



HAL
open science

Reconnaissance automatique des sillons corticaux

Matthieu Perrot

► **To cite this version:**

Matthieu Perrot. Reconnaissance automatique des sillons corticaux. Mathématiques [math]. École normale supérieure de Cachan - ENS Cachan, 2009. Français. NNT: . tel-00457072

HAL Id: tel-00457072

<https://theses.hal.science/tel-00457072>

Submitted on 16 Feb 2010

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

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

**THESE DE DOCTORAT
DE L'ECOLE NORMALE SUPERIEURE DE CACHAN**

Présentée par

Monsieur Matthieu PERROT

pour obtenir le grade de

DOCTEUR DE L'ECOLE NORMALE SUPERIEURE DE CACHAN

Domaine :

MATHEMATIQUES

Sujet de la thèse :

Reconnaissance Automatique des Sillons Corticaux

Thèse présentée et soutenue à Neurospin le 26 octobre 2009 devant le jury composé de :

Christian BARILLOT	Directeur de recherche - CNRS	Rapporteur
Olivier COULON	Chargé de recherche - CNRS/Marseille	Examineur
Jérôme LACAILLE	Expert en Algorithmes - SNECMA	Directeur de thèse
Jean-François MANGIN	Directeur du laboratoire LNAO / Neurospin / CEA	Directeur de thèse
Xavier PENNEC	Directeur de recherche - INRIA	Rapporteur
Jean RÉGIS	Professeur des universités-Praticien hospitalier - Marseille	Examineur
Denis RIVIÈRE	Ingénieur-Chercheur - LNAO / Neurospin / CEA	Encadrant
Alain TROUVÉ	Professeur - ENS Cachan	Président

Centre de Mathématiques et de Leurs Applications
(ENS CACHAN/UMR 8536)
61, avenue du Président Wilson, 94235 CACHAN CEDEX (France)

Remerciements

La motivation de départ d'une thèse n'est pas toujours très originale et résulte souvent d'une succession d'intersections d'opportunités et de hasards. Naturellement, c'est au travers de son développement et de son dénouement qu'un projet prend sa pleine mesure. Finalement, j'en retiendrai bien plus qu'une accumulation de savoir, puisque ce travail m'a permis de vivre une véritable aventure personnelle et humaine, qui ne m'aurait sans doute pas autant profité sans sa traditionnelle traversée du désert.

Une grande partie de la richesse et de l'intérêt que j'ai porté à mon sujet de thèse tenait au challenge que celui-ci représentait. Je tiens à remercier tout particulièrement mon directeur de thèse, Jean-François Mangin, de m'avoir estimé apte à relever ce défi, d'avoir su me transmettre sa vision structurelle de l'anatomie et une forme de passion, le meilleur moteur à mon opiniâtreté. Je remercie également mon deuxième directeur de thèse, Jérôme Lacaille, d'avoir su éveiller en moi un certain intérêt pour la formalisation mathématique, ce qui m'a permis de progresser dans ma démarche scientifique. Mon plus grand soutien au quotidien demeure mon encadrant, Denis Rivière, qui m'a apporté un appui technique essentiel et une oreille attentive à mes nombreuses réflexions. Je ne les remercierai jamais assez de la confiance qu'ils m'ont accordée qui m'a permis de proposer des solutions en rupture avec le passé.

Je tenais à souligner la chance que j'ai eu de connaître la naissance du centre Neurospin et la croissance rapide de ses laboratoires de recherche, qui a démultiplié les interactions scientifiques. Cette aventure n'aurait pas été aussi enrichissante sans les nombreuses rencontres qu'elle a suscitées, tout particulièrement celles nées dans le cadre de collaborations. Je pense notamment à Guillaume Auzias avec qui nous formions la *sulcal team* dernier rempart face aux méthodes icôniques, Marine Fouquet qui a eu le courage d'utiliser mes outils en avant première, Merlin Keller qui m'a permis de mettre un pied dans le monde de l'imagerie fonctionnelle, Alan Tucholka avec qui j'ai passé de nombreuses heures à labéliser manuellement des sillons.

J'ai eu la chance de partager mon bureau avec Irina Kezele, Pierre Fillard et Julien Lefèvre, trois post-doctorants que je souhaite vivement remercier pour avoir assouvi mon besoin incessant d'échange scientifique, m'avoir ouvert à d'autres thématiques et initié au monde de l'après-thèse. Je remercie également Benjamin Thyreau, qui m'a précédé au SHFJ et a largement influencé mon choix de stage de DEA qui m'a conduit à cette thèse, en ventant les mérites de cet environnement de recherche. J'ai une pensée particulière pour Philippe Cuicui, Édouard Duchesnay et Bertrand Thirion pour leur soutien et les échanges que nous avons eus et qui n'ont pas manqué d'alimenter mon inspiration scientifique. Sans les lister de peur d'en oublier, je souhaite remercier l'ensemble des stagiaires, thésards, post-doc, collaborateurs et permanents avec qui j'ai partagé mes moments de détente pendant ces quelques années et qui ont fait naître de véritables amitiés.

L'ensemble des membres de mon jury ont contribué, chacun à sa façon, à l'émergence de perspectives intéressantes qu'elles soient méthodologiques ou applicatives. Je tenais à les remercier de la richesse des échanges que nous avons pu avoir, en espérant fortement les voir se concrétiser dans l'avenir.

Et je terminerai cette liste par la personne la plus chère à mes yeux, ma compagne Élise qui a subi au quotidien mes passages à vide, mes absences répétées, mon travail nocturne. Elle a su rester à l'écoute et m'a apporté un regard extérieur éclairé sur des considérations didactiques ou artistiques.

Je tenais à rendre un hommage particulier à mon grand-père Roland et ma grande-tante Jacqueline qui nous ont quitté pendant ma thèse et à qui j'aurais aimé consacrer d'avantage de temps. Cette thèse leur est dédiée.

Résumé

La mise en évidence de biomarqueurs spécifiques de pathologies cérébrales à l'échelle d'une population reste extrêmement difficile compte tenu de la variabilité inter-individuelle de la topographie sulco-gyrale. Cette thèse propose de répondre à cette difficulté par l'identification automatique de 125 structures sulcales et leur mise en correspondance au travers des individus, à partir d'une base de 62 sujets labélisés manuellement. En s'appuyant sur la théorie des racines sulcales, les plissements corticaux sont découpés en entités élémentaires à labéliser.

Dans une première phase, l'approche structurelle proposée précédemment par Jean-François Mangin et Denis Rivière a été revisitée pour faire face aux nombreux descripteurs morphométriques impliqués dans le processus d'identification. Dans une deuxième phase, cette modélisation a été reconsidérée intégralement au profit d'un cadre Bayésien exploitant des informations localisées (positions ou directions) négligées jusqu'alors, autorisant ainsi des schémas d'optimisation efficace. Dans ce cadre, la normalisation des données est essentielle; cette question a été traitée sous la forme d'un processus de recalage affine global ou local à chaque sillon, de façon couplée au problème d'identification des sillons. Dans l'optique d'introduire plus d'information structurelle, une modélisation Markovienne traduisant une vue localisée de l'agencement entre plissements corticaux voisins a été introduite avec succès pour atteindre un taux de reconnaissance de plus de 86% pour chaque hémisphère.

Seules les configurations atypiques ou les structures anatomiques les plus variables présentent encore de réelles difficultés.

Abstract

The determination of specific biomarkers of brain pathologies at population scale is extremely difficult because of the huge inter-individual variability of the sulco-gyral topography. This thesis addresses this issue by automatically identifying 125 sulcal structures and pairing them through individuals, thanks to a manually labeled database of 62 subjects. Relying on the sulcal roots theory, cortical folds are split into elementary segments to be labeled.

In a first time, the structural approach proposed earlier by Jean-François Mangin and Denis Rivière has been revisited to manage the increasing amount of morphometric features involved in the identification process. In a second time, this model has been fully reviewed in favor of a Bayesian framework based on localized information (positions or directions) previously neglected, thus allowing effective optimization schemes. In this context, data normalization is essential; this issue has been considered through global or sulciwise local affine registration techniques, jointly to the sulcal identification. In order to introduce more structural informations, a Markovian model has been successfully introduced to reflect the local neighbored cortical folds organization. Finally, the overall recognition rate has reached 86 % for each hemisphere.

From now on, only atypical patterns or the most variable anatomical structures remain a real issue.

Table des matières

Introduction générale	2
1 Introduction	3
2 Modélisation «historique»	7
2.1 Une approche structurale	7
2.2 Discussion et résultats	16
I Modélisation : approche locale et considérations générales	21
3 Données : représentation	25
3.1 Représentations des sillons	26
3.2 Brainvisa : notre représentation	28
4 Modélisation : concepts de base	31
4.1 Modélisation Bayésienne indépendante	32
4.2 Comment modéliser la variabilité des sillons? $P(D L)$	38
4.3 A priori : $P(L)$	44
5 Résultats et principe d'évaluation	51
5.1 Mesures d'erreur	51
5.2 Comparaison de modèles	53
5.3 Stratégie d'évaluation	53
5.4 Méthodes locales	54
5.5 Mélange avec l'approche «historique»	57
6 Conclusion et perspectives	61
6.1 Conclusion	61
6.2 Perspectives	61
II Normalisation et sillons	65
7 Couplage : recalage et labélisation	69
7.1 Normalisation en tant que prétraitement	69
7.2 Techniques de recalage retenues	70
7.3 Recalage et labélisation joints	72
7.4 Recalage et estimation de modèle joints	74
7.5 A priori de recalage	76
7.6 Optimisation des paramètres de recalage	81
8 Résultats	89
8.1 Modèles	89
8.2 Résultats et discussion	90

9 Conclusion et perspectives	97
9.1 Conclusion	97
9.2 Perspectives	98
III Approches Markoviennes	101
10 Modélisation Markovienne	105
10.1 Présentation	105
10.2 Apprentissage supervisé	107
10.3 Inférence : optimisation de champs de Markov	109
11 Sillons corticaux et champ de Markov	115
11.1 Modèle	117
11.2 Graphe de relations	120
11.3 Potentiels de relations	123
12 Résultats	131
12.1 Approche	131
12.2 Expériences	132
13 Conclusion et perspectives	139
13.1 Conclusion	139
13.2 Perspectives	139
Conclusion générale	146
14 Conclusion générale	147
14.1 Résumé des contributions	147
14.2 Applications	148
14.3 Perspectives	149
IV Annexes	153
Notations	155
A Base de données de sillons	157
A.1 Motivations	157
A.2 Élaboration d'une base de données sulcales	158
A.3 Outils	159
A.4 Planches anatomiques	160
B Méthodes diverses	169
B.1 Iterative Closest Point (ICP)	169
B.2 <i>k-means</i> Riemannien de données directionnelles et axiales	173
B.3 Quelques distributions de probabilités	177
B.4 Alternatives et extensions au modèle SPAM	193

C	Développements informatiques	205
C.1	Introduction	205
C.2	Technologies utilisées	205
C.3	Commandes et fonctionnalités	206
	Bibliographie	211

Introduction générale

Introduction

Les neurosciences fascinent et font régulièrement les gros titres de la presse de vulgarisation scientifique. Des discours passionnés tenant plus du discours philosophique que scientifique décrivent la complexité du cerveau et son fonctionnement et laissent libre cours à de nombreuses interprétations fantaisistes, alors que cette science en est à ces balbutiements et peine à en décrire les fonctions les plus simples. Ce n'est sans doute que l'expression exacerbée de notre curiosité face à un domaine où tout est encore à découvrir. Cette science est une invitation à l'introspection puisque le cerveau est le siège des souvenirs, de la connaissance, de l'émotion, de l'intelligence et de la conscience; il est sans doute la meilleure caractérisation de notre individualité.

Si on prête à Hippocrate (4^{ème} siècle av. J.C), le père de la médecine occidentale, les propos indiquant le cerveau comme le siège de l'intellect, il a fallu attendre le 17^{ème} siècle pour que les premières études en neuroanatomie et en neurologie commencent réellement. Au 19^{ème} siècle, la phrénologie, alors très populaire, cherchait à mettre en évidence des liens entre la forme du crâne et le caractère d'un être humain. Finalement, la compréhension du fonctionnement du cerveau a connu un nouveau souffle avec le développement de la neuroimagerie et l'avènement de techniques d'imagerie non-invasives telle que l'IRM il y a tout juste 3 décennies. Plus récemment encore, l'étude de son fonctionnement est devenue possible *in vivo* grâce à l'IRM fonctionnelle qui détecte l'activité du cerveau en réponse à un stimulus par la mise en évidence de surconsommations d'oxygène dans le sang. À la même période l'IRM de diffusion s'est développée et devrait permettre à terme de révéler les faisceaux de fibres de la matière blanche transportant l'information entre des aires fonctionnelles distantes, en s'appuyant sur l'anisotropie de diffusion des molécules d'eau. Desormais l'avancée croissante de la puissance de calcul des ordinateurs modernes ou des grilles de calcul permet d'envisager des modélisations mathématiques ambitieuses de l'organisation cérébrale et de son fonctionnement.

Si les neurosciences tirent leur épingle du jeu ces dernières années, encouragées par l'état, c'est avant tout à cause du vieillissement de la population et des pathologies qui y sont liées. Aujourd'hui déjà la maladie d'Alzheimer touche des centaines de milliers de personnes. Demain, avec le «papi boom», c'est plusieurs millions de personnes qui seront concernées, plaçant les maladies du vieillissements au rang d'enjeu de santé publique. Les pathologies du cerveau guident une partie de la recherche en neurosciences. On y retrouve les maladies neurodégénératives comme la sclérose en plaques, la maladie de Parkinson, d'Huntington ou d'Alzheimer, ainsi que certaines pathologies psychiatriques comme l'autisme, la schizophrénie ou les troubles bipolaires. C'est dans ce contexte que se situe ce travail de thèse. Plus précisément, il s'intègre dans le cadre du projet PSYMARKER (ANR 2005), en association avec l'unité INSERM U.797 *Neuroimagerie en psychiatrie*, dont l'objectif est la mise en évidence de biomarqueurs développementaux caractéristiques de telles pathologies psychiatriques. Ces troubles semblent liés à des perturbations du développement cérébral qui sont susceptibles de s'exprimer au niveau des structures anatomiques (sillons, gyri, fibres de la matière blanche...) qui reflètent dans une certaine mesure des anomalies fonctionnelles.

Ce manuscrit sera limité à l'étude des sillons, au travers de la modélisation de leur variabilité inter-individuelle et leur identification automatique. Les sillons correspondent aux creux des plissements corticaux, les gyri représentent les bosses. Le cerveau humain présente des configurations

complexes très variables d'un individu à l'autre. Il s'agit d'une particularité propre à l'homme et à quelques mammifères comme les grands singes, mais dans une moindre mesure, ou les cétacés dont le cerveau est étonnamment gros et plissé. L'origine de ces plissements est la traduction d'un besoin évolutionniste (au sens de Darwin) visant à l'augmentation de la «puissance de calcul» cérébrale supportée par les neurones de la surface corticale, alors que le volume du crâne a peu évolué. Cette contrainte a favorisé le développement d'interactions à courte distance au sein des gyri. Le degré de plissement, la complexité de leur agencement et leur variabilité au travers des individus semblent corrélés avec l'intelligence prêtée à chaque espèce (si l'on omet la singularité des cétacés).

Dans le cas du cerveau humain, la complexité des plissements rend l'appariement et la comparaison de sujets difficile. La recherche en anatomie a permis d'identifier de nombreuses structures sulcales présentes chez tous les sujets. Mais, hormis les sillons primaires, premiers à apparaître au cours du développement, leur définition reste imprécise et demande une expertise longue et difficile menée principalement par des neuroanatomistes. Des logiciels d'assistance tentent de faciliter autant que possible cette opération fastidieuse [Rettmann 2005, Shattuck 2009, Rivière 2009]. L'approche proposée dans cette thèse vise à rendre cette opération entièrement automatique pour faire face à la taille croissante des bases de sujets étudiés. Pour rendre compte de la difficulté de cet entreprise, nous pourrions comparer ce problème à celui de la reconnaissance d'écriture ou de visage. Par exemple, il est très facile à un être humain d'effectuer ces opérations, mais après un long apprentissage dans le cas de la lecture. De même, la distinction de plusieurs dizaines de chiens de même race et même âge paraît une opération très difficile parce que nous manquons de points de repère. Pourtant un éleveur serait capable de tous les nommer sans mal parce qu'il a identifié avec l'habitude quelles caractéristiques morphologiques étaient discriminantes. Pour le cerveau humain le principe est identique. Même s'il est très complexe, il faut apprendre à le lire, à le décoder. Les sillons et leur agencement jouent alors le même rôle que les yeux, le nez, la bouche et leur emplacement relatif pour la reconnaissance de visage.

La connaissance de ces indices anatomiques permet donc d'associer des structures similaires chez plusieurs sujets. L'origine des plissements corticaux pourrait résulter en partie de la tension exercée par les faisceaux de fibres sur la surface corticale pendant le processus de développement [Essen 1997]. Les fibres servent de support aux réseaux fonctionnels en véhiculant l'information entre des aires fonctionnelles distantes. Ces deux informations cumulées suggèrent une organisation anatomo-fonctionnelle du cerveau où les plissements représenteraient dans une certaine mesure les délimitations de zones fonctionnellement homogènes. Elle a été constatée pour certaines aires fonctionnelles primaires (par exemple, le sillon central sépare les aires motrices des aires sensorielles). C'est pourquoi la normalisation de sujets selon des critères anatomiques reste un procédé couramment appliqué avant une étude de groupe sur des données fonctionnelles.

Pour résumer, ce sujet de cette thèse est motivé par l'intérêt montant que suscite l'analyse corticale dans de nombreuses applications : de l'étude du développement à la dégénérescence cérébrale en passant par la réduction de la variabilité anatomo-fonctionnelle et l'étude morphométrique de pathologies à visée de diagnostic précoce.

Plan

Ce travail de thèse est dans la lignée du projet initié il y a près de 15 ans par Jean-François Mangin pendant sa thèse [Mangin 1995a] et poursuivi par Denis Rivière ensuite [Rivière 2000]. Il suit une représentation structurelle de l'information corticale [Mangin 2005], c'est à dire s'appuyant sur des structures organisées. Il s'agit d'une représentation «haut niveau» qui permet de s'appuyer sur la terminologie neuroscientifique afin de décrire des structures et leurs interactions spatiales, mécaniques et fonctionnelles.

L'objectif de cette thèse est de proposer une méthode entièrement automatique pour l'iden-

tification des sillons corticaux. Le travail réalisé se positionne comme un travail méthodologique guidé par l'expertise anatomique. La méthodologie proposée ne dépend pas à proprement parler du choix de nomenclature utilisé pour définir les structures sulcales puisque l'estimation des modèles est supervisée; mais l'interprétation des résultats a été menée au regard de cette nomenclature guidant ainsi nos choix méthodologiques. Elle a permis de faire la part des choses entre variabilité inter-individuelle, état actuel de nos connaissances en anatomie et limitations intrinsèques aux modèles.

Introduction

L'introduction se poursuit par un chapitre résumant la stratégie mise en place par Denis Rivière [Rivière 2002] afin de répondre à cette problématique. Cette modélisation repose sur des informations «haut niveau» résumées selon une représentation structurée des sillons et de leur agencement. La variabilité des formes et des configurations est apprise par un ensemble d'experts dédiés chacun à une structure particulière ou à leur relation. L'identification des sillons repose finalement sur la combinaison des réponses des experts constituant un champ de Markov.

Dans le cadre de cette thèse, des améliorations ont été proposées visant à améliorer les performances locales des experts, rendre la méthode plus robuste et moins sujette au sur-apprentissage induit par le nombre restreint de sujets considérés relativement au nombre élevé d'attributs permettant de décrire les structures étudiées.

Ce modèle est ambitieux, mais sa complexité rend sa validation problématique, limitant actuellement l'interprétation des résultats et par conséquent nos propositions d'améliorations. C'est pourquoi la suite de cette thèse a été consacrée à établir de nouvelles bases, plus stables, sur lesquelles s'appuyer pour construire à terme un modèle tout aussi ambitieux mais mieux maîtrisé. Ce changement de direction a été initié à la moitié de cette thèse, mais nous avons préféré privilégier dans ce manuscrit les aspects les plus innovants qui ont donné lieu à des améliorations significatives. Nous avons également privilégié la cohérence du cœur du manuscrit au détriment de ce chapitre, ainsi certaines notions nécessaires à sa compréhension ont été déportées plus loin dans le manuscrit.

Partie I

Cette partie présente un résumé bibliographique des représentations de l'information sulcale et l'intérêt du choix retenu dans Brainvisa. Elle pose les bases de notre modélisation et des principes de validation. Le principe poursuivi est l'exploitation d'informations discriminantes locales, s'exprimant à l'échelle des plissements élémentaires à labéliser (appelés segments sulcaux). Cette information est injectée dans un modèle volontairement simplifié qui suppose l'indépendance statistique des segments entre eux; des méthodes d'estimation supervisée et non supervisée sont présentées. La variabilité sulcale y est représentée par le modèle SPAM et exprimée en termes de localisation des structures.

À partir de ce modèle introductif, deux autres modélisations plus poussées sont dérivées. Elles font l'objet des deux autres parties de cette thèse. Elles sont indépendantes, leur ordre de lecture importe donc peu. Elles apportent chacune des réponses complémentaires à la modélisation et à l'identification des structures sulcales.

Partie II

Le modèle présenté dans la partie précédente repose sur la localisation des structures sulcales dans un espace normalisé. Sa qualité dépend donc fortement du choix de la méthode de normalisation spatiale. Cette partie se propose de reconsidérer une telle approche en intégrant l'estimation des paramètres de normalisation à notre modélisation, que ce soit pendant l'estimation du modèle

à partir d'une base de données ou l'identification des sillons d'un sujet. Des considérations méthodologiques nous amèneront alors à reconsidérer ce problème comme une méthode de recalage contraintes par des labels «cachés» (inconnus) qui prend la forme d'une optimisation EM dans un cadre Bayésien.

Partie III

Dans les deux parties précédentes, les labels des segments étaient considérés indépendants les uns des autres, ce qui pouvait entraîner des incohérences locales de labélisation. Dans cette partie, cette restriction est levée au profit d'une modélisation Markovienne qui introduit une dépendance entre chaque segment sulcal et ses voisins proches. Des liens locaux impliquant des paires de segments en relation sont envisagés et caractérisés par leur direction ainsi que leur distance. Cette partie illustre notre volonté d'exploiter chaque niveau de représentation de l'information : des segments et liens entre segments ici, aux sillons et leur agencement dans le modèle présenté en introduction.

Annexes

Le premier chapitre est consacré à la description de la base de données sulcales de 62 sujets labélisés manuellement il y a tout juste un an et utilisés majoritairement dans les résultats présentés pour faciliter les comparaisons. Quelques planches anatomiques de cette base sont proposées afin d'apprécier la variabilité anatomique des plissements corticaux.

Le deuxième chapitre regroupe un ensemble de méthodes et modèles déportés en annexe pour ne pas nuire à la fluidité de la partie principale du manuscrit. Nous y présentons la méthode de recalage ICP, une méthode de *clustering* de données directionnelles et axiales, une collection de distributions de probabilités utilisées dans notre modélisation (notamment des distributions dédiées aux angles, axes, directions, rotations). Nous proposons également plusieurs extensions et approximations au modèle SPAM développées récemment.

Le dernier chapitre évoque le sujet du développement informatique au travers de quelques réflexions personnelles sur sa pratique dans un environnement de recherche. Nous y résumerons nos choix technologiques avec en toile de fond l'historique des grandes avancées de notre laboratoire en terme de développement informatique. Finalement une sélection de commandes est présentée donnant un aperçu de l'environnement logiciel mis en place pendant cette thèse.

Modélisation «historique»

Ce chapitre occupe une place particulière dans cette thèse et peut être sauté en première lecture puisque les modèles présentés dans la suite du manuscrit suivent une direction différente. Toutefois, ces travaux précèdent chronologiquement ceux présentés dans la suite. Ils expliquent pourquoi nous avons suivi une démarche différente ensuite, ce qui justifie sa place en introduction. On y présentera un résumé des travaux menés dans la lignée de la thèse de Denis Rivière [Rivière 2000] au cours de la première moitié de cette thèse. Dans la suite, nous ferons référence à cette approche sous le nom **modélisation «historique»**.

Ce chapitre présente tout d’abord une modélisation structurelle «haut-niveau» de la variabilité des structures sulcales et de leur agencement, suivie par la discussion des résultats obtenus. Des compléments méthodologiques et des résultats additionnels sont disponibles dans le rapport suivant [Perrot 2005]. Une partie de la méthodologie nécessaire à la compréhension de ce chapitre a été reportée plus loin dans le manuscrit pour ne pas nuire à la lecture du cœur de l’ouvrage, le lecteur sera invité à s’y reporter au besoin.

2.1 Une approche structurelle

L’approche ambitieuse suivie par ces travaux vise à créer un système de reconnaissance automatique des structures sulcales imitant le comportement d’un expert humain. Une telle expertise s’appuie essentiellement sur une reconnaissance locale de chaque sillon en s’appuyant sur sa position, son orientation, sa forme et son agencement par rapport aux structures sulcales environnantes.

Nous commencerons notre description par une description générale du modèle proposé et illustré en figure 2.1. Dans un deuxième temps, nous nous focaliserons sur la représentation des composantes locales.

2.1.1 Modélisation générale

Afin de mieux saisir et maîtriser la complexité de données aussi variables que les plissements corticaux (voir base de données en annexe A), le problème d’identification de ces structures a été décomposé en un ensemble de sous-problèmes locaux plus faciles à traiter. Avant d’aller plus loin dans la description, nous précisons juste que le problème de labélisation des sillons consiste dans notre cas à étiqueter un ensemble de plissements élémentaires (segments sulcaux) issus d’une étape de segmentation des volumes IRM (section 3.2).

2.1.1.1 Experts locaux

L’approche retenue se rapproche de la démarche suivie par un expert humain pour réaliser l’étiquetage manuel. En effet, pour identifier l’ensemble des labels des sillons d’un cerveau, il faut être capable de résoudre le problème localement au niveau de chacun des sillons. Toutefois, pour y parvenir, l’expertise humaine ne se limite pas à chaque structure isolée, mais la considère replacée dans son contexte en utilisant l’information des sillons avoisinant et les relations qui les lient les uns aux autres (embranchement, séparé par un gyrus, ...).

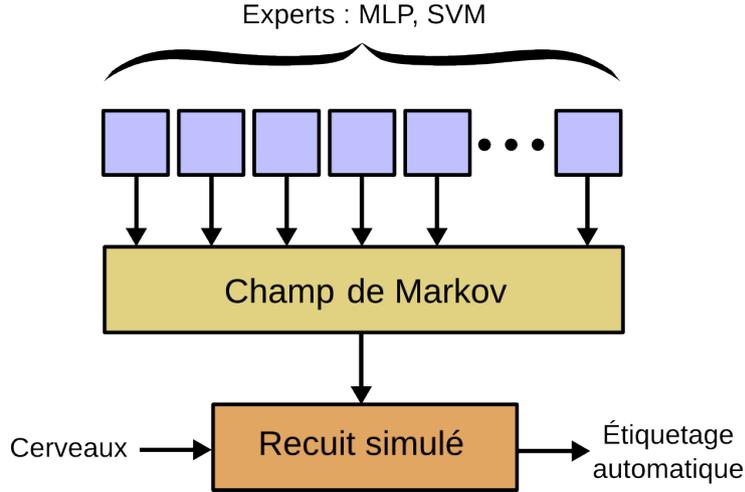


FIG. 2.1 – Le système de reconnaissance automatique des sillons se compose d’une assemblée de 261 experts capables chacun d’identifier un sillon ou une relation particulière au travers d’une mesure traduisant la fiabilité du résultat. L’ensemble des informations locales est ensuite combiné au sein d’un champ de Markov pour former une évaluation globale de l’étiquetage (équation 2.1). Cette énergie peut alors être minimisée pour réaliser l’étiquetage automatique d’un cerveau.

L’idée consiste finalement à disposer d’un ensemble d’experts artificiels, chacun dédié à un **sillon** ou à l’agencement d’une **relation** particulière (paire de sillons voisins, c’est à dire reliés par un embranchement ou séparés par un gyrus) pour évaluer la qualité d’une proposition de labélisation. Au niveau d’un seul expert de nombreuses ambiguïtés peuvent subsister, leur réponse ne doit donc pas être binaire, mais refléter au contraire un degré de satisfaction de la structure ainsi identifiée (que l’on peut réinterpréter comme une probabilité dans le modèle).

Un expert de sillon permet d’évaluer une collection de segments sulcaux portant le même label, noté ici l_i . Cet ensemble de structures est alors résumé sous la forme d’un vecteur descriptif d_i (voir section 2.1.2.1). L’évaluation de cette représentation par l’expert prend la forme d’une énergie notée $U_i(d_i)$. De la même façon, une relation (l_i, l_j) est représentée par un vecteur d_{ij} dont l’évaluation vaut $U_{ij}(d_{ij})$.

2.1.1.2 Combinaison : champ de Markov

Pour un étiquetage complet d’un hémisphère, impliquant plusieurs labels, chaque expert local est mis à contribution pour évaluer les labels le concernant. Ensuite, l’ensemble des réponses est combiné au sein d’un champ de Markov (présenté dans le chapitre 10 et illustré en figure 2.2) afin d’apporter une note globale à l’ensemble. L’énergie globale rattachée à une labélisation L donnée est évaluée comme étant la somme pondérée des potentiels locaux U_i calculés à partir de l’ensemble des vecteurs descriptifs $d = \{d_i; d_{ij}; (i, j) \in \mathcal{L}^2\}$ avec \mathcal{L} l’ensemble des labels possibles :

$$\begin{cases} U(d; L) &= \sum_i w_i U_i(d_i) + \sum_{i \sim j} w_{ij} U_{ij}(d_{ij}) \\ P(L|d) &= \frac{1}{Z} \exp(-U(d; L)) \end{cases} \quad (2.1)$$

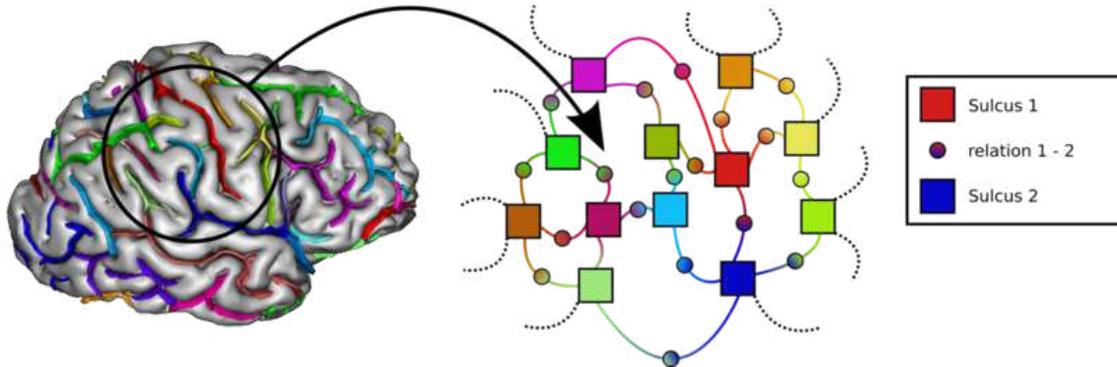


FIG. 2.2 – Illustration locale du champ de Markov sur un ensemble de sillons et leur voisinage. Les experts de sillons sont représentés par des carrés et les experts de relation par des ronds reliant deux carrés.

où Z est une constante de normalisation. Ici, chaque poids est proportionnel à la fiabilité de l'expert. Dans le cas où le rôle de l'expert est joué par un classifieur, cet indice de confiance est obtenu à partir du taux d'erreurs de classification évalué sur une base de données de test réservée à cet effet. Dans le cas où le rôle de l'expert est tenu par une méthode de régression, on utilise l'erreur absolue moyenne normalisée (pour les perceptrons) ou l'erreur quadratique moyenne normalisée (pour les ε -SVR). En notant e_i , l'erreur considérée, le poids w_i associé vaut alors $w_i = \max(0, 1 - 2e_i)$ en saturant les pourcentages d'erreur supérieurs à 50%. De cette façon, les experts ne dépassant pas le niveau de la chance n'interviennent pas dans le calcul de l'énergie globale $U(D; L)$.

La combinaison d'un ensemble d'experts n'est pas une tâche aisée et dépend sans doute de la nature des experts considérés. L'approche suivie ici consiste à normaliser chaque sortie U_i pour ne pas privilégier arbitrairement un expert selon la plage de valeur sur laquelle s'exprime cette sortie. Les valeurs sont normalisées entre -1 et 1 de telle façon que -1 corresponde à la satisfaction maximale de l'expert, 1 à l'inverse et 0 à une réponse neutre. Cette approche reste approximative puisqu'elle ne tient pas compte de la distribution des valeurs de sortie de chaque expert, mais on espère que le phénomène est négligeable devant la confiance (les poids w_i) qu'on leur accorde.

2.1.1.3 Cas particuliers

Dans la base d'apprentissage labélisée manuellement, si un plissement cortical est trop rare pour qu'on en estime correctement la variabilité, un label par défaut est utilisé, que l'on nommera «label inconnu». Dans le cadre de la labélisation automatique un expert implicite est rattaché à cette classe de rejet. L'évaluation des structures portant ce label donne toujours une énergie nulle, c'est à dire $U_0(d_0) = 0$ en notant 0 la classe rattachée à ce label. Nous verrons plus loin, qu'une telle énergie est neutre et privilégie l'utilisation de cette classe de rejet dès lors qu'aucun autre label n'a pu être identifié avec une fiabilité de plus de 50%. Dans le cas des experts de relation entre ce label «inconnu» et un autre label voisin, la même idée est poursuivie, donnant lieu à une énergie nulle $U_{0i}(d_{0i}) = 0$ pour tout label i en relation avec le label «inconnu».

Pour rentrer dans un cadre Markovien, seul un nombre limité de relations sont considérées (200 en pratique). Seuls les labels dont la mise en relation est constatée suffisamment souvent sur une base d'apprentissage, se voient attribués un expert dédié à leur reconnaissance. Dans le cas contraire, un expert implicite est utilisé. Sa réponse $U_{ij}(d_{ij}) = 1$ correspond à la pénalisation maximale (énergie

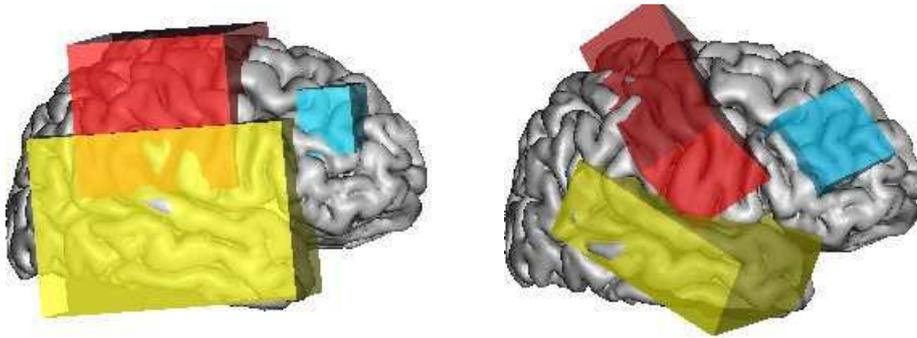


FIG. 2.3 – Boîtes englobantes de quelques sillons : sillon central (rouge), frontal intermédiaire (bleu), temporal supérieur (jaune). **Gauche** : les boîtes sont orientées par rapport au repère stéréotaxique de Talairach. **Droite** : le repère est relatif à chaque structure. Chaque boîte, calculée à partir d’une base d’apprentissage, délimite la localisation spatiale autorisée pour chaque label.

locale la plus haute) autorisée pour un expert, visant à interdire la présence de telles relations lors de la labélisation automatique.

Dans le calcul de l’énergie rattachée à une labélisation donnée, il est également souhaitable de tenir compte de la situation où une structure habituellement présente est absente. Dans ce cas l’expert explicite¹ correspondant ne peut pas être utilisé puisqu’il n’y a pas de descripteur associée à une structure anatomique manquante. Afin de privilégier la présence d’une telle structure, une énergie locale alternative (U_i et U_{ij} selon les cas) est utilisée. Elle pénalise, selon un procédé *ad hoc* [Rivière 2000], l’absence d’une structure à hauteur de son degré d’apparition constatée sur une base d’apprentissage.

Qu’il soit explicite ou implicite, il existe un expert par paire de labels possible. Mais seuls ceux correspondant à des labels en relation au sens du voisinage sont effectivement utilisés.

2.1.1.4 Reconnaissance automatique

La phase de reconnaissance consiste à présenter un hémisphère non labélisé (représenté sous la forme d’un graphe structuré de segments sulcaux : voir section 11.2.1) au modèle. Il s’agit ensuite de trouver la labélisation L qui maximise l’équation 2.1. L’espace des configurations possibles est très vaste et ne peut être parcouru intégralement, il s’agit donc d’optimiser cette quantité par une stratégie adéquate. Avant toute chose chaque label est contraint spatialement à une boîte englobante délimitant (de façon binaire) l’espace dans lequel on peut le trouver (et donc le proposer pendant la phase de labélisation automatique). Ces boîtes englobantes peuvent se réinterpréter comme des a priori spatiaux des labels. Afin de limiter la combinatoire le plus possible, il est nécessaire de réduire au maximum ces zones comme le montre la figure 2.3.

Afin d’optimiser l’énergie globale proposée, une méthode de recuit simulé (détaillée en section 10.3.2.2) légèrement modifiée a été utilisée. Il s’agit d’une méthode itérative basée sur la décroissance d’un paramètre de température permettant de modifier la forme globale de l’énergie optimisée et éviter en théorie le problème des minima locaux. Au cours du processus, des labélisations sont proposées et évaluées de la façon suivante :

- L’ensemble des segments portant le même label sont regroupés afin de calculer un vecteur descriptif et caractéristique de sa morphométrie. On agit de même pour les paires de labels en relation.

¹associé à un modèle de classification ou de régression et nécessitant un apprentissage.

- Chacun des vecteurs ainsi formé est évalué par l’expert correspondant (ou s’il n’existe pas on applique les règles par défaut décrites plus tôt). Celui-ci retourne une classe ou une valeur de régression déterminant dans quelle mesure le vecteur d’entrée proposé correspond au modèle de l’expert.
- Finalement les notes sont combinées en les pondérant par les taux de confiance estimés au préalable sur une base de test.

