normal aging. *NeuroImage : Clinical*, 3 :180–195.
- Novianingsih, K. and Hadiani, R. (2014). Modeling flight departure delay distributions. In *2014 International Conference on Computer, Control, Informatics and Its Applications (IC3INA)*, pages 30–34.
- Nusbaum, A. O., Tang, C. Y., Wei, T., Buchsbaum, M. S., and Atlas, S. W. (2000). Whole-brain diffusion MR histograms differ between MS subtypes. *Neurology*, 54(7) :1421–1427.
- Orrù, G., Pettersson-Yeo, W., Marquand, A. F., Sartori, G., and Mechelli, A. (2012). Using Support Vector Machine to identify imaging biomarkers of neurological and psychiatric disease : a critical review. *Neuroscience and Biobehavioral Reviews*, 36(4) :1140–1152.
- Park, H., Yang, J.-j., Seo, J., and Lee, J.-m. (2012). Dimensionality reduced cortical features and their use in the classification of Alzheimer's disease and mild cognitive impairment. *Neuroscience Letters*, 529(2) :123–127.
- Pearson, K. (1895). Contributions to the Mathematical Theory of Evolution. II. Skew Variation in Homogeneous Material. *Philosophical Transactions of the Royal Society of London Series A*, 186 :343–414.
- Pearson, K. (1905). "Das Fehlergesetz und Seine Verallgemeinerungen Durch Fechner und Pearson." A Rejoinder. *Biometrika*, 4(1/2) :169–212.
- Pellegrini, E., Ballerini, L., Hernandez, M. d. C. V., Chappell, F. M., González-Castro, V., Anblagan, D., Danso, S., Muñoz-Maniega, S., Job, D., Pernet, C., Mair, G., MacGillivray, T. J., Trucco, E., and Wardlaw, J. M. (2018). Machine learning of neuroimaging for assisted diagnosis of cognitive impairment and dementia : A systematic review. *Alzheimer's & Dementia : Diagnosis, Assessment & Disease Monitoring*, 10 :519–535.
- Pepe, M. S. and Thompson, M. L. (2000). Combining diagnostic test results to increase accuracy. *Biostatistics (Oxford, England)*, 1(2) :123–140.
- Petersen, R. C., Smith, G. E., Waring, S. C., Ivnik, R. J., Tangalos, E. G., and Kokmen, E. (1999). Mild cognitive impairment : clinical characterization and outcome. *Archives of Neurology*, 56(3) :303–308.
- Peyro Saint Paul, L., Debruyne, D., Bernard, D., Mock, D. M., and Defer, G. L. (2016). Pharmacokinetics and pharmacodynamics of MD1003 (high-dose biotin) in the treatment of progressive multiple sclerosis. *Expert Opinion on Drug Metabolism & Toxicology*, 12(3) :327–344.
- Pike, G. B., Stefano, N. d., Narayanan, S., Francis, G. S., Antel, J. P., and Arnold, D. L. (1999). Combined Magnetization Transfer and Proton Spectroscopic Imaging in the Assessment of Pathologic Brain Lesions in Multiple Sclerosis. *American Journal of Neuroradiology*, 20(5) :829–837.
- Pini, L., Pievani, M., Bocchetta, M., Altomare, D., Bosco, P., Cavedo, E., Galluzzi, S., Marizzoni, M., and Frisoni, G. B. (2016). Brain atrophy in Alzheimer's Disease and aging. *Ageing Research Reviews*, 30 :25–48.
- Pinson, C. (2015). Intérêt des séquences de transfert de magnétisation pour l'évaluation de l'activité des fistules périnéales dans la maladie de Crohn. page 103.
- Polman, C. H., Reingold, S. C., Banwell, B., Clanet, M., Cohen, J. A., Filippi, M., Fujihara, K., Havrdova, E., Hutchinson, M., Kappos, L., Lublin, F. D., Montalban, X., O'Connor, P., Sandberg-Wollheim, M., Thompson, A. J., Waubant, E., Weinshenker, B., and Wolinsky, J. S. (2011). Diagnostic criteria for multiple sclerosis : 2010 Revisions to the McDonald criteria. *Annals of Neurology*, 69(2) :292–302.

- Polman, C. H., Reingold, S. C., Edan, G., Filippi, M., Hartung, H.-P., Kappos, L., Lublin, F. D., Metz, L. M., McFarland, H. F., O'Connor, P. W., Sandberg-Wollheim, M., Thompson, A. J., Weinshenker, B. G., and Wolinsky, J. S. (2005). Diagnostic criteria for multiple sclerosis : 2005 revisions to the "McDonald Criteria". *Annals of Neurology*, 58(6) :840–846.
- Poupon, C. (1999). *Detection des faisceaux de fibres de la substance blanche pour l'étude de la connectivité anatomique cérébrale*. thesis, Paris, ENST.
- Preston, D. C. (2011). "MRI Basics" by David C Preston.
- Priyadharshini, Devi.S, A., and Askerunisa.A (2014). Kullback-Leibler Divergence Measurement for Clustering Based On Probability Distribution Similarity. *International Journal of Innovative Research in Science, Engineering and Technology*, 3(3).
- Promteangtrong, C., Kolber, M., Ramchandra, P., Moghbel, M., Houshmand, S., Schöll, M., Bai, H., Werner, T. J., Alavi, A., and Buchpiguel, C. (2015). Multimodality Imaging Approach in Alzheimer disease. Part I : Structural MRI, Functional MRI, Diffusion Tensor Imaging and Magnetization Transfer Imaging. *Dementia & Neuropsychologia*, 9(4) :318–329.
- Querbes, O. (2009). *Mesure de l'épaisseur corticale en IRM : application au diagnostic précoce individuel de la maladie d'Alzheimer et à la notion de réserve cognitive*. phd, Université de Toulouse, Université Toulouse III - Paul Sabatier.
- Querbes, O., Aubry, F., Pariente, J., Lotterie, J.-A., Démonet, J.-F., Duret, V., Puel, M., Berry, I., Fort, J.-C., Celsis, P., and Alzheimer's Disease Neuroimaging Initiative (2009). Early diagnosis of Alzheimer's disease using cortical thickness : impact of cognitive reserve. *Brain : A Journal of Neurology*, 132(Pt 8) :2036–2047.
- Racine, A. M., Brickhouse, M., Wolk, D. A., and Dickerson, B. C. (2018). The personalized Alzheimer's disease cortical thickness index predicts likely pathology and clinical progression in mild cognitive impairment. *Alzheimer's & Dementia : Diagnosis, Assessment & Disease Monitoring*, 10 :301–310.
- Rand, W. M. (1971). Objective Criteria for the Evaluation of Clustering Methods. *Journal of the American Statistical Association*, 66(336) :846–850.
- Rao, C. R. (1945). Information and accuracy attainable in the estimation of statistical parameters. *Bulletin of the Calcutta Mathematical Society*, 37 :81–91.
- Rebbah, S., Delahaye, D., Puechmorel, S., Maréchal, P., Nicol, F., Berry, I., and Initiative, F. A. D. N. (2019a). Classification of Multiple Sclerosis patients using a histogram-based K-Nearest Neighbors algorithm. In *Organization for Human Brain Mapping (OHBM2019)*.
- Rebbah, S., Delahaye, D., Puechmorel, S., Nicol, F., Marechal, P., and Berry, I. (2018a). A Combined MRI Biomarker Approach Using a Non-Standard Multiple Factor Analysis. In *2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, pages 1–6.
- Rebbah, S., Delahaye, D., Puechmorel, S., Nicol, F., Maréchal, P., and Berry, I. (2018b). Support Vector Machine-based classification of Alzheimer's Disease population using a combination of structural MRI biomarkers. Washington D.C., USA.
- Rebbah, S., Nicol, F., and Puechmorel, S. (2019b). The Geometry of the Generalized Gamma Manifold and an Application to Medical Imaging. *Mathematics*, 7(8) :674.
- Richert, N. D., Ostuni, J. L., Bash, C. N., Duyn, J. H., McFarland, H. F., and Frank, J. A. (1998). Serial whole-brain magnetization transfer imaging in patients with relapsing-remitting multiple sclerosis at baseline and during treatment with interferon beta-1b. *AJNR. American journal of neuroradiology*, 19(9) :1705–1713.