Les propositions de labélisation sont réalisées selon le schéma suivant. Chaque segment est considéré à tour de rôle dans un ordre aléatoire. Un label est proposé pour ce segment par échantillonnage de Gibbs.

Une fois toutes les 10 passes (une passe correspond à un parcourt de l’ensemble des segments du sujet considéré), une phase spéciale est intégrée au processus de recuit simulé. Elle propose le remplacement de l’ensemble des sillons portant un label donné par le label *inconnu*, d’après sa contribution à l’énergie globale. Ensuite, pour chaque label, on essaie de remplacer l’une après l’autre, chacune de ses composantes connexes² par le label *inconnu*. Finalement, on propose d’attribuer un même label à l’ensemble des segments d’une même composante connexe parmi les morceaux étiquetés *inconnu*. Chacune de ces trois étapes reste soumise au principe d’acceptation et de tirage propre au recuit simulé, prenant ainsi en compte les différences d’énergie impliquées par chaque changement de labels. Cette étape permet de réaliser des sauts (des modifications complexes) dans l’espace des configurations de labels et accélère le processus de convergence. Malheureusement cette approche ne vérifie par le principe de la balance détaillée (équilibre des probabilités de passage entre deux configurations, dans les deux sens) et ne garantit plus l’optimisation globale de l’énergie considérée.

2.1.2 Modélisation locale

Nous détaillerons dans cette section comment les sillons et les relations sont représentés. Nous présenterons ensuite comment modéliser et discriminer cette information selon un schéma d’apprentissage supervisé. Finalement une étape de réduction de dimension est suggérée afin d’améliorer la robustesse des experts locaux.

2.1.2.1 Représentation des données

Les sillons et les relations sont décrits par un vecteur descriptif résumant la localisation spatiale dans Talairach, l’orientation, la forme des structures au travers de descripteurs morphométriques et topologiques. Au niveau des relations, des mesures relatives ou de corrélations sont ajoutées. Une description complète et détaillée de l’estimation de ces descripteurs à partir d’un volume IRM est réalisée dans [Rivière 2000]. Nous nous contenterons ici de lister les descripteurs les plus explicites pour fixer les idées du lecteur.

Descripteurs de sillons :

- coordonnées des extrémités du sillon
- normale moyenne au sillon
- direction moyenne au sillon
- taille du sillon
- profondeur minimale et maximale
- nombre de plis corticaux
- distance maximale entre les plissements corticaux
- longueur de la connexion avec l’enveloppe corticale
- descripteurs de forme : moments invariants (12 premiers) [Mangin 2003b, Sun 2007]

²Ici, il s’agit de l’ensemble des segments reliés entre eux : embranchement ou plis de passage.

Descripteurs de relations : entre deux sillons : S_1 et S_2 .

- $\left\{ \begin{array}{l} \text{taille du sillon } S_1. \\ \text{nombre de plis corticaux de } S_1. \\ \text{nombre de plis corticaux de } S_1 \text{ impliqués dans la relation.} \\ \text{distance point de contact - extrémité de } S_1 \end{array} \right.$
- $\left\{ \begin{array}{l} \text{taille du sillon } S_2. \\ \text{nombre de plis corticaux de } S_2. \\ \text{nombre de plis corticaux de } S_2 \text{ impliqués dans la relation.} \\ \text{distance point de contact - extrémité de } S_2 \end{array} \right.$
- distance minimum entre les sillons.
- direction de la relation
- produit scalaire des directions de S_1 et S_2 .
- position du milieu du segment de contact entre les deux sillons.
- nombre de points de contact entre les deux sillons.
- nombre de plis de passage (section 11.2.1).
- profondeur maximale des plis de passage (section 11.2.1).

Remarque : les descripteurs présentés ci-dessus s'expriment à l'échelle d'un sillon entier ou d'une paire de sillons. Il est donc nécessaire de calculer ces descripteurs après avoir reconstitué le sillon en regroupant l'ensemble des segments sulcaux portant le même label. Puisque de nombreuses propositions de labélisation sont réalisées pendant l'algorithme d'optimisation (recuit simulé), le calcul de ces descripteurs doit rester rapide. C'est pourquoi, la plupart des descripteurs utilisés sont calculables comme une simple combinaison linéaire de descripteurs élémentaires précalculés s'exprimant localement au niveau des segments sulcaux. Dans la liste présentée, seuls les moments invariants ne vérifient pas cette contrainte et alourdissent considérablement les temps de calcul si on les considère. En pratique, ces descripteurs de forme n'ont pas amélioré les résultats dans nos expériences et n'ont donc pas été retenus dans le modèle.

2.1.2.2 Apprentissage Supervisé

Un expert dédié à un sillon ou une relation doit être capable d'évaluer si un vecteur de descripteurs correspond ou non à un label donné. Afin d'éviter des effets de pallier - dommageable à l'optimisation - dans l'énergie considérée, la réponse de l'expert ne doit pas être binaire mais indiquer à quel point l'expert a reconnu ou non ce sillon. L'approche suivie par le modèle proposé ici est discriminative : un expert doit être capable de distinguer un sillon particulier d'autres configurations de labels qu'on pourrait lui présenter. Compte tenu de la complexité du problème, la calibration des paramètres de chaque expert explicite est réalisée par un apprentissage supervisé à partir d'une base de données labélisées manuellement.

Dans le cadre de l'optimisation par recuit simulé, toutes les configurations, même les plus aléatoires et donc les moins crédibles peuvent apparaître au cours du processus. Afin de pénaliser ces situations auprès des experts concernés, l'approche suivie consiste à présenter des configurations similaires lors de la phase d'apprentissage. Nous appellerons ces situations des «contre-exemples». Pour un label donné, l'apprentissage des experts est réalisé de façon supervisée et vise à distinguer les véritables sillons (issus de la base d'apprentissage) portant ce label à une collection de contre-exemples (illustrés à gauche de la figure 2.4) représentatifs de configurations erronées que l'on peut rencontrer pendant la phase d'optimisation. Nous ne disposons pas directement de telles données. L'idée suivie dans ces travaux a été d'en générer à partir des sujets d'apprentissage. Ces faux sillons sont obtenus à partir des véritables sillons en échangeant, ajoutant et enlevant un certain nombre de segments portant le label considéré. En pratique, trois familles de contre-exemples sont utilisées

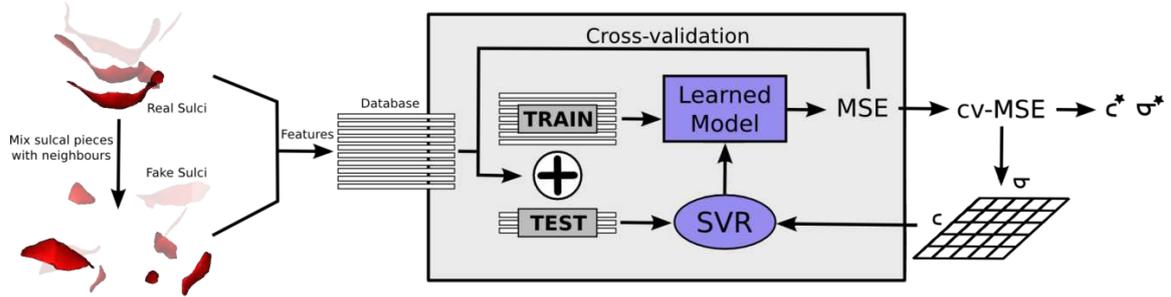


FIG. 2.4 – Illustration de la méthode d’apprentissage supervisée du modèle à partir de véritables sillons et de contre-exemples de sillons. Dans le cadre d’experts basés sur des SVR, les paramètres libres (C, σ) du modèle sont estimés par validation croisée.

afin de couvrir toutes les situations : des légères modifications aux changements plus radicaux.

L’apprentissage direct de la distinction entre véritables sillons et contre-exemples pose problème puisqu’on court le risque d’avoir généré des structures trop similaires aux données réelles. Pour limiter cette difficulté, on attache une distance ad hoc à chaque contre-exemples qui modélise à quel point il diffère du sillon d’origine dont il est issu. Dans ce cadre, au lieu de faire de la classification entre deux familles d’objets, on réalise plutôt une régression des vecteurs de descripteurs avec cette distance. Dans ce cadre, un véritable sillon a naturellement une valeur de distance nulle.

Cette partie du modèle demande un calibrage long et difficile, afin de doser la quantité de contre-exemples et définir les bons critères de génération. On ne peut exclure que cette étape soit sujette à un sur-apprentissage des paramètres de contrôle. Nos expériences ont montré que ces contre-exemples jouaient parfaitement leur rôle en début d’optimisation lorsqu’on est loin de la solution recherchée, et à mesure qu’on se rapproche des véritables exemples, ils perdent en efficacité. Nous avons mis en évidence que dans ce cas, il était possible de remplacer les contre-exemples d’un sillon particulier par l’ensemble des sillons voisins susceptibles de rentrer en compétition avec lui. Dans ce cas, on revient à un problème de classification classique, le degré de reconnaissance exprimé par l’énergie locale U_i ou U_{ij} est alors obtenu en calculant une probabilité a posteriori de classification ou en utilisant la fonction de décision du classifieur (la classe étant obtenue par seuillage de cette quantité).

2.1.2.3 Les experts

Plusieurs méthodes de classification ou de régression ont été explorées mais deux méthodes ont retenu notre attention. Il s’agit d’un réseau de neurones standard nommé «Perceptron Multi-Couches» (*Multi-Layer Perceptron*, MLP [Le Cun 1988]) et des machines à vecteur de support (*Support Vector Machine*, SVM [Vapnik 1998]). Afin de ne pas alourdir le propos nous ne détaillerons pas ces techniques d’apprentissage (se reporter à [Rivière 2000] pour le MLP et [Perrot 2005] pour le SVM).

Perceptron Multi-Couches (figure 2.5 à gauche). Ce modèle a été proposé initialement par Denis Rivière [Rivière 2002]. Il s’agit d’un réseau neuronal composé d’un ensemble d’unités de calcul élémentaires reliées entre elles par des liens pondérés. Ces unités forment un réseau organisé en couches reliant une entrée, représentant la donnée à apprendre (dans notre cas le vecteur de descripteurs) à une sortie (la classe ou la mesure de distance associée selon si on souhaite faire de la classification ou de la régression). La phase d’apprentissage consiste à optimiser ces poids de façon supervisée à partir d’un ensemble de couples d’entrée-sortie.

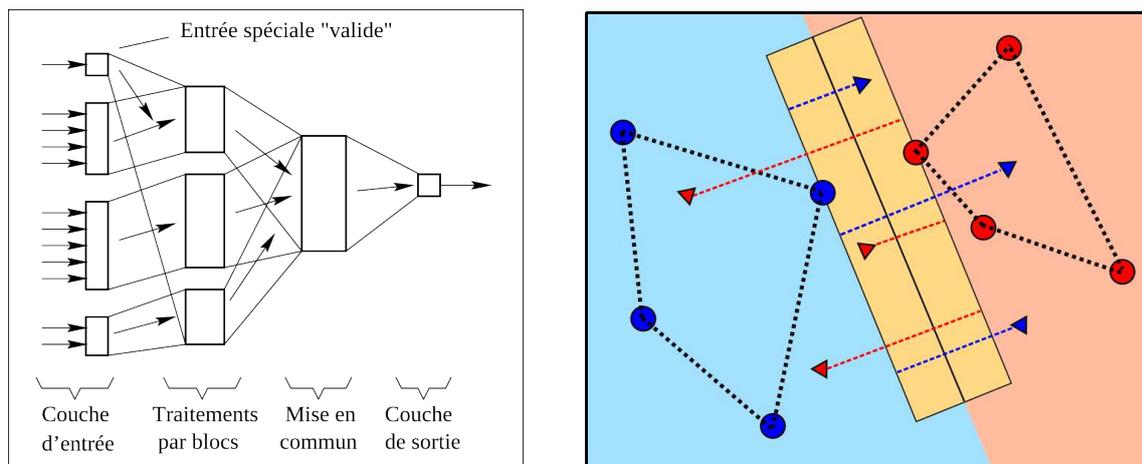


FIG. 2.5 – **Gauche** : illustration de la topologie utilisée pour le modèle MLP. **Droite** : illustration du principe de classifieur à large marge comme le SVM où quelques erreurs sont tolérées.

L'interprétation des paramètres internes (les poids) du modèle est assez difficile, même dans le cas de topologie très simple. On observe en pratique une cartographie de l'espace des données d'entrées par les valeurs de sortie, induite par les observations successives de données d'apprentissage. La nature hautement non-linéaires de ces modèles ne permet pas de prédire le comportement des zones non-explorées par l'apprentissage, ou les zones à la limite du domaine d'expression des descripteurs. L'estimation classique de ces poids par une méthode stochastique de rétropropagation du gradient ne permet pas de garantir un apprentissage robuste des paramètres dans ces zones. Pour éviter autant que possible ce problème, une base de test (fixe et composée de seulement 5 cerveaux à l'époque) a été utilisée pour stopper le processus d'estimation avant qu'il y ait sur-apprentissage. L'erreur moyenne de prédiction représente également le taux de confiance utilisé pour pondérer l'importance relative d'un expert dans l'énergie globale du système.

Dans notre cadre d'utilisation, une topologie particulière a été définie pour les modèles de sillons et une autre pour les modèles de relation. Son élaboration consistait à regrouper entre elles des descripteurs de même nature et hautement corrélés dans une première couche de neurones (figure 2.5 à gauche), comme les coordonnées spatiales par exemple.

Machine à vecteurs de support (figure 2.5 à droite). Ce modèle a été introduit au cours au début de cette thèse en remplacement des réseaux de neurone afin d'améliorer la robustesse de l'apprentissage. Il s'agit d'une méthode à noyau initialement consacrée à des tâches de classification (étendue par la suite au cas de la régression) qui vise à trouver un plan séparateur (dans l'espace de Hilbert correspondant) maximisant la séparabilité de deux classes (la largeur de la marge, représentée en jaune orangé sur la figure 2.5) de façon robuste. C'est à dire en évitant le sur-apprentissage, en autorisant des éléments mal classés (éléments placés du mauvais côté de la marge, représentés ici par des triangles). La tolérance aux erreurs est contrôlée par un paramètre réel noté C .

Que ce soit dans le cas de l'apprentissage à base de contre-exemples ou de sillons voisins, on s'attend à ce que cette classe entoure la classe des véritables sillons. Pour cette raison un noyau linéaire comme celui présenté dans l'exemple n'est pas adapté, on lui a préféré un noyau gaussien isotrope contrôlé par le paramètre d'échelle σ (l'écart-type de la Gaussienne) qui permet de modéliser théoriquement toutes les situations. En pratique, la fonction de décision du SVM s'appuie sur un sous-ensemble des données d'apprentissage : les vecteurs de support dont la contribution est exprimée par

un poids. Intuitivement, cette fonction prend la forme d'un champ de Gaussiennes : une Gaussienne par vecteur de support dont la taille est proportionnelle à leur poids. L'interprétation des paramètres du modèle semble beaucoup plus explicite que dans le cas des réseaux de neurones. Ainsi, les capacités de modélisation restent limitées par le nombre de données apprises. Si les effectifs des classes sont déséquilibrés (se vérifie en pratique), la classe la plus petite limite la complexité de la zone de séparation. D'ailleurs, il est nécessaire d'en tenir compte lors de l'apprentissage du modèle au niveau de l'estimation des paramètres internes du SVM (les poids des Gaussiennes) ou au niveau de l'étape de validation croisée pour optimiser les paramètres libres du modèle (C et σ).

Différences principales entre les deux modèles. Dans le cas de l'apprentissage du modèle MLP, le ratio du nombre de bons exemples par rapport au nombre de contre-exemples est de 1/10000 afin de bien modéliser leur répartition. Le modèle SVM quant à lui est plus contraint et généralise mieux que le précédent, le ratio tombe à seulement 1/10 pour des taux de reconnaissance globaux équivalents. La quantité de données traitées étant beaucoup plus faible, le processus d'apprentissage est plus rapide. Le temps gagné a été réutilisé pour estimer par validation croisée les paramètres libres du SVM.

L'obtention de résultats satisfaisants avec le modèle MLP n'a pu être obtenu qu'au prix de la construction minutieuse d'une topologie de réseau de neurones adaptée. Malheureusement cette organisation dépend fortement du choix des descripteurs fournis en entrée au réseau. Tout ajout ou retrait de descripteurs nécessiterait une étude similaire. Le modèle SVM ne nécessite aucune intervention manuelle et peut gérer un nombre d'attributs quelconque. Nous verrons dans la section suivante l'importance d'une telle flexibilité.

2.1.2.4 Ajout de descripteurs et réduction de dimension

Afin d'améliorer les capacités de reconnaissance du système exposé dans ce chapitre, une idée simple est de lui fournir plus d'information descriptive afin qu'il tire partie de celles qui sont discriminantes (ou de leur combinaison). L'extension du jeu de descripteurs a fait l'objet du stage de Master de Pauline Roca³ au cours duquel ont été introduites plusieurs mesures de l'épaisseur corticale (matière grise), la plus précise est basée sur la méthode du *fast-marching*. Dans le cadre d'études morphométriques, il est également envisagé d'étendre le nombre d'attributs, renforçant cette tendance.

Nous disposons d'un nombre assez restreint de sujets labélisés, 26 à l'époque, dont seulement 21 utilisés pendant la phase d'apprentissage, pour 27 descripteurs de sillons et 23 descripteurs de relation. Quand le nombre de descripteurs devient plus grand que le nombre de sujets considérés, on évoque le problème de la «malédiction de la dimension» (*curse of dimensionality*) qui reflète l'incapacité du modèle à représenter correctement la complexité des données sans plus de contraintes ou d'a priori. L'utilisation de méthodes multivariées ou non-linéaires peut être sujette à caution selon les situations (cela peut même survenir pour un nombre plus faible de dimensions, en présence de données peu structurées) et entraîner un sur-apprentissage. Les deux modèles considérés (MLP et SVR) sont sujets à ce problème.

Le jeu de descripteurs utilisé permet de décrire un objet structuré de façon compacte. Cette cohérence se traduit donc par des corrélations entre les descripteurs. Cet a priori peut être mis à profit pour contraindre l'estimation des paramètres des experts. Dans ce cas précis, il peut prendre la forme d'une méthode de réduction de dimension qui vise à trouver un sous-espace de dimension réduite dans lequel les données s'expriment. Un nombre limité de méthodes ont été envisagées. Pour un aperçu plus large du sujet le lecteur est invité à consulter l'article suivant [Fodor 2002]. On

³co-encadrée avec Denis Rivière.

distingue les **approches univariées** réalisant de la sélection de dimension, où l'intérêt de chaque descripteur est jugé séparément des autres, des **approche multivariées** réalisant de l'extraction de descripteurs en combinant les dimensions d'origines. On distingue également les approches linéaires et non-linéaires selon la nature du sous-espace recherché.

La réduction de dimension n'est pas une fin en soi, mais participe à améliorer la robustesse de l'expert concerné. Cette étape peut constituer un pré-traitement visant à simplifier la représentation d'un sillon ou d'une relation ou s'intégrer directement à l'algorithme de classification ou de régression. Les descripteurs n'intervenant pas dans le processus discriminatif peuvent alors être ignorés. Ce processus contribue ainsi à accélérer les calculs puisque la dimensionnalité des modèles est réduite et l'évaluation des descripteurs se limite à ceux sélectionnés. Il est probable que les descripteurs discriminants diffèrent d'un sillon à l'autre. Ainsi la réduction de dimension, permet de mieux comprendre les critères morphométriques impliqués dans la reconnaissance des structures sulcales ; la sélection de dimension indique directement les descripteurs retenus alors que dans le cas multivarié de nouvelles dimensions sont extraites résultant de la combinaison linéaire des descripteurs d'origine dont les contributions reflètent l'intérêt dans le processus de discrimination. Les méthodes utilisées sont présentées dans la section suivante accompagnée de résultats sur des données réelles.

Un intérêt croissant a été accordé à ces considérations avec l'augmentation du nombre de descripteurs, dont la mise en œuvre s'est concrétisée au début de cette thèse. La modélisation alternative proposée dans le cœur de l'ouvrage s'appuie sur un nombre encore limité de descripteurs, mais à terme elle fera également face à la même problématique.

2.2 Discussion et résultats

Nous utilisons ici le même protocole d'apprentissage et validation que celui utilisé dans [Rivière 2002]. Une base de données de 26 sujets manuellement labélisés est utilisée à hauteur de 21 sujets pour la phase d'apprentissage du modèle (dont 16 pour l'estimation locale des paramètres de chaque expert : base d'apprentissage, 5 pour évaluer un indice de confiance permettant de pondérer leur contribution dans l'énergie globale : base de test) et 5 sujets pour juger les capacités de généralisation du système de reconnaissance (base de généralisation). Les résultats présentés dans cette section seront limités aux hémisphères droits.

Aucune validation croisée n'a pu être utilisée dans ce contexte à cause de difficultés pratiques introduisant des interventions manuelles rendant la tâche très fastidieuse. Les taux de reconnaissance présentés dans cette section ne sont sans doute pas fiables et limitent leur interprétation. Ils ont toutefois permis de relever des tendances et des limitations, ce qui a orienté nos choix pour la suite.

2.2.1 MLP ou SVM ?

Nous rappelons ici quelques résultats principaux (pour des résultats plus complets, se reporter à [Perrot 2005]) justifiant l'intérêt de l'approche à base de SVR. Le modèle MLP à topologie fixée est comparé à plusieurs variantes basées sur des machines à vecteurs de support : SVC (classification : réponse binaire), SVC-proba (classification, réponse continue basée sur le calcul d'une probabilité a posteriori dans l'espace de Hilbert [Platt 2000, Lin 2003]), SVR (régression, réponse continue).

Les résultats locaux sont présentés en figure 2.6. Il s'agit d'histogrammes des taux de classification observés sur la base d'apprentissage et la base de test. Pour le modèle SVR comme il s'agit d'une méthode de régression, la classe est obtenue par seuillage de la sortie à une valeur intermédiaire valant 0.

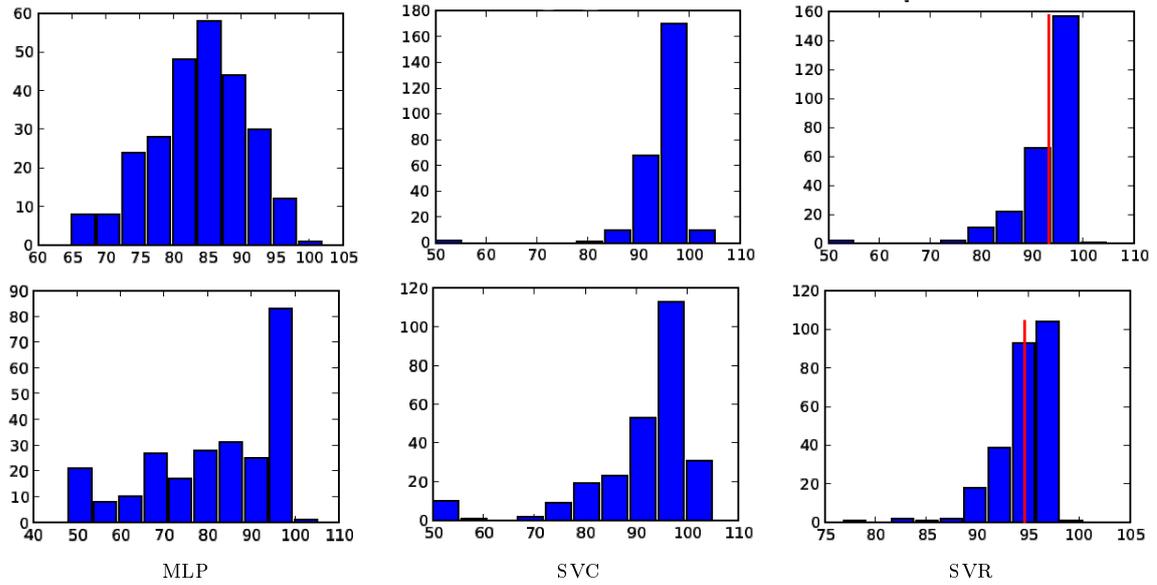


FIG. 2.6 – Histogrammes des taux de classification des 261 experts (sillons et relations). **abscisse** : taux de classification. **ordonnée** : nombre d’experts correspondant à un taux de classification donné. Les résultats sont pondérés de façon à rééquilibrer l’importance entre les deux classes considérées. De **Gauche** à **Droite** : modèles MLP, SVC et SVR. **Haut** : sur la base d’apprentissage de 16 sujets (environ 16 véritables sillons contre 160 contre-exemples). **Bas** : sur une base de test de 5 sujets (environ 5 véritables sillons contre 50 contre-exemples).

Au niveau local, les résultats favorisent largement les modèles SVC et SVR qui généralisent mieux que le modèle MLP. Les machines à vecteurs de support sont connues pour leur robustesse, comme l’illustrent les résultats obtenus. Les taux de reconnaissance obtenus sur la base d’apprentissage et la base de test diffèrent moins pour ces modèles que pour le MLP.

Les résultats globaux prennent la forme de taux d’erreurs de labélisation exprimant le pourcentage de structures reconnues en tenant compte de leur taille⁴ :

Taux d’erreurs	MLP	SVC	SVC-proba	SVR
Global	25.22 (7.27)	33.23 (4.99)	28.53 (6.34)	23.04 (5.06)
Apprentissage	19.90 (4.94)	30.42 (3.57)	25.64 (6.05)	21.10 (4.52)
Test	25.24 (7.20)	32.96 (6.66)	27.72 (6.02)	24.81 (5.24)
Généralisation	30.88 (4.97)	36.31 (4.11)	31.89 (5.40)	27.48 (2.34)

Les résultats globaux sont obtenus à partir des labélisations automatiques impliquant l’ensemble des experts de sillon et de relation combinés au sein d’un champ de Markov. Contrairement à ce que laissait présager les résultats de classification obtenus localement, le modèle MLP présente de meilleurs résultats que les modèles SVC (dont la réponse s’appuie sur la classe ou la probabilité a posteriori). Seul le modèle SVR obtient des taux d’erreurs comparables avec une légère amélioration des performances en généralisation, dont il est malheureusement difficile de juger de la significativité à partir du nombre restreint de sujets.

⁴Il s’agit de l’erreur E_{mass} décrite en section 5.1.

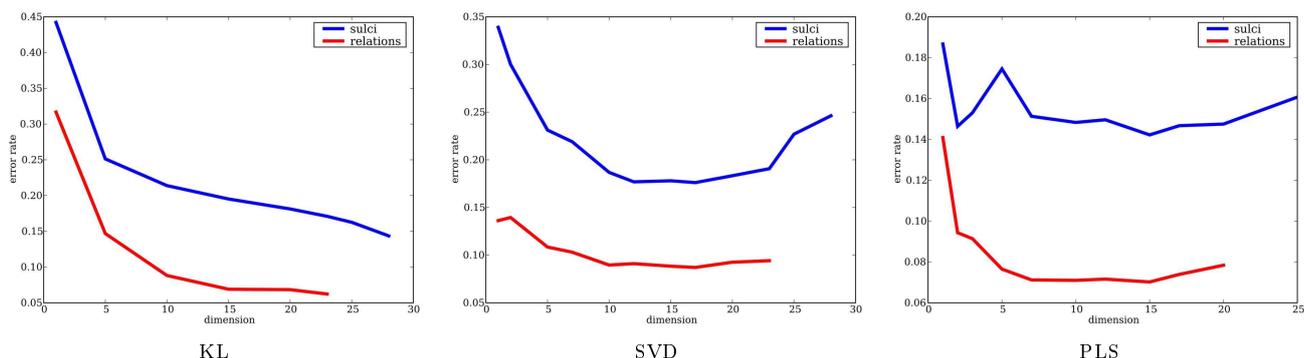


FIG. 2.7 – Taux d’erreurs de classification moyens (sur l’ensemble des experts de sillon en bleu et sur l’ensemble des experts de relation en rouge) en fonction du nombre de dimensions retenu par la méthode de réduction de dimension. De **gauche** à **droite** : méthode univariée basée sur la divergence de Kullback-Leibler (KL), SVD, PLS

2.2.2 Réduction de dimension

Dans le cadre des méthodes testées, la réduction de dimension intervient en tant que pré-traitement sur les vecteurs de descripteurs. L’intérêt de cette approche a été évalué sur les performances locales (erreurs quadratiques moyennes) et globales (erreurs de labélisation) induites par des experts basés sur un modèle SVR. Pour chaque approche de réduction de dimension, un modèle SVR a été appris sur les données d’apprentissage et ces paramètres C et σ optimisés par validation croisée. Le modèle MLP quant à lui nécessite la définition d’une topologie de réseau de neurones dépendantes de la nature des dimension et de leur nombre, ce qui le rend difficilement utilisable dans ce cadre.

Dans nos essais, des méthodes relativement basiques ont été introduites afin d’éprouver le concept et mettre en place les technologies logicielles dans le système de reconnaissance.

Nous avons utilisé deux approches univariées ad hoc. Dans les deux cas le principe est le même : évaluation indépendante de chaque descripteur qui donne lieu à une note indicatrice de ses capacités discriminantes. Les dimensions sont alors ordonnées dans l’ordre décroissant de ces notes ; seules les k premières sont conservées. La première est basée sur la divergence de Kullback-Leibler (KL) le long de chaque dimension des deux distributions considérées (véritables sillons et contre-exemples ici). La deuxième utilise le taux de classification de machines à vecteur de support (SVC dont les paramètres C et σ sont optimisés par validation croisée sur les données) dédiées chacune à une dimension particulière.

Concernant les méthodes multivariées, nous avons testé la décomposition en valeur singulière (SVD) qui tient compte des directions de dispersion des données, mais ignore l’information de classe. La méthode *Partial Least Squares* (PLS1 du livre [Hastie 2001]) permet d’introduire cette notion dans le processus d’extraction de dimension (fonctionne également dans le cas d’information continue comme dans notre situation). Cette approche construit itérativement de nouveaux attributs orthogonaux entre eux en cherchant celui qui maximise la corrélation avec les données (orthogonalisées par rapport au jeu d’axes déjà trouvé) et le vecteur de régression (partie inexpliquée par le jeu d’axes déjà trouvé).

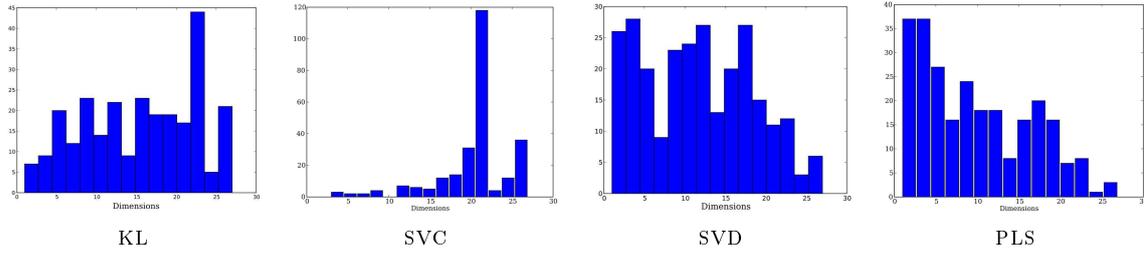


FIG. 2.8 – Histogramme du nombre de dimensions optimal sélectionné pour chacun des 261 experts. Plus l’histogramme est «tassé» sur la gauche meilleure est la «compression» des données. De **gauche** à **droite** : modèle univarié KL, modèle univarié basé sur un modèle SVC, SVD, PLS.

L’influence du nombre de dimensions sélectionnées ou extraites est illustrée en figure 2.7 pour 3 méthodes. Les résultats sont résumés sur l’ensemble des sillons d’une part et l’ensemble des relations d’autre part. Cette séparation est nécessaire puisque le nombre d’attributs diffère entre les deux catégories. Les méthodes SVD et PLS suivent le schéma attendu dans le cas d’un sur-apprentissage induit par un nombre de dimensions trop important. C’est à dire, des taux d’erreurs plus importants dans les extrêmes (nombre faible ou important de dimensions) avec un nombre de dimensions optimal entre les deux. La méthode basée sur la divergence KL ne suit pas ce schéma mais présente des taux de classification comparable à son minimum, lorsque le nombre total de dimension est sélectionné. Les méthodes SVD et PLS présentent un taux d’erreur minimal assez proche, mais la méthode PLS permet une meilleure «compression» des données.

Dans cette étude le nombre de dimensions a été fixé à la même valeur pour l’ensemble des experts de relation d’une part et des experts de sillon d’autre part. Bien entendu, le nombre idéal de dimensions permettant de représenter et discriminer l’information dépend du label ou de la paire de labels considérée. En effet, la variabilité anatomique inter-individuelle est plutôt inhomogène d’une structure à l’autre. Dans une deuxième expérience, nous avons cherché à optimiser pour chaque expert le nombre de dimensions extraites sur la base de sa capacité à prédire les valeurs mesurées sur une base de test de 5 sujets. La figure 2.8 indique pour les 4 méthodes de réduction de dimension la répartition de ce nombre optimal pour l’ensemble des experts. Cette mesure penche en faveur des méthodes multivariées et en particulier la méthode PLS qui démontre encore une fois ses capacités de «compression» de l’information. Les résultats en moyenne et les écarts-types correspondant sont les suivant : KL : 15.38 (7.04), SVC : 20.63 (4.67), SVD : 11.66 (6.76), PLS : 9.84 (6.63).

Les résultats obtenus localement ne semblent pas se reporter à l’échelle globale (labélisation d’un hémisphère entier) comme l’indique le tableau suivant :

Méthode	Type	Sélection	Nombre	Taux d’erreurs (apprentissage)	Taux d’erreur (test)	Taux d’erreur (généralisation)
AUCUNE	-	fixée	tout	21.10	24.81	27.48
PLS	multivariée	fixée	15	17.25	20.50	26.32
SVD	multivariée	fixée	15	15.76	26.45	27.28
KL	univariée	fixée	15	20.99	29.47	27.75
PLS	multivariée	optimisée	9.84 (6.63)	18.16	22.46	28.50
SVC	univariée	optimisée	20.63 (4.67)	18.54	25.88	28.43

Les taux d’erreurs présentés correspondent à l’erreur E_{mass} décrite en section 5.1. La meilleure méthode locale donne également lieu aux meilleurs résultats globaux en généralisation : 26.32% d’erreurs, en fixant le nombre de dimensions sélectionnées à 15 (meilleur résultat en moyenne sur la méthode PLS, correspond à 97% de la variance expliquée pour la méthode SVD). L’optimisation

du nombre de dimensions sélectionnées ne permet pas d'améliorer les résultats, alors que les performances de chacun des experts locaux ont été optimisées. Il est possible que les performances de la méthode soient limitées par le nombre très restreint de sujets utilisés pour l'optimisation de ce critère.

2.2.3 Critique

Comme l'indiquent les expériences présentées plus tôt, une amélioration nette au niveau des capacités de reconnaissance des experts locaux ne présentent pas de la qualité des labélisations à l'échelle d'un hémisphère entier. Plusieurs facteurs peuvent expliquer ce phénomène :

- Le nombre restreint de sujets ne permet pas une bonne estimation des paramètres du modèle et ne permet pas de juger avec précision les performances du système en généralisation. Une validation croisée serait une bonne solution, mais n'est pas envisageable sans un remaniement profond de la structure logicielle du système de reconnaissance, ou de très nombreuses interventions manuelles.
- Le problème de discrimination traité par les experts vise à différencier de véritables sillons de contre-exemples de sillons. Ces données sont artificielles et générées à partir des données réelles. Elles couvrent une large diversité de situations aberrantes susceptibles d'apparaître pendant le processus de labélisation automatique. Par conséquent, rien n'indique que les améliorations constatées concernent les situations les plus ambiguës (les configurations proches des véritables solutions).

Si l'idée générale de ce système de reconnaissance de sillons n'est pas à mettre en cause, sa mise en œuvre actuelle pose des difficultés d'interprétation. Il est très difficile d'identifier les points faibles de la méthode afin de les améliorer. De la même façon l'identification du flux d'information pertinent menant à la solution n'est pas aisée. Dans la suite, le cœur du manuscrit présente de nouveaux modèles dont l'élaboration étagée (en augmentant la complexité du modèle petit à petit) a été guidée par une validation plus précise des résultats.

Première partie

Modélisation : approche locale et
considérations générales

Table des Matières

3	Données : représentation	25
3.1	Représentations des sillons	26
3.2	Brainvisa : notre représentation	28
4	Modélisation : concepts de base	31
4.1	Modélisation Bayésienne indépendante	32
4.1.1	Approche supervisée	33
4.1.2	Approche semi-supervisée	35
4.2	Comment modéliser la variabilité des sillons ? $P(D L)$	38
4.2.1	Segments sulcaux et sillons	39
4.2.2	Modélisation au niveau des segments	40
4.2.3	Approche locale : SPAMS	42
4.3	A priori : $P(L)$	44
4.3.1	Fréquence	45
4.3.2	Dirichlet	46
4.3.3	Multinomiale	48
4.3.4	Approche Markovienne	49
5	Résultats et principe d'évaluation	51
5.1	Mesures d'erreur	51
5.2	Comparaison de modèles	53
5.3	Stratégie d'évaluation	53
5.4	Méthodes locales	54
5.5	Mélange avec l'approche «historique»	57
5.5.1	Modèle	57
5.5.2	Expérimentations et discussion	59
6	Conclusion et perspectives	61
6.1	Conclusion	61
6.2	Perspectives	61

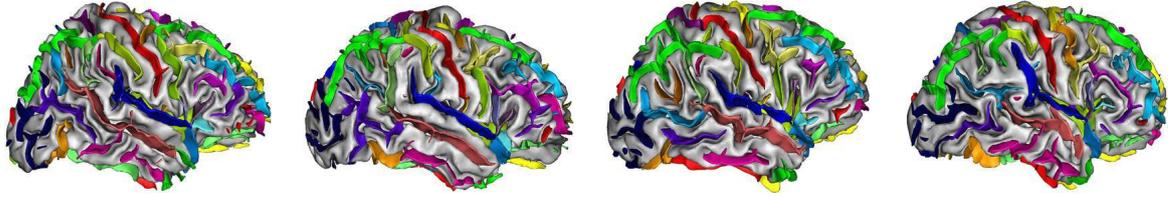


FIG. 2.9 – Variabilité sulcale : 4 sujets labélisés manuellement selon une nomenclature basée sur la théorie des racines sulcales.

Le cerveau humain présente une surface extrêmement plissée où se dessine des agencements de plis complexes (voir figure 2.9), dont l'organisation macroscopique semble suivre des règles communes, au delà des particularités individuelles. On s'attend à ce que l'étude de la variabilité des structures corticales permette de mettre en évidence des variations anatomiques atypiques, ce qui pourrait suggérer la présence d'anomalies développementales ou l'évolution de dégénérescences. L'expression des variations du cortex s'exprime dans une mesure importante sur les seules structures sulcales, qu'une expertise adéquate permet de mettre en correspondance d'un individu à l'autre. Cette identification manuelle exploite des informations diverses, allant de la position, l'orientation, la forme, à l'agencement de ces structures.

Chronologiquement, cette thèse a débuté comme la suite logique de travaux menés précédemment par Denis Rivière dans le cadre d'une modélisation Markovienne de la variabilité sulcale en vue de réaliser la labélisation automatique de telles structures sur de nouveaux sujets. Une phase d'étude approfondie du modèle proposé a permis d'en mesurer les limites, freinant désormais considérablement son évolution. Quelques défauts principaux sont à rappeler, il s'agit d'un modèle très complexe, lourd à mettre en œuvre et donc difficile à valider. Par conséquent l'interprétation des résultats s'en trouve limitée de telle manière qu'il est assez difficile d'identifier quelles parties du modèle contribuent significativement au processus de labélisation. Dans cette partie, nous avons donc reconsidéré notre sujet d'étude dans sa globalité afin de déterminer quelles informations avaient été négligées à tort dans la modélisation précédente.

Cette présentation débute par une revue des représentations possibles des données sulcales. Il s'agira alors d'identifier l'intérêt et les difficultés liées à chaque approche. Nous détaillerons ensuite le processus d'extraction des segments sulcaux représentant le support unitaire de la labélisation. Nous proposerons alors une modélisation de la variabilité de localisation spatiale des structures sulcales basée sur une approche à noyaux (modèle SPAM). Nous l'utiliserons pour inférer des labels sur de nouveaux sujets grâce à une première formalisation du problème où chaque label porté par un segment est considéré comme indépendant des autres. Elle servira de base aux solutions suggérées plus loin dans les parties II et III, en suivant une logique incrémentale : une succession de modèles emboîtés de plus en plus complexes. Une série de résultats démontrera finalement le potentiel de la prise en compte d'informations localisées, ce qui nous conduira naturellement à reconsidérer la façon dont nos données sont normalisées.

Données : représentation

Fonctions et structures semblent être fortement liées au niveau cortical. S'il s'agit d'un fait avéré dans le cas de certaines fonctions primaires bien connues (système moteur, visuel ou sensitif par exemple), cela reste à vérifier à une échelle plus globale. Dans le cas où l'hypothèse se vérifie, l'identification de structures corticales et leur appariement d'un individu à l'autre permettrait sans doute de réduire la variabilité anatomo-fonctionnelle. De ce point de vue, l'information pertinente s'exprime notamment au niveau de la surface corticale, qu'il s'agisse de l'interface matière grise-LCR (liquide céphalo-rachidien) ou matière grise-matière blanche. Ces deux représentations sont complémentaires puisqu'à elles deux elles permettent de calculer l'épaisseur de la matière grise; épaisseur, qui présente des spécificités locales qu'il serait envisageable d'exploiter pour identifier des structures anatomiques aux propriétés spécifiques. Certains travaux visant à l'identification de structures sulcales utilisent directement des informations localisées sur la surface qui nécessitent donc une mise en correspondance surfacique. Notamment, [Fischl 2004] se base sur une paramétrisation surfacique sur laquelle en chaque point est mesurée la fréquence d'apparition des différentes structures, la courbure moyenne et la convexité moyenne. Quant à lui, [Behnke 2003] utilise un recalage élastique par la méthode d'HAMMER [Shen 2001] sur un atlas volumique de gyri [Kabani 1998] qui sert également à définir une description anatomique de chaque zone surfacique à labéliser selon son profil gyral (fréquence d'apparition des labels de l'atlas sur la zone d'intérêt).

L'information corticale est souvent résumée aux seules informations sulcales (relatives au sillons : les creux de la surface corticale) ou gyrales (relatives aux gyri : les bosses). D'un point de vue purement anatomique, les structures gyrales sont plus complexes et moins bien définies que les sillons. En effet, un sillon se présente souvent comme un amas de petits plissements reliés entre eux, ou tout du moins très proches les uns des autres, alors qu'un gyrus n'a pas de frontière bien délimitée puisque la plupart des gyri sont connectés entre eux. D'ailleurs, en pratique, un gyrus est souvent défini par rapport aux sillons le délimitant. Sous l'angle d'informations issues d'IRM anatomiques, les sillons restent une façon efficace de représenter l'essentiel de l'information anatomique.

Par contre, considéré sous l'angle d'informations issues de l'IRM de diffusion ou fonctionnelle, on peut s'attendre à une conclusion différente. En effet, en IRM de diffusion, les fibres de la matière blanche semblent se répartir de manière préférentielle en éventail dans les gyri [Essen 1997]. Sous ces conditions, l'activité fonctionnelle y est sans doute également concentrée, là où le nombre de connexions neuronales est le plus concentré.

Le travail présenté dans cette thèse a été initié uniquement avec des données anatomiques, ce qui nous a fait préférer la représentation sulcale. En début de 3^{ème} année de cette thèse, un effort particulier a été mené pour élargir la base de données anatomiques sulcales (explications détaillées en annexe A) avec le souci d'y inclure également des informations issues de l'IRM de diffusion et de l'IRM fonctionnelle. Il sera sans doute nécessaire, lors de futurs travaux dépassant le cadre actuel de cette thèse, d'associer ces informations complémentaires reposant sur des supports différents (gyri et sillons) qu'il faudra alors mettre en relation.

3.1 Représentations des sillons

Dans le contexte de la cartographie cérébrale (*brain mapping*), les sillons ont une place privilégiée puisqu'ils représentent de bons candidats pour définir des amers anatomiques entre individus. Ils ont donc suscité un intérêt particulier qui se retrouve dans la diversité des représentations présentes dans la littérature. Ces approches se déclinent selon leur degré d'approximation, leur fiabilité à reproduire la complexité de la structure sulcale. On distingue les approches basées sur un ensemble de points où seules certaines parties critiques de la forme de la structure sont conservées, sur des lignes sulcales où les embranchements et les interruptions sont négligées, sur des segments sulcaux (plissements élémentaires) représentés implicitement sur la surface corticale par les berges entourant le sillon ou explicitement par une structure représentant le pli en lui-même sous la forme d'une surface médiane. Nous allons détailler dans la suite ces différentes approches.

La représentation la plus basique des sillons corticaux et sans doute la plus économique en terme de compression de données est réalisée par un nuage de points. Cette représentation a été popularisée par le modèle de distribution de points (*Points Distribution Model*, PDM [Taylor 1992, Caunce 1998, Lohmann 2000]) qui consiste (1) à réaliser une analyse Procrustes généralisée [Gower 1975] afin d'aligner un ensemble de points labélisés entre eux, suivi par (2) une analyse de la structure de covariance de l'ensemble des sujets, de telle façon que chaque sujet soit représenté par un grand vecteur concaténant l'ensemble de ses points descriptifs. Cette approche reste limitée puisqu'elle impose à chaque sujet d'être décrit par un nombre de points constant et prédéfini, ce qui est très difficile à garantir en pratique. Toutefois, l'étude du développement des plissements corticaux suggère que certains points singuliers généralement profondément enfouis dans les sillons seraient présents et localisés de la même façon chez tous les individus. Il s'agit notamment des racines sulcales [Régis 2005] (à la base de notre nomenclature de labels de sillons, voir annexe A), des cavités sulcales (*sulcal pits*) [Lohmann 2008, Im 2009] si l'on considère les points les plus profonds des plissements, ou des bassins sulcaux [Lohmann 2000] si la zone d'intérêt est étendue à tout un volume avoisinant et délimité par des singularités de courbure au niveau des lignes de fond de sillon. L'étude multi-échelle de la courbure de la surface corticale semble révéler certains de ces plissements primordiaux [Cachia 2003a].

L'anatomie corticale telle qu'on peut la découvrir dans des ouvrages de médecine se résume généralement par une présentation simplifiée où les sillons y sont représentés par de grandes lignes plutôt rectilignes pour les sillons primaires et éventuellement par quelques lignes entrecoupées pour les sillons secondaires, les tertiaires étant généralement peu traités. Cette schématisation caricaturale des plissements corticaux témoigne de la compréhension encore très limitée du phénomène de plissement et des *patterns* qu'il engendre. L'omni-présence de la représentation des sillons sous la forme de lignes sulcales dans les milieux académiques explique en partie pourquoi elle est aussi populaire en neuroimagerie.

Cette fois encore, il s'agit d'une représentation très concise de l'information sulcale et relativement facile à manipuler, mais beaucoup plus réaliste que la précédente (tout du moins tant qu'on ne sera pas capable d'identifier avec certitude les racines sulcales). On retrouve cette représentation dans de nombreuses études concernant l'extraction et l'identification de lignes sulcales (il s'agit le plus souvent de lignes continues ininterrompues) [Thompson 1997, Lohmann 1998, Royackkers 1999, Tao 2002, Fillard 2007a, Lui 2006, Kao 2007, Tu 2007, Kobashi 2007, Mani 2008, Auzias 2009, Hurdal 2008, Shi 2008, Tosun 2008, Shattuck 2009]. Cette représentation s'intègre facilement à des méthodes de recalage dérivées du procrustes [Tao 2002] ou basées sur les courants [Durrleman 2007, Auzias 2009] ou comme contraintes dans une représentation conforme (*conformal mapping*) [Lui 2006]. Dans le cadre de ces études, les auteurs se concentrent sur les plus grands sillons (généralement moins d'une dizaine par hémisphère) aux formes les plus simples (primaires

et secondaires). Dans ce cadre, cette représentation peut sembler raisonnable, mais elle néglige notamment les interruptions qui sont fréquentes sur la plupart des sillons secondaires : post-central, pré-central, temporaux, frontaux. Sur les structures primaires, les interruptions sont rares mais existent néanmoins chez certains sujets sains. De la même façon, les ramifications, les branches secondaires sont rarement prises en compte alors qu’elles sont omniprésentes, prenant la forme d’une sorte de bruit résiduel de plissement.

Les lignes sulcales demeurent encore une approximation populaire, mais l’on peut déplorer le manque d’attention accordée à l’influence de cette approximation sur les sillons où l’hypothèse se vérifie le moins. De plus, cette représentation présente un problème évident de validation, puisqu’il est plutôt difficile d’établir des règles strictes avec un sens anatomique (la plupart du temps en cas d’interruption par un gyrus la ligne se poursuit le long de celui-ci) permettant de garantir la cohérence du tracé de ces lignes sulcales réalisé par des experts humains. Dans le cadre de l’extraction automatique de telles structures, l’évaluation de la qualité d’un tracé automatique consiste à comparer ce dernier avec le tracé manuel selon des critères ad-hoc qui ne permettent pas de quantifier avec justesse les performances réelles de ces méthodes. Notre représentation, comme nous le verrons plus loin, répond plutôt bien à cette question.

Les parois d’un sillon ne présentent pas forcément un profil similaire à celui du fond de ce même sillon. Afin de capter cette information dans le sens de la profondeur tout en maintenant une représentation simple et pratique à manipuler, un sillon peut être représenté par un ruban sulcal (*sulcal ribbon*) proposé par [Vaillant 1997] (en se basant sur une méthode de contour actif ou *snake*), par [Goualher 1997] sous la forme de B-splines, par [Zeng 1999] grâce à une méthode de *level-set* et par [Zhou 1999] basée sur une technique d’extraction nommée *voxel coding*. Ces rubans corticaux se présentent généralement comme la surface médiane entre les deux berges définissant un sillon cortical. Les formulations proposées intègrent la possibilité d’interrompre le ruban en 2 sous-rubans si le sillon est partagé en plusieurs plis séparés par des gyri, voire même par d’autres sillons. Cette représentation a essentiellement été utilisée afin de définir une paramétrisation (2D : dans la longueur et la profondeur) des sillons généralement utilisée pour réaliser des statistiques de forme, créer des sillons moyens ou servir de contrainte dans un recalage ([Collins 1998] basé sur [Goualher 1997] et [Chui 2003] basé sur [Zeng 1999]).

Les dernières représentations qu’on retrouve dans la littérature intègrent l’ensemble de l’information anatomique portée par les structures sulcales, y compris les interruptions, les branches secondaires, les embranchements et les variations de courbure dans le sens de la profondeur du plissement. Il est intéressant de constater que l’ensemble des articles utilisant cette représentation vise à labéliser précisément ces structures anatomiques que ce soit manuellement ou automatiquement. On distingue deux représentations principales : les représentations surfaciques corticales [Behnke 2003, Rettmann 2005, Vivodtzev 2005, Yang 2009] et les représentations volumiques de l’interface entre les 2 parois avoisinant les sillons [Mangin 1995a, Goualher 1998, Klein 2005, Rivière 2002, Perrot 2009]. Chaque sillon est alors représenté par une ou plusieurs régions sulcales (selon la terminologie de [Rettmann 2005]) pour la première catégorie et par un ou plusieurs segments sulcaux (ou *sulcal pieces* terme que nous avons introduit dans [Perrot 2008]). Ces approches partagent les mêmes avantages et inconvénients. Chacune de ces structures macroscopiques ne peut porter qu’un seul et unique label de sillon ce qui limite considérablement le nombre de solutions possibles. De plus, on constate en pratique que les structures extraites sont plutôt représentatives de l’anatomie corticale et ce quelquesoit la méthode employée. Les lignes sulcales ou les rubans corticaux en comparaison ne sont que des approximations dont la simplicité de représentation masque la plupart des difficultés ; le sens anatomique de telles représentations est souvent discutable et biaise favorablement les résultats en leur faveur puisque la qualité des appariements ne tient pas compte des parties les plus variables : les branches. Au contraire les approches basées sur une parcellisation des structures

sulcales (segments ou régions) reposent sur des structures calculées une fois pour toutes dont la signification et la qualité n'a généralement pas besoin d'être remise en cause. La seule difficulté commune rencontrée est qu'il est difficile de garantir qu'un seul et unique label sulcal s'exprime sur chaque parcelle sans les surdécouper. Finalement, dans ce cadre, la qualité d'une labélisation est naturellement déterminée par le pourcentage de structures bien étiquetées (en pondérant éventuellement les résultats par la taille des structures).

Les régions et les segments sulcaux diffèrent toutefois sur un point essentiel. Dans le cas de l'approche corticale l'information est utilisée directement sur la surface du cortex, y compris les asymétries de courbure le long des deux parois ou les variations d'ouverture du sillon. Dans le cas des segments sulcaux, on obtient une représentation moyenne (entre les deux berges du sillon) de la courbure locale et l'information d'ouverture a disparu puisqu'on considère une structure médiale aux parois constituant le sillon, ce qui devrait en faire une représentation moins sensible à l'âge des sujets (puisque l'ouverture des sillons augmente avec l'âge).

Ces multiples représentations posent néanmoins un problème sérieux de comparaison des résultats, puisqu'elles ne sont pas directement équivalentes. En effet, il n'existe aucune procédure standard permettant de convertir une représentation en une autre. Le problème de conversion est d'ailleurs plutôt mal posé dans le cas des sillons aux formes arborescentes où plusieurs représentations approchées sous forme de lignes sulcales pourraient coexister. Néanmoins, un effort devrait sans doute être réalisé dans cette direction afin de permettre la comparaison des différentes méthodes de labélisation de la littérature.

3.2 Brainvisa : notre représentation

Ce travail de thèse s'inscrit dans une succession de travaux réalisés dans le cadre du projet Brainvisa [Rivière 2009]. Jusqu'à présent, nous nous sommes contenté d'utiliser la représentation des structures sulcales fournie par la chaîne de traitement anatomique. Les plissements corticaux y sont représentés par une collection de segments sulcaux comme décrits dans la section précédente. À la lumière des différentes conclusions appararissant tout au long du manuscrit, il ne semble pas nécessaire de remettre en cause ce choix de représentation, ou alors uniquement par curiosité. À une exception notable, au niveau de l'insula, qui reste une structure particulière où réside une dizaine de sillons qui demeurent pour certains mieux représentés directement sur la surface corticale.

Le déroulement de la chaîne de traitement anatomique a été décrit originellement dans [Mangin 1995b] et plus récemment dans [Mangin 2003a], les descriptions les plus détaillées et bien illustrées se trouvent dans [Rivière 2000, Rivière 2002]. Les détails de ces opérations ne sont pas nécessaires à la compréhension de la suite de ce document, nous nous contenterons donc ici d'un résumé rapide des étapes principales, en insistant sur les points importants. Les solutions proposées dans cette chaîne de traitement sont autant que possible insensibles au choix de la séquence IRM, ce qui a permis le traitement de plusieurs milliers de cerveaux en près de 15 ans.

L'extraction des données corticales d'un sujet sain adulte peut être réalisée à partir d'un volume IRM pondéré en T1 qui présente un bon contraste permettant de séparer matière grise, matière blanche et LCR. La chaîne de traitement anatomique a surtout été utilisée sur des sujets adultes mais depuis quelques années, plusieurs études ont étendu son cadre d'application aux données d'enfants, de bébés, de prématurés, et même de fœtus. Des images pondérées en T2 ou une combinaison des contrastes T1 et T2 seraient alors nécessaire dans le cas de données fœtales [Cachia 2003a], ou d'enfants nés avant terme [M. Radoš 2006].

Avant toute chose, une étape classique de correction de biais est appliquée afin de tenir compte des inhomogénéités du champ magnétique durant l'acquisition. Cette étape optimise un champ lisse

multiplicatif à l'image d'origine, qui minimise l'entropie de la distribution des niveaux de gris de l'image corrigée.

L'histogramme de l'image est étudié selon une analyse espace-échelle (*scale-space*) [Mangin 1998] afin de détecter les modes correspondant à la matière grise et à la matière blanche. Une estimation de la moyenne et de la variance en intensité est ainsi obtenue pour la matière grise et la matière blanche.

Les résultats de l'étape précédente sont utilisés pour segmenter le cerveau. Cette étape s'appuie sur une modélisation Markovienne permettant de régulariser spatialement la solution et sur des opérations de morphologie mathématique qui permettent de conserver les gyri les plus fins [Mangin 1998]. Une succession de traitements morphologiques classiques permettent de séparer la matière blanche en 5 parties : les deux hémisphères cérébraux, le tronc cérébral et les hémisphères du cervelet [Mangin 1996].

Finalement, la matière grise et le LCR sont séparés de la matière blanche. L'opération garantit que l'interface gris/blanc ainsi extraite est de topologie sphérique [Mangin 1995c] en se basant sur une succession d'opérations homotopiques élémentaires appliquées à une boîte englobant le cerveau. Cet enchaînement est conduit par la minimisation d'une énergie issue d'un modèle Markovien de type Ising.

Finalement, la matière grise associée au LCR est squelettisée [Mangin 1995d] avec le même souci d'homotopie [Mangin 1995c]. Cette opération est guidée par une méthode de «montée des eaux» (*watershed*) réalisée sur la courbure locale moyenne de cette structure qui met en évidence les lignes de crêtes correspondant à l'emplacement des sillons corticaux. Ce squelette a l'apparence d'un moule en négatif du cerveau, composé d'une enveloppe visible de l'extérieur et d'une série de branches rentrant dans les sillons. Les sillons sont alors extraits de cette structure d'après des considérations de topologie discrète.

En l'état, les sillons ainsi découpés peuvent représenter plusieurs labels sulcaux, puisque certains plissements peuvent se rejoindre et ne faire plus qu'un. En se basant sur la nomenclature issue des racines sulcales [Régis 2005] (mais il en serait sûrement de même avec une définition différente des structures sulcales), ce type de situation arrive régulièrement. De plus certains plissements secondaires peuvent se retrouver enfouis dans un plissement plus gros. Ainsi, le squelette comporte des ramifications secondaires qu'il est nécessaire de séparer. Finalement, une dernière opération de sur-segmentation est réalisée afin de garantir autant que possible qu'un label au plus s'exprime sur une telle structure sulcale. Cette étape étudie les points de l'interface gris/blanc dont la courbure Gaussienne est négative, ainsi que les minima locaux de la profondeur géodésique (marquant la trace d'un gyrus atrophié) qu'elle utilise comme graines dans un processus de parcellisation dans une opération de «montée des eaux» (*watershed*).

Les structures finalement obtenues (appelées surfaces simples dans [Rivière 2000]) seront désignées sous le terme de «segments sulcaux» dans la suite de cette thèse. Il s'agit de structures élémentaires, sans trous, sans branches, aux variations de courbure légères ; ils constituent l'unité de base de notre labélisation. Dans l'ensemble des modèles présentés par la suite, on ne pourra attribuer qu'un unique label à un segment sulcal donné. Le problème de labélisation est donc double : il s'agit de reconstituer des structures tout en leur donnant des noms. Afin de découpler la difficulté de chacune de ces tâches, une expérience a été réalisée à cet effet. Elle consistait en la reconstitution artificielle des sillons d'après leurs véritables labels de telle façon que chaque label ne peut être porté que par un unique segment sulcal. Dans de telles conditions, lorsque la segmentation est supposée connue, la labélisation automatique des structures sulcales devient quasi-triviale (le taux de reconnaissance global du modèle SPAM depuis Talairach est inférieur à 5%). Cette expérience démontre que la difficulté principale du processus de labélisation réside dans la reconstitution des sillons à partir des segments sulcaux portant le même label. Si l'on imagine disposer d'une segmentation idéale où chaque segment sulcal correspond directement à un sillon, de telle façon qu'un label ne correspond qu'à une seule structure, alors la labélisation des structures corticales n'est plus

une difficulté. L'étape de segmentation est donc cruciale, un compromis reste à définir entre sur-segmentation permettant de bien séparer les structures d'une part et limiter le nombre de segment afin de simplifier le problème de labélisation d'autre part.

Aucune amélioration n'a été apportée à cette procédure d'extraction des sillons durant cette thèse. Toutefois, il semble désormais inévitable de retravailler certaines étapes afin de renforcer la robustesse de l'opération qui réclame ponctuellement des réajustements manuels. De plus, nous avons constaté que la qualité des données sulcales situées en zone basale (sous le lobe temporal et occipital) aurait beaucoup à gagner d'une meilleure segmentation gris/blanc. Par suite, la modélisation de ces structures et leur labélisation automatique en bénéficierait également.

Modélisation : concepts de base

Nous supposerons dans ce chapitre (cette restriction sera levée dans la partie II) que tous les sujets étudiés vivent dans un espace commun, dont les coordonnées 3D sont liées dans une certaine mesure à des informations anatomiques pertinentes. On peut considérer que les données ont été pré-traitées par une normalisation (affine ou non) garantissant un alignement raisonnable des structures sulcales au travers de l'ensemble des sujets. En effet, nous verrons au cours de cette partie I que de nombreuses informations descriptives de la variabilité sulcale et discriminatives (permettant de distinguer les sillons entre eux) sont plutôt sensibles à la qualité de cette première étape, telles que des informations de position ou de direction des structures. Lors de nos premiers essais, l'espace commun utilisé était celui défini par Talairach [Talairach 1988] qui fournit un alignement limité (au départ, il a été mis au point pour aligner les structures sous-corticales) mais suffisant en première approximation (voir la section 7.1 pour quelques suggestions de normalisations alternatives).

La modélisation proposée précédemment (section 2) s'appuyait essentiellement sur un modèle Markovien basé sur des potentiels adhoc, évalués par des experts locaux, chacun dédié à un sillon particulier ou à 2 sillons voisins. Pour un hémisphère donné plus de 5000 descripteurs (27 descripteurs par sillons et 23 par relations pour le modèle le plus utilisé) étaient utilisés pour caractériser des informations de forme, de taille, de position ou d'orientation. Les études (voir section 2.2) menées alors ont montré qu'il était plutôt difficile de savoir quels descripteurs contribuaient le plus à discriminer les labels entre eux. Après avoir longuement testé des versions dégradées de ce système, nous en sommes venu à la conclusion simple que nous manquions de point de comparaison avec des méthodes naïves standards. De telles méthodes seraient loin d'être aussi bien adaptées au problème que ne l'est le système «historique», mais elles seraient beaucoup plus faciles à comprendre et à évaluer de façon précise sur des données réelles (point qui fait cruellement défaut à l'approche «historique» et qui a freiné de nombreuses fois l'interprétation des résultats). Une expérience simple que nous évoquerons plus loin (voir chapitre 5), nous a convaincu que certaines informations positionnelles relativement discriminantes étaient sous-exploitées dans le modèle «historique», sans doute noyées dans le bruit engendré par 5000 descripteurs.

L'approche «historique» est localement discriminative : elle cherche à modéliser une fonction qui permet de distinguer chaque structure des autres. Plus précisément, dans ce cadre, chaque expert local est dédié à la discrimination d'un label donné vis-à-vis d'autres labels ou de contre-exemples de ce label (candidats non-représentatifs de la classe), sans qu'aucune description directe de chaque catégorie soit nécessaire. Cette approche est partagée par les travaux de [Shi 2007, Shi 2008, Shi 2009] dans le cadre de l'extraction de lignes sulcales (nous y reviendrons dans le chapitre 11). Puisque le but premier de ce travail est de labéliser les sillons du cortex, une telle approche peut remplir ce rôle (nous proposerons une alternative en ce sens dans la section 10.1.2). Toutefois, rien n'interdit d'utiliser une approche générative qui décrit avec attention les structures associées aux labels et leurs variations sur une base de données. C'est en confrontant les structures sulcales aux différents modèles sulcaux possibles que la discrimination s'établit. Dans un cadre Bayésien, cette phase se traduit par l'inférence des labels en maximisant la probabilité a posteriori adéquate. Par contre, une attention particulière doit alors être accordée à la définition des différents modèles utilisés puisqu'on ne contrôle pas directement la discrimination des structures. L'approche générative comporte un avantage certain en terme de compréhension : les modèles manipulés, les paramètres estimés sont plus en lien avec les structures et plus faciles à confronter à nos propres connaissances en anatomie.

Le choix des modèles, ainsi que l'introduction d'a priori issus de l'expertise d'un anatomiste est alors possible. Pour illustrer cette idée, il est souvent plus simple (même pour un expert) de décrire et limiter le champ de variation d'un sillon, que de décrire la frontière permettant de le distinguer d'un autre sillon. Finalement, n'oublions pas que la difficulté essentielle de la labélisation des sillons réside dans la reconstitution des segments les composant. Ce qui signifie que nous ne chercherons pas à déterminer si un plissement donné correspond intégralement à tel ou tel sillon, mais uniquement s'il en fait partie. L'approche discriminative peut être alors mise en défaut, puisqu'elle sera amenée à évaluer des propositions de sillons reconstitués (en aglutinant tous les segments portant le même label) qui n'ont aucune réalité anatomique (c'est la raison pour laquelle l'approche « historique » utilisait des « faux » sillons afin de modéliser ces situations). L'approche générative, quant à elle, modélise individuellement la variabilité associée à chaque label ; chaque ajout ou retrait de segment dans la composition d'un sillon a donc un effet direct sur son évaluation par le modèle associé, puisque ces modifications le rapprochent ou l'éloignent du cœur de cette distribution. Dans le cas d'une approche discriminative, l'interprétation n'est pas immédiate puisqu'elle se fait conditionnellement à une deuxième classe (contres-exemples, autres labels...). La qualité des réponses d'un tel modèle dépend alors fortement de cette classe puisqu'elle permet de délimiter le domaine d'expression du label considéré.

La représentation choisie pour décrire les données impose certaines restrictions naturelles sur le modèle (voir section 3.1 pour une description détaillée de la représentation choisie ainsi que des représentations alternatives). En effet, pour un sujet donné, les structures à labéliser sont représentées par un ensemble de plissements élémentaires (chacun décrit en pratique par un ensemble de voxels contigus dans l'espace, sans trous, sans branche, ni courbure trop élevée) nommés *segments sulcaux*. Ces plissements sont idéalement de nature sulcale ou sous-sulcale, puisqu'une telle structure est sensé prendre au plus un label, alors qu'il faut un ou plusieurs segments pour composer un sillon. Pour un sujet donné, à partir de sa seule IRM T1, l'ensemble des pré-traitements de la chaîne de traitement anatomique de Brainvisa [Rivière 2009] permet d'obtenir une segmentation que l'on notera \mathcal{E} . Cette information est supposée connue et fixée, c'est à dire que l'on ne remettra pas en cause ce découpage élémentaire a posteriori pendant la phase d'inférence des labels. Dans l'idéal, ce découpage est un bon compromis entre 1) finesse de découpage permettant de n'avoir qu'un seul label par segment, 2) qualité de l'information anatomique locale disponible sur chaque segment de manière indépendante et accessoirement 3) combinatoire du problème. En effet, plus le découpage est fin, meilleures seront les séparations entre les structures, plus l'information anatomique locale s'amenuise et plus la dimensionalité du problème (nombre de configurations possibles) augmente, rendant sa résolution plus difficile. Nous ne rentrons pas plus dans ce débat dans cet ouvrage, puisque nous nous sommes contenté d'utiliser le paramétrage par défaut de l'étape de segmentation, qui est suffisant en première approximation (cette restriction correspond juste à un choix stratégique, elle sera sans doute levée dans de futurs travaux).

Nous allons maintenant rentrer plus dans les détails de la modélisation en proposant, tout d'abord, une modélisation formelle du problème sur laquelle repose l'essentiel des modèles à venir. De façon plus concrète, nous détaillerons certaines spécialisations d'intérêt par la suite, concernant la modélisation sulcale locale et l'a priori sur la labélisation. Les deux grandes parties (parties II et III) de cette thèse à suivre proposeront chacune des raffinements à ce premier modèle selon deux directions complémentaires.

4.1 Modélisation Bayésienne indépendante

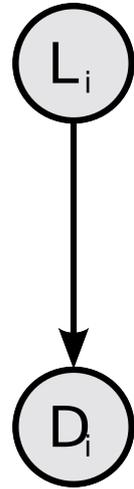
Les connaissances en anatomie du cortex que nous avons à notre disposition (expérience propre, littérature, expertise par de véritables anatomistes) ne sont pas suffisantes pour définir à elles seules une modélisation de la variabilité sulcale. Par contre, cette connaissance a priori prend tout son sens

confrontée à des données réelles. On se propose ici d'utiliser un ensemble de données d'apprentissage noté \mathcal{A} sur lequel l'ensemble des informations potentiellement utiles à notre étude est considéré connu; notamment ces données ont été labélisées manuellement (la base de données utilisée est décrite en annexe A) afin de décrire les plissements corticaux avec plus de 60 labels par hémisphère (63 à gauche, 62 à droite). À partir de là, les variations de ces structures (en termes de forme, taille, longueur, position, ...) peuvent être mises en corrélation avec les labels afin d'estimer l'ensemble des paramètres d'un modèle représentant cette variabilité. Nous avons d'ailleurs fait le choix délibéré de limiter fortement l'introduction de tels a priori de telle manière qu'il n'y ait pas de spécificité liée à un label particulier. Cette restriction peut sembler assez draconienne puisque la compréhension du cortex est loin d'être homogène : certains sillons sont plutôt bien connus (la plupart des sillons primaires) alors que d'autres le sont beaucoup moins (sillons secondaires et tertiaires), voire presque pas (les sillons du lobe occipital notamment). Il s'agit de notre part, d'une règle de prudence qui s'est imposée d'elle-même : au départ par méconnaissance de l'anatomie corticale en ce qui me concerne, et ensuite par prise de conscience de l'étendue de la variabilité corticale (vision, sans aucun doute, encore partielle puisque tirée de l'étude d'à peine quelques centaines de cerveaux de type caucasien), et encore, on ne parle ici que d'individus sains âgés de 25 à 35 ans.

À partir de cette base d'apprentissage, notre but est donc d'être capable de généraliser leurs labélisations manuelles individuelles à de nouveaux sujets à partir des seules informations issues de l'étape de segmentation. Deux approches permettent de répondre à cette attente : l'approche supervisée qui exploite uniquement les données d'apprentissage labélisées et l'approche semi-supervisée qui tire également partie de l'ensemble des données non-labélisées que l'on a à notre disposition.

4.1.1 Approche supervisée

Pour rendre la présentation un peu plus concrète, nous allons prendre comme angle d'attaque l'utilisation pratique qui va être faite de ce modèle : la labélisation des sillons d'un nouveau sujet. Considérons alors que nous disposons d'un modèle \mathcal{M} (dont les détails ne sont pas précisés pour le moment) estimé à partir d'une base d'apprentissage \mathcal{A} et que l'on cherche à labéliser les sillons d'un sujet t issu d'une base de test \mathcal{T} . Comme indiqué plus tôt, nous considérons que ce sujet a subi un premier recalage vers un espace commun à tous les sujets. Cette transformation est alors connue et sera négligée dans les équations à suivre. La labélisation L recherchée se base sur des informations notées D issues de l'IRM T1 du sujet (nous décrirons la nature possible de telles informations ainsi que la manière dont on peut en modéliser la variabilité en section 4.2). Les labels inconnus ainsi que les données structurelles mesurées sont exprimés sur l'ensemble \mathcal{E} des segments sulcaux du sujet (\mathcal{E}_l désignera la restriction de cet ensemble aux segments portant le label l), de telle sorte que $L = \{L_i\}_{i \in \mathcal{E}}$ et $D = \{D_i\}_{i \in \mathcal{E}}$, chaque segment i avec ses propres informations locales D_i (dédites uniquement des caractéristiques du segment i afin que ces informations puissent être considérées comme indépendantes d'un segment à l'autre) et un unique label associé L_i . On considérera dans un premier temps, que chaque segment et par conséquent chaque labélisation locale L_i est indépendante de ses voisins (cette restriction est levée dans la partie III), ce qui est évidemment faux mais permet de simplifier à l'extrême les schémas d'optimisation et d'obtenir une première estimation de la labélisation en des temps records. Cette supposition s'appliquera ou non sur $P(L|\mathcal{M})$ selon l'a priori choisi (voir section 4.3). Dans un contexte Bayésien, l'inférence d'une telle labélisation est basée sur une simple application de la règle de Bayes (pour alléger les expressions probabilistes, les variables aléatoires ainsi que les valeurs concrètes qu'elles peuvent prendre seront confondues



modèle
graphique

s'il n'y a pas d'ambiguïté manifeste : par exemple, $L = l$ deviendra simplement L :

$$\begin{aligned}
P(\text{Labélisation}|\text{Données}) &= P(L|D; \mathcal{M}) \\
&\propto P(L|\mathcal{M})P(D|L; \mathcal{M}) \\
&\propto P(L|\mathcal{M}) \prod_{l \in \mathcal{L}} P(\{D_i\}_{i \in \mathcal{E}_l} | \{L_i = l\}_{i \in \mathcal{E}_l}; \mathcal{M}) \\
&\propto P(L|\mathcal{M}) \prod_{i \in \mathcal{E}} P(D_i|L_i; \mathcal{M})
\end{aligned} \tag{4.1}$$

où l'on a supposé l'indépendance conditionnelle des données vis-à-vis des labels (hypothèse considérée tout au long de ce manuscrit) et l'indépendance des segments entre eux. Ici, $P(L|\mathcal{M})$ représente l'a priori sur la labélisation. C'est à dire, la chance d'avoir une telle labélisation avant même d'avoir observé les données D propres au sujet considéré. En effet, certaines labélisations sont hautement improbables quelqu'ait le sujet (par exemple, une labélisation où le label «sillon central» est absent est sans doute impossible sur des données réelles). Dans notre cas, les labélisations des segments sulcaux sont considérées comme étant indépendantes. Si l'on considère en plus une indépendance au niveau des segments pour l'a priori $P(L|\mathcal{M}) = \prod_{i \in \mathcal{E}} P(L_i|\mathcal{M})$ (ce que nous avons fait le plus souvent en pratique), cette formule correspond alors à la fois aux risques Bayésiens MAP (*Maximum A Posteriori*) et MPM (*Marginal Posterior Mode*) qui seront détaillés plus loin (voir section 10.3.1). Ce qui signifie que chaque problème de labélisation locale (au niveau d'un segment sulcal) a la même importance aux yeux de ce modèle, la labélisation optimale l^* vérifie alors :

$$l^* = \arg \max_l P(L = l|D; \mathcal{M}) \tag{4.2}$$

$$= \left\{ \arg \max_{l_i} P(D_i|L_i = l_i; \mathcal{M})P(L_i = l_i|\mathcal{M}) \right\}_{i \in \mathcal{E}} \tag{4.3}$$

Grâce aux indépendances statistiques supposées, l'étape de labélisation consiste donc simplement en un ensemble de labélisations locales au niveau de chaque segment. Pour un segment donné, on mesure alors la vraisemblance $P(D_i|L_i = l_i; \mathcal{M})$ qu'il soit issu de tel ou tel modèle dédié à la représentation d'un sillon particulier. Finalement, le label correspondant à la réponse la plus probable est conservée. De plus, les probabilités postérieures locales $P(L_i|D_i; \mathcal{M})$ peuvent être calculées par une simple normalisation :

$$P(L_i = l_i|D_i; \mathcal{M}) = \frac{P(D_i|L_i = l_i; \mathcal{M})P(L_i = l_i|\mathcal{M})}{\sum_{l_i \in \mathcal{L}} P(D_i|L_i = l_i; \mathcal{M})P(L_i = l_i|\mathcal{M})} \tag{4.4}$$

où \mathcal{L} correspond à l'ensemble des labels possibles. Cette expression permet de mesurer à quel point le modèle est confiant dans le choix de sa labélisation.