- Richert, N. D., Ostuni, J. L., Bash, C. N., Leist, T. P., McFarland, H. F., and Frank, J. A. (2001). Interferon beta-1b and intravenous methylprednisolone promote lesion recovery in multiple sclerosis. *Multiple Sclerosis (Houndmills, Basingstoke, England)*, 7(1) :49–58.
- Richiardi, J., Gschwind, M., Simioni, S., Annoni, J.-M., Greco, B., Hagmann, P., Schlupe, M., Vuilleumier, P., and Van De Ville, D. (2012). Classifying minimally disabled multiple sclerosis patients from resting state functional connectivity. *NeuroImage*, 62(3) :2021–2033.
- Ridha, B. H., Symms, M. R., Tozer, D. J., Stockton, K. C., Frost, C., Siddique, M. M., Lewis, E. B., MacManus, D. G., Boulby, P. A., Barker, G. J., Rossor, M. N., Fox, N. C., and Tofts, P. S. (2007). Magnetization transfer ratio in Alzheimer disease : comparison with volumetric measurements. *AJNR. American journal of neuroradiology*, 28(5) :965–970.
- Robert, P. and Escoufier, Y. (1976). A Unifying Tool for Linear Multivariate Statistical Methods : The RV-Coefficient. *Journal of the Royal Statistical Society Series C*, 25(3) :257–265.
- Rocca, M. A., Iannucci, G., Rovaris, M., Comi, G., and Filippi, M. (2003). Occult tissue damage in patients with primary progressive multiple sclerosis is independent of T2-visible lesions—a diffusion tensor MR study. *Journal of Neurology*, 250(4) :456–460.
- Ropele, S., Schmidt, R., Enzinger, C., Windisch, M., Martinez, N. P., and Fazekas, F. (2012). Longitudinal magnetization transfer imaging in mild to severe Alzheimer disease. *AJNR. American journal of neuroradiology*, 33(3) :570–575.
- Ross, S. (2009). *Introduction to Probability and Statistics for Engineers and Scientists*. Elsevier.
- Rovaris, M., Agosta, F., Sormani, M. P., Inglese, M., Martinelli, V., Comi, G., and Filippi, M. (2003). Conventional and magnetization transfer MRI predictors of clinical multiple sclerosis evolution : a medium-term follow-up study. *Brain*, 126(10) :2323–2332.
- Rovaris, M., Bozzali, M., Iannucci, G., Ghezzi, A., Caputo, D., Montanari, E., Bertolotto, A., Bergamaschi, R., Capra, R., Mancardi, G. L., Martinelli, V., Comi, G., and Filippi, M. (2002a). Assessment of normal-appearing white and gray matter in patients with primary progressive multiple sclerosis : a diffusion-tensor magnetic resonance imaging study. *Archives of Neurology*, 59(9) :1406–1412.
- Rovaris, M., Filippi, M., Falautano, M., Minicucci, L., Rocca, M. A., Martinelli, V., and Comi, G. (1998). Relation between MR abnormalities and patterns of cognitive impairment in multiple sclerosis. *Neurology*, 50(6) :1601–1608.
- Rovaris, M., Filippi, M., Minicucci, L., Iannucci, G., Santuccio, G., Possa, F., and Comi, G. (2000). Cortical/subcortical disease burden and cognitive impairment in patients with multiple sclerosis. *AJNR. American journal of neuroradiology*, 21(2) :402–408.
- Rovaris, M., Gass, A., Bammer, R., Hickman, S. J., Ciccarelli, O., Miller, D. H., and Filippi, M. (2005). Diffusion MRI in multiple sclerosis. *Neurology*, 65(10) :1526–1532.
- Rovaris, M., Iannucci, G., Falautano, M., Possa, F., Martinelli, V., Comi, G., and Filippi, M. (2002b). Cognitive dysfunction in patients with mildly disabling relapsing-remitting multiple sclerosis : an exploratory study with diffusion tensor MR imaging. *Journal of the Neurological Sciences*, 195(2) :103–109.
- Ruiz, E., Ramírez, J., Górriz, J. M., Casillas, J., and Alzheimer’s Disease Neuroimaging Initiative (2018). Alzheimer’s Disease Computer-Aided Diagnosis : Histogram-Based Analysis of Regional MRI Volumes for Feature Selection and Classification. *Journal of Alzheimer’s disease : JAD*, 65(3) :819–842.
- Río, J., Auger, C., and Rovira, (2017). MR Imaging in Monitoring and Predicting Treatment Response in Multiple Sclerosis. *Neuroimaging Clinics of North America*, 27(2) :277–287.