Dans les équations précédentes, \mathcal{M} correspond à l'ensemble des paramètres ajustant le modèle à une base d'apprentissage donnée. Ces paramètres incluent l'a priori sur la labélisation $P(L)$ et l'expression de la vraisemblance $P(D|L)$ (l'adéquation entre les structures sulcales représentées par D et la variabilité constatée de chaque label). Les données sulcales sont plutôt complexes et loin d'être bien comprises, c'est pourquoi ces modèles ne sont pas définis a priori mais estimés

empiriquement à partir d'une base d'apprentissage \mathcal{A} .

$$\begin{aligned}
& P(\text{Modèle} | \text{Données d'apprentissage}) \\
&= P(\mathcal{M} | \{D_a L_a\}_{a \in \mathcal{A}}) \\
&\propto \prod_{a \in \mathcal{A}} \underbrace{P(D_a | L_a; \mathcal{M}_s)}_{\substack{\text{modèles génératifs} \\ \text{sulcaux}}} \underbrace{P(L_a | \mathcal{M}_p)}_{\substack{\text{a priori sur} \\ \text{la labélisation}}} \underbrace{P(\mathcal{M})}_{\substack{\text{a priori sur} \\ \text{le modèle}}} \tag{4.5}
\end{aligned}$$

où chaque indice a correspond à une donnée spécifique au sujet a . Ainsi, l'optimisation est composée de 3 parties : les modèles génératifs de sillons, l'a priori sur la labélisation et l'a priori sur le modèle. Dans cette expression, \mathcal{M}_s et \mathcal{M}_p correspondent respectivement aux paramètres spécifiques aux modèles sulcaux et à l'a priori sur la labélisation. Dans le cadre des études que nous avons menées, l'a priori sur le modèle consiste uniquement à fixer la valeur de certains hyperparamètres non spécifiés ici, ce qui se traduit formellement par des distributions Dirac, chacune centrée sur la valeur d'intérêt. Ainsi, \mathcal{M}_s et \mathcal{M}_p peuvent être estimés séparément. De la même manière, les paramètres \mathcal{M}_l du modèle spécifique à un label l de sillon donné peuvent être optimisés indépendamment les uns des autres :

$$\prod_{a \in \mathcal{A}} P(D_a | L_a; \mathcal{M}_s) = \prod_{a \in \mathcal{A}} \prod_{i \in \mathcal{E}_a} P(D_{a,i} | L_{a,i}; \mathcal{M}_s) \tag{4.6}$$

$$= \prod_{l \in \mathcal{L}} \left[\prod_{a \in \mathcal{A}} \prod_{i \in \mathcal{E}_{a,l}} P(D_{a,i} | L_{a,i} = l; \mathcal{M}_l) \right] \tag{4.7}$$

où \mathcal{E}_a désigne l'ensemble des segments sulcaux du sujet a et $\mathcal{E}_{a,l}$ au même ensemble restreint au segment labélisé manuellement avec le label l . Ce qui signifie tout simplement que l'estimation de l'ensemble des paramètres dédiés à un label l donné se fait sur la base de l'ensemble des informations anatomiques définies par les segments sulcaux portant ce label.

Concernant l'a priori sur la labélisation, il vise à modéliser la variabilité induite par la segmentation, indifféremment des données structurelles D . Les approches envisagées sont détaillées dans la section 4.3.

4.1.2 Approche semi-supervisée

Dans la section précédente, seule les données de la base d'apprentissage \mathcal{A} étaient exploitées pour construire le modèle. De façon implicite nous considérons alors le modèle comme étant indépendant des données de la base de test. L'approche semi-supervisée combine l'approche supervisée basée sur des données labélisées et l'approche non-supervisée (regroupant notamment les méthodes de *clustering*). Elle vise à tirer partie de l'ensemble des informations disponibles en exploitant la structuration des informations entre elles à l'échelle d'une base de données et en les confrontant aux quelques données labélisées (voir figure 4.1). L'approche semi-supervisée présentée dans cette section est proposée pour son intérêt méthodologique et parce qu'elle est déjà fonctionnelle mais elle n'a pas encore été éprouvée sur des données réelles.

Que l'on considère l'approche supervisée ou semi-supervisée, il s'agit en général de méthode *inductive* qui vise à proposer un modèle robuste répondant à une réalité plus générale qui dépasse le cadre des données disponibles. Lorsqu'une méthode se focalise uniquement sur un ensemble de données de test \mathcal{T} particulier, l'exploitation semi-supervisée de l'ensemble de ces données pour inférer des labels constitue une méthode *transductive* [Vapnik 1998] qui ne prétend pas alors résoudre le problème dans sa globalité, mais uniquement sur le jeu de données non labélisé.

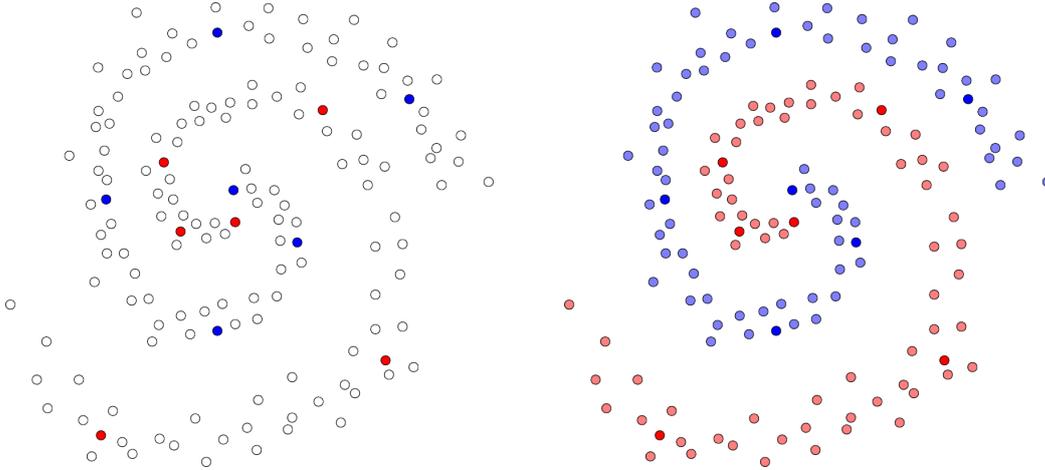


FIG. 4.1 – Apport d’une méthode semi-supervisée sur un cas d’école. À gauche : les données labélisées sont réparties en 2 classes : les points rouge et les points bleus. Les points blancs sont non labélisés et font partie de la base de test. Les points labélisés seuls ne semblent pas présenter de structure. À droite : résultat attendu avec une méthode semi-supervisée en utilisant de simples informations de voisinage. Cet exemple démontre ainsi que le modèle que l’on cherche à estimer doit être capable de s’adapter à la nature des données de test. Dans le cadre de la labélisation de sillons, on pense que les meilleurs modèles obtenus uniquement à partir des données d’apprentissage remplissent cette contrainte et sont représentatifs des données de test. Dans ce cas, les données de test permettraient de raffiner l’estimation des paramètres du modèle.

Dans le cas semi-supervisé l’estimation du modèle \mathcal{M} doit se faire en tenant compte des informations issues de la segmentation D_t des sujets t de la base de test \mathcal{T} . Puisque les labélisations L_t respectives de ces sujets sont inconnues, cette approche vérifie :

$$\begin{aligned}
 & P(\text{Modèle} | \text{Données d'apprentissage et de test}) \\
 &= P(\mathcal{M} | \{D_a L_a\}_{a \in \mathcal{A}} \{D_t\}_{t \in \mathcal{T}}) \\
 &= \sum_{\{L_t = l_t\}_{t \in \mathcal{T}}} P(\mathcal{M} | \{D_a L_a\}_{a \in \mathcal{A}} \{D_t L_t = l_t\}_{t \in \mathcal{T}}) P(\{l_t\}_{t \in \mathcal{T}}) \\
 &= \sum_{L_{\mathcal{T}} = l_{\mathcal{T}}} P(\mathcal{M} | D_{\mathcal{A}} L_{\mathcal{A}} D_{\mathcal{T}} L_{\mathcal{T}}) P(L_{\mathcal{T}} = l_{\mathcal{T}})
 \end{aligned} \tag{4.8}$$

où l’on somme sur toutes les labélisation possibles sur l’ensemble des sujets de la base de test. Dans la dernière ligne, nous avons introduit une série de notations afin d’en alléger considérablement la formulation ainsi que d’autres à venir. Ainsi, $\{D_t\}_{t \in \mathcal{T}}$ devient $D_{\mathcal{T}}$, l’ensemble des informations issues de la segmentation de tous les sujets de la base de test. De la même manière, $\{L_t = l_t\}_{t \in \mathcal{T}}$ devient $L_{\mathcal{T}} = l_{\mathcal{T}}$, où $l_{\mathcal{T}}$ correspond à un ensemble de labélisations, une pour chaque sujet de la base de test.

En tant que telle, l’optimisation directe de l’expression 4.8 semble plutôt difficile. Il s’agit en fait d’un problème classique d’inférence avec données cachées (les labels inconnus des sujets de la base de test). Dans ce contexte, l’algorithme EM (*Expectation-Maximization* [Dempster 1977]) répond à ce problème par une approche itérative. Il part d’une initialisation à définir et raffine le résultat initial au fur et à mesure des itérations. Cette approche permet d’atteindre un maximum local de la probabilité dont la qualité dépend fortement de la régularité de la fonction optimisée et de l’initialisation proposée. Dans notre cas, l’initialisation des paramètres du modèle, notée $\mathcal{M}^{(0)}$, la

plus naturelle et qui plus est de bonne qualité consisterait à prendre le résultat de l'estimation du modèle obtenu par la méthode supervisée présentée dans la section précédente. Cette méthode consiste à itérer le processus suivant :

$$\mathcal{M}^{(n)} = \arg \max_{\mathcal{M}} Q \left(\mathcal{M} \middle| \mathcal{M}^{(n)} \right) \quad (4.9)$$

où $Q \left(\mathcal{M} \middle| \mathcal{M}^{(n)} \right)$ correspond à une mesure d'espérance (étape *Expectation*) que l'on cherche à maximiser (étape *Maximization*). Dans la suite, nous noterons $C_x(\mathcal{M})$ des termes indépendents de \mathcal{M} , qui ne jouent donc aucun rôle dans la maximisation de $Q \left(\mathcal{M} \middle| \mathcal{M}^{(n)} \right)$. Détaillons maintenant ce terme :

$$\begin{aligned} Q \left(\mathcal{M} \middle| \mathcal{M}^{(n)} \right) &= \mathbb{E}_{L_{\mathcal{T}}} \left[\log P(D_{\mathcal{A}} L_{\mathcal{A}} D_{\mathcal{T}} L_{\mathcal{T}}; \mathcal{M}) \middle| D_{\mathcal{A}} L_{\mathcal{A}} D_{\mathcal{T}} \mathcal{M}^{(n)} \right] \\ &= \sum_{L_{\mathcal{T}}=l_{\mathcal{T}}} P \left(L_{\mathcal{T}} = l_{\mathcal{T}} \middle| D_{\mathcal{A}} L_{\mathcal{A}} D_{\mathcal{T}}; \mathcal{M}^{(n)} \right) \log [P(\mathcal{M} \middle| D_{\mathcal{A}} L_{\mathcal{A}} D_{\mathcal{T}} L_{\mathcal{T}} = l_{\mathcal{T}})] + C_1(\mathcal{M}) \end{aligned}$$

Or, les labels des données de test ne dépendent qu'indirectement des données d'apprentissage :

$$= \sum_{L_{\mathcal{T}}=l_{\mathcal{T}}} P \left(L_{\mathcal{T}} = l_{\mathcal{T}} \middle| D_{\mathcal{T}}; \mathcal{M}^{(n)} \right) \log \left[\frac{P(D_{\mathcal{T}} \middle| L_{\mathcal{T}} = l_{\mathcal{T}}; \mathcal{M}) P(L_{\mathcal{T}} = l_{\mathcal{T}} \middle| \mathcal{M})}{P(D_{\mathcal{A}} \middle| L_{\mathcal{A}}; \mathcal{M}) P(L_{\mathcal{A}} \middle| \mathcal{M}) P(\mathcal{M})} \right] + C_2(\mathcal{M})$$

En posant $\mathcal{B} = \mathcal{A} \cup \mathcal{T}$ et $w_{l_{\mathcal{T}}}^{(n)} = P \left(L_{\mathcal{T}} = l_{\mathcal{T}} \middle| D_{\mathcal{T}}; \mathcal{M}^{(n)} \right)$:

$$= \underbrace{\sum_{L_{\mathcal{T}}=l_{\mathcal{T}}} w_{l_{\mathcal{T}}}^{(n)} \log [P(D_{\mathcal{B}} \middle| L_{\mathcal{B}}; \mathcal{M}_s) P(\mathcal{M}_s)]}_{\text{modèles génératifs sulcaux}} + \underbrace{\sum_{L_{\mathcal{T}}=l_{\mathcal{T}}} w_{l_{\mathcal{T}}}^{(n)} \log [P(L_{\mathcal{B}} \middle| \mathcal{M}_p) P(\mathcal{M}_p)]}_{\text{a priori sur la labélisation}} + C_2(\mathcal{M})$$

Introduisons deux nouvelles notations pour désigner ces deux termes :

$$= Q_s \left(\mathcal{M}_s \middle| \mathcal{M}_s^{(n)} \right) + Q_p \left(\mathcal{M}_p \middle| \mathcal{M}_p^{(n)} \right) + C_2(\mathcal{M}) \quad (4.10)$$

Cette maximisation se décompose donc, comme dans le cas supervisé, en 2 termes indépendants : le premier consacré aux modèles génératifs dédiés à la reconnaissance d'un sillon particulier et le second à l'a priori sur la labélisation. Détaillons maintenant le premier. En posant $w_{l_t}^{(n)} = P \left(L_t = l_t \middle| D_t; \mathcal{M}_s^{(n)} \right)$, il s'exprime de la façon suivante :

$$\begin{aligned} Q_s \left(\mathcal{M}_s \middle| \mathcal{M}_s^{(n)} \right) &= \underbrace{\sum_{t \in \mathcal{T}} \sum_{L_t=l_t} w_{l_t}^{(n)} \log [P(D_t \middle| L_t = l_t; \mathcal{M}_s)]}_{(1)} + \log P(D_{\mathcal{A}} \middle| L_{\mathcal{A}}; \mathcal{M}_s) + \log P(\mathcal{M}_s) \\ &\text{car } \sum_{L_{\mathcal{T}}=l_{\mathcal{T}}} w_{l_{\mathcal{T}}}^{(n)} = 1. \end{aligned}$$

$$\text{Or (1)} = \sum_{t \in \mathcal{T}} \sum_{i \in \mathcal{E}_t} \sum_{L_{t,i}=l_{t,i}} P \left(L_{t,i} = l_{t,i} \middle| D_{t,i}; \mathcal{M}_s^{(n)} \right) \log [P(D_{t,i} \middle| L_{t,i} = l_{t,i}; \mathcal{M}_s)]$$

puisque les labels des segments sont considérés comme indépendants dans cette parti I, où $l_{t,i}$ correspond au label du segment i du sujet t .

$$= \sum_{l \in \mathcal{L}} \left[\sum_{t \in \mathcal{T}} \sum_{i \in \mathcal{E}_t} P \left(L_{t,i} = l \middle| D_{t,i}; \mathcal{M}_s^{(n)} \right) \log [P(D_{t,i} \middle| L_{t,i} = l; \mathcal{M}_l)] \right] \quad (4.11)$$

avec \mathcal{M}_l l'ensemble des paramètres du modèle dédié au label l . Or, concernant la base d'apprentissage :

$$\log P(D_{\mathcal{A}}|L_{\mathcal{A}}; \mathcal{M}_s) = \sum_{l \in \mathcal{L}} \left[\sum_{a \in \mathcal{A}} \sum_{i \in \mathcal{E}_{a,l}} \log [P(D_{a,i}|L_{a,i} = l; \mathcal{M}_l)] \right] \quad (4.12)$$

avec $\mathcal{E}_{a,l}$ l'ensemble des segments du sujet a restreints à ceux labélisés manuellement avec le label l . Donc finalement :

$$Q_s(\mathcal{M}_s | \mathcal{M}_s^{(n)}) = \sum_{l \in \mathcal{L}} \left[\sum_{b \in \mathcal{B}} \sum_{i \in \mathcal{E}_b} w_{L_{b,i}=l}^{(n)} \log [P(D_{b,i}|L_{b,i} = l; \mathcal{M}_l)] \right] + \log P(\mathcal{M}_s) \quad (4.13)$$

$$\text{avec } w_{L_{b,i}=l}^{(n)} = \begin{cases} P(L_{t,i} = l | D_{t,i}; \mathcal{M}_s^{(n)}) & \text{si } b \in \mathcal{T} \\ 1 & \text{si } b \in \mathcal{A} \text{ et } l = l_{b,i} \\ 0 & \text{sinon} \end{cases} ,$$

où $l_{b,i}$ correspond au label manuel du segment i du sujet b . L'estimation des paramètres associés à un label donné revient alors à tenir compte de l'ensemble des données d'apprentissage de la même façon que dans le cas supervisé et à prendre en compte l'ensemble des données issues de la base de test à hauteur des poids $w_{L_{b,i}=l}^{(n)}$. C'est à dire en tenant compte des incertitudes liées à leur labélisation automatique.

Concernant la partie de Q consacrée à l'a priori sur la labélisation, seuls les plus simples (basés sur la fréquences d'apparition des labels ou estimés par une loi de Dirichlet, voir section 4.3) vérifiant des propriétés d'indépendances fortes entre les labels, tel que $P(L_b | \mathcal{M}_p) = \prod_{l \in \mathcal{L}} P(\{L_{b,i}\}_{i \in \mathcal{E}_{b,l}} | \mathcal{M}_p)$ apportent des simplifications suffisantes permettant un calcul direct de l'expression (en suivant le même raisonnement que pour Q_s) :

$$Q_p(\mathcal{M}_p | \mathcal{M}_p^{(n)}) = \sum_{l \in \mathcal{L}} \left[\sum_{b \in \mathcal{B}} \sum_{i \in \mathcal{E}_b} w_{L_{b,i}=l}^{(n)} \log [P(L_{b,i} = l | \mathcal{M}_p)] \right] + \log P(\mathcal{M}_p) \quad (4.14)$$

Jusqu'à présent nous n'avons évoqué que l'estimation du modèle sans se préoccuper de ce qui nous intéresse avant toute chose, la labélisation des sillons des sujets de la base de test. On a pu voir au cours du processus que des poids $w_{L_{b,i}=l}^{(n)}$ venaient pondérer l'estimation des paramètres. Ils correspondent en fait aux probabilités a posteriori des labélisations associées. À l'issue du processus d'itération de l'algorithme EM, il nous suffit de sélectionner la labélisation la plus probable pour chaque sujet de la base de test. Ce qui revient à utiliser le même processus que dans le cas supervisé, si ce n'est que le modèle utilisé a bénéficié d'informations provenant de la base de test.

4.2 Comment modéliser la variabilité des sillons ? $P(D|L)$

La modélisation proposée fait apparaître un terme de vraisemblance $P(D|L)$ qui relie la labélisation L d'un sujet à un ensemble de données D issues de la chaîne de traitement anatomique de Brainvisa [Rivière 2009]. Deux problèmes se posent alors : quelles sont les informations pertinentes, permettant de caractériser les labels ? et comment modéliser de façon simple et robuste ces informations ? Le problème traité dans cette section est loin d'être clos et est susceptible d'évoluer en fonction des informations auxquelles nous aurons accès dans le futur (informations issues d'autres modalités par exemple).

La modélisation «historique» proposée précédemment se rapproche par certains aspects des *weak classifiers*. Un ensemble de classifieurs, chacun avec des performances limitées, incapables de résoudre individuellement le problème posé, permet d'élaborer un meta-classifieur performant issu de leur combinaison. La même approche a été suivie pour la description des sillons (ou des relations entre sillons dans ce modèle), puisque de nombreux descripteurs servent à caractériser chacun

des sillons. Chaque descripteur à lui seul est également «faible», il ne permet pas de résoudre le problème. C’est alors leur combinaison multivariée qui rend le processus de labélisation possible en donnant lieu à une discrimination bien meilleure. En réalité, les choses sont un petit peu plus compliquées en pratique. Si le nombre de données est beaucoup plus grand que le nombre de descripteurs considérés, tout se passe bien; mais dans le cas contraire (cas rencontré en pratique sur les données réelles), l’estimation robuste de modèles multivariés demande des précautions certaines afin de se prémunir contre la malédiction des grandes dimensions (*curse of dimensionality*). Plus le nombre de descripteurs est grand plus il y a de chance que certains descripteurs soient discriminants uniquement par hasard. Le bruit lié à l’acquisition ou à la segmentation et les imprécisions de mesures liées au calcul des descripteurs eux-mêmes participent à ce phénomène. L’ajout inconsidéré de nombreux descripteurs rend difficile leur exploitation et peut finir par nuire à leur potentiel discriminant, même si une réduction de dimensions automatique a été utilisée au préalable. Ces méthodes restent aveugles à la nature des données et ne peuvent pas toujours distinguer le bruit du reste avec aussi peu d’information. Seule une préselection des attributs guidée par des connaissances anatomiques et l’étude approfondie de leur robustesse aux sources d’imprécisions évoquées plus tôt, peut à ce stade faire la différence. En effet, si on applique le principe précédent, mais qu’on remplace les descripteurs «faibles» par d’autres déjà discriminants (s’ils existent), l’approche reste valide, mais le support est plus fiable et plus facile à exploiter. Les études de sélection automatique d’attributs menées dans le cadre du modèle «historique» ont révélé des difficultés méthodologiques indéniables qui nous ont convaincu qu’une sélection manuelle et une amélioration individuelle des attributs contribueraient plus efficacement à améliorer la qualité du modèle (tout du moins dans un premier temps).

La suite de cette section se décompose de la façon suivante. Tout d’abord, nous détaillerons quelles informations restent pertinentes à l’échelle d’un segment. Ensuite nous verrons comment ces informations peuvent être prises en compte pour l’estimation d’un modèle génératif ou pour déterminer la vraisemblance qu’un segment corresponde à un label donné. Et pour finir, nous détaillerons le modèle SPAM que nous avons largement privilégié par rapport à d’autres par sa capacité à traduire la variabilité de localisation des sillons (des compléments et extensions diverses se trouvent en annexe B.4).

4.2.1 Segments sulcaux et sillons

Le segment sulcal est l’unité élémentaire de notre labélisation. C’est l’équivalent des pixels d’une image. Un sillon est alors défini comme étant l’ensemble des segments portant le même label. Certaines informations caractéristiques d’un label particulier ne s’expriment qu’à l’échelle du sillon entier une fois reconstitué. De telles mesures sont la longueur, la surface, le volume, des descripteurs de forme globaux (moments invariants [Mangin 2003b, Sun 2007], invariants géométriques [Hurdal 2008], réponses à des filtres de Haar [Tu 2007]), ou toutes mesures locales moyennées à l’échelle du sillon : centre de gravité, direction moyenne, profondeur moyenne, épaisseur corticale moyenne... Lorsque l’information est traitée à l’échelle d’un sillon entier, l’évaluation pour un sillon l donné de l’exactitude de sa labélisation est mesurée grâce à : $P(\{D_i\}_{i \in \mathcal{E}_l} | \{L_i = l\}_{i \in \mathcal{E}_l})$, qui nécessite des stratégies d’optimisation complexes et coûteuses comme le recuit simulé (voir section 10.3.2.2) où chaque proposition de labélisation implique sa traduction sous forme de descripteurs (cette étape est parfois coûteuse). Bien entendu toutes ces mesures ne sont pas transposables à l’échelle des segments sulcaux, non pas qu’on ne puisse les calculer, mais que leur spécificité à représenter un sillon particulier peut être sérieusement réduite voire annulée. Par exemple, les informations de taille n’ont pas de signification particulière à ce niveau de représentation compte tenu de l’étendue de la variabilité de la segmentation des sillons en segments. Au contraire certaines informations restent fiables sur la plupart des segments. Nous avons choisi de focaliser nos efforts en priorité sur ces données car elles donnent lieu à des schémas d’optimisation plus simples et plus rapides. Ce constat justifie

notre proposition de modélisation locale indépendante qui est le sujet de cette partie I, puisque chaque segment i peut prétendre être labélisé individuellement à partir de ses seules informations locales : $P(D_i|L_i)$ que l'on peut précalculer une fois pour toutes.

Au niveau des segments seules quelques informations demeurent intéressantes, à noter principalement sa position, son orientation et l'épaisseur corticale (d'autres modalités permettront sans doute de compléter cette liste). Même à l'échelle d'un unique voxel ces informations ont toujours du sens (quitte à utiliser le voisinage dans le cas de la direction). À quelques détails près, l'information de profondeur pourrait rentrer dans cette catégorie mais elle comporte plusieurs défauts : elle est assez variable, notamment au niveau des extrémités et la segmentation peut entraîner des coupures dans le sens de la hauteur perturbant fortement l'utilisation de cette mesure. L'épaisseur corticale semble être un bon candidat, mais nous n'avons pas pris le temps de nous y intéresser dans le cadre de ce modèle, seules les informations de position et de direction ont été considérées en nous concentrant tout particulièrement sur la première catégorie où les modélisations les plus évoluées et efficaces ont été mises au point en pratique.

4.2.2 Modélisation au niveau des segments

Il existe plusieurs façons d'exploiter de telles informations à l'échelle des segments en fonction de la granularité de l'information et de leur fiabilité. Pour une information donnée (position par exemple), on distinguera 3 façons de représenter l'information : **(1)** par une mesure moyenne (ou globale), **(2)** par une mesure moyenne accompagnée d'une mesure de variance ou **(3)** par un ensemble de mesures locales (par exemple, l'ensemble des positions des voxels composant le sillon). L'exploitation de ces informations doit permettre d'élaborer un modèle capable d'évaluer la vraisemblance qu'un segment soit rattaché à un label donné. La contribution de chaque segment devrait sans doute être pondérée par sa taille pour rétablir l'équilibre entre les petits et les grands segments. Dans le cas de données de position nous avons utilisé le centre de gravité de chaque segment (la moyenne des positions des voxels le composant) que nous avons modélisé par une Gaussienne 3D avec une matrice de covariance pleine (6 paramètres : 3 axes principaux et facteurs d'échelle le long de ces axes). Dans le cas de données de direction (ou axiale lorsque le sens importe peu), nous avons utilisé des distributions de Kent B.3.1.3 ou de Bingham B.3.1.5 (approximation d'une Gaussienne à des données directionnelles et axiales respectivement).

Nous allons maintenant revenir sur l'apprentissage supervisé des paramètres de ce modèle génératif dans les 3 cas de figure évoqués plus tôt. Nous noterons x_i une telle donnée (par exemple le centre de gravité du segment) rattachée au segment i et ϕ les paramètres du modèle génératif considéré (seuls les segments portant un des labels fixé arbitrairement seront considérés pour l'exemple) sur lequel nous ne considérons pas d'a priori particulier (nous ometterons donc toute référence à $P(\phi)$).

Dans le cas **(1)**, l'estimation de ϕ par maximum de vraisemblance s'exprime ainsi sous sa forme logarithmique :

$$E_1 = \sum_{i \in \mathcal{E}} w_i \log P_{\mathcal{M}}(x_i | \phi) \quad (4.15)$$

où \mathcal{E} désigne l'ensemble des segments considérés, w_i d'éventuelles pondérations pour prendre en compte la taille des différents segments et la notation $P_{\mathcal{M}}(\cdot)$ indique qu'il s'agit de la vraisemblance exprimée par le modèle considéré.

Pour le cas **(2)**, en lieu et place de chaque x_i nous avons une valeur x_i et une matrice de covariance, qui permet de proposer une modélisation Gaussienne de la mesure empirique locale à chaque segment i , qu'on notera $P_e(x|S=i)$ où S est une variable aléatoire indicatrice du segment considéré. Elles forment une fois réunies sur l'ensemble des segments de la base de données, une probabilité empirique des données x considérées : $P_e(x) = \sum_{i \in \mathcal{E}} P_e(x|S=i)P_e(S=i)$, où $P_e(S=$

$i) = w_i$ correspond à l'a priori du segment i . Cette fois, l'optimisation des paramètres ϕ peut être réalisé en minimisant la divergence de Kullback-Leibler KL entre $P_e(x)$ et $P_{\mathcal{M}}(x|\phi)$:

$$KL(P_e(x), P_{\mathcal{M}}(x|\phi)) = \int_x P_e(x) \log \frac{P_e(x)}{P_{\mathcal{M}}(x|\phi)} dx \quad (4.16)$$

où l'intégrale s'exprime sur l'ensemble des positions spatiales dans notre exemple. Ce qui revient à maximiser l'énergie suivante :

$$\begin{aligned} E_2 &= \int_x P_e(x) \log P_{\mathcal{M}}(x|\phi) dx \\ &= \int_x \left[\sum_{i \in \mathcal{E}} P_e(x|S=i) P_e(S=i) \right] \log P_{\mathcal{M}}(x|\phi) dx \end{aligned} \quad (4.17)$$

Par analogie avec l'énergie du cas **(1)** (équation 4.15), le cas **(2)** le généralise et revient à pondérer chaque valeur de x possible à hauteur de la probabilité $P_e(x)$ dans un maximum de vraisemblance classique.

Le cas **(3)** consiste à remplacer l'expression précédente de $P_e(x|S=i)$ (modélisé par une Gaussienne) par une mesure empirique : $P_e(x|S=i) = \sum_{x_{ij} \in \mathcal{E}_i} w_{ij} \delta(x_{ij})$ où $\delta(x_{ij})$ correspond à une distribution de Dirac centrée en x_{ij} (la j ième mesure locale du segment i), w_{ij} correspond à l'information locale portée par la distribution de Dirac tel que $\sum_{x_{ij} \in \mathcal{E}_i} w_{ij} = 1$ et \mathcal{E}_i le découpage en zones d'intérêt locales. Nous verrons dans la section suivante comment tirer partie des informations locales par une approche alternative, non paramétrique (à partir d'estimateurs à noyaux) cette fois, dont l'optimisation se base sur un critère aux moindres carrés avec contraintes de régularisation [Girosi 1995], plutôt que la divergence de Kullback-Leibler.

Ces 3 représentations participent à la même logique et diffèrent par la façon dont sont résumées les informations à l'échelle d'un segment entier. Toutefois, une question d'importance demeure lorsque l'on cherche à labéliser un segment, puisque qu'on peut imaginer y exploiter les 3 mêmes représentations.

Le cas **(1)** ne pose pas de problème car la donnée moyenne manipulée correspond à ce que le modèle a appris et sait générer.

Dans les deux autres cas, un ensemble de mesures locales (l'ensemble des positions des voxels dans notre exemple) caractérisent chaque segment considéré. Il s'agit alors d'évaluer la probabilité jointe d'un ensemble de mesures noté x (représentées par la distribution empirique $P_e(x|S=i)$). Par analogie, avec l'étape d'estimation des paramètres (équation 4.17), la vraisemblance de cet ensemble de mesures sur le segment i selon le modèle \mathcal{M} est définie de la façon suivante :

$$P_{\mathcal{M}}(P_e(x|S=i)|\phi) \propto \exp\left(\mathbb{E}_{P_e}[\log P_{\mathcal{M}}(x|\phi)|S=i]\right) = \exp\left(\int_x P_e(x|S=i) \log P_{\mathcal{M}}(x|\phi) dx\right)$$

Ce qui se simplifie dans le cas **(3)** par :

$$\begin{aligned} &= \exp\left(\sum_{x_{ij} \in \mathcal{E}_i} w_{ij} \log P_{\mathcal{M}}(x|\phi)\right) \\ &= \prod_{x_{ij} \in \mathcal{E}_i} P_{\mathcal{M}}(x|\phi)^{w_{ij}} = \prod_{x_{ij} \in \mathcal{E}_i} P_{\mathcal{M}}(x|\phi)^{\frac{1}{|\mathcal{E}_i|}} \end{aligned}$$

puisque chaque information locale à un segment à la même importance. (4.18)

Cette règle induite de la divergence de Kullback-Leibler fait l'hypothèse implicite que chaque mesure x est indépendante des autres. De plus, nous avons résumé l'ensemble de l'information empirique

comme une mixture de mesures élémentaires. La variabilité des mesures entre elles ont donc été négligé ici. Pour aller plus loin on pourrait donc modéliser la distribution suivie par l'ensemble de ces mesures empiriques $P_e(x_i)$ (représentée chacune par une somme pondérée de Dirac). L'utilisation d'une modélisation de type Gaussien s'exprimant sur de telles données, permettrait de coder des covariabilités entre les mesures des différents segments. Dans le cas des données de localisation spatiale, cela permettrait sans doute d'incorporer des informations liées à la forme ou à l'orientation des segments dans le modèle.

4.2.3 Approche locale : SPAMs

Après une étape de normalisation affine standard vers un template ou vers un espace commun bien choisi, les sillons corticaux d'un groupe de sujets se retrouvent plus ou moins alignés les uns aux autres. C'est la raison pour laquelle la variabilité sulcale peut être modélisée dans une certaine mesure au travers de la probabilité de présence spatiale d'un sillon. Une façon très simple de l'exprimer et de considérer une mesure s'exprimant au niveau de chaque voxel et mesurant la probabilité que ce voxel appartienne à un sillon donné. À cette fin, nous nous sommes intéressés au modèle SPAM (*Statistical Probabilistic Anatomy Map* [Evans 1994]), tout d'abord utilisé dans le contexte de la modélisation sulcale dans l'étude menée par [Goualher 1998] qui possède de nombreux points communs avec la modélisation proposée dans cette partie I. Le principe de cette approche est de compter pour chaque label combien de sujets possèdent un sillon passant par une position donnée. Ces fréquences voxel à voxel sont notées pour chaque label l , par une fonction $f_l(x)$ qui dépend de la position spatiale 3D x . Ces mesures sont fiables seulement si de nombreux sujets sont utilisés. Avec la base de données que nous avons à notre disposition (comportant 62 sujets, voir annexe A pour plus d'informations), ce procédé permet l'obtention de modèles plutôt bien définis pour la plupart des labels. Pour les sillons les plus variables, beaucoup plus de sujets seraient nécessaires. En l'état, la carte spatiale des fréquences peut présenter des irrégularités induites par les empreintes individuelles des zones les plus variables ou des sujets *outliers* (voir figure 4.2). Pour surpasser cette difficulté et rendre l'estimation des fréquences plus robustes nous suggérons d'introduire une incertitude spatiale artificielle pendant la phase d'estimation des fréquences. Pour y parvenir un noyau Gaussien isotrope K_σ de variance σ^2 est utilisé pour lisser la carte des fréquences. La largeur du noyau σ a été fixée a priori à $2mm$ et devrait être étudiée avec attention dans le futur. Cette valeur n'est d'une part pas trop grande, permettant ainsi de conserver des modèles SPAM assez fins, et d'autre part pas trop petite pour atténuer les traces spécifiques à un individu particulier dans la carte de fréquences. Finalement, la probabilité de trouver une structure donnée portant le label l à l'emplacement x s'écrit :

$$P_{spam}(x|L=l) = \frac{(K_\sigma * f_l)(x)}{\sum_{x \in \Omega} (K_\sigma * f_l)(x)} \quad (4.19)$$

où $K_\sigma * f$ est la notation exprimant le produit de convolution entre f et K_σ . Ici, la constante de normalisation est calculée sur l'intégralité de l'espace Ω , qui correspond en théorie à \mathbb{R}^3 , l'espace des positions 3D dans son ensemble. En pratique, une boîte englobante est utilisée par label pour limiter le domaine de définition de chaque vraisemblance. Cette opération permet d'accélérer l'estimation du modèle et limite la mémoire nécessaire à sa définition. Si chaque label utilisait une boîte englobante de la taille d'un hémisphère, l'ensemble des modèles (nécessaires pendant l'étape d'inférence sur un sujet à labéliser) ne tiendrait pas en mémoire. Les modèles de SPAMs sont codés actuellement sous la forme de volumes 3D, centrés chacun sur le cœur de la distribution et limités par des vraisemblances négligeables (au delà de 10^{-15}), où chaque voxel contient la vraisemblance correspondant à sa position respective. Le facteur de normalisation permet de transformer un ensemble de fréquences en un véritable modèle génératif dédié à un sillon particulier, de telle façon que l'intégrale de la probabilité sur Ω vaille 1.

La formulation du modèle SPAM (équation 4.19) est aussi connu dans le cadre des estimateurs de densité comme étant l'expression d'une méthode de fenêtre de Parzen [Parzen 1962] ou plus généralement comme une estimation par une méthode à noyau. Cette interprétation ouvre la voie à d'autres estimateurs plus robustes que nous évoquerons en annexes (voir annexes B.4.2)

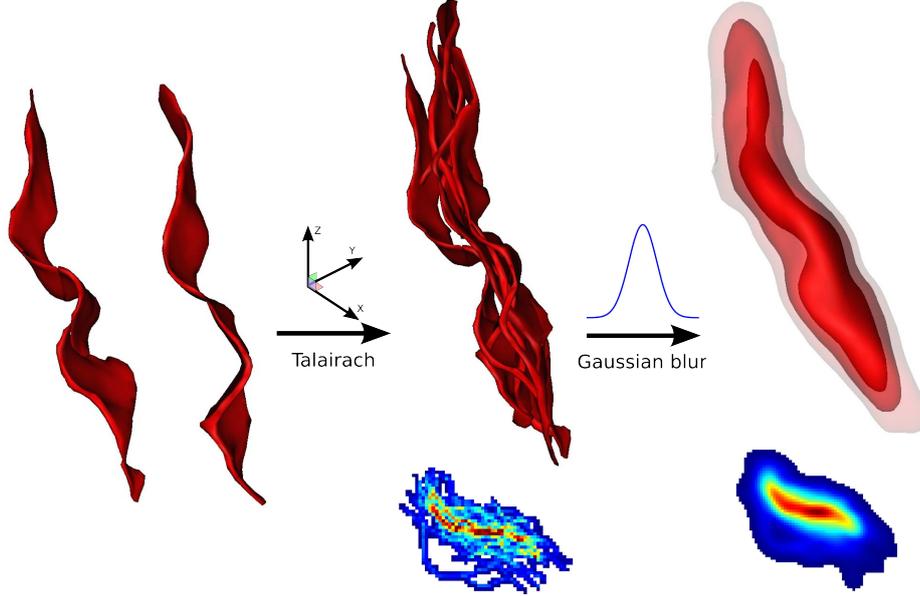


FIG. 4.2 – À label fixé, l'estimation du modèle de SPAM est réalisée en pratique simplement en comptant pour chaque voxel le nombre de sujets possédant une structure de ce label passant par ce voxel. Cette carte de mesure est alors lissée spatialement par un noyau Gaussien K_σ pour rendre la mesure plus fiable. La distribution est finalement alors obtenue par normalisation.

L'unité de base de notre labélisation est le segment sulcal (voir section 3.2). Comme nous l'avons montré dans la section précédente, pendant la phase d'inférence, on est amené à juger la vraisemblance qu'un ensemble de voxels constituant un segment soit issu du modèle SPAM correspondant à un label donné. Il s'agit donc de mesurer la probabilité jointe d'un ensemble de positions spatiales. La véritable réponse à cette question demeure difficile et nécessiterait sans doute une modélisation plus complexe (plus de paramètres et donc encore plus de précautions pour un apprentissage robuste) que celle présentée ici. Toutefois, concernant l'inférence, nous ne serons amené à considérer que des structures de plissement réelles, on peut donc restreindre le cadre d'application de l'expression de cette probabilité jointe à ce type de données uniquement. Ce qui signifie alors que ce modèle ne sera pas utilisable pour générer un segment sulcal qui ait un sens anatomique, mais uniquement pour évaluer un véritable segment. Nous avons alors choisi de négliger les dépendances statistiques spatiales entre les voxels composant chaque segment. Finalement, la probabilité de trouver le label l sur un segment composé de l'ensemble de voxels \mathcal{X} s'exprime de la façon suivante (qui n'est pas sans rappeler la probabilité sur les mesures empiriques évoquées plus tôt dans l'équation 4.18) :

$$P_{spam}(\mathcal{X}|l) = \prod_{x \in \mathcal{X}} P_{spam}(x|l)^{\frac{1}{|\mathcal{X}|}} \quad (4.20)$$

avec $|\mathcal{X}|$ le nombre de voxels du segment considéré qui se comporte comme un facteur de normalisation. Afin de comprendre la nécessité de cette normalisation géométrique, considérons 2 segments sulcaux, un faisant 2 fois la taille de l'autre, et une distribution multinomiale fictive contrôlant les fréquences d'apparition des labels de telle manière qu'elles soient invariantes spatialement. Sous

ces conditions, les vraisemblances calculées sur le plus grand segment valent le carré de celles calculées sur le plus petit. Dans les 2 cas, l'ordre des probabilités est préservé. Par contre, plus le nombre de voxels d'un segment est important plus le contraste entre les différents labels possibles se renforce. Sans normalisation, il ne faut qu'une dizaine de voxels pour que toutes les probabilités de labélisation locale a posteriori $P(L_i|D_i)$ soient nulles, sauf une qui vaut alors 1. En d'autres termes, le modèle et donc la labélisation serait sensible à une échelle arbitraire (un sujet représenté à deux résolutions différentes devrait induire des labélisations et des probabilités a posteriori très similaires aux détails près liés aux différences de discrétisation), ce qui n'est pas acceptable. De plus, d'autres modèles étudiés plus loin (normalisation jointe à la labélisation, voir partie II) nécessitent des probabilités a posteriori comparables sur l'ensemble des segments puisque celles-ci agissent pour pondérer la contribution de chaque segment à un processus de recalage. Les mêmes poids apparaissent également dans le cadre de l'apprentissage semi-supervisé (voir équation 4.13).

À l'échelle d'un segment, l'expression de la probabilité jointe que nous venons de proposer ne correspond plus à un modèle génératif. En fait, la nature du découpage d'un sillon en segment n'est pas modélisée ici. Par conséquent, notre modèle est plutôt sensible à la qualité de la segmentation des plissements en segments élémentaires.

L'estimation des SPAMS est réalisée à partir d'un espace commun à l'ensemble des sujets étudiés. L'ensemble des voxels composant les sillons de chaque sujet est exprimé dans leur référentiel respectif. Sans plus de précautions, la transformation d'un segment vers le repère commun entraînerait d'inévitables problèmes d'échantillonnage qui nuiraient sans doute à la qualité de l'estimation de la carte de fréquences. Un voxel de l'espace de départ est transformé vers l'espace d'arrivée où il se retrouve généralement à cheval sur plusieurs voxels de l'espace d'arrivée. Dans cette thèse, nous avons considéré des transformations affines ou affines par morceaux. Si on représente un sillon par un volume binaire (1 pour un voxel appartenant à la structure, 0 sinon), l'application de la transformation conjointement à une interpolation linéaire des valeurs permet de répartir de façon cohérente la contribution de chaque voxel de la structure dans l'espace commun. D'un autre côté, pendant la phase d'inférence, pour des raisons de performance, le calcul des vraisemblances est réalisé en arrondissant les coordonnées spatiales une fois transformées. Cette approximation ne pose pas de problèmes particuliers puisqu'un seul segment est impliqué à chaque fois, alors que dans le cas de l'estimation du modèle, de telles approximations se seraient cumulées sur l'ensemble des sujets.

4.3 A priori : $P(L)$

Le but de l'a priori sur la labélisation est de contraindre l'espace des labélisations possibles sans même tenir compte d'informations spécifiques au sujet que l'on cherche à labéliser, ormis éventuellement les informations liées au découpage des plissements en segments sulcaux. Précédemment, nous avons noté \mathcal{E} , l'ensemble des segments composant les plissements d'un sujet donné. Par analogie, nous noterons E l'ensemble des informations liées à ce découpage, comprenant la partition des sillons en segments et au besoin des informations de voisinage et de connexité entre ces segments (utiles pour une approche Markovienne). En toute rigueur, nous devrions écrire $P(L|E)$ plutôt que $P(L)$, de même que sur une grande majorité des probabilités présentées dans cette thèse. Cette information est sans doute sous-exploitée pour le moment car nous ne lui avons pas accordé une grande attention, principalement parce qu'elle est beaucoup moins informative (et moins fiable) que le reste du modèle (c'est à dire $P(D|L)$). En effet, ce dernier s'appuie principalement sur les spécificités individuelles du sujet à labéliser qu'il confronte à la variabilité inter-individuelle.

Cette étape de segmentation vise à séparer les plissements susceptibles de porter des labels différents, en se basant sur un surdécoupage en segments sulcaux élémentaires. À l'issue de ce procédé détaillé en section 3.2, de nombreuses découpes réalisées au cours de cette étape de segmentation n'ont pas de signification anatomique réelle (leur répartition sur la base de données utilisée est très

variable), il s'agit de faux positifs, le procédé visant à minimiser le risque d'oubli, et représente une limitation indéniable du modèle actuel. L'a priori $P(L|E)$ est donc naturellement très sensible à la variabilité de E (sur lequel nous n'avons pas de modèle génératif complet rappelons-le), entraînant principalement un découpage non-homogène des plissements de telle manière que les tailles des segments sulcaux varient au moins d'un rapport de 1 à 20. Les seules informations dont on dispose sont alors la répartition des labels sur les segments, y compris leur agencement les uns par rapport aux autres et le nombre de segments portant chaque label. Les seules informations pertinentes ici semblent être les fréquences relatives d'apparition des labels sur les segments, ou de manière plus large les fréquences d'apparition conjointes de paires de labels voisins. L'ensemble des modélisations de $P(L|E)$ permettant de prendre en compte ces informations gagnerait à être étendu à des mesures idéalement insensibles à la variabilité de E (ce qui ne veut pas dire indépendant de E , puisque la connaissance de E reste nécessaire à l'application du modèle sur un sujet particulier).

Un premier pas dans cette direction consiste à considérer conjointement aux labels des informations de taille. Les volumes et longueurs des segments sulcaux sont de bons candidats. Il n'est pas encore clair pour nous quelle serait la mesure la plus pertinente à notre application, à savoir la plus discriminante. Dans ce cas, le modèle considéré est légèrement différent de celui présenté précédemment, nous allons ici séparer les informations de taille notées T du reste des informations potentielles portées par la variable D . Ainsi, la phase de labélisation revient à maximiser en L l'expression suivante :

$$P(L|DT; E) \propto P(D|L; E)P(T|L; E)P(L|E) \quad (4.21)$$

Ici $P(D|LE)$ ne doit pas exploiter d'informations liées à la taille des structures considérées. Les modélisations proposées pour cette composante (voir section 4.2) vérifient cette contrainte ; notamment dans le cas des SPAMs, on applique une normalisation par le nombre de voxels de chaque structure évaluée (voir équation 4.20) qui a cet effet. Dans cette reformulation du problème, $P(T|L; E)$ permet de modéliser l'ensemble des tailles des sillons ainsi définies par la labélisation L . Comme dans le cas de l'a priori $P(L|E)$, nous ne considérons pas directement les valeurs absolues des tailles, mais de façon relative les unes aux autres, ce qui permet de ne pas tenir compte de la variabilité induite par les différences d'échelle entre les sujets.

Ainsi, $P(T|L; E)$ et $P(L|E)$ reposent respectivement sur des informations de fréquences d'apparition des labels au niveau des voxels (dans le cas où l'information de taille est représentée par le volume) et au niveau des segments sulcaux. Les modèles proposés dans la suite pourront alors s'appliquer indifféremment à ces deux probabilités. De manière pratique les modèles à suivre seront présentés sur l'a priori $P(L|E)$, mais seront transposables sans aucune difficulté, par un simple jeu d'écriture, à $P(T|L; E)$.

Dans la suite de cette section, les modèles seront présentés par ordre croissant de complexité. En pratique, lors de nos tests sur des données réelles, le modèle le plus simple a été majoritairement utilisé.

4.3.1 Fréquence

Ce modèle fait 2 hypothèses très fortes. L'a priori local relatif à un segment sulcal particulier est considéré comme étant indépendant des autres segments : c'est à dire : $P(L) = \prod_{i \in \mathcal{E}} P(L_i)$. Dans le cas où la même hypothèse a été faite sur $P(D|L)$ (ce qui est le cas dans les modèles des parties I et II, il ne paraît pas déraisonnable de le faire également sur l'a priori (bien qu'en réalité cette hypothèse soit fausse). La seconde hypothèse pourrait être implicite, elle stipule que les segments sont indifférenciés aux yeux de cet a priori (conséquence logique de la première hypothèse). Ce qui signifie que quelquesoit le segment i considéré, $P(L_i = l | \mathcal{M}_p) = \pi_l$ où π_l ne dépend que du label considéré et vérifie $\sum_{l \in \mathcal{L}} \pi_l = 1$. Pour des raisons pratiques, nous noterons $\pi = \{\pi_l\}_{l \in \mathcal{L}}$. Ainsi, $P(L_i = l | \mathcal{M}_p) = P(L_i = l | \pi) = \pi_l$. π correspond donc à l'ensemble des paramètres nécessaires

à l'estimation de l'a priori. L'estimateur par maximum de vraisemblance de ce vecteur à partir d'une base d'apprentissage (une approche semi-supervisée est possible en suivant la formulation de l'équation 4.14) s'écrit de la façon suivante :

$$\pi^* = \arg \max_{\pi} \prod_{a \in \mathcal{A}} \prod_{i \in \mathcal{E}_a} P(L_{a,i} = l_{a,i} | \pi) = \arg \max_{\pi} \prod_{a \in \mathcal{A}} \prod_{i \in \mathcal{E}_a} \pi_{l_{a,i}} = \arg \max_{\pi} \prod_{l \in \mathcal{L}} \pi_l^{|\mathcal{E}_l|} \quad (4.22)$$

avec $|\mathcal{E}_l|$ le nombre de segments sulcaux portant le label l .

En prenant le logarithme de l'équation précédente, le problème à résoudre devient :

$$\begin{cases} \pi^* = \arg \max_{\pi} \sum_{l \in \mathcal{L}} |\mathcal{E}_l| \log(\pi_l) \\ \text{s.t.} \quad \sum_{l \in \mathcal{L}} \pi_l = 1 \end{cases} \quad (4.23)$$

La contrainte est introduite par l'intermédiaire du Lagrangien λ :

$$E = \sum_{l \in \mathcal{L}} |\mathcal{E}_l| \log(\pi_l) + \lambda \left(1 - \sum_{l \in \mathcal{L}} \pi_l \right) \quad (4.24)$$

La valeur optimale annule les dérivées partielles :

$$\begin{aligned} \frac{\partial E}{\partial \pi_l} &= \frac{|\mathcal{E}_l|}{\pi_l} - \lambda = 0 \\ \text{ainsi,} \quad \pi_l &= \frac{|\mathcal{E}_l|}{\lambda} = \frac{|\mathcal{E}_l|}{\sum_{l \in \mathcal{L}} |\mathcal{E}_l|} \end{aligned} \quad (4.25)$$

puisque π_l est normalisé.

Ainsi, l'hypothèse d'indépendance statistique supposée plus tôt mène à une formulation très simple et intuitive de l'a priori. $P(L_i = l | \pi) = \pi_l$ correspond finalement à la fréquence empirique (ce qui lui doit son nom) d'apparition de ce label l sur l'ensemble des segments de la base d'apprentissage.

4.3.2 Dirichlet

Afin de percevoir une limitation de l'approche précédente, imaginons une situation pas complètement réaliste mais qui met bien en avant le problème. Dans notre exemple, nous nous intéressons à un segment i particulier sur lequel 2 labels l_1 et l_2 sont en concurrence de telle sorte que leurs a priori vérifient $\pi_{l_1} = \pi_{l_2}$, de la même manière les vraisemblances vérifient $P(D_i | L_i = l_1) = P(D_i | L_i = l_2)$. Vis-à-vis de cette situation, les labels sont donc aussi probable l'un que l'autre sur ce segment. Par contre, si le choix du label l_1 entraîne une non-représentation du label l_2 sur le sujet considéré, alors que celui-ci représente un sillon forcément présent, on aurait sans doute envie de revoir notre jugement. En effet, le modèle précédent à une vision uniquement locale du problème. Cet exemple démontre également que c'est à l'a priori de pénaliser si besoin l'absence d'un sillon, ou au contraire sa sur-représentation.

La distribution de Dirichlet (les détails concernant l'estimation des paramètres sont reportés en annexe B.3.2.2) s'exprime sur un simplexe : un ensemble de variables sommant à 1. Les fréquences d'apparition des labels sur un cortex vérifient cette propriété. À partir d'une base d'apprentissage, la fréquence de chaque label est caractérisée par sa moyenne et sa variance inter-sujets. L'estimation supervisée des paramètres de cette distribution tient compte finement des moments d'ordre 1 et très grossièrement des moments d'ordre 2 (puisque un seul paramètre y est consacré). La distribution généralisée de Dirichlet (détails en annexe B.3.2.3) répond à cette attente et permet de tenir compte

de l'intégralité des moments d'ordre 1 et 2 avec la même précision. Il est difficile de savoir si les 62 sujets qui constituent notre base de données sont suffisants pour obtenir une estimation robuste de l'ensemble des paramètres de cette dernière distribution (qui en comporte environ 2 par labels, soit 120 environ).

Le choix de l'un de ces a priori impose des contraintes certaines lors de l'étape d'inférence des labels sur un nouveau sujet. En effet, l'ensemble des variables $\{L_i\}_{i \in \mathcal{E}}$ dépendent alors les unes des autres. Dans ce contexte, l'étape de labélisation a été menée dans nos expériences en utilisant la méthode ICM (*Iterated Conditional Modes* [Besag 1986]) ou celle du recuit simulé. Cette méthode n'est peut-être pas optimale mais peut être suffisante compte tenu de la grande régularité de ces a priori (il n'y a qu'un seul mode centré sur les fréquences moyennes). Les détails de l'optimisation retenue dépassent le cadre de cette section et sont reportés en section 10.3.2.1.

La modélisation indépendante proposée dans la section précédente comporte des avantages certains, puisqu'elle permet une inférence simplifiée des labels. On peut imaginer une approche qui tente de lier cet avantage à la modélisation plus fine apportée par les distributions de Dirichlet. Nous proposons donc une approche en 2 temps, où la variabilité des fréquences d'apparition des labels est tout d'abord estimé par une distribution de Dirichlet (ou une distribution de Dirichlet généralisée). Dans un second temps, cette modélisation est utilisée comme a priori pour contraindre l'estimation des fréquences indépendantes. Cette approche permet, en fait d'exploiter le modèle plus élaboré pour contraindre l'estimation d'un modèle plus simple, ce qui devrait permettre une estimation plus robuste des paramètres du 2ième modèle. Dans ce cas, l'estimation des paramètres contrôlant l'a priori à partir d'une base d'apprentissage \mathcal{A} , revient à trouver le modèle \mathcal{M}_p maximisant la quantité suivante :

$$P(\mathcal{M}_p | \{L_a\}_{a \in \mathcal{A}}) = P(\pi \delta | \{L_a\}_{a \in \mathcal{A}}) \quad (4.26)$$

où π a la même signification que dans la section précédente, δ est l'ensemble des paramètres contrôlant la distribution de Dirichlet (ou sa version généralisée). Ces paramètres sont donc optimisés en 2 parties :

$$\begin{cases} \delta^* = \arg \max_{\delta} P(\delta | \pi \{L_a\}_{a \in \mathcal{A}}) \propto P(\{L_a\}_{a \in \mathcal{A}} | \delta) \\ \pi^* = \arg \max_{\pi} P(\pi | \delta^* \{L_a\}_{a \in \mathcal{A}}) \propto P(\{L_a\}_{a \in \mathcal{A}} | \pi) P(\pi | \delta^*) \end{cases} \quad (4.27)$$

La première ligne correspond donc à l'optimisation du modèle de Dirichlet et la deuxième à l'apprentissage du modèle indépendant contraint par le modèle de Dirichlet.

Dans le cas de la distribution de Dirichlet de paramètre α , $\delta = \alpha$ et $P(\pi | \alpha) \propto \prod_{l \in \mathcal{L}} \pi_l^{\alpha_l - 1}$. L'estimation de π revient alors à maximiser l'énergie E suivante (qui généralise l'équation 4.24 en y ajoutant l'a priori de Dirichlet) :

$$E = \sum_{l \in \mathcal{L}} |\mathcal{E}_l| \log(\pi_l) + \sum_{l \in \mathcal{L}} (\alpha_l - 1) \log \pi_l + \lambda \left(1 - \sum_{l \in \mathcal{L}} \pi_l \right) \quad (4.28)$$

où λ correspond encore une fois au Lagrangien intégrant la contrainte $\sum_{l \in \mathcal{L}} \pi_l = 1$. Par annulation des dérivées, finalement :

$$\pi_l^* = \frac{|\mathcal{E}_l| + \alpha_l - 1}{\lambda} = \frac{|\mathcal{E}_l| + \alpha_l - 1}{\sum_{l \in \mathcal{L}} [|\mathcal{E}_l| + \alpha_l - 1]} \quad (4.29)$$

Quand $\alpha_l = 1$, on retrouve la solution du cas sans a priori.

Pour la distribution de Dirichlet généralisée, on a $\delta = (\alpha, \gamma)$, tel que $P(\pi|\alpha\gamma) \propto \prod_{l \in \mathcal{L}} \pi_l^{\alpha_l - 1} (1 - \pi_1 - \dots - \pi_l)^\gamma$. Cette fois l'énergie à maximiser s'exprime de la façon suivante :

$$E = \sum_{l \in \mathcal{L}} |\mathcal{E}_l| \log(\pi_l) + \sum_{l \in \mathcal{L}} [(\alpha_l - 1) \log \pi_l + \gamma_l \log(1 - \pi_1 - \dots - \pi_l)] + \lambda \left(1 - \sum_{l \in \mathcal{L}} \pi_l\right) \quad (4.30)$$

Cette fois, la solution optimale vaut :

$$\pi_l^* = \frac{|\mathcal{E}_l| + \alpha_l - 1 - \gamma_l}{\lambda} = \frac{|\mathcal{E}_l| + \alpha_l - 1 - \gamma_l}{\sum_{l \in \mathcal{L}} [|\mathcal{E}_l| + \alpha_l - 1 - \gamma_l]} \quad (4.31)$$

L'approche variationnelle permettrait de proposer des alternatives à cette approche. Elle consisterait à estimer les paramètres de la distribution $P(L|\pi)$ possédant les propriétés d'indépendances recherchées de façon à ce qu'elle soit le plus proche possible (au sens des distributions : divergence de KullBack-Leibler) de la distribution $P(L|\alpha)$ plus complexe.

4.3.3 Multinomiale

La formulation des distributions de Dirichlet impose une autre contrainte plutôt handicapante. En effet, sur les bords du simplexe, c'est à dire dès qu'une des variables d'intérêt s'annule (dans notre cas la fréquence d'un des labels) la vraisemblance associée vaut zéro. Par conséquent, définies telles quelles, ces distributions forcent artificiellement la présence de tous les labels sur le sujet à labéliser, rendant l'absence de certains sillons rares impossibles. Cette situation est tout à fait normale et met simplement en avant la confusion que nous avons fait entre fréquence d'apparition des labels (ce que modélise le modèle) et la probabilité de générer une configuration de labels donnée à partir de ces labels (ce qu'on cherche à évaluer sur un sujet à labéliser). En effet, si un label a une fréquence faible, il a peu de chance d'être généré selon ce modèle et la probabilité qu'il soit absent est d'ailleurs élevée.

Une première réponse à ce problème est apportée par la distribution multinomiale, qui permet d'étendre la modélisation à base de fréquence de la section 4.3.1 en tenant compte du nombre d'éléments portant un label donné. Posons alors $L|\pi \sim$ multinomiale de paramètres π , où π a la même signification que dans la section 4.3.1, c'est à dire l'ensemble des fréquences d'apparition des labels mesurées sur une base d'apprentissage. L'a priori s'écrit alors :

$$P(L|\pi) = P(\{|\mathcal{E}_l|\}_{l \in \mathcal{L}}|\pi) = \frac{|\mathcal{E}|!}{\prod_{l \in \mathcal{L}} |\mathcal{E}_l|!} \prod_{l \in \mathcal{L}} \pi_l^{|\mathcal{E}_l|} \quad (4.32)$$

où $|\mathcal{E}|$ et $|\mathcal{E}_l|$ corespondent respectivement au nombre total de segments (ou de voxels) et au nombre de segments (ou de voxels) portant le label l pour le sujet considéré. Cette distribution mesure la probabilité d'obtenir une labélisation donnée suite à $|\mathcal{E}|$ tirages de labels selon le vecteur de probabilité π .

Naturellement, une autre réponse peut être apportée au problème précédent, consistant à étendre la modélisation de Dirichlet selon le même principe. Cette approche associe une distribution multinomiale à une distribution de Dirichlet et porte le nom de distribution de Pólya multivariée [Minka 2003] (ou Dirichlet-Multinomiale). Cette fois encore, la distribution de Dirichlet intervient afin de modéliser l'incertitude sur le vecteur π des fréquences. En notant α , les paramètres de la distribution de Dirichlet, l'a priori devient finalement :

$$P(L|\alpha) = \int_{\pi} P(L|\pi)P(\pi|\alpha)d\pi = \frac{\Gamma(\sum_l \alpha_l)}{\Gamma(\sum_l |\mathcal{E}_l| + \alpha_l)} \prod_{l \in \mathcal{L}} \frac{\Gamma(|\mathcal{E}_l| + \alpha_l)}{\Gamma(\alpha_l)} \quad (4.33)$$

Ici, $P(L|\pi)$ a la même formulation que dans l'équation 4.32, il s'agit d'une distribution multinomiale, alors que $P(\pi|\alpha)$ correspond à la loi de Dirichlet de paramètre α . Les paramètres α régissent également la distribution de Pólya multivariée. Ils sont alors plus difficiles à estimer que dans le cas de la distribution de Dirichlet. Toutefois, il existe des schémas itératifs pour y parvenir [Minka 2003] ou des approximations [Madsen 2005].

Cette distribution permet donc de gérer les sillons absents de façon appropriée. Par contre, il n'est plus question ici de normaliser nos mesures par le nombre de segments (ou de voxels) pour effacer la variabilité due aux différences d'échelle. Le meilleur moyen reste sans doute de reconsidérer cette distribution face à la modélisation proposée dans la partie II, puisqu'alors cet a priori aura été estimé dans un espace normalisé. Dans ce cadre, la valeur de l'a priori $P(L)$ pour un sujet donné dépendra alors de paramètres de normalisation (implicites dans la formulation actuelle).

4.3.4 Approche Markovienne

La partie III étant consacrée à l'utilisation d'une telle modélisation à l'échelle du modèle complet (et non uniquement à l'a priori comme ici), une majorité de détails non nécessaires à la compréhension de l'idée présentée ici est reportée plus loin.

Dans les modélisations précédentes, aucune information spatiale n'était exploitée. Si on relie entre eux les segments labélisés selon un graphe de voisinage (défini dans notre cas en suivant le procédé de la section 11.2.1, en reliant les segments qui se touchent ou qui se font face, séparés par un gyrus), cette simple information supplémentaire permettrait de définir à quelle fréquence certains labels peuvent être en relation l'un avec l'autre (c'est à dire voisins dans le graphe). Une approche Markovienne classique pourrait prétendre répondre à cette modélisation : le modèle généralisé de Potts [Boykov 1998] (présenté plus récemment par [Blanchet 2005]). Dans sa formulation la plus générale, il dépend du vecteur de paramètres a et de la matrice de paramètres B modélisant les interactions entre labels voisins. Dans ce cadre, l'a priori sur la labélisation s'exprime :

$$P(L|aB\mathcal{G}) = \frac{1}{Z(a, B)} \exp \left(\sum_{i \sim j \in \mathcal{G}} l_i^t B l_j + \sum_{i \in \mathcal{E}} a^t l_i \right) \quad (4.34)$$

où \mathcal{G} correspond au graphe de voisinage des segments, $Z(a, B)$ est une constante de normalisation dépendant des paramètres et de la structure du graphe. Ici, l_i est un vecteur binaire, où chaque entrée correspond à un label particulier. Le vecteur est rempli de zéros sauf au niveau de la ligne concernant le label du segment i où il y a un 1. Il s'agit d'une astuce d'écriture, puisqu'ainsi $l_i^t B l_j = B_{l_i, l_j}$ et $a^t l_i = a_{l_i}$.

Cette formulation comporte un nombre très important de paramètres, de l'ordre du carré du nombre de sillons. En l'état, cette distribution semble très difficile à estimer à partir des 62 sujets que comporte notre base de données. Évidemment de nombreux labels ne sont jamais en contact les uns avec les autres au travers du graphe de voisinage, on pourrait donc envisager d'introduire des a priori codant cette information. Malheureusement, il resterait alors encore beaucoup trop de paramètres. L'introduction d'a priori supplémentaires est donc nécessaire. Une solution simple envisagée à ce problème et de remplacer la matrice B par $(1 - \lambda)B + \lambda \mathbb{I}$ où \mathbb{I} est une matrice de la même taille que B et ne contenant que des 1. Le paramètre λ permet alors de doser l'importance accordée aux paramètres de B (si $\lambda = 0$ la matrice B d'origine est utilisée, si $\lambda = 1$ alors on ne tient plus compte de B).

L'estimation des paramètres d'un tel modèle est très complexe. Nous discuterons de quelques approximations standards dans la partie III consacrée au champ de Markov en section 10.2.

Précédemment, nous avons fait part de notre volonté d'étendre l'a priori de labélisation du niveau «segments» au niveau «voxels». L'application de cette généralisation sur le modèle de Potts

généralisé n'a pas de réponse directe évidente, puisqu'il remet en cause la structure même du graphe de voisinage sur lequel la distribution repose. En l'état, ce modèle n'a donc finalement pas été retenu.

Résultats et principe d'évaluation

Ce chapitre présente un ensemble de critères permettant d'apprécier la qualité des labélisations proposées par un modèle et les confronter à celles obtenues avec d'autres approches. Cette étape est cruciale et a été régulièrement raffinée afin de mieux identifier des erreurs systématiques qui reflètent généralement une faiblesse (ou un manque d'information) du modèle étudié. Cette approche nous a naturellement poussés à proposer des solutions dans ce sens (nous y reviendrons en partie II et III).

Nous détaillons ici les résultats obtenus avec un modèle basique Gaussien sur les centres de gravité des segments qui nous servira de base de comparaison pour juger de la qualité du modèle SPAM. Nous illustrerons alors rapidement l'intérêt de la prise en compte d'informations directionnelles dans le modèle. Finalement, un modèle hybride, mélange de l'«approche historique» et du modèle SPAM, est proposé comme une première tentative de couplage entre informations de localisation et informations structurelles. La question des a priori n'est pas abordée ici et sera envisagée plus loin dans le chapitre 12.

5.1 Mesures d'erreur

Selon le type de représentation utilisée pour caractériser les plissements sulcaux, la façon de mesurer les erreurs peut différer. Concernant la représentation à base de segments sulcaux, nous indiquerons dans cette section les principales mesures (globales et locales) que nous avons utilisées tout au long de ce travail de thèse; seule l'erreur nommée E_{mass} était utilisée dans les travaux précédents.

Au fur et à mesure de l'avancée de nos travaux (y compris ceux présentés dans les parties à suivre), nous nous sommes intéressés à des mesures de plus en plus sensibles permettant une analyse détaillée de la qualité de nos modèles et de leurs labélisations. En effet, plus les modèles sont fiables, plus il est difficile de les comparer puisqu'un nombre limité d'erreurs est commis; de cette façon, certaines améliorations ou détériorations deviennent exceptionnelles. Ainsi, mesurer la significativité des résultats dans de telles conditions n'est pas toujours évident.

Les mesures à suivre évaluent chaque sujet individuellement selon la qualité de sa labélisation (labels automatiques) inférée depuis un modèle sulcal en la comparant à une vérité de terrain connue (labels manuels). Avant tout, il est nécessaire d'introduire quelques notations : $FP(l)$, $FN(l)$, $VP(l)$ correspondent respectivement, pour un label l donné, aux mesures de faux-positifs (segments labélisés automatiquement avec le label l alors que leur véritable label est différent), faux négatifs (segments labélisés manuellement avec le label l alors que leur label automatique est différent) et vrais positifs (segments où le label manuel et automatique correspondent au label l). Pour chacune de ces mesures, la valeur retenue est la taille cumulée (nombre de voxels dans l'espace du sujet) de chaque segment sulcal vérifiant les conditions spécifiées plus tôt.

Historiquement (utilisée essentiellement dans les travaux basés sur l'approche «historique» [Rivière 2002, Perrot 2008]) la première mesure d'erreurs E_{mass} pénalise uniquement la proportion totale de faux positifs (ou de faux négatifs) :

$$E_{\text{mass}} = \frac{\sum_{l \in \mathcal{L}} FP(l)}{\sum_{l \in \mathcal{L}} FP(l) + VP(l)} = \frac{\sum_{l \in \mathcal{L}} FN(l)}{\sum_{l \in \mathcal{L}} FP(l) + VP(l)} \quad (5.1)$$

Ensuite, cette mesure globale peut être étendue afin de caractériser des changements locaux concernant chacun des labels. Du point de vue d'un label l donné, il est nécessaire de pénaliser les segments labélisés à tort avec ce label (faux positifs) et ceux labélisés à tort par un autre label (faux négatifs) :

$$E_{\text{local}}(l) = \frac{FP(l) + FN(l)}{FP(l) + FN(l) + VP(l)} \quad (5.2)$$

Cette mesure permet de mettre en relief certaines améliorations locales qui pourraient être noyées dans le bruit à l'échelle d'un hémisphère complet.

La mesure locale précédente est une bonne base pour obtenir une mesure globale plus sensible que la mesure E_{mass} . Nous avons tout d'abord proposé l'erreur E_{mean} suivante (utilisée dans [Perrot 2008]) :

$$E_{\text{mean}} = \sum_{l \in \mathcal{L}} \frac{FP(l) + FN(l)}{FP(l) + FN(l) + VP(l)} \quad (5.3)$$

qui permet de focaliser artificiellement la mesure globale sur les petites structures. Cette mesure n'est pas complètement satisfaisante puisqu'elle ne tient pas compte des erreurs partagées entre les labels comme nous allons le voir.

En adaptant l'indice de similarité (*Similarity Index*, SI) proposé par [Yang 2007] à notre représentation des plissements, nous avons défini la mesure suivante :

$$E_{SI} = \sum_{l \in \mathcal{L}} w_l \frac{FP(l) + FN(l)}{FP(l) + FN(l) + 2VP(l)} \quad (5.4)$$

où $w_l = \frac{s_l}{\sum_{l \in \mathcal{L}} s_l}$ avec $s_l = FN(l) + VP(l)$ la véritable taille du sillon portant le label l .

Chaque composante de la somme diffère de la mesure locale $E_{\text{local}}(l)$ sur deux points. Tout d'abord, les vrais positifs $VP(l)$ comptent 2 fois plus que les faux positifs $FP(l)$ et faux négatifs $FN(l)$ afin de rééquilibrer l'importance des erreurs partagées par plusieurs labels. En effet, chaque segment en trop pour un label correspond à un segment manquant pour un autre. Deuxièmement, chaque composante de la somme a une valeur comprise entre 0 et 1 qui est pondérée à hauteur de la taille réelle du sillon correspondant. De ce fait, les grands sillons comptent plus que les petits sillons, mais en contrepartie, chaque voxel labélisé à la même importance. De façon pratique, cette nouvelle mesure globale semble plus draconienne que la précédente puisqu'elle est toujours plus élevée que cette dernière en pratique (ce qui n'est pas forcément le cas en théorie).

Toutes ces mesures dépendent fortement du choix des labels automatiques. Parfois, plusieurs labels sont en compétition sur un segment i , leurs probabilités a posteriori $P(L_i|D_i)$ sont alors presque égales. Dans ce cas, le label finalement retenu est plutôt arbitraire. D'autre part, deux modèles peuvent fournir des labélisations assez proches alors que leurs probabilités a posteriori sont très différentes, ce qui signifie que les deux modèles sont en accord sur la labélisation mais que l'un des deux est plus confiant que l'autre sur la qualité de la labélisation proposée. Pour répondre à ces besoins, nous proposons une nouvelle mesure locale basée sur les mesures de probabilité a posteriori locales de la labélisation. Pour un label l donné, nous calculons la valeur moyenne de cette mesure sur l'ensemble des segments portant réellement ce label :

$$E_{\text{post}}(l) = \frac{\sum_{i \in \mathcal{E}_l} s_i P(L_i = l | D_i)}{\sum_{i \in \mathcal{E}_l} s_i} \quad (5.5)$$

où \mathcal{E}_l est l'ensemble des segments sulcaux labélisés manuellement avec le label l et s_i le nombre de voxels du segment i .

5.2 Comparaison de modèles

Pour évaluer si un modèle apporte des améliorations par rapport à un autre, le premier critère pertinent est la différence entre les taux d'erreurs de ces deux modèles, qu'ils soient locaux ou globaux. Plus les différences sont faibles moins cette mesure est fiable. En effet, il est probable que certaines différences minimales soient seulement dues au hasard alors qu'il n'y a pas de tendance particulière à l'échelle de l'ensemble des sujets.

Afin de préciser nos résultats, dans ces cas tangents, nous avons utilisé un test statistique de Wilcoxon [Wilcoxon 1945] non signé (appelé *two-tailed Wilcoxon* ou encore *two-sided Wilcoxon*) afin de mesurer la significativité des améliorations et détériorations des résultats observés. Ce test permet d'obtenir une mesure nommée *p-value* qui correspond à la probabilité d'obtenir une telle répartition des mesures pour les deux modèles alors même qu'on suppose que la distribution de ces différences est symétrique et centrée en zéro. Cela permet en quelque sorte de mesurer avec quelle probabilité une telle configuration pourrait être obtenue par hasard et donc ne serait pas significative.

Il s'agit d'un test de rang (on mesure donc en réalité l'entrelacement entre les 2 distributions de mesure) que nous avons préféré au test de *Student* (le fameux *t-test*) puisque les données à comparer ne sont pas forcément Gaussiennes (taux d'erreurs entre 0 et 1). Nous n'avons pas rencontré de situation où le test de *Student* était en contradiction flagrante avec le test de Wilcoxon dans nos expérimentations. Le test de Wilcoxon a alors été utilisé pour mesurer la significativité de telles différences, d'une part à l'échelle d'un hémisphère où les erreurs globales sont alors comparées entre elles, et d'autre part à l'échelle de chaque sillon où un test est réalisé par sillon. Dans ce dernier cas, une correction de Bonferroni pour les comparaisons multiples est appliquée. Elle consiste à multiplier chaque *p-value* obtenue par le nombre de tests réalisés, on parlera alors de *p-value* corrigée. De cette façon, l'espérance du nombre d'erreurs d'interprétation (penser qu'une situation observée est significative alors qu'elle est en fait due au hasard) devient indépendante du nombre de tests réalisés. Dans tous les cas (local ou global), on considérera qu'une mesure est significative si la *p-value* obtenue est en deçà d'un seuil fixé à 5.10^{-2} .

5.3 Stratégie d'évaluation

Pour noter et comparer les modèles étudiés, il est nécessaire d'évaluer leurs capacités de généralisation. Toutes les études réalisées ont été faites en suivant l'approche supervisée décrite plus tôt. Ainsi, nous disposons d'une base d'apprentissage suffisamment grande pour estimer nos modèles et d'une base de test permettant de tester l'efficacité du modèle sur des sujets non considérés pendant la phase d'apprentissage. Dans le cadre de nos travaux sur l'approche «historique» [Rivière 2002], la base de données de l'époque (26 sujets) était découpée en une base d'apprentissage (21 sujets) et une base de test (5 sujets). Ce schéma n'était pas satisfaisant puisque trop peu de données étaient utilisées pour estimer le modèle ou en mesurer la qualité. De plus, les résultats étaient assez sensibles à la répartition des sujets dans les deux bases.

À partir de cette partie I et jusqu'à la fin du manuscrit (hormis les résultats liés à l'approche «historique» dont ceux de la section 5.5), une nouvelle stratégie de validation a été choisie pour répondre à ce problème, il s'agit d'utiliser un schéma de validation croisée. Dans nos études nous avons utilisé la variante *leave-one-out* que nous allons décrire. Pour chaque sujet, une base de données d'apprentissage lui est rattachée. Elle comporte l'ensemble des sujets de la base de données sauf le sujet considéré qui constitue alors la base de test. Ainsi, pour chaque hémisphère, 62 modèles (autant que de sujets) sont estimés à partir de chaque base d'apprentissage. Ensuite, chacun des 62 sujets est labélisé à partir de son modèle respectif. Par ce truchement, les résultats sont beaucoup plus fiables et interprétables que les précédents. L'estimation complète d'un modèle peut être très

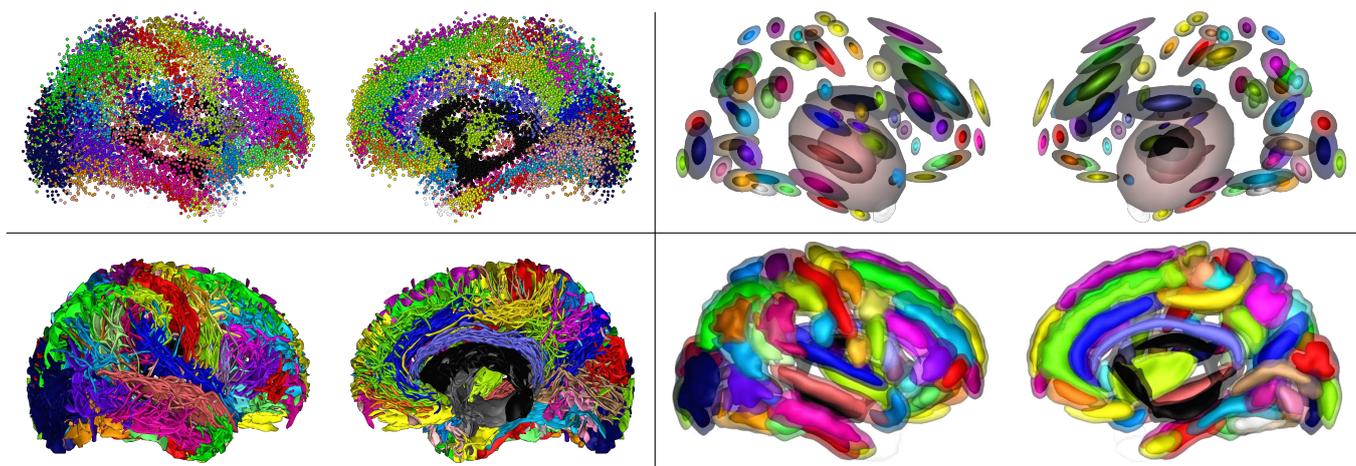


FIG. 5.1 – Modélisation de la variabilité sulcale sur 62 sujets (illustration sur hémisphères droits). **Haut.** Gauche : centres de gravité des segments. Droite : modélisation Gaussienne de ces données. **Bas.** Gauche : segments sulcaux d'une douzaine de sujets. Droite : modèles SPAMs estimés à partir des 62 sujets.

longue; aux vues du nombre de modèles évalués, il a été nécessaire d'avoir recours à des méthodes de calcul sur grille (*grid computing*, voir annexe C.3.4).