- Samper-González, J., Burgos, N., Bottani, S., Fontanella, S., Lu, P., Marcoux, A., Routier, A., Guillon, J., Bacci, M., Wen, J., Bertrand, A., Bertin, H., Habert, M.-O., Durrleman, S., Evgeniou, T., and Colliot, O. (2018). Reproducible evaluation of classification methods in Alzheimer's disease : Framework and application to MRI and PET data. *NeuroImage*, 183 :504–521.
- Santos, A. C., Narayanan, S., de Stefano, N., Tartaglia, M. C., Francis, S. J., Arnaoutelis, R., Caramanos, Z., Antel, J. P., Pike, G. B., and Arnold, D. L. (2002). Magnetization transfer can predict clinical evolution in patients with multiple sclerosis. *Journal of Neurology*, 249(6) :662–668.
- Schaefer, A. J., Johnson, E. L., Kleywegt, A. J., and Nemhauser, G. L. (2005). Airline Crew Scheduling Under Uncertainty. *Transportation Science*, 39(3) :340–348.
- Schmierer, K., Scaravilli, F., Altmann, D. R., Barker, G. J., and Miller, D. H. (2004). Magnetization transfer ratio and myelin in postmortem multiple sclerosis brain. *Annals of Neurology*, 56(3) :407–415.
- Schmitzer, B. and Schnörr, C. (2013). Object Segmentation by Shape Matching with Wasserstein Modes. In Heyden, A., Kahl, F., Olsson, C., Oskarsson, M., and Tai, X.-C., editors, *Energy Minimization Methods in Computer Vision and Pattern Recognition*, Lecture Notes in Computer Science, pages 123–136. Springer Berlin Heidelberg.
- Scholkopf, B. and Smola, A. J. (2001). *Learning with Kernels : Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, Cambridge, MA, USA.
- Schuff, N., Capizzano, A. A., Du, A. T., Amend, D. L., O'Neill, J., Norman, D., Kramer, J., Jagust, W., Miller, B., Wolkowitz, O. M., Yaffe, K., and Weiner, M. W. (2002). Selective reduction of N-acetylaspartate in medial temporal and parietal lobes in AD. *Neurology*, 58(6) :928–935.
- Sedel, F., Papeix, C., Bellanger, A., Touitou, V., Lebrun-Frenay, C., Galanaud, D., Gout, O., Lyon-Caen, O., and Tourbah, A. (2015). High doses of biotin in chronic progressive multiple sclerosis : a pilot study. *Multiple Sclerosis and Related Disorders*, 4(2) :159–169.
- Shima, H. (2013). Geometry of hessian structures. In Nielsen, F. and Barbaresco, F., editors, *Geometric Science of Information*, pages 37–55, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Shonk, T. K., Moats, R. A., Gifford, P., Michaelis, T., Mandigo, J. C., Izumi, J., and Ross, B. D. (1995). Probable Alzheimer disease : diagnosis with proton MR spectroscopy. *Radiology*, 195(1) :65–72.
- Smith, A. D. (2002). Imaging the progression of Alzheimer pathology through the brain. *Proceedings of the National Academy of Sciences*, 99(7) :4135–4137.
- Song, S.-K., Kim, J. H., Lin, S.-J., Brendza, R. P., and Holtzman, D. M. (2004). Diffusion tensor imaging detects age-dependent white matter changes in a transgenic mouse model with amyloid deposition. *Neurobiology of Disease*, 15(3) :640–647.
- Song, S.-K., Sun, S.-W., Ramsbottom, M. J., Chang, C., Russell, J., and Cross, A. H. (2002). Dysmyelination revealed through MRI as increased radial (but unchanged axial) diffusion of water. *NeuroImage*, 17(3) :1429–1436.
- Stacy, E. W. (1962). A generalization of the gamma distribution. *The Annals of Mathematical Statistics*, 33(3) :1187–1192.
- Stacy, E. W. and Mihram, G. A. (1965). Parameter estimation for a generalized gamma distribution. *Technometrics*, 7(3) :349–358.
- Steenwijk, M. D., Geurts, J. J. G., Daams, M., Tijms, B. M., Wink, A. M., Balk, L. J., Tewarie, P. K., Uitdehaag, B. M. J., Barkhof, F., Vrenken, H., and Pouwels, P. J. W. (2016). Cortical atrophy patterns in multiple sclerosis are non-random and clinically relevant. *Brain : A Journal of Neurology*, 139(Pt 1) :115–126.

- Steffen-Smith, E. A., Sarlls, J. E., Pierpaoli, C., Shih, J. H., Bent, R. S., Walker, L., and Warren, K. E. (2014). Diffusion Tensor Histogram Analysis of Pediatric Diffuse Intrinsic Pontine Glioma. *BioMed Research International*, 2014.
- Stejskal, E. O. and Tanner, J. E. (1965). Spin Diffusion Measurements : Spin Echoes in the Presence of a Time-Dependent Field Gradient. *The Journal of Chemical Physics*, 42(1) :288–292.
- Stern, Y. (2002). What is cognitive reserve? Theory and research application of the reserve concept. *Journal of the International Neuropsychological Society : JINS*, 8(3) :448–460.
- Strozyk, D., Blennow, K., White, L. R., and Launer, L. J. (2003). CSF Abeta 42 levels correlate with amyloid-neuropathology in a population-based autopsy study. *Neurology*, 60(4) :652–656.
- Su, J. Q. and Liu, J. S. (1993). Linear Combinations of Multiple Diagnostic Markers. *Journal of the American Statistical Association*, 88(424) :1350–1355.
- Su, Z., Zeng, W., Wang, Y., Lu, Z.-L., and Gu, X. (2015). Shape Classification Using Wasserstein Distance for Brain Morphometry Analysis. *Information Processing in Medical Imaging : Proceedings of the ... Conference*, 24 :411–423.
- Tambasco, N., Nigro, P., Romoli, M., Simoni, S., Parnetti, L., and Calabresi, P. (2015). Magnetization transfer MRI in dementia disorders, Huntington's disease and parkinsonism. *Journal of the Neurological Sciences*, 353(1-2) :1–8.
- Thambisetty, M., Wan, J., Carass, A., An, Y., Prince, J. L., and Resnick, S. M. (2010). Longitudinal changes in cortical thickness associated with normal aging. *NeuroImage*, 52(4) :1215–1223.
- Thompson, A. J., Banwell, B. L., Barkhof, F., Carroll, W. M., Coetzee, T., Comi, G., Correale, J., Fazekas, F., Filippi, M., Freedman, M. S., Fujihara, K., Galetta, S. L., Hartung, H. P., Kappos, L., Lublin, F. D., Marrie, R. A., Miller, A. E., Miller, D. H., Montalban, X., Mowry, E. M., Sorensen, P. S., Tintoré, M., Traboulsee, A. L., Trojano, M., Uitdehaag, B. M. J., Vukusic, S., Waubant, E., Weinshenker, B. G., Reingold, S. C., and Cohen, J. A. (2018). Diagnosis of multiple sclerosis : 2017 revisions of the McDonald criteria. *The Lancet Neurology*, 17(2) :162–173.
- Tievsky, A. L., Ptak, T., and Farkas, J. (1999). Investigation of apparent diffusion coefficient and diffusion tensor anisotropy in acute and chronic multiple sclerosis lesions. *AJNR. American journal of neuroradiology*, 20(8) :1491–1499.
- Tofts, P. S., Davies, G. R., and Dehmshki, J. (2004a). Histograms : Measuring Subtle Diffuse Disease. In *Quantitative MRI of the Brain*, pages 581–610. John Wiley & Sons, Ltd.
- Tofts, P. S., Davies, G. R., and Dehmshki, J. (2004b). *Quantitative MRI of the Brain : Principles of Physical Measurement*. John Wiley & Sons, Ltd, first edition edition.
- Tohka, J. (2014). Partial volume effect modeling for segmentation and tissue classification of brain magnetic resonance images : A review. *World Journal of Radiology*, 6(11) :855–864.
- Tortorella, C., Viti, B., Bozzali, M., Sormani, M. P., Rizzo, G., Gilardi, M. F., Comi, G., and Filippi, M. (2000). A magnetization transfer histogram study of normal-appearing brain tissue in MS. *Neurology*, 54(1) :186–193.
- Tourbah, A. and Berry, I. (2000). Contribution des techniques de résonance magnétique à la SEP. *Pathol Biol*, 48 :151–61.
- Tourbah, A., Lebrun-Frenay, C., Edan, G., Clanet, M., Papeix, C., Vukusic, S., De Sèze, J., Debouverie, M., Gout, O., Clavelou, P., Defer, G., Laplaud, D.-A., Moreau, T., Labauge, P., Brochet, B., Sedel, F., and Pelletier, J. (2016a). MD1003 (high-dose biotin) for the treatment of progressive multiple sclerosis : A randomised, double-blind, placebo-controlled study. *Multiple Sclerosis (Houndmills, Basingstoke, England)*, 22(13) :1719–1731.