5.4 Méthodes locales

Les méthodes dites «locales» regroupent l'ensemble des modèles évalués à partir de caractéristiques locales mesurables à l'échelle d'un segment sulcal (l'unité élémentaire de labélisation). Nous nous sommes principalement intéressés à des informations de localisation et de direction. Les résultats présentés dans cette section suivent la stratégie d'évaluation par validation croisée proposée dans la section 5.3. La modélisation proposée considère que chaque segment peut être labélisé séparément à partir d'un modèle appris d'une base de données de 62 sujets (section A) de façon supervisée. Tous les essais présentés ici ont été réalisés avec l'a priori $P(L)$ (section 4.3.1) basé sur la fréquence d'apparition des labels sur les segments de la base d'apprentissage.

Dans le cas des positions, le tout premier modèle étudié consistait à modéliser la variabilité de localisation des centres de gravité des segments (figure 5.1, en haut à gauche) par une Gaussienne multivariée (figure 5.1, en haut à droite). Ce modèle est sans doute le modèle génératif le plus simple que l'on peut envisager. Il permet ainsi de mesurer une qualité de labélisation minimale que tout autre modèle plus complexe se doit de dépasser. Ce modèle est à l'origine de l'essentiel des travaux présentés dans ce manuscrit, puisqu'il nous a convaincu de l'intérêt d'une approche descriptive et explicative. En effet ce modèle, dans des conditions d'évaluation similaires à celles utilisées pour l'approche «historique», a donné des résultats légèrement meilleurs (avoisinant 20% d'erreurs) que ce dernier selon l'erreur E_{mass} . Ces résultats ont été confirmés par la suite par une validation croisée sur la base de données de 62 sujets. Bien entendu, de nombreuses erreurs intolérables demeurent possibles puisque la modélisation proposée reste assez limitée. En effet, la variabilité de la localisation est estimée globalement à l'échelle d'un sillon; toutes les variations locales le long du sillon ne sont

pas représentées alors qu’elles sont inhomogènes en réalité (notamment les extrémités des sillons sont généralement plus variables que le centre du sillon, de plus les sillons ne sont pas tous rectilignes).

Diverses réponses apparaissent dans la littérature, essentiellement basées sur une approximation de la forme des sillons (lignes et rubans corticaux : *splines* ou *B-splines* [Goualher 1997]) par des formes simples que l’on peut ensuite paramétriser et mettre en correspondance au travers de cette paramétrisation. Dans notre cas, nous travaillons sur des données réelles non simplifiées en tenant compte de toute la complexité des formes que peuvent présenter les sillons avec des ramifications, des embranchements et des coupures. Le modèle SPAM (section 4.2.3) n’est pas sensible à ces contraintes et a déjà été testé avec succès [Goualher 1998] sur des données proches des nôtres, il s’est donc imposé naturellement pour représenter la variabilité de localisation locale des structures sulcales (figure 5.1, bas). Ce modèle a permis d’améliorer les résultats de labélisation de façon très significative (p -values¹ sur E_{SI} : $8.1e^{-11}$ à gauche et $1.4e^{-9}$ à droite) pour un taux d’erreurs moyen avoisinant les 17% (voir tableau plus loin). Malgré ces résultats très encourageants, des erreurs grossières subsistent : dédoublement de sillons (si 2 structures se retrouvent de part et d’autre d’un modèle SPAM) impliquant en particulier les sillons central, pré-central et post-central, morceaux manquants même en plein milieu d’une structure (le modèle ne l’interdit pas puisque la labélisation est locale, chaque segment étant traité indépendamment des autres).

Les parties des sillons les plus profondes sont considérées comme plus stables [Lohmann 2008] à l’échelle d’une population de sujets. La chaîne de traitement anatomique de Brainvisa nous fournit, associé à la représentation sous forme de segments sulcaux, l’ensemble des voxels traçant la ligne du fond de chaque plissement élémentaire. Cette information a été utilisée à la place des segments sulcaux pour évaluer son potentiel de discrimination des structures sulcales en utilisant un modèle de SPAMS. Contrairement à nos attentes, il s’est avéré qu’en l’état ces lignes n’étaient pas suffisamment reproductibles au niveau des branches et des extrémités où ces lignes peuvent remonter jusqu’à la surface. Les taux d’erreurs constatés sur les fonds de sillons se sont avérés environ 5% plus élevés qu’en utilisant les segments sulcaux. Nous restons persuadé qu’une méthode d’estimation plus robuste que le modèle SPAM actuellement proposé (voir annexe B.4 pour plusieurs suggestions en ce sens), associée sans doute à une meilleure définition des fonds des sillons, pourrait améliorer les résultats.

Certaines structures sulcales peuvent être proches spatialement et susceptibles d’être confondues sur cet unique critère. Dans certaines situations, il est pourtant facile de les distinguer à condition de repérer à quelle grande zone corticale ces structures appartiennent. En considérant la forme générale d’un hémisphère on peut distinguer 3 zones principales : la face externe de forme convexe (comprenant la majorité des sillons), la face interne de forme planaire qui fait face à l’autre hémisphère et la zone sous le cerveau, plutôt planaire en frontal et concave en parieto-occipital. Les zones frontalières entre ces zones présentent le plus souvent une courbure assez forte permettant de les distinguer. Nous avons envisagé tout d’abord une méthode projective basée sur des techniques de *RayTracing* utilisés en synthèse d’images afin de segmenter le cortex en 3 zones et contraindre ainsi les labels autorisés dans chacune des zones. Cette approche s’est avéré peu fiable dès que la zone séparatrice ne présentait pas une courbure suffisamment marquée. Nous avons donc abandonné cette méthode (que nous ne détaillerons pas plus) au profit d’une modélisation de la variabilité des normales à l’enveloppe convexe.

L’enveloppe convexe (figure 5.2 à gauche) du cerveau est calculée par une dilatation morphologique de l’interface gris/blanc de telle façon que la trace des sillons soit fortement estompée. Cette représentation caractérise la forme générale du cerveau et les informations de courbure sont portées par les normales à cette surface. Afin de différencier les structures sulcales voisines mais appartenant à des zones différentes, nous avons modélisé pour chaque label la variabilité de ces normales par

¹test de Wilcoxon, voir détails en section 5.2

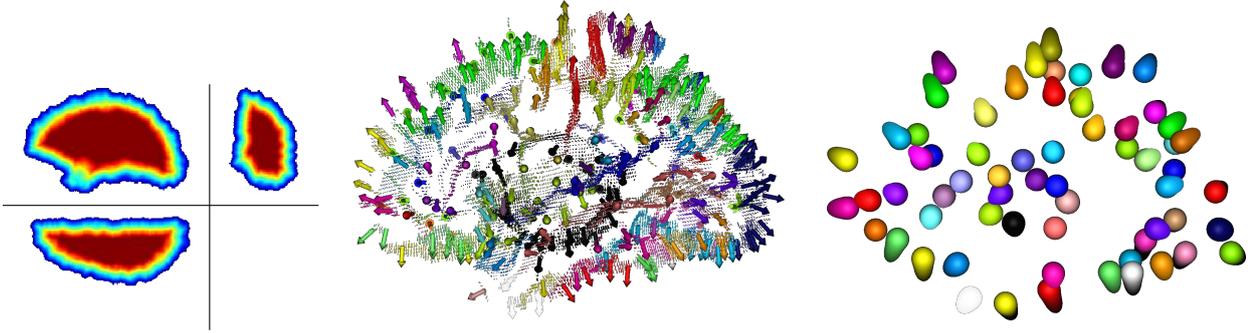


FIG. 5.2 – Modélisation de la normale à l'enveloppe d'une hémisphère (hémisphère gauche). Gauche : distance depuis l'enveloppe. Centre : gradient de la carte de distance : normales par voxel et par sillon. Droite : modèles de Kent estimés sur 62 sujets.

une distribution de Kent (équivalent d'une Gaussienne multivariée sur des données directionnelles, voir annexe B.3.1.3). Nous avons également essayé d'utiliser une distribution de Von Mises Fisher B.3.1.2, l'équivalent d'une Gaussienne isotrope sur données directionnelles. Ce modèle directionnel n'était pas suffisamment informatif et n'améliorait pas les résultats par rapport à un modèle de SPAM seul. Pour des raisons de robustesse, l'estimation n'est pas calculée sur les seules normales à l'enveloppe. On procède alors en plusieurs étapes : 1) dilatation morphologique des structures sulcales, 2) carte de distance à l'enveloppe (calculée ici par la méthode du chanfrein) uniquement dans les structures sulcales ainsi dilatées, 3) champs vectoriel du gradient de cette carte de distance. Pour chaque sillon, on obtient donc un ensemble de vecteurs ; leur norme servira alors à pondérer leur contribution dans l'estimation du modèle Kent. Cette approche permet de tenir compte d'une éventuelle inclinaison (par rapport à la normale) des sillons dans leur profondeur. La légère dilatation rend le procédé plus robuste sachant que les structures sulcales ont une épaisseur de l'ordre du voxel rendant les calculs de gradient imprécis sans cette opération. La figure 5.2 illustre ces opérations.

L'ensemble des taux d'erreurs globaux (et leur écart-type respectif entre parenthèse) obtenus par validation croisée sur 62 sujets sont présentés dans le tableau suivant :

		Gauss.	Gauss. + Kent	<i>p-values</i>	SPAM	SPAM + Kent	<i>p-values</i>
E_{SI}	Gauche	21.76 (5.06)	21.24 (5.02)	0.061	17.55 (5.93)	17.08 (4.12)	0.023
	Droite	20.00 (4.07)	19.92 (4.45)	0.37	16.83 (3.77)	16.52 (3.76)	0.33
E_{mass}	Gauche	21.27 (5.02)	21.11 (4.99)	-	16.64 (3.69)	16.54 (3.84)	-
	Droite	19.62 (3.93)	16.52 (3.76)	-	15.96 (3.66)	16.10 (3.61)	-

où les p -values² mesurent la significativité des différences de performance avec et sans le modèle Kent sur les normales.

Avant toute chose pour placer une base de comparaison, considérons le taux d'erreurs qu'obtiendrait une méthode de labélisation au hasard ou chaque label serait tiré uniformément parmi les labels possibles de chacun des segments sulcaux. En pratique les labels étant limités à des boîtes englobantes, seul 5 à 10 labels sont en concurrence sur chacun des segments. Le taux d'erreur du hasard se situe donc entre 80 et 90%.

²test de Wilcoxon, voir détails en section 5.2

Le modèle de Kent seul, concernant les normales, n'est discriminant qu'aux frontières entre les 3 grandes zones composant la surface corticale. On obtient des résultats de l'ordre de 60% d'erreurs en l'utilisant seul. Les modèles positionnels (modèle à base de Gaussiennes sur les centres de gravité des segments, nommé *Gauss.* dans le tableau, et modèle SPAM) sont plus discriminants, mais tirent légèrement profit du modèle de Kent notamment sur les hémisphères gauches. Le gain est plutôt faible (de l'ordre de 0.5%, à la limite de la significativité pour le modèle Gaussien et légèrement significatif pour les SPAM). Concernant les hémisphères droits les différences constatées ne sont pas significatives. L'intérêt de l'information directionnelle reste encore assez limitée en l'état.

5.5 Mélange avec l'approche «historique»

Les informations de localisation permettent de réaliser une labélisation déjà très satisfaisante des sillons. Comme nous l'avons vu précédemment, cette approche laisse néanmoins passer des erreurs que l'on ne peut tolérer. Elles s'expliquent en partie par un manque d'informations contextuelles permettant d'assurer la cohérence de la labélisation entre des segments sulcaux voisins. Nous avons alors envisagé différentes façon de mélanger [Perrot 2008] le modèle SPAM avec le modèle «historique» qui comporte des informations structurelles complémentaires. Nous utiliserons ici la variante basée sur des machines à vecteurs de support que nous noterons MSVR (*Markov-SVR*). Le modèle «historique» comporte également certaines informations de localisation (position du centre de gravité et extrémités de chaque sillon) que nous n'avons pas enlevées pour des raisons techniques, entraînant une légère redondance d'information et une dépendance entre les 2 modélisations.

5.5.1 Modèle

La méthode la plus simple et la plus rapide (nommée SPAM init) reste sans doute d'utiliser le modèle de SPAMs pour fournir une solution initiale et la raffiner ensuite en réalisant une descente d'énergie par le modèle «historique». Il s'agit d'une approche gloutonne, considérant les segments sulcaux au hasard les uns après les autres et choisissant pour chaque segment le label qui minimise le plus l'énergie considérée. Les résultats obtenus avec le modèle SPAM montrent que les probabilités a posteriori locales des labels sont majoritairement nulles sauf sur un ensemble très restreint de segments. Précédemment [Rivière 2002], l'espace des labélisations possibles était limité par une modélisation très grossière de la localisation de chaque label en délimitant leur présence au sein d'une boîte englobante. Le modèle SPAM permet de restreindre de façon plus précise l'emplacement des labels. Ainsi, sur un segment donné, seuls les labels dont la probabilité a posteriori est supérieure à 0.01 sont autorisés lors du processus de labélisation. À partir de cette base, nous avons étudié 3 types d'optimisation : la méthode ICM (section 10.3.2.1) qui plonge dans le premier minimum local de l'énergie rencontrée, le recuit simulé avec échantillonnage de Gibbs commençant à basse température (section 10.3.2.2), ou la méthode MPM (section 10.3.1) qui suit un schéma MCMC proche d'un recuit simulé à température fixée. L'approche MPM est suivi par l'application de la méthode ICM pour atteindre le minimum local de l'énergie le plus proche. Ces approches considèrent toutes que l'initialisation est suffisamment bonne et proche de la solution optimale pour qu'un nombre limité de changements locaux de labels soit nécessaire.

Sous une condition d'indépendance des différentes composantes (localisation, forme, organisation locale), l'ensemble de ces données D est décomposée sur les segments sulcaux, sillons et paires de sillons : $D = \{D_p, D_s, S_{st}\}$. On se propose de coupler la formulation des vraisemblances de chacun des modèles en une seule probabilité :

$$P(L|D) \propto P(L) \prod_{i \in \text{segments sulcaux}} P(D_i|L_i) \prod_{s \in \text{sillons}} P(D_s|L_s) \prod_{s \ t \in \text{paire de sillons}} P(D_{st}|L_{st})$$

		Modèles seuls		Mélange de modèles		
		SPAM	Msvr + Gibbs	SPAM init Msvr + ICM	SPAM init Msvr + MPM	SPAM potentiel Msvr + Gibbs
E_{mass}	Apprentissage	10.33	19.81	11.39	14.18	9.90
	Test	21.42	26.70	21.76	20.46	19.14
E_{mean}	Apprentissage	21.82	32.12	22.09	27.11	19.17
	Test	43.08	44.72	41.71	39.46	38.04

TAB. 5.1 – Taux d'erreurs sur une base d'hémisphères droits : 21 sujets d'apprentissage et 5 sujets de test, confrontant l'approche SPAM à l'approche «historique». De gauche à droite : modèles seuls : SPAM, MSVR (variante du modèle «historique» basée sur des machines à vecteurs de support, chapitre 2), mixture de modèles, avec "SPAM init" : initialisation et limitation des labels possibles en se basant sur le modèle SPAM, "SPAM potentiel" : intégration complète du modèle de SPAM en tant que potentiel dans le champs de Markov de l'approche «historique», "MSVR + *mode*" : ici *mode* précise la méthode d'optimisation utilisée pour inférer les labélisations.

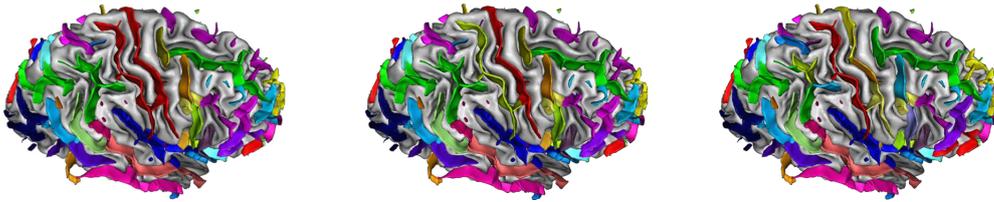


FIG. 5.3 – Erreurs de labélisation typiques. **Gauche** : dédoublement du sillon central avec le modèle à base de SPAMs : invraisemblance anatomique mais les erreurs sont limitées spatialement. **Centre** : labélisation manuelle. **Droite** : décalage de toutes les structures vers l'arrière du cerveau avec le modèle «historique» : la cohérence globale des structures est préservée mais les erreurs sont nombreuses.

où nous avons utilisé l'a priori basé sur la fréquence d'apparition des labels sur les segments pour $P(L)$ (section 4.3.1).

Chacun des deux modèles apporte ainsi sa contribution à l'évaluation globale d'une labélisation. Certains sillons sont bien localisés spatialement (sillons temporaux, sillon olfactif, insula, ...) alors que d'autres présentent une forme bien définie (fissure cingulaire, sillon central ou calleux, ...), parfois seules les informations contextuelles fournies par le voisinage permettent leur reconnaissance (sillons frontal intermédiaire, occipital, ...). Les différents types d'information devraient alors être pondérées afin de s'ajuster aux particularités locales. Nous avons suggéré dans [Perrot 2008] d'aplatir chaque vraisemblance à hauteur du taux d'erreurs par label constaté sur une base d'apprentissage non utilisée jusqu'à présent. En toute logique, cette approche a posteriori n'est pas nécessaire dans un cadre probabiliste si les modèles ont été estimés de façon robuste et couplés (ce qui n'est pas le cas ici, nous reviendrons sur cette notion dans la partie consacrée au champ de Markov III). En effet, la forme des distributions est alors plus ou moins piquée selon la pertinence du critère modélisé, l'ajustement devrait donc se faire de lui-même.

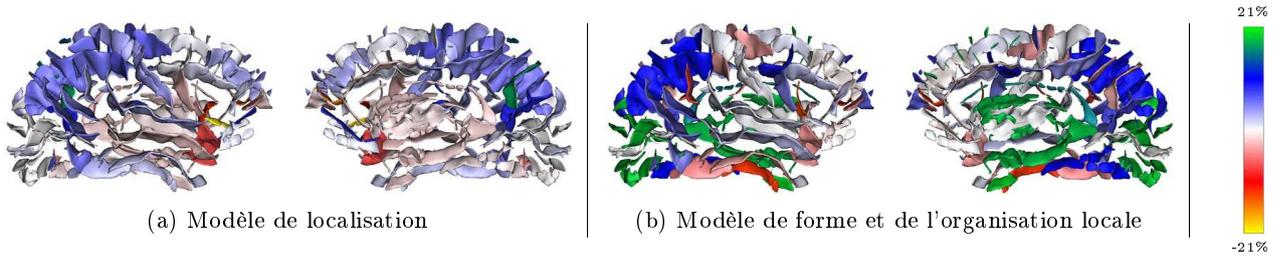


FIG. 5.4 – Différences de taux d’erreurs locaux ($E_{\text{local}}(l)$, voir équation 5.2) entre chacun des modèles pris individuellement ((a) SPAM, (b) MSVR) et le modèle couplé (SPAM+MSVR) présenté sur un sujet arbitraire.

5.5.2 Expérimentations et discussion

Pour pouvoir se confronter au modèle «historique» les expérimentations ont été menées sur la même base de données que dans [Rivière 2002], à savoir, 26 hémisphères droits labélisés manuellement où 21 sont utilisés pour l’apprentissage du système (16 pour apprendre chaque expert local : MLP ou SVR, 5 pour noter et pondérer chaque modèle local en fonction de sa pertinence dans l’énergie globale à optimiser) et 5 pour évaluer les capacités de généralisation de la méthode. Dans cette section uniquement, aucune validation croisée n’a été utilisée, les résultats sont donc à considérer avec précaution puisqu’un changement dans la répartition des bases d’apprentissage et de test peut influencer sur les résultats. L’ensemble des informations considérées dans D sont calculées après une normalisation vers l’espace de Talairach.

Le modèle de SPAM est plutôt rapide à apprendre et donne de meilleurs résultats en moyenne que le modèle «historique». Les taux d’erreurs ne sont pas toujours en accord avec l’appréciation visuelle que pourrait faire un expert humain (ce qui ne veut pas dire que les chiffres ont tort d’ailleurs, ils ne représentent pas la même chose, voir figure 5.3). Cette impression est due principalement au fait que l’œil humain tolère difficilement des échanges locaux de labels (un segment mal labélisé au milieu d’un sillon par exemple) alors qu’il est moins regardant au déplacement cohérent d’un ensemble de structures à condition que chaque structure conserve une forme de sillon. Nos mesures d’erreurs pénalisent plus la deuxième situation que l’on rencontre fréquemment (même si ce n’est que localement) avec l’approche «historique» que la première qui correspond aux erreurs obtenues avec les SPAMs.

Finalement, des informations de haut niveau restent requises, même si les taux d’erreurs constatés sur le modèle à base de SVR sont supérieurs. En effet, chaque labélisation inférée présente des formes et des *patterns* plus consistants. Les résultats apportés par le mélange des deux approches sont reportés dans le tableau 5.1 et indiquent une tendance à l’amélioration des résultats. Les meilleurs résultats du modèle MSVR sont obtenus avec l’optimisation basée sur un échantillonnage de *Gibbs* et un recuit simulé, l’optimisation MPM à une température $T = 1$ étant légèrement moins performante. L’approche ICM est beaucoup moins bonne si on part d’une initialisation au hasard. On constate dans toutes les situations que le modèle SPAM améliore considérablement les résultats du modèle MSVR, même s’il n’est utilisé qu’en tant que simple initialisation. Par contre, les résultats du modèle SPAM ne sont améliorés que dans le cas de l’intégration la plus poussée avec une optimisation par recuit simulé. Dans ce cas, le gain est moins net à l’échelle globale. L’erreur E_{mean} du tableau 5.1 permet de mieux mettre en relief des améliorations impliquant des petites structures puisque la contribution de chaque label à l’erreur est normalisée. Plus précisément encore, la figure 5.4 permet de constater quelques détériorations isolées, de nombreux sillons où le couplage des deux modèles n’apporte rien et tout autant de sillons présentant des améliorations nettes.

Cette expérimentation présente une première tentative de réconciliation entre le nouveau modèle proposé et des informations contextuelles. L'approche présentée ici souffre des défauts de l'approche «historique», notamment sa lourdeur empêchant une validation plus poussée (pas de validation croisée possible ici) et une inhomogénéité dans la nature des différents potentiels du champ de Markov empêchant une normalisation correcte des différents termes. En effet, pour les SPAM, le potentiel n'est pas borné car issu d'une distribution de probabilité et l'autre correspond à un potentiel ad-hoc basé sur la sortie bornée (entre -1 et 1) d'une fonction de décision. En l'occurrence, il semblerait que le modèle SPAM ait été privilégié artificiellement, le modèle Markovien ne venant qu'apporter des ajustements à la solution proposée. Nous verrons plus loin dans la partie III, une modélisation Markovienne différente menant à une meilleure interaction des différents termes.

Conclusion et perspectives

6.1 Conclusion

Nos analyses ont fait ressortir que certaines informations (positionnelles ou directionnelles principalement) pouvaient être déterminées avec précision à l'échelle des segments sulcaux. Nous avons tenté dans cette partie d'exploiter au maximum cette information locale au travers d'une modélisation basique où chaque segment est considéré indépendamment des autres. Nous avons proposé dans ce cadre plusieurs modèles locaux permettant d'exploiter de telles informations locales. Notre attention a été retenue par le modèle SPAM qui fournit une représentation robuste de la variabilité de la localisation des plissements corticaux. Nous avons d'ores et déjà envisagé des extensions à ce modèle pour mieux modéliser des variations locales de cette variabilité (voir annexe B.4). Un ensemble d'a priori a été envisagé permettant de contraindre la labélisation. Une attention particulière leur a été accordée afin de s'abstraire autant que possible de la segmentation en segments sulcaux. Cette modélisation a été traduite dans un cadre supervisé (basé sur une base de données intégralement labélisée) et semi-supervisé (exploitant également un ensemble de données non labélisées) qui permettra l'exploitation prochaine de bases de données de taille conséquente sans nécessiter une labélisation manuelle dont l'expertise est autant rare que difficile.

Cette approche locale nous a permis de simplifier au maximum les schémas d'optimisation pour la phase de labélisation et d'atteindre une rapidité de traitement inégalée : moins de 10s pour la labélisation d'un hémisphère. Les résultats obtenus sur notre base de données ont démontré l'intérêt et l'efficacité d'une telle approche, notamment avec le modèle SPAM qui surpasse le modèle «historique» [Rivière 2002] en terme de taux d'erreurs. Comme nous l'avons indiqué plus tôt, ces mesures ne prennent pas en compte la cohérence des plissements qui reste meilleure dans le cas du modèle «historique». On retrouvera naturellement le modèle proposé dans cette partie au cœur des modélisations à venir dans la suite de ce manuscrit.

Cette modélisation est évidemment sensible aux incertitudes de localisation des structures induites par la normalisation des sujets vers un espace commun. Nous réponderons à cette limitation dans la partie II. Les SPAMs sont très sensibles à la qualité de la segmentation des plissements corticaux en segments sulcaux. Ainsi, plus un segment sulcal est petit, plus l'information devient locale et moins elle est pertinente sur le plan anatomique. L'introduction d'informations contextuelles (tirées des segments voisins) pourraient résoudre le problème et assurer une plus grande cohérence à la labélisation. Nous proposerons en partie III une modélisation Markovienne permettant d'inclure de telles informations.

6.2 Perspectives

De nombreuses améliorations semblent encore possibles à l'échelle des segments sulcaux. Nous avons eu l'occasion de constater que les modèles SPAMs utilisés pour caractériser la variabilité positionnelle des sillons codent également dans une certaine mesure des informations de forme. Sans même considérer une étude de la forme des plissements à l'échelle d'un sillon entier (en regroupant les segments portant le même label), nous pensons que les informations de forme restent sous exploitées.

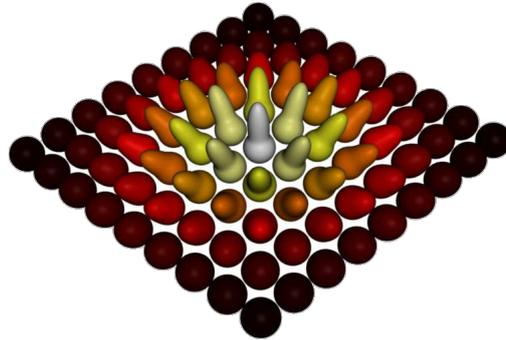


FIG. 6.1 – Exemple de représentation souhaitée pour caractériser la variabilité locale d’un paramètre directionnel. Ici chaque distribution locale (à un voxel) correspond à une distribution de Von Mises Fisher sur une donnée directionnelle. Chaque structure sulcale serait alors caractérisée par une direction par voxel de la structure et une distribution directionnelle uniforme partout ailleurs. L’estimation de chacune de ces distributions serait alors réalisée en pondérant la contribution de chaque donnée directionnelle par un noyau Gaussien par analogie avec le modèle SPAM.

Dans la partie II, nous proposerons une méthode qui permet de s’affranchir des différences locales de position et d’orientation des structures (par une transformation rigide ou affine) afin de mieux caractériser leurs formes.

À label fixé, le modèle SPAM modélise de façon robuste la probabilité qu’une structure passe par un voxel donné. Jusqu’à présent, nous avons considéré que ces mesures de probabilité étaient indépendantes d’un voxel à l’autre (si l’on omet la normalisation par le nombre de voxels). Nous pourrions alors envisager une prise en compte de dépendances locales (de type Markovien par exemple). Ceci reviendrait à considérer la probabilité qu’une structure passe par plusieurs voxels voisins. Une telle approche devrait permettre de prendre en compte des informations directionnelles locales. Nous suggérons en annexe B.4, en tant que première ébauche, une modélisation Gaussienne basée sur une représentation iconique des sillons, permettant de prendre en compte une structure de covariance entre les voxels. Cette représentation est fortement dépendante du choix de l’espace de représentation des données (y compris son orientation et son niveau de résolution). Nous envisageons désormais une représentation alternative non sujette à de telles limitations, où les structures (segments, sillons) sont représentés par des mesures de probabilités [Glaunes 2004]. Il serait alors envisageable d’étendre la modélisation précédente en utilisant une norme adaptée aux distributions.

La modélisation de la covariabilité locale des structures devrait se traduire par la prise en compte d’informations directionnelles locales. L’étude que nous avons menée sur les mesures directionnelles à l’échelle des segments est l’équivalent directionnel de la modélisation Gaussienne proposées sur les centres de gravité des segments. Cette mesure est bien définie puisque calculée sur de nombreux éléments, mais ne permet pas de distinguer certaines variations locales. De la même façon que le modèle SPAM répond à cette limitation, on pourrait imaginer étendre cette notion aux directions puisque certains sillons présentent des profils particuliers. Si chaque structure est représentée par un ensemble de vecteurs localisés, cela reviendrait à estimer une distribution sur des champs de vecteurs. On peut imaginer l’estimer par un ensemble de modèles locaux comme le suggère la figure 6.1 où chaque estimation locale est réalisée de manière robuste en pondérant la contribution des vecteurs par un noyau centré sur le voxel considéré. Précédemment nous avons suggéré l’utilisation de mesures de probabilité pour représenter et modéliser les informations positionnelles. Ces notions devraient s’étendre naturellement à la représentation sous forme de courants [Glaunès 2006] dans le cas directionnel. Cette idée peut être étendue aux 3 directions privilégiées caractérisant un sillon

(direction, normale à la surface, normale à l'enveloppe convexe) que l'on modéliserait alors par une distribution de type *Matrix* Von-Mises-Fisher (voir section B.3.1.6) permettant de modéliser l'orientation locale du repère formé par ces 3 vecteurs. Ces axes pourraient définir dans une certaine mesure les directions de courbure locales d'une structure. Finalement, une extension naturelle serait de modéliser la variabilité des tenseurs de courbures locaux. Avec une telle représentation, on pousserait au maximum la prise en compte d'informations locales. Chaque voxel d'une structure serait alors défini pleinement par 5 paramètres (3 paramètres pour la rotation définissant les 2 directions de courbure et 2 paramètres pour les courbures le long de ces directions) dont l'apprentissage nécessiterait une vigilance accrue pour limiter le sur-apprentissage.

La plupart des estimateurs à noyaux, le modèle SPAM et ses extensions (section B.4) y compris, sont contrôlés par un paramètre d'échelle. Dans le cas des SPAM la largeur σ du noyau Gaussien a été fixée arbitrairement à $2mm$, mesure sans doute suffisante sur les grands sillons et largement sous-évaluée pour des structures plus variables (branches du sillon temporal supérieur par exemple). Plusieurs stratégies sont possibles afin de mieux caractériser cette échelle d'étude. La première consisterait à trouver sa valeur optimale pour chaque label, couplée à l'étape d'estimation du modèle, ce qui reviendrait à maximiser la probabilité suivante :

$$P(\mathcal{M}\sigma|\{D_a L_a\}_{a\in\mathcal{A}}) \quad (6.1)$$

à la place de l'équation 4.5 proposée précédemment. L'optimisation de σ pourrait nécessiter une estimation du modèle SPAM par *leave-one-out* sur la base d'apprentissage \mathcal{A} pour éviter une sous-estimation du paramètre. Si l'on souhaite une prise en compte complète de ce paramètre dans un cadre Bayésien, il est nécessaire d'intégrer l'incertitude sur ce paramètre dans le modèle afin d'en tenir compte lors de l'inférence des labels en maximisant la probabilité suivante :

$$\begin{aligned} P(L|D; \mathcal{M}) &\propto P(L|\mathcal{M})P(D|L; \mathcal{M}) \\ &= P(L|\mathcal{M}) \int_{\sigma} \underbrace{P(D|L\sigma; \mathcal{M})}_{\substack{\text{modèles génératifs} \\ \text{sulcaux}}} \underbrace{P(\sigma|\mathcal{M})}_{\text{a priori}} d\sigma \end{aligned} \quad (6.2)$$

où l'a priori sur le paramètre d'échelle est déterminé sur la base d'apprentissage \mathcal{A} :

$$\begin{aligned} P(\sigma|\mathcal{M}) &= P(\sigma|\{D_a L_a\}_{a\in\mathcal{A}}; \mathcal{M}) \\ &= \frac{P(\sigma) \prod_{a\in\mathcal{A}} P(D_a|L_a\sigma; \mathcal{M})}{\int_{\sigma} P(\sigma) \prod_{a\in\mathcal{A}} P(D_a|L_a\sigma; \mathcal{M}) d\sigma} \end{aligned} \quad (6.3)$$

Cette formulation nécessite l'estimation de SPAMs à diverses valeurs de σ permettant l'échantillonnage des intégrales apparaissant dans les formules. Cette approche construit alors un modèle multi-échelle sans doute plus robuste que celui que nous avons présenté dans cette partie.

Les structures sulcales les plus simples sont caractérisées par une forme générale de ruban perturbée localement par de petites branches secondaires quasi-orthogonales au plissement principal. Ces plis sont plus variables, plus petits et moins représentatifs de la forme des sillons (ils n'apparaissent pas dans la représentation visuelle des SPAMs). Les mêmes remarques s'appliqueront aux modèles locaux directionnels suggérés. Ils sont donc plutôt mal modélisés en l'état et nécessiteraient sans doute un traitement particulier. Il faudrait sans doute introduire des sous-classes au sein de chaque label afin de les distinguer dans la modélisation.

Deuxième partie

Normalisation et sillons

Table des Matières

7	Couplage : recalage et labélisation	69
7.1	Normalisation en tant que prétraitement	69
7.2	Techniques de recalage retenues	70
7.3	Recalage et labélisation joints	72
7.4	Recalage et estimation de modèle joints	74
7.5	A priori de recalage	76
7.5.1	Approche globale	76
7.5.2	Approche locale	77
7.5.3	Modélisation	78
7.6	Optimisation des paramètres de recalage	81
7.6.1	Cas non paramétrique : modèle SPAM	81
7.6.2	Cas paramétrique : approche Riemannienne	82
8	Résultats	89
8.1	Modèles	89
8.2	Résultats et discussion	90
9	Conclusion et perspectives	97
9.1	Conclusion	97
9.2	Perspectives	98
9.2.1	Méthodes de recalage	98
9.2.2	Remarques diverses	99

En neuroimagerie, de nombreux défis méthodologiques et neuroscientifiques impliquent des études de groupe qui tentent de résumer, répéter ou généraliser des résultats individuels à l'échelle d'un groupe de sujet. Les techniques de cartographie cérébrale (*Brain Mapping*) ont été largement explorées en ce sens et visent entre autres à identifier et appairer des structures communes entre des individus. Elles couvrent un spectre très large d'applications, de la détection d'amers fiables pour guider des opérations chirurgicales, à la compréhension des fonctions cérébrales, en passant par la détection de troubles neurodéveloppementaux. La qualité et la significativité des résultats issues d'une étude de groupe dépendent en grande partie de la qualité de cette cartographie.

On distingue principalement deux approches permettant de définir une cartographie cérébrale à l'échelle d'un groupe de sujets : les techniques de recalage et celles visant à l'identification de structures (labélisation ou segmentation). Chacune de ces approches possède son angle d'attaque privilégié qui repose sur son propre *background* méthodologique. Ceci explique sans doute pourquoi la combinaison des deux approches se limite la plupart du temps à la succession de l'une puis de l'autre¹.

Dans ce contexte, les méthodes de recalage visent à normaliser l'ensemble des données étudiées vers un repère commun par l'intermédiaire d'une transformation qui garde un sens anatomique (transformations affines, non-linéaires contraintes ou difféomorphiques). Pour une revue beaucoup plus détaillée des techniques de recalage en neuroimagerie, le lecteur est invité à consulter les articles suivants : [Maintz 1998] et [Gholipour 2007]. Nous concentrerons notre description rapide aux études sur les structures sulcales.

À l'origine, la plupart des méthodes de recalage se basaient sur des méthodes iconiques, guidées par les mesures d'intensité des voxels des volumes IRM. Les approches surfaciques ont fait ensuite leur apparition [Fischl 2004, Tosun 2008] facilitant l'exploitation multi-échelle d'informations telles que la courbure ou la profondeur corticale². Ces informations ne suffisent pas à lever les ambiguïtés locales entre des plissements voisins aux propriétés géométriques similaires. C'est pourquoi de nombreuses méthodes de recalage sont contraintes par des structures corticales identifiées qu'elles tentent d'appairer entre les sujets. Dans le cas des sillons, ces contraintes s'expriment sous la forme d'une ensemble de points [Counce 1998, Chui 1999, Lohmann 2000, Chui 2000], le long de la surface corticale [Thompson 2000, Van Essen 2005, Joshi 2005, Lui 2007] (souvent représentée par une projection plane ou sphérique) ou encore définies de façon volumique [Collins 1995, Collins 1998, Cachier 2003, Auzias 2009]. Dans la plupart de ces références, ces structures sont identifiées manuellement par des experts en anatomie, ce qui réclame beaucoup de rigueur et de temps même pour des personnes qualifiées. Par contre une telle approche devient impossible avec l'augmentation du nombre de données à traiter.

L'identification automatique de telles structures anatomiques résulte de processus de segmentation et de classification. Dans le cas des structures corticales (par opposition aux structures sous-corticales), ces étapes se succèdent généralement, en commençant par l'extraction des sillons, suivie par leur labélisation. On distingue les approches graphiques, représentant l'information anatomique sous forme structurée [Rivière 2002, Shi 2007, Shi 2009, Yang 2007, Yang 2009] (graphe de descripteurs, section 11), celles basées sur la localisation spatiale des structures depuis un repère commun [Goualher 1998, Lohmann 2000, Perrot 2008] et les méthodes surfaciques exploitant en plus des informations de courbure [Tosun 2003, Fischl 2004, Yeo 2008b].

Une autre famille de méthodes, à la frontière entre la normalisation et l'identification de structures, consiste à déformer (par une méthode de recalage) un atlas de labels sur chaque sujet d'étude et à en reporter ensuite les labels. La dernière décennie a montré un intérêt croissant pour les atlas probabilistes [Mazziotta 1995, Thompson 1997, Joshi 2004, Ashburner 2005, De Craene 2005,

¹les deux sens existent comme nous le verrons plus en détail dans la suite.

²sorte de distance à l'enveloppe cérébrale.

[Allasonnière 2007, Shattuck 2008], qui prennent mieux en compte la variabilité inter-individuelle, avec une attention particulière sur les atlas non-biaisés (ne dépendant pas du choix arbitraire d'un sujet de référence).

Finalement, la précision d'un atlas ou l'identification de structures anatomiques reposent sur la qualité du co-recalage réalisé [Lyttelton 2007]. Réciproquement, de meilleures normalisations sont obtenues à partir d'un atlas plus fiable ou à partir d'atlas anatomiques comme les sillons. Très peu d'articles font référence à de telles interconnexions et en tirent avantage. Dans le contexte de la labélisation de sillons, [Vaillant 1999] explore une stratégie itérative visant à reconsidérer et raffiner selon une approche hiérarchique l'identification de telles structures grâce à un recalage. Récemment, [Ashburner 2005, Pohl 2006, Yeo 2008a] ont proposé des formulations Bayésiennes relativement similaires et répondant assez bien à ce problème par l'estimation jointe de paramètres d'un recalage et de l'identification de structures (par segmentation ou classification). Ces articles suivent une formulation Bayésienne standard mais avec des différences significatives dans la représentation de la variabilité anatomique, les méthodes de recalage, la formulation et l'estimation des différents a priori. En effet, [Ashburner 2005] contraint les paramètres de transformation par un a priori Gaussien dont la covariance est dérivée du Laplacien du champ de déformation ; [Yeo 2008a] utilise, pour chaque arrête du maillage cortical, un a priori Gaussien proportionnel à la longueur de l'arrête considérée, alors que [Pohl 2006] estime cet a priori de façon empirique à partir d'une base d'apprentissage. Dans cette partie, nous allons suivre cette dernière approche en raffinant la façon dont les paramètres de recalage sont représentés et contraints.

Couplage : recalage et labélisation

Dans ce chapitre nous étudions la place d'un procédé de normalisation dans le cadre de la modélisation de la variabilité sulcale ainsi que l'identification de telles structures chez un individu test. Tout d'abord, la normalisation est considérée, comme il est d'usage, en tant que pré-traitement fixé une fois pour toutes avant toute analyse complémentaire. Ensuite, ce schéma standard est remis en cause au profit d'une approche réalisant la normalisation et les analyses en même temps. Nous proposerons alors deux techniques de recalage assez simples basées sur des transformations affines ou rigides. Ensuite, ce schéma est formalisé dans un cadre Bayésien permettant de raffiner l'estimation d'un modèle sulcal ou une labélisation corticale. Finalement, nous détaillerons l'ensemble des a priori utilisés permettant de contraindre le processus.

7.1 Normalisation en tant que prétraitement

Nous envisageons dans cette section une étape de normalisation précédant toute autre analyse et visant à définir un espace commun aux informations anatomiques étudiées. Il s'agit du schéma classique suivi dans le cadre des études de groupe qui permet de mettre en correspondance et comparer les structures similaires.

Le cas qui nous intéresse est le suivant. Il s'agit de spécifier d'une part un espace commun pour modéliser la variabilité sulcale de nos données d'apprentissage (à label connu) et d'autre part transformer vers ce même espace des données de test aux structures sulcales non-identifiées. Une méthode de recalage conforme à ce contexte serait aveugle aux labels sulcaux afin de prendre en compte les données de test. Par conséquent, seules les données directement accessibles depuis les intensités des volumes IRM (y compris les approches volumiques, ou celles basées sur des représentations surfaciques ou des champs de courbure) ou depuis d'autres modalités (faisceaux de fibres de la matière blanche, activations fonctionnelles) devraient être considérées. Cet éventail de méthodes de recalage comprend des approches linéaires (ou affines) et d'autres non-linéaires. Les méthodes non-linéaires sont séduisantes par bien des aspects mais il est difficile d'empêcher des confusions locales en cas d'ambiguïtés. Dans notre cas, plus de 60 labels sont considérés (y compris des petits plissements), de telle façon que cette situation n'est pas rare et peut être dommageable à l'identification de ces structures. Notre expertise reste limitée sur les méthodes de recalage non-linéaires et leur étude prématurée. Nous la réservons donc à de futurs travaux dans lesquels sera présentée une telle méthode de recalage intégrée au cadre méthodologique à suivre dans ce chapitre.

Nous listons ici l'ensemble des méthodes de recalage affine que nous avons testées afin de juger de la qualité de nos modèles en comparaison.

Premièrement, nous avons utilisé une variante du système de coordonnées de Talairach [Talairach 1988] tel qu'il est défini dans la chaîne de traitement anatomique de la suite logicielle Brainvisa [Rivière 2009]. Une transformation affine permet de passer de l'espace du sujet à l'espace de Talairach. Elle est définie manuellement en identifiant les commissures antérieure et postérieure ainsi qu'un point du plan interhémisphérique. Ensuite le plan interhémisphérique et les facteurs d'échelle appliqués dans le référentiel défini par le plan sont automatiquement trouvés en faisant correspondre les bords du cerveau du sujet avec les trois axes de l'atlas de Talairach. Ce système de

coordonnées a été originellement conçu pour aligner des structures sous-corticales profondes (thalamus, putamen, noyaux caudés...) qui sont plutôt bien localisées spatialement. Au niveau cortical, la variabilité est beaucoup plus importante, mais éventuellement sur-estimée puisque le référentiel de Talairach ne lui est pas adapté.

Nous avons également essayé le logiciel Baladin [Ourselin 2000] dédié initialement au recalage multimodal de volumes IRM, restreint ici à sa version affine. Cette méthode est relativement fiable puisqu'elle se base sur une stratégie multi-échelle de mise en correspondance par bloc (*block matching*) afin de capturer des déplacements à petite et grande échelle. De plus, le critère de similarité utilisé, le coefficient de corrélation, est optimal dans ce cas d'utilisation. Finalement, la transformation affine résulte d'une optimisation basée sur des estimateurs robustes (*least trimmed squares*) ignorant les données trop marginales (*outliers*). Cette méthode a été inventée au départ pour reconstruire des volumes en 3D à partir d'une collection de séquences 2D non recalées, mais s'est avérée plutôt efficace pour le recalage de volumes entiers.

En comparaison, nous avons également testé une des méthodes de recalage les plus standards en neuroimagerie : la version affine du logiciel SPM8b [Ashburner 2007]. Le principe général de cette méthode est de trouver les paramètres de recalage qui minimisent l'information mutuelle entre les 2 images à mettre en correspondance. La formulation du problème est Bayésienne et utilise des a priori prenant en considération la variabilité d'une tête humaine (visible de façon assez distincte sur le *template* considéré) en termes de taille et de forme.

Le co-recalage de plus de 2 sujets d'apprentissage avec Baladin et SPM a été mené en recalant tous les sujets considérés sur une image cible (un *template* ou une image moyenne). Nous avons utilisé un des *template* les plus utilisés dans la communauté, à savoir le *template* MNI nommé ICBM152 [Mazziotta 2001] qui résulte de la moyenne de 152 sujets recalés par des transformations affines sur le *template* MNI305 [Evans 1993]. Tant qu'on se restreint à des transformations affines ou rigides, les *templates* de la littérature sont relativement similaires les uns aux autres : assez flous au niveau des zones corticales à cause de l'importance de la variabilité inter-individuelle, seule une forme grossière de chaque hémisphère et du crâne peut être réellement discernée.

Finalement, nous nous sommes également intéressé à la version affine du logiciel ANIMAL [Collins 1994, Collins 1995] (utilisé dans le cadre de la construction du *template* MNI305), mais nous n'avons malheureusement pas réussi à obtenir d'amélioration de la qualité de l'alignement sulcal. Nous ne détaillerons donc pas plus cette expérience.

7.2 Techniques de recalage retenues

Aucune des 3 méthodes standards de normalisation décrites dans la section précédente (Talairach, SPM, Baladin), ne peuvent réellement prétendre fournir une transformation optimale vis-à-vis de l'alignement des structures corticales. En effet, il se peut qu'elles ne soient pas en mesure de distinguer des plissements corticaux voisins les uns des autres. Par conséquent, la prise en compte d'information anatomique, au travers des modèles sulcaux comme les SPAMS (section B.4) dans le processus de recalage devrait permettre d'introduire des contraintes levant certaines ambiguïtés. À l'inverse, le choix de l'espace de référence est primordial afin de réduire autant que possible la variabilité anatomique de notre modélisation et faciliter sa confrontation avec d'autres sujets. Chacune des deux informations considérées (transformation vers un espace commun et identification de structures anatomiques) présente un intérêt pour l'autre. Le cadre Bayésien nous permet de proposer une formulation permettant de répondre de façon élégante à cette question, mais à condition de retravailler la phase d'estimation des paramètres de recalage, comme nous le verrons plus loin. Pour la partie consacrée au recalage, l'idée principale est de remplacer le *template* IRM utilisé par les approches iconiques décrites précédemment par un modèle probabiliste des sillons qui peut être vu comme un atlas probabiliste des structures sulcales.

En suivant ces idées, nous allons considérer dans la suite deux méthodes principales de recalage : la première, **l’approche globale** comporte une unique transformation affine, la seconde, **l’approche locale** comprend une transformation affine locale par label.

La transformation globale est définie par seulement deux variables $(A_g, t_g) = \theta_g$, où A_g correspond à une matrice non singulière et t_g un vecteur de translation. Ici, l’indice g indique qu’il s’agit de l’approche globale. A_g peut être décomposée en 3 composantes plus facilement compréhensibles par l’intermédiaire d’une décomposition en valeurs singulières (SVD) : $A_g = U_g D_g V_g^t$, avec U_g et V_g deux matrices de rotation et D_g une matrice diagonale contenant des facteurs d’échelle. V_g^t définit le plan le long duquel les facteurs d’échelle sont appliqués (on s’attend à ce que le processus d’optimisation aligne les axes principaux du cerveau avec les axes canoniques). $U_g V_g^t$ représente alors la partie rigide de la transformation A_g . Une transformation affine est finalement contrôlée par 12 paramètres scalaires (3 pour chaque rotation, 3 pour la mise à l’échelle et 3 pour la translation). Par conséquent, il est également intéressant de considérer des transformations rigides (sans paramètre d’échelle) qui sont définies par seulement 6 paramètres (3 pour la rotation et 3 pour la translation), donc plus contraintes et plus faciles à optimiser. Nous verrons par la suite (section 7.5) que les facteurs d’échelle demandent une attention particulière (utilisation d’a priori pour contraindre le processus) pour empêcher la divergence du processus d’optimisation.

Précédemment dans cette section, nous avons laissé de côté les méthodes de recalage non-linéaires dans le cas où les sillons sont inconnus. Les modèles sulcaux étudiés au cours de cette thèse donnent de sérieux indices sur l’identification des plissements corticaux. Ainsi, dans le cadre de la méthodologie envisagée (coestimation de la normalisation spatiale et de la labélisation : section 7.3) les approches non-linéaires pourraient être reconsidérées à condition d’être intégrées avec précaution en choisissant des a priori appropriés. Cette dernière décennie, ne nombreuses recherches se sont focalisées sur des techniques de recalage difféomorphique [Avants 2004, Joshi 2004, Arsigny 2005, Beg 2005, Ashburner 2007, ?, Auzias 2009] qui garantissent l’obtention d’une transformation non-linéaire, inversible et spatialement lisse. Ces méthodes peuvent avoir un intérêt dans ce contexte, mais ajoutent un surcoût calculatoire non négligeable qu’on doit également prendre en considération. Des approximations seront sans doute nécessaires afin d’obtenir une méthode efficace. Nous laissons ces considérations pour de futurs travaux, bien conscient qu’il s’agit d’une étape désormais incontournable. Nous précisons toutefois que de telles contraintes (inversibilité, continuité de la transformation) peuvent être excessives si l’application visée est la labélisation de structures anatomiques. Dans le cas où les sillons ainsi identifiés devraient être utilisés dans un processus de recalage contraint, il serait sans doute plus efficace d’intégrer cette étape dans l’approche jointe proposée afin de prendre en compte l’incertitude liée à la labélisation.

Bien sûr cet argument ne tient que si l’application envisagée ne comporte pas de processus de recalage ultérieur, auquel cas il serait sans doute plus efficace de l’intégrer directement dans l’approche proposée dans cette partie.

Un premier pas est fait dans la direction des techniques de recalage non-linéaires qui étend naturellement l’approche rigide globale précédente. Plutôt que d’utiliser une seule transformation affine ou rigide, une collection de transformations rigides locales sont considérées : une par sillon, notée par les paramètres suivant : $\theta_s = \{\theta_{s,l}\}_{l \in \mathcal{L}} = \{(R_{s,l}, t_{s,l})\}_{l \in \mathcal{L}}$. Ici, l’indice s indique que les transformations sont rattachées à un sillon particulier et l indique leurs labels respectifs. Nous avons limité notre étude aux transformations rigides, principalement afin d’éviter le surcoût calculatoire de l’approche affine alors que peu de profits sont à espérer si l’on en croit les résultats très proches obtenus avec les méthodes affines et rigides globales.

7.3 Recalage et labélisation joints

On suppose dans cette section que nous avons déjà à notre disposition un modèle \mathcal{M} de la variabilité sulcale¹ qui a été estimé à partir d'une base d'apprentissage. Il a pu être estimé à partir d'un espace commun défini a priori (quelques méthodes standards de normalisation affine sont décrites en section 7.1) ou estimé de façon conjointe avec les autres paramètres du modèle (voir section 7.4). Dans cette section, toute référence au modèle \mathcal{M} restera implicite puisque toutes les probabilités sont définies conditionnellement au modèle.

Notre but premier est d'identifier la meilleure labélisation sulcale l possible d'un sujet test (c'est à dire, non observé pendant la phase d'apprentissage). En même temps, nous recherchons la meilleure normalisation, régie par les paramètres θ (limitée dans nos expériences à une transformation affine globale ou à une collection de transformations affines locales : une par sillon, voir la section 7.2 pour plus de détails) afin de maximiser la mise en correspondance entre le modèle et le sujet non labélisé. Dans un cadre Bayésien, cette formulation s'écrit :

$$l^*, \theta^* = \arg \max_{l, \theta} P(L = l \Theta = \theta | D) \quad (7.1)$$

La résolution directe de cette équation est très difficile puisque tous les paramètres doivent être optimisés en même temps alors que l'expression à optimiser présente de nombreux maxima locaux. Cette probabilité peut être intégrée selon certains paramètres afin d'en simplifier l'expression. Par exemple, les probabilités marginales suivantes : $P(L|D)$ et $P(\Theta|D)$ résultent d'un tel processus et devraient donner des résultats plutôt proches. En effet maximiser l'une des marginales ($P(L|D)$ ou $P(\Theta|D)$), et utiliser ce maximum (respectivement l^* et θ^*) pour maximiser l'autre composante (respectivement, θ^* maximisant $P(\Theta|DL = l^*)$ ou l^* maximisant $P(L|D\Theta = \theta^*)$) revient sensiblement au même que l'étude directe de $P(L\theta|D)$ puisque les deux quantités sont fortement corrélées. Idéalement, l'étude de $P(L|D)$ est à préférer puisqu'elle considère le problème d'optimisation comme un problème de labélisation (le sujet qui nous intéresse de prime abord) prenant en compte toute la variabilité d'une normalisation inconnue θ :

$$P(L|D) \propto \int P(D|L\Theta = \theta)P(L|\Theta = \theta)P(\theta)d\theta \quad (7.2)$$

Les paramètres θ régissant la transformation s'expriment dans un espace continu puisqu'il s'agit dans notre cas de la composition de rotations, facteurs d'échelle et de vecteurs de translation. Par conséquent, nous devrions estimer une labélisation pour chaque valeur possible de θ pour estimer les probabilités a posteriori $P(\Theta = \theta|LD)$, essentielles à l'optimisation. En l'état, c'est impossible puisque l'espace considéré est continu et de très grande dimension. une évaluation de l'intégrale par des techniques d'échantillonnage coûteuses (de type Markov Chain Monte Carlo MCMC par exemple) serait sans doute nécessaire.

Finalement, l'autre probabilité marginale que nous avons évoqué $P(\Theta|D)$ semble plus facile à manipuler et, comme nous allons le voir, est propice à des simplifications mathématiques dans le cadre d'une modélisation de la variabilité sulcale basée sur une indépendance des segments sulcaux (partie I) et modélisation Markovienne entre les segments sulcaux voisins (partie III). Cette fois, cette mesure correspond à l'optimisation de paramètres θ contrôlant la normalisation du sujet considéré contrainte par des amers corticaux dont l'ensemble de la variabilité est prise en compte. Par conséquent, cette méthode conduit à une méthode de recalage robuste contrainte par des labels cachés (inconnus mais inférés à partir des modèles sulcaux). Ensuite, la labélisation est optimisée

¹Le cas de l'estimation du modèle SPAM est décrit en section 4.2.3.

en fixant les paramètres de normalisation aux meilleurs paramètres θ^* obtenus précédemment.

$$\left\{ \begin{array}{l} \theta^* = \arg \max_{\theta} P(\Theta = \theta | D) \\ \quad = \arg \max_{\theta} P(\theta) \sum_{L=l} P(L = l | \theta) P(D | L = l; \theta) \\ l^* = \arg \max_l P(L = l | D; \Theta = \theta^*) \\ \quad = \arg \max_l P(D | L = l; \Theta = \theta^*) P(L = l | \Theta = \theta^*) \end{array} \right. \quad (7.3)$$

L'optimisation de paramètres alors que des variables demeurent cachées correspond à un problème bien connu, traité généralement par l'algorithme Expectation Maximization (EM [Dempster 1977]). Cette méthode part d'une initialisation bien choisie des paramètres à optimiser. Dans notre cas, ces paramètres sont donnés par une première normalisation, telle que celles décrites en section 7.1 par exemple. L'approche jointe proposée avec recalage affine global peut être utilisée en tant qu'initialisation d'une autre méthode de recalage moins grossière (l'approche affine locale par morceau par exemple). La méthode EM est itérative et améliore à chaque étape la probabilité à maximiser. Premièrement, elle vise à calculer les probabilités a posteriori $P(L = l | D \Theta)$ pour chaque labélisation l possible. C'est à dire, $|\mathcal{L}|^{|\mathcal{E}|}$ combinaisons, avec $|\mathcal{L}|$ le nombre de labels à notre disposition et $|\mathcal{E}|$ le nombre de segments sulcaux du sujet considéré, qui correspond à environ 10^{1000} pour 62 labels et 250 segments sulcaux. C'est la raison pour laquelle, nous rechercherons des simplifications mathématiques pour rendre les calculs réalisables. Deuxièmement, ces probabilités a posteriori sont utilisées en tant que poids afin de raffiner l'estimation précédente des paramètres de normalisation. Finalement, le processus converge vers un maximum local de la probabilité d'intérêt. Le schéma décrit ci-dessus, se résume de la façon suivante :

$$\theta^{(n+1)} = \arg \max_{\theta} Q(\theta | \theta^{(n)}) \quad (7.4)$$

avec

$$\begin{aligned} Q(\theta | \theta^{(n)}) &= \mathbb{E}_L [\log(P(DL\theta)) | D\theta^{(n)}] \\ &= \sum_{L=l} w_l^{(n)} \log [P(D|L=l\theta)P(L=l|\theta)P(\theta)] \end{aligned} \quad (7.5)$$

où $\mathbb{E}_L [\cdot]$ correspond à l'espérance en L sachant l'ensemble des informations inconnues et les probabilités a posteriori s'écrivent $w_l^{(n)} = P(L=l | D\theta^{(n)})$. Dans la formule ci-dessus, la somme est réalisée sur toutes les labélisations l possibles.

Maintenant, nous allons décliner cet algorithme pour les approches globale affine et globale rigide. Plutôt que supposer de fortes dépendances statistiques entre les données observées D , nous utilisons une hypothèse classique d'indépendance conditionnelle des données $\{D_i\}_{i \in \mathcal{E}}$ sachant la labélisation L qui permet de simplifier considérablement le problème. D'ailleurs, comme nous l'avons détaillé plus tôt, l'a priori sur la labélisation $P(L)$ ne dépend pas de la transformation θ . Nous supposons également dans cette partie l'indépendance statistique de l'a priori de chaque labélisation locale : $P(L) = \prod_{i \in \mathcal{E}} P(L_i)$. De cette façon, l'algorithme EM dans le cadre du recalage global proposé s'écrit :

$$Q_g(\theta_g | \theta_g^{(n)}) = \log [P(\theta_g)] + \sum_{i \in \mathcal{E}} \sum_{L_i=l} P(L_i = l | D_i; \theta_g^{(n)}) \log [P(D_i | L_i = l; \theta_g) P(L_i = l)] \quad (7.6)$$

Ici, l'étape de calcul d'espérance (E) consiste en l'évaluation des probabilités a posteriori locales $P(L_i=l | D_i; \theta_g^{(n)})$ pour chaque segment sulcal et pour chaque label, connaissant la valeur courante des paramètres de normalisation $\theta_g^{(n)}$. Dans le cas étudié, il s'agit donc d'appliquer le schéma décrit en section 4.2.3. Ces probabilités a posteriori mesurent le degré de confiance du modèle en une

labélisation donnée. Ainsi, elles jouent le rôle de poids pendant la phase de maximisation (M). Puisque les modèles sulcaux sont plutôt localisés spatialement, sur chaque segment sulcal la plupart des probabilités a posteriori sont nulles. Seulement quelques labels dirigent réellement l'étape de recalage. Ces poids agissent comme des liens (ou des ressorts) entre les segments et les modèles sulcaux dont la force est proportionnelle à la valeur.

Les lecteurs familiers avec les modèles de mixture auront sûrement reconnu en ce schéma, une optimisation relativement standard de paramètres. Ici, les modèles sulcaux (par exemple les modèles SPAMs) sont mélangés ensemble. L'unique subtilité étant que les paramètres du modèles sont fixes et définis à un recalage affine près.

En partant à nouveau de l'équation 7.5, si l'on considère que la transformation θ_s est composée d'un ensemble de transformations locales indépendantes $\theta_{s,l}$ (où l'indice s indique que chaque transformation est dédiée à un sillon particulier), une par label l (dans notre cas, une transformation rigide par label est considérée), l'algorithme EM s'écrit :

$$Q_s \left(\theta_s \middle| \theta_s^{(n)} \right) = \log [P(\theta_{s,l})] + \sum_{l \in \mathcal{L}} \sum_{i \in \mathcal{E}} P \left(L_i = l \middle| D_i; \theta_{s,l}^{(n)} \right) \log [P(D_i | L_i = l; \theta_{s,l}) P(L_i = l)] \quad (7.7)$$

avec $P(\theta_s) = \prod_{l \in \mathcal{L}} P(\theta_{s,l})$.

Cette fois, chacune des évaluations consacrées à un label particulier peut être décorrélée de telle façon que leur optimisation peut être réalisée séparément :

$$\theta_s^* = \left\{ \arg \max_{\theta_{s,l}} \sum_{i \in \mathcal{E}} P \left(L_i = l \middle| D_i \theta_{s,l}^{(n)} \right) \log [P(D_i | L_i \theta_{s,l}) P(\theta_{s,l})] \right\}_{l \in \mathcal{L}} \quad (7.8)$$

Puisque les recalages locaux par label sont supposés indépendants alors qu'en réalité ils ne le sont pas, les détails du processus d'optimisation peuvent être contre-intuitifs. Le processus part d'une initialisation que l'on peut supposer commune pour tous les labels pour simplifier (les paramètres $\theta_{s,l}^{(0)}$ sont égaux et indépendant de l). Une nouvelle fois, les probabilités a posteriori fournissent des poids pour diriger l'étape de recalage, mais cette fois, il en résulte une transformation différente pour chaque optimisation (une par label). Virtuellement, nous obtenons autant de cerveaux transformés (associés à leur transformation affine respective) que de labels puisque les labels restent inconnus. Dans l'étape E à suivre, la vraisemblance d'un label donné est obtenue en mélangeant chaque cerveau transformé avec le modèle sulcal correspondant. Les probabilités a posteriori sont alors obtenues après normalisation de la façon suivante :

$$P(L_i = l | D_i; \theta_s^{(n)}) = \frac{P \left(D_i | L_i = l; \theta_{s,l}^{(n)} \right) P(L_i = l)}{\sum_{l_i \in \mathcal{L}} P \left(D_i | L_i = l_i; \theta_{s,l_i}^{(n)} \right) P(L_i = l_i)} \quad (7.9)$$

puisque $P \left(D_i | L_i = l_i; \{\theta_{s,l}^{(n)}\}_{l \in \mathcal{L}} \right) = P \left(D_i | L_i = l_i; \theta_{s,l_i}^{(n)} \right)$. C'est à dire que la vraisemblance d'un segment sulcal i d'avoir un label l dépend uniquement de la transformation rattachée à ce label.

Que ce soit pour l'approche globale ou locale, une fois qu'une valeur optimale θ^* a été atteinte pour les paramètres de recalage, une labélisation est inférée comme précédemment (section 4.2.3) mais cette fois après avoir appliqué la transformation induite par les paramètres θ^* :

$$l^* = \{l_i^*\}_{i \in \mathcal{E}} = \left\{ \arg \max_{l_i} P(D_i | L_i = l_i; \Theta = \theta^*) P(L_i = l_i | \Theta = \theta^*) \right\}_{i \in \mathcal{E}} \quad (7.10)$$

7.4 Recalage et estimation de modèle joints

La section précédente indique comment corriger les imprécisions de la localisation spatiale de structures anatomiques fournies induites par une première étape de normalisation, alors même que

ces structures sont identifiées. Des imprécisions identiques interviennent lors de l'évaluation du modèle, c'est pourquoi des améliorations similaires sont à espérer en les corrigeant. Maintenant, nous allons reconsidérer l'estimation du modèle sulcal \mathcal{M}^2 (l'équation 4.5 décrit le cas général de l'inférence des paramètres du modèle lorsque l'espace normalisé est connu) en le raffinant par l'optimisation du choix de l'espace normalisé commun. À partir d'une base d'apprentissage \mathcal{A} (dans ce travail, nous avons utilisé la base de données décrite en annexe A), cet espace est décrit par un ensemble de paramètres de recalage inconnus θ_a (qui déplace le sujet a de son propre espace à l'espace commun). Dans ce contexte, pour chaque sujet d'apprentissage a , les labels $L_{a,i}$ de ses segments sulcaux i sont connus. Ainsi, la formulation Bayésienne idéale de ce problème est de trouver le meilleur jeu de paramètres m^* et les paramètres de recalage θ_a pour chaque sujet d'apprentissage à partir des seules données anatomiques D_a :

$$m^*, \{\theta_a^*\}_{a \in \mathcal{A}} = \arg \max_{m, \{\theta_a\}_{a \in \mathcal{A}}} P(\mathcal{M} = m \{\Theta_a = \theta_a\}_{a \in \mathcal{A}} | \{D_a L_a\}_{a \in \mathcal{A}}) \quad (7.11)$$

Une approximation utilisée fréquemment dans le cadre de la création de *template* anatomiques consiste à partir d'une initialisation bien choisie et ensuite alterner deux étapes : la création du *template* à partir de sujets normalisés à partir d'une version précédente du *template* et ensuite recalculer chaque sujet sur ce *template* et ainsi de suite. Cette idée peut facilement s'adapter à notre situation. À partir d'un espace de référence (obtenu par une première normalisation par exemple, quelques méthodes sont proposées en section 7.1), représenté par un ensemble de transformations $\{\theta_a^{(0)}\}_{a \in \mathcal{A}}$ (une par sujet a), une première estimation du modèle \mathcal{M} est réalisée. Ensuite, chaque sujet est «recalé» sur le modèle au sens de notre formulation Bayésienne (voir ci-dessous) qui détermine une version raffinée de l'espace commun précédent sur lequel le modèle est réestimé à nouveau et ainsi de suite. Cette approche suggère de séparer et d'alterner l'optimisation du modèle \mathcal{M} et des paramètres de recalage $\{\theta_a\}_{a \in \mathcal{A}}$:

$$\begin{cases} m^{(n)} = \arg \max_m P(\mathcal{M} = m | \{D_a L_a\}_{a \in \mathcal{A}}; \{\Theta_a = \theta_a^{(n)}\}_{a \in \mathcal{A}}) \\ \{\theta_a^{(n+1)}\}_{a \in \mathcal{A}} = \arg \max_{\{\theta_a\}_{a \in \mathcal{A}}} P(\{\Theta_a = \theta_a\}_{a \in \mathcal{A}} | \{D_a L_a\}_{a \in \mathcal{A}}; \mathcal{M} = m^{(n)}) \end{cases} \quad (7.12)$$

Le modèle \mathcal{M} de localisation spatiale considéré est constitué d'un sous-modèle \mathcal{M}_l par label l qui peuvent être estimés séparément. Ainsi, pour un sous-modèle \mathcal{M}_l donné, seuls les segments sulcaux liés à ce même label l sont considérés (que l'on notera : $L_{a,i} = l$) :

$$\begin{cases} m_l^{(n)} = \arg \max_{m_l} \prod_{\substack{a \in \mathcal{A} \\ i \in \mathcal{E}_{a,l}}} P(D_{a,i} | L_{a,i} = l \Theta_a = \theta_a^{(n)}; \mathcal{M} = m) P(L_{a,i} = l | \mathcal{M} = m) \\ \theta_a^{(n+1)} = \arg \max_{\theta_a} P(D_a | L_a \Theta_a = \theta_a; \mathcal{M} = m^{(n)};) P(\Theta_a = \theta_a | \mathcal{M} = m^{(n)}) \end{cases} \quad (7.13)$$

où $\mathcal{E}_{a,l}$ correspond à l'ensemble des segments sulcaux du sujet a restreint à ceux dont le véritable label est l . Nous avons omis ici l'a priori $P(\mathcal{M} = m)$, puisque seul le modèle SPAM en utilise un et celui-ci consiste uniquement à fixer un paramètre interne (largeur du noyau utilisé, section 4.2.3). Une fois encore, nous considérons que l'a priori sur la labélisation $P(L_{a,i} = l | \mathcal{M} = m)$ est indépendant des paramètres de recalage, de telle façon qu'il peut être calculé une fois pour toutes (section 4.3).

Maintenant, considérons l'optimisation d'un sujet donné a (les références à l'indice a , θ_a et $m^{(n)}$ sont omises ici puisqu'il n'y a pas d'ambiguïté possible) et prenons le logarithme de la deuxième

²L'estimation du modèle SPAM depuis un espace normalisé fixé est décrit en section 4.2.3.

ligne de l'équation 7.13 :

$$\begin{aligned} \log [P(D|L\Theta; \mathcal{M}) P(\Theta|\mathcal{M})] &= \sum_{i \in \mathcal{E}} \log [P(D_i|L_i\Theta; \mathcal{M}) P(\Theta|\mathcal{M})] \\ &= \sum_{i \in \mathcal{E}} \sum_{L_i=l} \delta_{l,l_i} \log [P(D_i|L_i=l\Theta; \mathcal{M}) P(\Theta|\mathcal{M})] \end{aligned} \quad (7.14)$$

où δ_{l,l_i} correspond au symbole de Kronecker qui vaut 0 si $l \neq l_i$ (les labels sont différents) et 1 si $l = l_i$ (les labels sont les mêmes). Ici, l correspond à l'itérateur de la somme, alors que l_i est le véritable label (n'oublions pas que les labels sont connus sur la base d'apprentissage). Après réécriture de l'expression optimisée, nous pouvons reconnaître l'étape M de l'équation 7.5 pour la labélisation et le recalage joint. En d'autres termes, avec des poids bien choisis (probabilités a posteriori mesurant l'incertitude du modèle ou la traduction des certitudes d'un étiquetage manuel) le même algorithme peut être utilisé dans les deux situations.

Finalement, ce schéma itératif donne un maximum local de la probabilité exprimée dans l'équation 7.11, proche de l'optimum si l'initialisation est suffisamment bonne, ce qui semble être le cas dans notre situation.

7.5 A priori de recalage

L'approche jointe proposée pour la labélisation des plissements corticaux (section 7.3) et le raffinement du modèle correspondant (section 7.4) incluent des étapes de recalage. Les deux cas sont contraints par des a priori sur les paramètres de recalage θ qui limitent l'étendue des transformations possibles. Dans ce manuscrit, nous nous sommes restreint à une transformation affine globale et une collection de transformations affines locales (une par label). Le choix de l'a priori doit correspondre à la nature particulière de ces mesures. Comme nous l'avons indiqué plus tôt, de telles transformations sont caractérisées par une matrice inversible A et un vecteur de translation t . Les difficultés sont assez différentes entre les deux techniques de normalisation proposées, nous allons donc maintenant les détailler.

7.5.1 Approche globale

Premièrement, à propos de l'approche globale, seule une transformation affine contrôle le recalage, un nombre assez limité de paramètres peut donc être optimisé. Commençons par le cas rigide pour faciliter l'explication. Le processus itératif (EM) part d'une solution initiale proche de la solution recherchée, c'est à dire en plutôt bonne adéquation avec le modèle (au sens des vraisemblances manipulées), par conséquent les grands déplacements sont déjà pénalisés, ainsi dans ce cas aucun a priori n'est réellement nécessaire. Aucune dérive du processus n'a été constaté en pratique; c'est à dire que le modèle se raffine petit à petit, centré sur sa position initiale.

Maintenant l'argument ne tient plus pour le cas des transformations affines puisque l'adéquation entre un sujet et le modèle est mesurée à un facteur d'échelle près. Pour éviter des phénomènes de divergence du processus induits par une tendance artificielle du modèle à se rétrécir sur lui-même d'une itération à l'autre alors qu'il se raffine, il est nécessaire de contraindre ces composantes autour de l'échelle d'origine des données. En effet, la partie linéaire d'une transformation affine s'écrit : $A_g = U_g D_g V_g^t$ où U_g et V_g correspondent à deux matrices de rotation et D_g à une matrice diagonale contenant les facteurs d'échelle. Nous n'avons pas vraiment d'a priori sur le choix des directions (définies par V_g^t) sur lesquelles appliquer les facteurs d'échelle. Ainsi, seule la matrice D_g peut être contrainte de façon raisonnable. Puisque les facteurs d'échelle sont des réels strictement positifs, nous avons choisi de les contrôler chacun à l'aide d'une distribution gamma $\Gamma(\cdot, \cdot)$ définie par son

paramètre de forme k et son paramètre d'échelle s , en fixant la moyenne de la distribution avec la contrainte : $ks = 1$, c'est à dire que l'échelle arbitraire retenue est l'échelle originale des données :

$$P(\theta_g) = P(D_g) = \prod_{i \in \{1,2,3\}} P(D_{g,i}) \quad \text{où chaque } D_{g,i} \sim \Gamma\left(k, \frac{1}{k}\right)$$

$$\text{ainsi, } \log[P(\theta_g)] = (k-1) \sum_{i \in \{1,2,3\}} \log[D_{g,i}] - kD_{g,i} + C(k) \quad (7.15)$$

où $D_{g,i}$ correspond au paramètre d'échelle de la i ème direction et $C(k)$ une constante ne dépendant que du paramètre k (qui peut être ainsi ignorée pendant la phase d'optimisation puisque k est fixé). Dans nos expériences, nous avons utilisé $k = 1600$ ce qui correspond à une distribution proche d'une Gaussienne centrée autour de 1 avec un écart-type égal à $\frac{1}{\sqrt{k}}$, c'est à dire 0.025 ici.

7.5.2 Approche locale

Dans le cas de l'approche locale, on considère une transformation rigide par label sulcal, au lieu de transformations affines afin d'accélérer et stabiliser l'optimisation. Si le besoin s'en faisait ressentir, il serait possible d'introduire des facteurs d'échelle comme nous l'avons présenté plus tôt dans le cas de l'approche globale. Précédemment, nous avons montré que les modèles sulcaux agissaient comme une contrainte globale sur le processus de recalage. Localement, les choses sont différentes. En effet, nous étudions près de 60 labels, ainsi la plupart des sillons les plus petits ont des formes relativement similaires puisqu'il s'agit de plissements élémentaires (à l'échelle des racines sulcales). Sans aucune contrainte l'approche locale peut être amenée à mettre en correspondance les structures anatomiques et les modèles qui se ressemblent le plus en termes de forme alors qu'ils sont spatialement très éloignés ou plutôt différents en termes d'orientation. La mise au point d'a priori appropriés s'appuie sur la nature spécifique des transformations rigides. Dans notre cas d'étude, nous nous sommes restreint à des a priori indépendants d'un label à l'autre, c'est à dire vérifiant : $P(\theta) = \prod_{l \in \mathcal{L}} P(\theta_l)$.

Les a priori sur les paramètres de recalage apparaissent pendant la phase d'estimation du modèle à partir d'une base d'apprentissage et la labélisation d'un sujet test.

Pour le premier cas, l'a priori doit être fixé a priori puisque nous n'avons aucune autre information à notre disposition. Par conséquent, les a priori doivent être assez peu informatifs afin de limiter uniquement les situations les plus improbables (par exemple, des translations de plus de 10 cm, des rotations d'un angle de π) mais tolérer une large gamme de transformations qui peuvent apparaître au sein des données d'apprentissage.