- Tourbah, A., Lebrun-Frenay, C., Edan, G., Clanet, M., Papeix, C., Vukusic, S., Seze, J. D., Debouverie, M., Gout, O., Clavelou, P., Defer, G., Laplaud, D., Moreau, T., Labauge, P., Brochet, B., Sedel, F., and Pelletier, J. (2016b). 24-Month-Treatment with-MD1003 (High Doses of Biotin) in Progressive Multiple Sclerosis : Results of the MS-SPI Trial Extension Phase (S49.004). *Neurology*, 86(16 Supplement) :S49.004.
- Toussaint, G. (1974). Bibliography on estimation of misclassification. *IEEE Transactions on Information Theory*, 20(4) :472–479.
- Toussaint, N., Souplet, J.-C., and Fillard, P. (2007). MedINRIA : Medical Image Navigation and Research Tool by INRIA.
- Tozer, D. J., Davies, G. R., Altmann, D. R., Miller, D. H., and Tofts, P. S. (2006). Principal component and linear discriminant analysis of T1 histograms of white and grey matter in multiple sclerosis. *Magnetic Resonance Imaging*, 24(6) :793–800.
- Tozer, D. J. and Tofts, P. S. (2003). Removing spikes caused by quantization noise from high-resolution histograms. *Magnetic Resonance in Medicine*, 50(3) :649–653.
- Traboulsee, A., Dehmeshki, J., Peters, K. R., Griffin, C. M., Brex, P. A., Silver, N., Ciccarrelli, O., Chard, D. T., Barker, G. J., Thompson, A. J., and Miller, D. H. (2003). Disability in multiple sclerosis is related to normal appearing brain tissue MTR histogram abnormalities. *Multiple Sclerosis (Houndmills, Basingstoke, England)*, 9(6) :566–573.
- Tu, Y., Ball, M. O., and Jank, W. S. (2008). Estimating Flight Departure Delay Distributions—A Statistical Approach With Long-Term Trend and Short-Term Pattern. *Journal of the American Statistical Association*, 103(481) :112–125.
- Tuzel, O., Porikli, F., and Meer, P. (2008). Pedestrian Detection via Classification on Riemannian Manifolds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(10) :1713–1727.
- van Buchem, M. A., Grossman, R. I., Armstrong, C., Polansky, M., Miki, Y., Heyning, F. H., Boncoeur-Martel, M. P., Wei, L., Udupa, J. K., Grossman, M., Kolson, D. L., and McGowan, J. C. (1998). Correlation of volumetric magnetization transfer imaging with clinical data in MS. *Neurology*, 50(6) :1609–1617.
- van der Flier, W. M., van den Heuvel, D. M. J., Weverling-Rijnsburger, A. W. E., Bollen, E. L. E. M., Westendorp, R. G. J., van Buchem, M. A., and Middelkoop, H. A. M. (2002). Magnetization transfer imaging in normal aging, mild cognitive impairment, and Alzheimer’s disease. *Annals of Neurology*, 52(1) :62–67.
- van Waesberghe, J. H., Kamphorst, W., De Groot, C. J., van Walderveen, M. A., Castelijns, J. A., Ravid, R., Lycklama à Nijeholt, G. J., van der Valk, P., Polman, C. H., Thompson, A. J., and Barkhof, F. (1999). Axonal loss in multiple sclerosis lesions : magnetic resonance imaging insights into substrates of disability. *Annals of Neurology*, 46(5) :747–754.
- van Walderveen, M. A., Barkhof, F., Hommes, O. R., Polman, C. H., Tobi, H., Frequin, S. T., and Valk, J. (1995). Correlating MRI and clinical disease activity in multiple sclerosis : relevance of hypointense lesions on short-TR/short-TE (T1-weighted) spin-echo images. *Neurology*, 45(9) :1684–1690.
- Vanquin, L. (2015). *Biomarqueurs de la morphologie du cortex cérébral par imagerie par résonance magnétique (IRM) anatomique : application à la maladie d’Alzheimer*. thesis, Université Toulouse III - Paul Sabatier.
- Vemuri, P. and Jack, C. R. (2010). Role of structural MRI in Alzheimer’s disease. *Alzheimer’s Research & Therapy*, 2(4) :23.
- Vågberg, M., Lindqvist, T., Ambarki, K., Warntjes, J. B. M., Sundström, P., Birgander, R., and Svenningsson, A. (2013). Automated determination of brain parenchymal fraction in multiple sclerosis. *AJNR. American journal of neuroradiology*, 34(3) :498–504.

- Wagner, M. W., Narayan, A. K., Bosemani, T., Huisman, T. A. G. M., and Poretti, A. (2016). Histogram Analysis of Diffusion Tensor Imaging Parameters in Pediatric Cerebellar Tumors. *Journal of Neuroimaging : Official Journal of the American Society of Neuroimaging*, 26(3) :360–365.
- Wasserman, L. (2018). Optimal Transport and Wasserstein Distance.
- Wasserstein, L. N. (1969). Markov processes over denumerable products of spaces describing large systems of automata. *Prob. Inf. Transmission*, 5.
- Webb, A. R. and Copsey, K. D. (2011). Introduction to Statistical Pattern Recognition. In *Statistical Pattern Recognition*, pages 1–32. John Wiley & Sons, Ltd.
- Weill, B. and Batteux, F. (2003). *Immunopathologie et réactions inflammatoires*. De Boeck Supérieur. Google-Books-ID : W6PXmDcywPOC.
- Werring, D. J., Brassat, D., Droogan, A. G., Clark, C. A., Symms, M. R., Barker, G. J., MacManus, D. G., Thompson, A. J., and Miller, D. H. (2000). The pathogenesis of lesions and normal-appearing white matter changes in multiple sclerosis : a serial diffusion MRI study. *Brain : A Journal of Neurology*, 123 (Pt 8) :1667–1676.
- Werring, D. J., Clark, C. A., Barker, G. J., Thompson, A. J., and Miller, D. H. (1999). Diffusion tensor imaging of lesions and normal-appearing white matter in multiple sclerosis. *Neurology*, 52(8) :1626–1632.
- Weston, P. S., Simpson, I. J., Ryan, N. S., Ourselin, S., and Fox, N. C. (2015). Diffusion imaging changes in grey matter in Alzheimer’s disease : a potential marker of early neurodegeneration. *Alzheimer’s Research & Therapy*, 7(1).
- Wilson, M., Morgan, P., Lin, X., Turner, B., and Blumhardt, L. (2001). Quantitative diffusion weighted magnetic resonance imaging, cerebral atrophy, and disability in multiple sclerosis. *Journal of Neurology, Neurosurgery, and Psychiatry*, 70(3) :318–322.
- Wolff, S. D. and Balaban, R. S. (1989). Magnetization transfer contrast (MTC) and tissue water proton relaxation in vivo. *Magnetic Resonance in Medicine*, 10(1) :135–144.
- Wolz, R., Julkunen, V., Koikkalainen, J., Niskanen, E., Zhang, D. P., Rueckert, D., Soininen, H., Lötjönen, J., and Initiative, t. A. D. N. (2011). Multi-Method Analysis of MRI Images in Early Diagnostics of Alzheimer’s Disease. *PLOS ONE*, 6(10) :e25446.
- Wottschel, V. (2017). *Supervised machine learning in multiple sclerosis : applications to clinically isolated syndromes*. Doctoral, UCL (University College London).
- Xu, N., Sherry, L., and Laskey, K. B. (2008). Multifactor Model for Predicting Delays at U.S. Airports. *Transportation Research Record*, 2052(1) :62–71.
- Xu, T., Fang, Y., Rong, A., and Wang, J. (2015). Flexible combination of multiple diagnostic biomarkers to improve diagnostic accuracy. *BMC Medical Research Methodology*, 15.
- Ye, D. H., Pohl, K. M., and Davatzikos, C. (2011). Semi-supervised Pattern Classification : Application to Structural MRI of Alzheimer’s Disease. In *2011 International Workshop on Pattern Recognition in NeuroImaging*, pages 1–4.
- Yin, J. and Tian, L. (2014). Optimal linear combinations of multiple diagnostic biomarkers based on Youden index. *Statistics in Medicine*, 33(8) :1426–1440.
- Young, R., Babb, J., Law, M., Pollack, E., and Johnson, G. (2007). Comparison of region-of-interest analysis with three different histogram analysis methods in the determination of perfusion metrics in patients with brain gliomas. *Journal of magnetic resonance imaging : JMRI*, 26(4) :1053–1063.

- Yu, C. S., Lin, F. C., Liu, Y., Duan, Y., Lei, H., and Li, K. C. (2008). Histogram analysis of diffusion measures in clinically isolated syndromes and relapsing-remitting multiple sclerosis. *European Journal of Radiology*, 68(2) :328–334.
- Zhu, X. (2006). *Semi-Supervised Learning Literature Survey*.
- Zhu, X., Schuff, N., Kornak, J., Soher, B., Yaffe, K., Kramer, J. H., Ezekiel, F., Miller, B. L., Jagust, W. J., and Weiner, M. W. (2006). Effects of Alzheimer disease on fronto-parietal brain N-acetyl aspartate and myo-inositol using magnetic resonance spectroscopic imaging. *Alzheimer Disease and Associated Disorders*, 20(2) :77–85.
- Zieger, S., Weik, N., and Nießen, N. (2018). The influence of buffer time distributions in delay propagation modelling of railway networks. *Journal of Rail Transport Planning & Management*, 8(3) :220–232.
- Zivadinov, R., Uher, T., Hagemeyer, J., Vaneckova, M., Ramasamy, D. P., Tyblova, M., Bergsland, N., Seidl, Z., Dwyer, M. G., Krasensky, J., Havrdova, E., and Horakova, D. (2016). A serial 10-year follow-up study of brain atrophy and disability progression in RRMS patients. *Multiple Sclerosis (Houndmills, Basingstoke, England)*, 22(13) :1709–1718.

Annexes

Estimation de la courbure corticale

La courbure peut se définir comme étant le caractère plus ou moins incurvé d'un objet. En 2D, la courbure correspond à la variation du vecteur tangent à la courbe au voisinage d'un point. En trois dimensions, la courbure donne une indication sur le degré d'éloignement de la surface au plan. Comme il y a une infinité de directions possibles en 3D, l'indice de courbure en un point d'une surface doit être exprimé par rapport à un repère particulier de telle sorte qu'il soit un invariant spécifique de la surface et non du repère utilisé.

On considère une surface Σ définie localement par l'équation $z = h(x, y)$. On peut définir cette surface par une représentation paramétrique de la forme :

$$\sigma : (x, y) \mapsto \sigma(x, y) = \begin{pmatrix} x \\ y \\ z = h(x, y) \end{pmatrix}$$

En trois dimensions, le calcul de la courbure est assez complexe car la fonction décrivant l'équation des interfaces du cortex $z = h(x, y)$ est inconnue. On considère alors que la surface est suffisamment régulière au voisinage d'un point $M(x_M, y_M, z_M)$. Ainsi ses dérivées partielles sont définies et continues au voisinage de ce point. La surface admet donc un développement limité à l'ordre 2 de $z = h(x, y)$ en M :

$$\begin{aligned} z = h(x, y) = & \frac{1}{2} \frac{\partial^2 h}{\partial x^2}(x_M, y_M)(x - x_M)^2 + \frac{1}{2} \frac{\partial^2 h}{\partial y^2}(x_M, y_M)(y - y_M)^2 \\ & + \frac{\partial^2 h}{\partial x \partial y}(x_M, y_M)(x - x_M)(y - y_M) + \frac{\partial h}{\partial x}(x_M, y_M)(x - x_M) \\ & + \frac{\partial h}{\partial y}(x_M, y_M)(y - y_M) + h(x_M, y_M) \end{aligned}$$

$$\begin{aligned} z = & c_1(x - x_M)^2 + c_2(y - y_M)^2 + c_3(x - x_M)(y - y_M) + c_4(x - x_M) \\ & + c_5(y - y_M) + h(x_M, y_M) \end{aligned}$$

En définissant un voisinage centré au point M et on obtient :

$$\begin{aligned} z = h(x, y) = & \frac{1}{2} h_{xx} x^2 + \frac{1}{2} h_{yy} y^2 + h_{xy} xy + h_x x + h_y y \\ = & c_1 x^2 + c_2 y^2 + c_3 xy + c_4 x + c_5 y \end{aligned}$$

On obtient la courbure S au point M en estimant les coefficients polynomiaux $c_{1...5}$.

$$S = \frac{(1 + c_4^2)c_2 - 2c_4c_5c_3 + (1 + c_5^2)c_1}{2(1 + c_4^2 + c_5^2)^{\frac{3}{2}}}$$

On effectue l'estimation de la courbure en un point des surfaces corticales en modélisant localement la surface du voisinage de ce point par une fonction

polynomiale quadratique.

On définit un point M de la surface dont on veut estimer la courbure et V un voisinage de ce point. On choisit V centré sur M, V étant un cube de côté n, $n \geq 5$. On définit ensuite Λ , l'intersection de la surface avec le voisinage V. Pour tout point de Λ , on a :

$$z = h(x, y) = c_1x^2 + c_2y^2 + c_3xy + c_4x + c_5y + c_6$$

où $c_1 \dots c_5$ sont des coefficients à déterminer ($c_6 = 0$ car V est centré sur M) et (x, y, z) est un repère local orthonormé centré en M et défini pour que z soit le vecteur unitaire normal à la surface corticale et orienté vers l'extérieur du cortex pour la surface corticale externe et vers l'intérieur du cortex pour la surface corticale interne.

On doit donc résoudre un système de la forme :

$$\Pi C^t = Z$$

Ce système est résolu au sens des moindres carrés et consiste à déterminer C^t qui minimise la norme euclidienne des résidus. On obtient alors :

$$C^t = (\Pi^t \Pi)^{-1} \Pi^t Z$$

On obtient la courbure (notée S) au point M en estimant les coefficients polynomiaux $c_1 \dots c_5$ et donc les dérivées partielles de $z = h(x, y)$:

$$S = \frac{(1 + h_x^2)h_{yy} - 2h_xh_yh_{xy} + (1 + h_y^2)h_{xx}}{2(1 + h_x^2 + h_y^2)^{3/2}} = \frac{(1 + c_4^2)c_2 - 2c_4c_5c_3 + (1 + c_5^2)c_1}{2(1 + c_4^2 + c_5^2)^{\frac{3}{2}}} \quad (5.35)$$

Sélection des caractéristiques pour la classification de la population de la maladie d'Alzheimer

Le tableau 5.6 représente, pour chaque classification, les variables optimales.

Variables	AD/HC	HC/sMCI	HC/pMCI	AD/sMCI	AD/pMCI	sMCI/pMCI
moy_EC	×	×	×	×	×	×
med_EC	×	×	×	×		
EC.Q10	×	×	×	×	×	×
EC.Q25	×	×	×	×	×	×
EC.Q75	×	×	×	×		×
EC.Q90	×	×	×	×		
EC.PL						
EC.PH	×	×		×		
moy_GyrExt				×		
med_GyrExt	×			×		
Q10_GyrExt					×	×
Q25_GyrExt						×
Q75_GyrExt	×					
Q90_GyrExt	×					
GyrExt.PL						
GyrExt.PH						
moy_SilExt				×	×	
med_SilExt		×			×	
Q10_SilExt					×	
Q25_SilExt					×	×
Q75_SilExt		×	×		×	×
Q90_SilExt						×
SilExt.PL						
SilExt.PH						
moy_GyrInt						
med_GyrInt						
Q10_GyrInt						
Q25_GyrInt						
Q75_GyrInt						
Q90_GyrInt			×			
GyrInt.PL						
GyrInt.PH						
moy_SilInt						
med_SilInt			×			
Q10_SilInt					×	×
Q25_SilInt						
Q75_SilInt						
Q90_SilInt						
SilInt.PL			×			
SilInt.PH						

Tableau 5.6 – Les variables optimale, sélectionner à l'aide du gain d'information pour chaque classification. Ces variables sont intégrées dans les classifieurs SVMs.

Mesures de dissimilarité

Dissimilarités	Familles	Formules
Euclidienne L_2	Minkowski_Lp	$\sqrt{\sum_{i=1}^d P_i - Q_i ^2}$
Manhattan L_1	Minkowski_Lp	$\sum_{i=1}^d P_i - Q_i $
Chebyshev L_∞	Minkowski_Lp	$\max_i P_i - Q_i $
Sorensen	Minkowski_Lp	$\frac{\sum_{i=1}^d P_i - Q_i }{\sum_{i=1}^d P_i + Q_i }$
Soergel	Minkowski_Lp	$\frac{\sum_{i=1}^d P_i - Q_i }{\sum_{i=1}^d \max(P_i, Q_i)}$
Lorentzian	Minkowski_Lp	$\sum_{i=1}^d \ln(1 + P_i - Q_i)$
Kulczynski	Minkowski_Lp	$\frac{\sum_{i=1}^d P_i - Q_i }{\sum_{i=1}^d \min(P_i, Q_i)}$
Gower	Minkowski_Lp	$\frac{1}{d} \sum_{i=1}^d P_i - Q_i $
Canberra	Minkowski_Lp	$\sum_{i=1}^d \frac{ P_i - Q_i }{P_i + Q_i}$
Intersection	Intersection	$\sum_{i=1}^d \min(P_i, Q_i)$
Non-intersection	Intersection	$\frac{1}{2} \sum_{i=1}^d P_i - Q_i $
Tanimoto	Intersection	$\frac{\sum_{i=1}^d (\max(P_i, Q_i) - \min(P_i, Q_i))}{\max(P_i, Q_i)}$
Motyka	Intersection	$\frac{\sum_{i=1}^d \min(P_i, Q_i)}{\sum_{i=1}^d (P_i + Q_i)}$
Czekanowski	Intersection	$\frac{\sum_{i=1}^d P_i - Q_i }{\sum_{i=1}^d (P_i + Q_i)}$
Wavehedges	Intersection	$\sum_{i=1}^d \frac{ P_i - Q_i }{\max(P_i, Q_i)}$
Kulczynski_s	Intersection	$\frac{\sum_{i=1}^d \min(P_i, Q_i)}{\sum_{i=1}^d P_i - Q_i }$
Jaccard	Inner Product	$\frac{\sum_{i=1}^d P_i^2 + \sum_{i=1}^d Q_i^2 - \sum_{i=1}^d P_i Q_i}{\sum_{i=1}^d (P_i - Q_i)^2}$
Dice	Inner Product	$\frac{\sum_{i=1}^d (P_i - Q_i)^2}{\sum_{i=1}^d P_i^2 + \sum_{i=1}^d Q_i^2}$
Inner product	Inner Product	$\sum_{i=1}^d P_i Q_i$
Cosine	Inner Product	$\frac{\sum_{i=1}^d P_i Q_i}{\sqrt{\sum_{i=1}^d P_i^2} \sqrt{\sum_{i=1}^d Q_i^2}}$
Kumar-Hassebrook	Inner Product	$\frac{\sum_{i=1}^d P_i^2 + \sum_{i=1}^d Q_i^2 - \sum_{i=1}^d P_i Q_i}{\sum_{i=1}^d P_i - Q_i }$
Ruzicka	Inner Product	$\frac{\sum_{i=1}^d \min(P_i, Q_i)}{\sum_{i=1}^d \min(P_i, Q_i)}$
Moyenne harmonique	Inner Product	$2 \sum_{i=1}^d \frac{P_i Q_i}{P_i + Q_i}$
Squared chord	Fidelity	$\sum_{i=1}^d (\sqrt{P_i} - \sqrt{Q_i})^2$
Matusita	Fidelity	$\sqrt{2 - 2 \sum_{i=1}^d \sqrt{P_i Q_i}}$
Hellinger	Fidelity	$2 \sqrt{1 - \sum_{i=1}^d \sqrt{P_i Q_i}}$
Bhattacharyya	Fidelity	$-\ln \sum_{i=1}^d \sqrt{P_i Q_i}$
Fidelity	Fidelity	$\sum_{i=1}^d \sqrt{P_i Q_i}$
Euclidienne quadratique	L2 quadratique	$\sum_{i=1}^d (P_i - Q_i)^2$
Chi quadratique	L2 quadratique	$\sum_{i=1}^d \frac{(P_i - Q_i)^2}{P_i + Q_i}$
Probabilistic symmetric	L2 quadratique	$2 \sum_{i=1}^d \frac{(P_i - Q_i)^2}{P_i + Q_i}$
Pearson	L2 quadratique	$\sum_{i=1}^d \frac{(P_i - Q_i)^2}{Q_i}$
Neyman	L2 quadratique	$\sum_{i=1}^d \frac{(P_i - Q_i)^2}{P_i}$
Additive symmetric	L2 quadratique	$\sum_{i=1}^b \frac{(P_i - Q_i)^2 (P_i + Q_i)}{P_i Q_i}$
Divergence	L2 quadratique	$2 \sum_{i=1}^d \frac{(P_i - Q_i)^2}{(P_i + Q_i)^2}$
Clark	L2 quadratique	$\sqrt{\sum_{i=1}^b \frac{(P_i - Q_i)^2}{(P_i + Q_i)^2}}$
Kullback-Leibler	Shannon entropie	$\sum_{i=1}^d P_i \ln \frac{P_i}{Q_i}$
Jensen-shannon	Shannon entropie	$\frac{1}{2} \left[\sum_{i=1}^d P_i \left(\ln \frac{2P_i}{P_i + Q_i} \right) + \sum_{i=1}^d Q_i \left(\ln \frac{2Q_i}{P_i + Q_i} \right) \right]$
Topsoe	Shannon entropie	$\sum_{i=1}^d \left[P_i \left(\ln \frac{2P_i}{P_i + Q_i} \right) + Q_i \left(\ln \frac{2Q_i}{P_i + Q_i} \right) \right]$
Jensen difference	Shannon entropie	$\sum_{i=1}^d \left[\frac{P_i \ln P_i + Q_i \ln Q_i}{2} - \left(\frac{P_i + Q_i}{2} \right) \ln \left(\frac{P_i + Q_i}{2} \right) \right]$
Jeffreys	Shannon entropie	$\sum_{i=1}^d (P_i - Q_i) \ln \frac{P_i}{Q_i}$
K-Divergence	Shannon entropie	$\sum_{i=1}^d P_i \ln \frac{2P_i}{P_i + Q_i}$
Average	Combinaisons	$\frac{\sum_{i=1}^d P_i - Q_i + \max_i P_i - Q_i }{2}$
Taneja	Combinaisons	$\sum_{i=1}^d \left(\frac{P_i + Q_i}{2} \right) \ln \left(\frac{P_i + Q_i}{2 \sqrt{P_i Q_i}} \right)$
Kumar-Johnson	Combinaisons	$\sum_{i=1}^d \frac{(P_i^2 - Q_i^2)^2}{2(P_i Q_i)^{3/2}}$

Tableau 5.7 – Tableau représentant les formules des 45 différentes mesures de dissimilarité, réparties en 7 familles.
P et Q représentent des distributions

Les régions d'intérêt du cortex

	Zone name	Brodmann areas
1	SensMotor_L	1 2 3 4 5 6
2	Frontal_L	8 9 10 11 12 25 46 47
3	Broca_L	44 45
4	Audit_L	41 42
5	Mes_Tem_L	27 28 34 35 36 48
6	Parietal_L	39 40 7
7	Temp_Lat_L	15 20 21 22 37 38
8	Cing_Post_L	23 29 30 31
9	Cing_Ant_L	24 32 33
10	Occ_L	17 18 19
11	Insula_audit_L	13 14 43
12	SensMotor_R	49 50 51 52 53 54
13	Frontal_R	56 57 58 59 60 73 94 95
14	Broca_R	92 93
15	Audit_R	89 90
16	Mes_Tem_R	75 76 82 83 84 96
17	Parietal_R	87 88 55
18	Temp_Rat_R	63 68 69 70 85 86
19	Cing_Post_R	71 77 78 79
20	Cing_Ant_R	72 80 81
21	Occ_R	65 66 67
22	Insula_audit_R	61 62 91

Tableau 5.8 – Définition des 22 régions d'intérêt du cortex. Ces 22 ROI correspondent à 11 zones de l'hémisphère droit ("R") et les 11 zones correspondantes de l'hémisphère gauche ("L").

Variables	AD/HC
EC_moy	x
SensMotor_L	
Frontal_L	
Broca_L	
Audit_L	
Mes_Tem_L	x
Parietal_L	
Temp_Lat_L	x
Cing_Post_L	x
Cing_Ant_L	x
Occ_L	
Insula_Audit_L	
SensMotor_R	
Frontal_R	
Broca_R	
Audit_R	
Mes_Tem_R	x
Parietal_R	
Temp_Lat_R	x
Cing_Post_R	x
Cing_Ant_R	x
Occ_R	
Insula_Audit_R	

Tableau 5.9 – Les 10 variables les plus discriminantes pour la classification AD/HC d'après la mesure du gain d'information.

Contributions scientifiques

REVUES INTERNATIONALES

- Rebbah, S., Nicol, F., & Puechmorel, S. (2019). The Geometry of the Generalized Gamma Manifold and an Application to Medical Imaging. *Mathematics*, 7(8), 674.
<https://doi.org/10.3390/math7080674>

CONFÉRENCES INTERNATIONALES

- Rebbah, S., Delahaye, D., Puechmorel, S., Maréchal, P., Nicol, F., Berry, I., & For Alzheimer's Disease Neuroimaging Initiative. (2019, June). Classification of Multiple Sclerosis patients using a histogram-based K-Nearest Neighbors algorithm. Presented at the *Organization for Human Brain Mapping (OHBM 2019)*, Rome, Italy.
https://files.aievolution.com/hbm1901/abstracts/50707/M442_REBBAH.pdf
- Rebbah, S., Delahaye, D., Puechmorel, S., Nicol, F., Maréchal, P., & Berry, I. (2018, October). Support Vector Machine-based classification of Alzheimer's Disease population using a combination of structural MRI biomarkers. Presented at the *ISMRM Workshop on Machine Learning Part II*, Washington D.C., USA.
<https://hal-enac.archives-ouvertes.fr/hal-01998187>
- Rebbah, S., Delahaye, D., Puechmorel, S., Nicol, F., Marechal, P., & Berry, I. (2018, October). A Combined MRI Biomarker Approach Using a Non-Standard Multiple Factor Analysis. In *11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, Beijing, China. *IEEE* (pp. 1–6).
<https://doi.org/10.1109/CISP-BMEI.2018.8633261>

CONFÉRENCES NATIONALES

- Rebbah, S., & Puechmorel, S. (2019, May). The geometry of the generalized gamma manifold with an application in medical imaging. Presented at the *Société de Mathématiques Appliquées et Industrielles (SMAI 2019)*, Morbihan.
- Rebbah, S., Delahaye, D., Puechmorel, S., Nicol, F., Marechal, P., & Berry, I. (2018, June). Aggregation of MRI biomarkers in multiple sclerosis clinical trials using geometric PCA. Presented at *ARSEP - Spain FRENCH MS MEETING*, Paris.
- Rebbah, S., Chauveau, N., Berry, I., & Aubry, F. (2017, February). Vector generalized linear model applied to cortical thickness in neurodegenerative disease follow-up. Presented at *ARSEP - 12th annual MRI workshop*, Paris.

Résumés

Distance entre distributions : application à l'Imagerie Médicale et à l'Aéronautique

Dans le domaine médical, au cours des deux dernières décennies, un nombre croissant de méthodes d'analyse d'images quantitatives ont été développées dont l'analyse par régions d'intérêt, l'analyse voxel à voxel et l'analyse d'histogrammes. Cette dernière est largement utilisée dans la cadre de la recherche sur la sclérose en plaques afin de quantifier les changements pathologiques diffus particulièrement présents dans cette maladie. Un inconvénient de cette approche est que l'ensemble des informations incluses dans l'histogramme n'est pas exploité; seules des mesures arbitraires sont choisies pour décrire l'histogramme; incluant la moyenne, la médiane, les centiles... Ainsi dans un premier lieu, nous avons proposé d'intégrer dans un classifieur toute l'information incluse dans l'histogramme et non pas seulement quelques descripteurs locaux, dans le but d'améliorer les performances de classification des populations de sclérose en plaques (groupes dans un essai thérapeutique et in fine groupes de pronostics différents). Par la suite, étant donné que l'histogramme est une estimation trop simpliste d'une distribution de probabilité, nous présentons l'une des applications possibles de la géométrie de l'information sur les distributions de probabilité et démontrons l'intérêt de l'utilisation de la géométrie non-euclidienne dans le contexte de la classification des populations de la maladie d'Alzheimer.

Nous avons notamment fait l'analogie avec le domaine de l'aéronautique, plus précisément dans l'étude des retards aéroportuaires. En effet, l'analyse actuellement réalisée se situe au niveau macroscopique et fournit un indicateur de retard moyen, sans tenir compte des mécanismes intermédiaires pouvant conduire au retard final. Ainsi, dans le cadre de la classification des retards aéroportuaires, tout comme dans les applications médicales, nous avons remplacé l'indicateur moyen par un modèle statistique paramétrique plus complet : les distributions de probabilité.

Distance between distributions : application to Medical Imaging and Aeronautics

In the medical field, over the past two decades, a growing number of quantitative image analysis have been developed including regions of interest analysis, voxel-by-voxel analysis and histogram analysis. The latter is widely use in Multiple Sclerosis research to quantify the diffuse pathological prominent in this disease. A disadvantage of this approach is that all the information included in the histogram is not exploited; only arbitrary measures are chosen to describe the histogram; including the average, the median, the percentiles... Thus, first, we proposed to integrate in a classifier all the information included in the histogram and not just some local descriptors, in order to improve the classification performance of the Multiple Sclerosis populations (i.e. groups in therapeutic trials and in fine groups at different prognosis). Thereafter, given that the histogram is an overly simplistic estimate of a probability distribution, we present one of the possible applications of information geometry on probability distributions and we demonstrate the interest of using non-Euclidean geometry in the context of the Alzheimer's disease population classification. Furthermore, we have made the analogy with the field of aeronautics, specifically in the study of flight delays. Indeed, the analysis currently carried out is at a macroscopic level and provides an indicator of average delay, without considering the intermediate mechanisms that may lead to the final delays. Thus, in the clustering of airport delays, we have replaced the average indicator with a more complete parametric statistical model : Distributions.