Dans le deuxième cas, les a priori de recalage peuvent soit être fixés a priori comme dans le cas précédent, ou être appris de façon empirique. Nous avons choisi cette deuxième solution puisque la variabilité sulcale est plutôt inhomogène et peut dépendre du choix de la base de données (ce qui peut être utile dans le cas d'étude de groupe si les sujets présentent des spécificités anatomiques communes). Ainsi, nous utilisons une base d'apprentissage pour estimer les paramètres contrôlant la forme des a priori. Ensuite, ils joueront le rôle de contraintes dans le processus de labélisation de données test. Finalement, de tels a priori dépendent des données d'apprentissage et devraient s'écrire $P(\theta_l | \{\theta_{a,l}\}_{a \in \mathcal{A}})$ avec $\theta_{a,l}$ les paramètres régissant la transformation concernant le sillon l du sujet a . Maintenant que ce point a été précisé, cette dépendance est rendue implicite dans les notations afin de les simplifier. Pendant la phase d'estimation du modèle, chaque sujet d'apprentissage a été co-recalé avec les autres vers un espace commun. Ces transformations sont utilisées pour définir l'étendue des transformations autorisées pour contraindre la labélisation de futurs sujets tests.

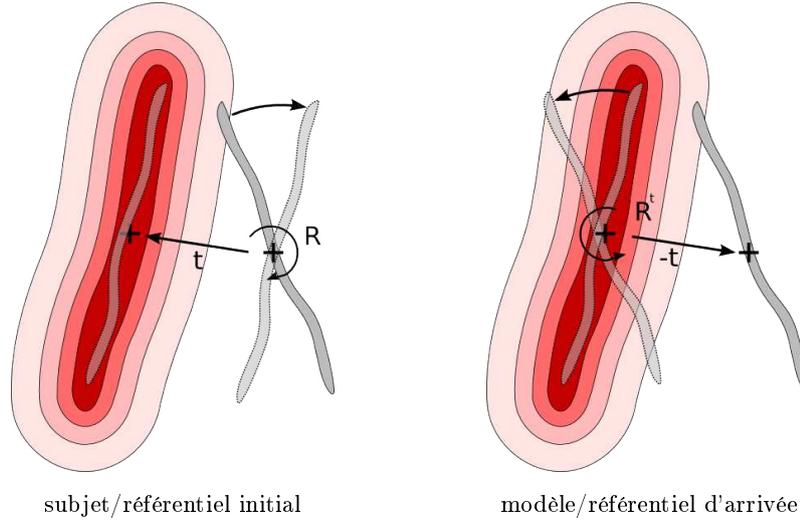


FIG. 7.1 – Deux référentiels optimaux pour les sillons considérés avec leur translation minimale respective. Ici, les formes suivant un dégradé du rouge au blanc suggèrent un modèle SPAM artificiel en 2D, alors que les formes grisées représentent un sillon à différentes étapes du processus de recalage rigide. **Gauche** : le sillon est d’abord tourné par la rotation R autour de son centre et ensuite déplacé du vecteur t . **Droit** : il s’agit de la transformation inverse depuis le référentiel précédent déplacé du vecteur t (qui se situe plus ou moins au niveau du centre du modèle) ; ainsi, le sillon considéré se trouve initialement au niveau du modèle, puis il est tourné autour de son centre (ce centre est défini en section 7.5.3.1) d’une rotation R^t et déplacé du vecteur $-t$.

7.5.3 Modélisation

Que ce soit à l’échelle globale ou locale, et qu’il s’agisse d’une transformation affine ou rigide, il est nécessaire d’étudier plus en détail la formulation de cette transformation afin de proposer des a priori adaptés. Nous étudierons ici le cas rigide (le cas affine ne comporte pas de difficultés supplémentaires) en notant R la matrice de rotation et t le vecteur de translation. La transformation du point x en un point y s’écrit : $y = R \cdot (x - g) + (t + R \cdot g - g) + g = R \cdot x_g + t_g + g$ où g correspond à un point de référence (le centre de la rotation). Ainsi, selon la position du point g , la variabilité de la composante de translation t_g peut être très différente. Le minimum de cette variabilité est atteint lorsque le référentiel est centré dans l’espace du sujet (figure 7.1, gauche) ou dans l’espace du modèle (figure 7.1, droite).

Nous allons maintenant étudier comment évaluer la position optimale de g afin de réduire la variabilité des translations. Ensuite nous détaillerons le choix des a priori pour chacune des composantes.

7.5.3.1 Optimisation des centres de rotation

Pour un sujet donné et une transformation rigide associée (composée d’une matrice R_a et d’un vecteur de translation t_a) globale ou dédiée à un sillon l particulier, la transformation d’un point x en un point y exprimé dans le référentiel g s’écrit : $y = R_a \cdot (x - g) + (R_a \cdot g + t_a - g) + g$. Ainsi, la transformation inverse s’écrit : $x = R_a^t \cdot (y - g) + (R_a^t \cdot (g - t_a) - g) + g$ où $(R_a^t \cdot (g - t_a) - g)$ correspond à la translation exprimée dans le référentiel g . La valeur idéale pour g est celle qui réduit la variabilité de ces translations. Nous suggérons donc de minimiser l’énergie suivante :

$$E = \sum_{a \in \mathcal{A}} \|R_a^t \cdot (g - t_a) - g\|^2 \quad (7.16)$$

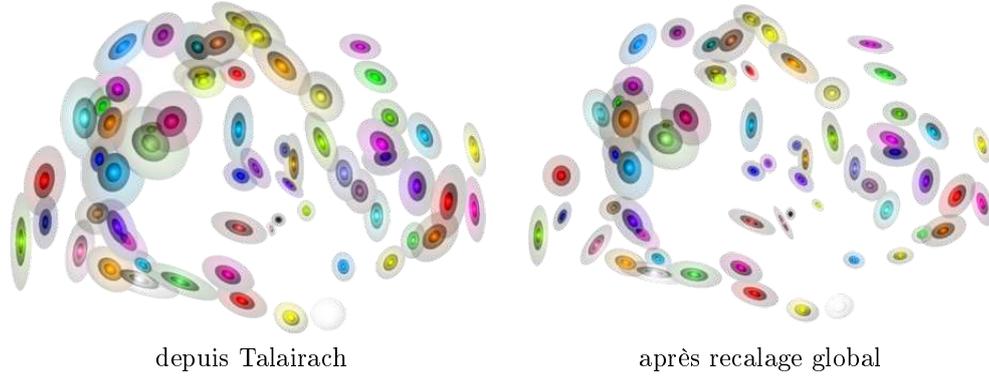


FIG. 7.2 – Représentation visuelle d’une collection d’a priori Gaussien pour les translations. Chaque label l considéré est centré sur son référentiel g_l . Ce graphique permet d’apprécier la différence d’estimation depuis l’espace de Talairach ou depuis un référentiel commun optimisé par recalage rigide.

Tout d’abord, si toutes les matrices vérifient : $R_a = Id$, E ne dépend plus de g , ainsi tous les référentiels g se valent. Dans le cas général, il nous faut étudier les dérivées de cette énergie :

$$\frac{\partial E}{\partial g} = 4|\mathcal{A}|g - 2 \sum_{a \in \mathcal{A}} [(R_a + R_a^t) \cdot g + (R_a - Id)^t \cdot t_a] = \vec{0}$$

$$\text{Ainsi, } g^* = \left[\sum_{a \in \mathcal{A}} (S_a + S_a^t) \right]^{-1} \left[\sum_{a \in \mathcal{A}} S_a^t \cdot t_a \right] \quad (7.17)$$

où $S_a = Id - R_a$ et $|\mathcal{A}|$ est le nombre de sujets.

En fait g^* est aussi la solution de l’énergie suivante :

$$E = \sum_{a \in \mathcal{A}} \|R_a \cdot g + t_a - g\|^2 \quad (7.18)$$

où chaque composante de la somme est le carré de la norme des translations exprimées depuis l’espace de référence à l’espace commun, c’est à dire la direction contraire de celle utilisée dans l’énergie précédente. Ainsi, les deux directions donnent le même point de référence optimal g . Mais à l’optimum, la première énergie proposée est plus faible que la seconde, indiquant une variabilité plus faible des translations et par suite des a priori plus piqués.

7.5.3.2 Modélisation des a priori

Une fois les référentiels g_l déterminés pour chaque label l , on peut discuter du choix des a priori. À partir de tels référentiels, l’a priori sur les translations $P(t_l)$ peut être modélisé simplement par une Gaussienne 3D anisotrope :

$$t_l \sim \mathcal{N}(\mu_l, \Sigma_l) \quad (7.19)$$

où μ_l correspond à la moyenne de la Gaussienne et Σ_l sa matrice de covariance. Les paramètres sont fixés a priori pendant la phase d’estimation du modèle : $\mu_l = \vec{0}$ et $\Sigma_l = (10 * I)^2$, c’est à dire, une déviation standard d’un centimètre. Ensuite, pendant la phase de labélisation, les Gaussiennes sont estimées à partir des translations mesurées sur les données d’apprentissage. Les spécificités anatomiques de ces a priori peuvent être appréciées en figure 7.2.

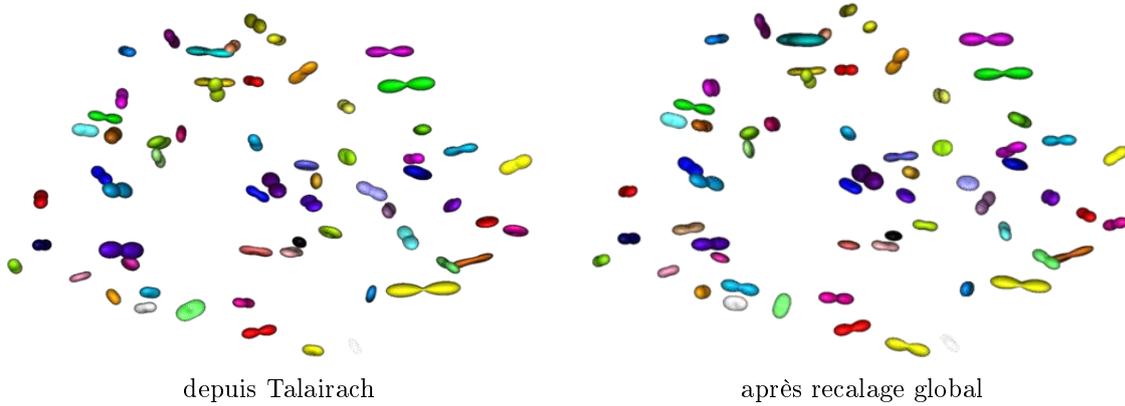


FIG. 7.3 – Représentation visuelle des modèles de Bingham estimés sur les axes de rotation. Le reste de la légende est la même que celle de la figure 7.2.

Les matrices de rotation 3D constituent un sous-espace courbe de l’espace vectoriel des matrices 3×3 qui n’est donc pas lui-même un espace vectoriel. La distribution Gaussienne classique ne peut pas s’appliquer directement aux éléments de la matrice. La généralisation idéale serait sans doute d’exprimer un modèle Gaussien directement sur la variété Riemannienne correspondante [Pennecc 1997, Boisvert 2008] en travaillant localement sur l’espace tangent à la variété au travers des cartes logarithmiques et exponentielles. Une autre solution, sans doute moins gourmande en temps de calcul, consiste à utiliser une distribution Matrix Von Mises que nous avons étudiée tardivement (annexe B.3.1.6), mais cette solution ne fait qu’approcher la distribution idéale spécifiée plus tôt.

Pour commencer, nous proposons une approximation ad hoc, qui permet de se faire une idée des spécificités anatomiques locales en comparant visuellement l’ensemble des a priori dédiés à un label particulier. L’approche proposée simplifie également l’estimation et l’évaluation des a priori en exploitant une paramétrisation standard, nommée «vecteur-rotation» où chaque matrice de rotation est représentée par un vecteur 3D w fortement lié aux considérations Riemanniennes évoquées plus tôt. Dans ce contexte, une matrice de rotation R s’écrit :

$$\begin{aligned} R &= \exp(J(w)) \\ &= I + \frac{\sin(\alpha)}{\alpha} J(w) + \frac{(1 - \cos(\alpha))}{\alpha^2} J(w)^2 \end{aligned} \quad (7.20)$$

où $\alpha = \|w\|$. La deuxième ligne de cette équation s’appelle la formule de Rodrigues et permet de passer facilement de la représentation vectorielle à la représentation matricielle de la rotation et évite ainsi l’évaluation coûteuse de l’exponentielle de matrice $\exp(J(w))$. Ici, $\exp(\cdot) = \text{Exp}_I(\cdot)$, c’est à dire la carte exponentielle au point I (la matrice identité 3×3). Ici, la matrice $J(w)$ vérifie : $J(w)^t = -J(w)$ et s’écrit :

$$J(w) = \begin{pmatrix} 0 & -w_z & w_y \\ w_z & 0 & -w_x \\ -w_y & w_x & 0 \end{pmatrix} \quad \text{avec } w = (w_x, w_y, w_z) \quad (7.21)$$

et est reliée au produit vectoriel puisque $J(w) \cdot v = w \times v$. Finalement, l’ensemble des rotations 3D est une variété tri-dimensionnelle (puisque’on peut la paramétrer par un vecteur de dimension 3) plongée dans l’espace à 9 dimensions des matrices 3×3 . Cette paramétrisation est simple à comprendre puisque la direction du vecteur w , valant $\frac{w}{\|w\|}$, représente l’axe de rotation alors que sa

norme $\alpha = \|w\|$ correspond à l'angle de la rotation en radians autour de cet axe. Cette représentation est unique aussi longtemps que α est différent de π ou $-\pi$ (le *cut locus* de cette paramétrisation). Par conséquent, pour un label donné l , l'a priori sur la rotation R_l peut être séparé en 2 parties :

$$P(R_l) = P(w_l) \approx P\left(\frac{w_l}{\alpha_l}; \alpha_l\right) \approx P\left(\frac{w_l}{\alpha_l}\right) P(\alpha_l) \quad (7.22)$$

où $\alpha_l = \|w_l\|$ et w_l correspond à la représentation vecteur-rotation de R_l .

Dans le cas particulier de la transformation identité, l'angle α_l est nul et la direction associée est indéterminée (que l'on notera par le symbole suivant : ?) . Dans ce cas, il est nécessaire d'intégrer la mesure de probabilité associée au modèle directionnel sur l'ensemble des directions possibles :

$$P(w = ?|A) = \mathbb{E}_{S(2)} [P(w|A)] \quad (7.23)$$

De façon similaire, pendant la phase d'estimation du modèle, il est nécessaire de prendre en compte ces situations particulières. Le cas de la distribution Bingham finalement retenue pour représenter la composante directionnelle est précisé en annexe B.3.1.5.

Dans la littérature, on trouve un certain nombre d'approximations de modèles Gaussiens à ce type particulier de données (voir annexe B.3.1 pour une revue détaillée). Puisque le calcul de l'angle α provient d'une norme, toutes les rotations s'expriment dans le sens direct autour du vecteur w . Pour un label donné, nous nous attendons à un ensemble de rotations proches les unes des autres ; il devrait donc en être de même pour les axes de rotation. Par contre, les directions sont regroupées seulement à un signe près puisque des rotations de signes opposés sont représentées (au sens des «vecteurs-rotations») par des vecteurs opposés puisque α est toujours positif. Ainsi, seule la partie axiale des rotations a été modélisée. On confondra donc w et $-w$ d'une part et l'angle de rotation est modélisé à un signe près d'autre part. Finalement, la variabilité de ces axes est modélisée par une distribution Bingham [Bingham 1974] (section B.3.1.5) et celle des angles par une distribution Von Mises [Mardia 2000] (section B.3.1.1). Des représentations visuelles de ces distributions se trouvent en figure B.4. Les distributions Bingham, caractérisant l'orientation privilégiée de l'axe de rotation de chacun des labels peuvent être représentées visuellement pour mieux apprécier les spécificités anatomiques locales (voir figure 7.3).

7.6 Optimisation des paramètres de recalage

Dans cette section, nous discuterons de la façon dont a été menée l'optimisation des paramètres dans le cas concret du modèle SPAM dont l'évaluation et la validité sont étudiées dans le chapitre suivant. Nous présenterons ensuite une approche uniquement adaptée à une modélisation paramétrique de la variabilité sulcale que nous détaillerons sur quelques modèles présentés en section 4.2.2, qui donne lieu à une optimisation très efficace.

7.6.1 Cas non paramétrique : modèle SPAM

Dans le cadre des techniques de recalage étudiées dans cette partie, il est nécessaire d'appliquer des transformations rigides ou affines aux segments sulcaux des sujets étudiés. Par définition, une telle transformation, composée des paramètres A (partie linéaire) et t (le vecteur de translation), s'intègre de la façon suivante dans le modèle de vraisemblance :

$$P(D_i = \mathcal{X}|L = l; \Theta = At) \triangleq P_{spam}(\{A \cdot x + t\}_{x \in \mathcal{X}}|L = l) \quad (7.24)$$

où les coordonnées de \mathcal{X} sont exprimées dans l'espace du sujet, ainsi $\{Ax + t\}_{x \in \mathcal{X}}$ s'exprime dans l'espace du modèle SPAM. Ici, θ est un meta-paramètre regroupant tous les paramètres contrôlant le recalage (liés à un label l particulier dans le cas de l'approche de normalisation locale).

Le modèle SPAM est non-paramétrique puisqu'il est actuellement représenté par un volume 3D de probabilités locales issues d'un estimateur à noyaux. Ainsi, un schéma d'optimisation basé sur une descente de gradient risque de s'avérer limité à cause des approximations liées à la discrétisation du gradient. Par conséquent, nous avons reporté notre attention sur une méthode réputée efficace mais aveugle à la formulation de la fonctionnelle à minimiser. C'est à dire, une méthode qui nécessite uniquement l'évaluation à la demande de la fonction en certains points de test. À cet effet, nous avons choisi d'utiliser la méthode de Powell [Powell 1964] afin de couvrir l'étendue de l'espace des paramètres de recalage. Le paramètre angulaire des rotations est contraint entre $-\pi$ et π puisque le paramètre est cyclique. Qu'il s'agisse d'une transformation affine ou rigide, les rotations sont représentées au travers de leur paramétrisation «vecteur-rotation». Ainsi la méthode de Powell suggère des propositions de transformations en parcourant l'espace des paramètres le long d'un ensemble d'axes (chacun combinaison linéaire des axes canoniques) qui sont raffinés itérativement au cours du processus d'optimisation en fonction des améliorations.

7.6.2 Cas paramétrique : approche Riemannienne

En partie I, nous avons présenté plusieurs modèles locaux permettant de représenter la variabilité sulcale de façon assez grossière, mais qui combinés au sein d'un modèle de mélange (par exemple, mixture de Gaussiennes pour représenter les modèles SPAMS en section B.4.1.2) révèlent tout leur intérêt. L'étude des modèles de mélange est reportée à des travaux ultérieurs mais ne devrait pas poser de difficultés. Nous envisageons ici, les modèles suivants pour représenter les vraisemblances $P(D_i|L = l\theta)$: modèle Gaussien (sur les centres de gravité ou l'ensemble des voxels des segments sulcaux), Bingham (données axiales : axe principal dans la longueur des sillons par exemple), Kent (données directionnelles : vecteur allant du fond du sillon jusqu'à l'enveloppe cérébrale).

Dans le cas paramétrique, nous disposons d'une formulation analytique des modèles sulcaux. Dans ce cadre, une optimisation plus efficace que l'algorithme de Powell proposé dans le cas des SPAMS est envisageable pour les paramètres de recalage θ . Comme nous l'avons indiqué plus tôt (section 7.5.3.2), les transformations affines ou rigides forment un espace courbe, l'optimisation de telles quantités ne peut être réalisée par des méthodes d'optimisation classiques. Par exemple, une matrice de rotation R vérifie les propriétés suivantes : $\det(R) = 1$ et $R^t R = R R^t = I$ où I correspond à la matrice identité. La moyenne classique matricielle de deux rotations R_1 et R_2 : $\frac{R_1 + R_2}{2}$ ne vérifie pas forcément les propriétés énoncées ci-dessus. La géométrie différentielle fournit un ensemble d'outils permettant de définir une moyenne qui reste sur la variété considérée tout en suivant l'idée intuitive qu'on se fait d'une telle moyenne³. En réalité ce calcul revient à minimiser une énergie (équation B.15) en restant sur la variété mathématique considérée. La figure 7.4 donne une idée intuitive du déroulement d'un algorithme de descente de gradient sur une telle variété. Dans le cadre de la labélisation jointe à la normalisation, l'optimisation des paramètres de recalage θ qui nous intéresse s'exprime également comme la minimisation d'une énergie. Nous suggérons dans la suite une minimisation le long de la variété considérée suivant l'approche de Newton-Raphson lorsque la matrice Hessienne est positive et une descente de gradient dans le cas contraire. Une autre stratégie, s'appuyant uniquement sur les expressions des dérivées premières aurait également permis d'éviter cette difficulté par l'utilisation d'une méthode de Gauss-Newton [Pennec 1997].

7.6.2.1 Principe général

Nous nous restreindrons ici à l'étude de l'approche rigide globale, qui sous l'hypothèse d'indépendance des transformations locales entre les labels, est plus générale que l'approche locale

³Pour plus de détails sur le calcul d'une moyenne dans ce cadre, nous reportons le lecteur à l'annexe B.2.2 où un tel calcul est décrit pour des données directionnelles et axiales.

(équation 7.6 et 7.8). L'extension aux matrices affines suivrait le même principe en considérant la décomposition en valeur singulière décrite précédemment (section 7.2).

Chaque étape ($n + 1$) de maximisation de l'algorithme EM, revient à minimiser en $\theta = (R, t)$ l'expression E suivante :

$$\begin{aligned} E &= -\frac{1}{2} \log [P(\theta_g)] - \frac{1}{2} \sum_{i \in \mathcal{E}} \sum_{L_i=l} P(L_i = l | D_i; \theta_g^{(n)}) \log [P(D_i | L_i = l; \theta_g)] \\ &= -\frac{1}{2} \log [P(\theta_g)] - \frac{1}{2} \sum_{i \in \mathcal{E}} \sum_{L_i=l} p_{il} \log [P(D_i | L_i = l; \theta_g)] \end{aligned} \quad (7.25)$$

en notant $p_{il} = P(L_i = l | D_i; \theta_g^{(n)})$ afin d'alléger les notations.

La variété considérée ici est le groupe $SE(3)$ (Special Euclidien), c'est à dire l'espace des matrices 4×4 homogène : des matrices M composées d'une rotation $R \in SO(3)$ et d'un vecteur de translation $t \in \mathbb{R}^3$:

$$M = \begin{bmatrix} & & & \\ & R & & t \\ & & & \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (7.26)$$

Afin de contraindre l'optimisation envisagée à la variété, la méthode standard utilisée en géométrie

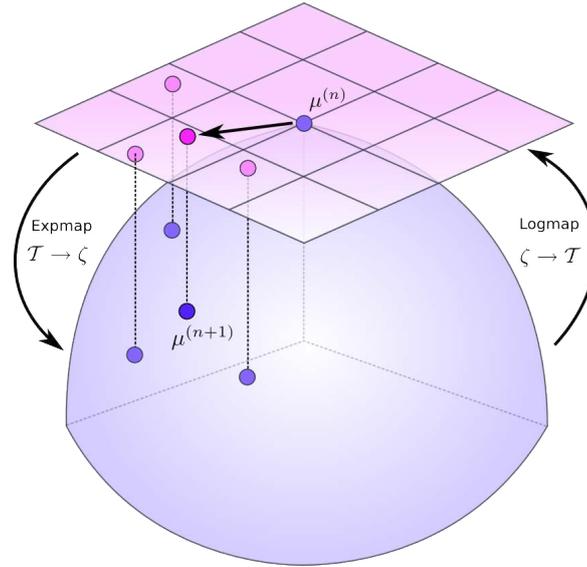


FIG. 7.4 – Exemple de minimisation d'énergie sur une variété. Ici, il s'agit de calculer la moyenne de 3 points sur une variété notée ζ et représentée ici en violet sous la forme d'une sphère. Le plan tangent au point $\mu^{(n)}$ est représenté en rose. Les points étudiés sont alors projetés grâce à la carte logarithmique sur le plan tangent ζ en rose où des opérations euclidiennes classiques peuvent être menées. La solution est mise à jour dans le plan tangent, la carte exponentielle permet alors d'avoir l'expression $\mu^{(n+1)}$ de cette solution sur la variété. À l'itération suivante, le plan tangent considéré sera pris au point $\mu^{(n+1)}$.

différentielle est d'approcher localement la variété par un espace euclidien appelé *espace tangent* (la figure 7.4 illustre ce concept sur une variété prenant la forme d'une sphère). On notera \mathcal{T}_x l'espace

tangent au point x et ζ l'espace courbe (la variété). L'astuce est donc de réaliser les calculs dans l'espace tangent et de les reporter ensuite sur la variété. Cette procédure nécessite des fonctions permettant de passer d'un espace à l'autre et réciproquement. Ce rôle est tenu par la carte logarithmique ($\zeta \rightarrow \mathcal{T}$) et exponentielle ($\mathcal{T} \rightarrow \zeta$) dont les expressions sont connues dans le cas de $SE(3)$.

Le passage entre le point $M_\zeta = \begin{pmatrix} R & t \\ 0 & 1 \end{pmatrix}$ de la variété et le point $M_\mathcal{T} = \begin{pmatrix} J(w) & v \\ 0 & 1 \end{pmatrix}$ de l'espace tangent au point M_0 , s'écrit [Murray 1994, Fletcher 2003] :

$$\begin{cases} \text{Log}_{M_0}(M_\zeta) = \log(M_0^{-1}M_\zeta) & \text{où} & \log(M_\zeta) = \log \begin{pmatrix} R & t \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} J(w) & A^{-1}t \\ 0 & 0 \end{pmatrix} \\ \text{Exp}_{M_0}(M_\mathcal{T}) = M_0 \exp(M_\mathcal{T}) & \text{où} & \exp(M_\mathcal{T}) = \exp \begin{pmatrix} J(w) & v \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} R & Av \\ 0 & 1 \end{pmatrix} \end{cases} \quad (7.27)$$

avec $A = I + \frac{1-\cos(\alpha)}{\alpha^2} J(w) + \frac{\alpha-\sin(\alpha)}{\alpha^3} J(w)^2$ et $A^{-1} = I - \frac{1}{2} J(w) + \frac{2\sin(\alpha)-\alpha(1+\cos(\alpha))}{2\alpha^2 \sin(\alpha)} J(w)^2$. Ici, w correspond à la représentation de R dans son espace tangent et sont reliés par la formule de Rodrigues et vérifie : $R = \exp(J(w))$ (équation 7.20). Les notations $\text{Log}_{M_0}(\cdot)$ et $\text{Exp}_{M_0}(\cdot)$ correspondent aux cartes logarithmiques et exponentielles, $\log(\cdot)$ et $\exp(\cdot)$ aux fonctions logarithmes et exponentielles standards pour les matrices.

Avec ces éléments en main, on conçoit facilement l'utilisation d'algorithme de type descente de gradient pour minimiser l'énergie qui nous intéresse. Dans notre cas, nous avons utilisé la méthode de Newton-Raphson adaptée au contexte Riemannien [Taylor 1994]. La méthode se déroule comme suit. À partir d'une initialisation bien choisie des paramètres, la méthode réalise une optimisation itérative. Dans l'espace tangent du point d'étude courant, l'énergie est alors approchée par une quadratique, ce qui permet de raffiner la solution en avançant dans la direction du gradient dans l'espace tangent. Pour simplifier les formulations, dans l'espace tangent les données sont représentées sous forme vectorielle : $M_\mathcal{T}$ est représenté par le vecteur $m_\mathcal{T} = [w, v]$. À partir d'une solution initiale $M_\zeta^{(0)}$, le schéma d'optimisation est le suivant :

$$\begin{cases} \text{En } M_\zeta^{(n)}, \text{ évaluer le gradient } g \text{ et la matrice Hessienne } H \text{ en dérivant selon les paramètres } w \text{ et } v. \\ \Delta m_\mathcal{T} \leftarrow -H^{-1}g & \text{pas dans l'espace Tangent} \\ M_\zeta^{(n+1)} \leftarrow M_\zeta^{(n)} \exp(\Delta m_\mathcal{T}) & \text{retour à la variété} \end{cases} \quad (7.28)$$

Dans les sections à suivre, nous détaillerons les expressions des dérivées premières et secondes en w et v pour les différents modèles considérés. Les expressions de ces dérivées pour la matrice de rotation R y apparaissent à plusieurs reprise, nous en détaillons donc le calcul ici. Une matrice de rotation R , peut être vue comme une fonction $R(w)$ d'une paramétrisation d'un plan tangent placé en un point R_0 arbitraire : $R(w) = R_0 \exp(J(w))$. L'approximation euclidienne portée par le plan tangent est exacte au niveau de la matrice R_0 , c'est à dire en $w = 0$. Dans le processus d'optimisation, on fera donc coïncider le point de calcul du plan tangent à la valeur courante de la solution. La matrice courante R recherchée vérifient donc :

$$\begin{aligned} \left. \frac{\partial R}{\partial w_i} \right|_{w=0} &= R_0 \left. \frac{\partial \exp(J(w))}{\partial w_i} \right|_{w=0} = R_0 J(e_i) \\ \left. \frac{\partial^2 R}{\partial w_i \partial w_j} \right|_{w=0} &= R_0 \left. \frac{\partial^2 \exp(J(w))}{\partial w_i \partial w_j} \right|_{w=0} = \frac{1}{2} R_0 (J(e_i)J(e_j) + J(e_j)J(e_i)) \end{aligned} \quad (7.29)$$

en notant w_i et w_j des composantes du vecteur w , et e_i correspond au i ème vecteur canonique de \mathbb{R}^3 . Précédemment, nous avons vu que $t = Av$ et A dépend de w , ce qui donne :

$$\begin{aligned} \left. \frac{\partial t}{\partial w_i} \right|_{w=0} &= \left. \frac{\partial Av}{\partial w_i} \right|_{w=0} = 0 \\ \left. \frac{\partial^2 t}{\partial w_i \partial w_j} \right|_{w=0} &= \left. \frac{\partial^2 Av}{\partial w_i \partial w_j} \right|_{w=0} = \frac{1}{6} (J(e_i)J(e_j) + J(e_j)J(e_i))v \end{aligned} \quad (7.30)$$

7.6.2.2 Modèle Gaussien

Dans le cas du modèle Gaussien, à chaque label l est rattaché un tel modèle de moyenne y_l et de matrice de covariance inverse S_l . La vraisemblance générique s'écrit donc : $P(D_i = x_i | L_i = l; \theta_g) \propto \exp(-\|y_l - Rx_i - t\|_{S_l}^2)$ avec $\theta_g = (R, t)$ et x_i le centre de gravité du segment sulcal i (cette expression peut s'étendre facilement à plusieurs points par segment). Dans ce contexte, l'énergie E (équation 7.25), s'écrit (en omettant l'a priori) :

$$\begin{aligned} E &= \frac{1}{2} \sum_{i \in \mathcal{E}} \sum_{L_i=l} p_{il} \|y_l - Rx_i - t\|_{S_l}^2 \\ &= \frac{1}{2} \sum_l \left\| (Y_l - RX - T) P_l^{\frac{1}{2}} \right\|_{S_l}^2 \end{aligned} \quad (7.31)$$

en posant $X = [x_i; i \in \mathcal{E}]$, $P_l = \text{diag}[p_{il}; i \in \mathcal{E}]$, $V = v\mathbb{I}^t$, $T = AV = t\mathbb{I}^t$ et $Y_l = y_l\mathbb{I}^t$ avec \mathbb{I}^t un vecteur de taille $|\mathcal{E}|$ ne contenant que des 1.

Le problème d'optimisation ci-dessus correspond en fait à une généralisation de la méthode Procrustes [Johnson 1966, Schönemann 1966] dans le cas où il s'agit de trouver la transformation rigide qui apparie (au sens des moindres carrés) deux ensembles de points de même taille. Classiquement, l'hypothèse d'un bruit Gaussien isotrope similaire en chaque point est retenu. Nous présentons ici l'extension de cette méthode à un bruit Gaussien potentiellement anisotrope et dépendant des points considérés.

Voici maintenant la liste des calculs des dérivées premières pris en $w = 0$ et $v = 0$.

$$\begin{aligned} \left. \frac{\partial E}{\partial v} \right|_{\substack{v=0 \\ w=0}} &= \left. \frac{1}{2} \sum_l \frac{\partial}{\partial v} \left[\bar{p}_l \|Av\|_{S_l}^2 - 2 \text{Tr}(\mathbb{I}^t P_l (Y_l - RX)^t S_l Av) \right] \right|_{\substack{v=0 \\ w=0}} \\ &= \sum_l AS_l [RX \bar{p}_l - \bar{p}_l y_l] \Big|_{w=0} = \sum_l S_l [R_0 X \bar{p}_l - \bar{p}_l y_l] \end{aligned} \quad (7.32)$$

où $\bar{p}_l = \mathbb{I}^t P_l \mathbb{I}$ et $\vec{p}_l = P_l \mathbb{I}$.

$$\begin{aligned} \left. \frac{\partial E}{\partial w_i} \right|_{\substack{v=0 \\ w=0}} &= \left. \frac{1}{2} \sum_l \frac{\partial}{\partial w_i} \left[\text{Tr}(R^t S_l R X P_l X^t) - 2 \text{Tr}(R^t S_l (Y_l - T) P_l X^t) \right] \right|_{\substack{v=0 \\ w=0}} \\ &= \sum_l \left\langle (R_0 X - Y_l) P_l X^t, \left. \frac{\partial R}{\partial w_i} \right|_{w=0} \right\rangle_{S_l} \end{aligned} \quad (7.33)$$

Et les dérivées secondes :

$$\begin{aligned} \left. \frac{\partial^2 E}{\partial v \partial w_i} \right|_{\substack{v=0 \\ w=0}} &= \sum_l \frac{\partial}{\partial v} \text{Tr} \left[X \vec{p}_l v^t A^t S_l \left. \frac{\partial R}{\partial w_i} \right|_{w=0} \right] \Big|_{v=0} \\ &= A^t \sum_l S_l \left. \frac{\partial R}{\partial w_i} X \vec{p}_l \right|_{w=0} = \sum_l S_l \left. \frac{\partial R}{\partial w_i} \right|_{w=0} X \vec{p}_l \end{aligned} \quad (7.34)$$

$$\begin{aligned}
\left. \frac{\partial^2 E}{\partial w_j \partial w_i} \right|_{\substack{v=0 \\ w=0}} &= \sum_l \left. \frac{\partial}{\partial w_j} \left\langle RXP_l X^t, \frac{\partial R}{\partial w_i} \right\rangle_{S_l} \right|_{\substack{v=0 \\ w=0}} \\
&= \sum_l \left\langle (R_0 X - Y_l) P_l X^t, \left. \frac{\partial^2 R}{\partial w_j \partial w_i} \right|_{w=0} \right\rangle_{S_l} + \left\langle \left. \frac{\partial R}{\partial w_j} \right|_{w=0} X P_l^{\frac{1}{2}}, \left. \frac{\partial R}{\partial w_i} \right|_{w=0} X P_l^{\frac{1}{2}} \right\rangle_{S_l}
\end{aligned} \tag{7.35}$$

$$\left. \frac{\partial^2 E}{\partial v \partial v} \right|_{\substack{v=0 \\ w=0}} = \sum_l \bar{p}_l S_l \tag{7.36}$$

7.6.2.3 Bingham

Ici, un modèle de Bingham (détaillé en section B.3.1.5) est rattaché à chaque label l , la vraisemblance générique s'écrit donc : $P(D_i = x_i | L_i = l; \theta_g) \propto \exp(-\|Rx_i\|_{S_l}^2)$ avec S_l une matrice symétrique définie positive, x_i représente une information axiale (par exemple un des axes principaux du segment sulcal considéré). L'énergie E (équation 7.25), s'écrit alors (en omettant l'a priori) :

$$\begin{aligned}
E &= \frac{1}{2} \sum_{i \in \mathcal{E}} \sum_{L_i=l} p_{il} \|Rx_i\|_{S_l}^2 \quad \text{avec } \|x_i\|^2 = 1 \\
&= \frac{1}{2} \sum_l \left\| RXP_l^{\frac{1}{2}} \right\|_{S_l}^2
\end{aligned} \tag{7.37}$$

en utilisant les mêmes notations introduites dans le cadre du modèle Gaussien.

Comme il s'agit d'une information axiale elle ne dépend pas des facteurs de translation t et son pendant Riemannien v . Par conséquent : $\frac{\partial E}{\partial v} = \frac{\partial^2 E}{\partial w \partial v} = 0$. Voici maintenant la liste des calculs des autres dérivées premières et secondes pris en $w = 0$ et $v = 0$:

$$\left. \frac{\partial E}{\partial w_i} \right|_{\substack{v=0 \\ w=0}} = \sum_l \left\langle R_0 X P_l X^t, \left. \frac{\partial R}{\partial w_i} \right|_{w=0} \right\rangle_{S_l} \tag{7.38}$$

$$\left. \frac{\partial^2 E}{\partial w_j \partial w_i} \right|_{\substack{v=0 \\ w=0}} = \sum_l \left\langle R_0 X P_l X^t, \left. \frac{\partial^2 R}{\partial w_j \partial w_i} \right|_{w=0} \right\rangle_{S_l} + \left\langle \left. \frac{\partial^2 R}{\partial w_j} \right|_{w=0} X P_l^{\frac{1}{2}}, \left. \frac{\partial^2 R}{\partial w_i} \right|_{w=0} X P_l^{\frac{1}{2}} \right\rangle_{S_l} \tag{7.39}$$

7.6.2.4 Kent

Cette fois, il s'agit du modèle de Kent (détaillé en section B.3.1.3), dont l'expression est : $P(D_i = x_i | L_i = l; \theta_g) \propto \exp(\kappa_l \langle Rx, \gamma_{1l} \rangle + 2\beta_l (\langle Rx, \gamma_{2l} \rangle^2 - \langle Rx - \gamma_{3l} \rangle^2))$ avec $\gamma_{1l}, \gamma_{2l}, \gamma_{3l}$ trois vecteurs de norme 1, κ_l et β_l 2 réels. L'énergie E (équation 7.25), s'écrit alors (en omettant l'a priori) :

$$\begin{aligned}
E &= \frac{1}{2} \sum_{i \in \mathcal{E}} \sum_{L_i=l} p_{il} \left[\kappa_l \langle Rx, \gamma_{1l} \rangle + 2\beta_l (\langle Rx, \gamma_{2l} \rangle^2 - \langle Rx - \gamma_{3l} \rangle^2) \right] \\
&= \frac{1}{2} \sum_l \kappa_l \langle G_{1l} P_l X^t, R \rangle + \beta_l \text{Tr} [G_{l\Delta_{23}} RXP_l X^t R]
\end{aligned} \tag{7.40}$$

avec $G_{l1} = g_{l1}\mathbb{I}^t$ et $G_{l\Delta_{23}} = g_{l2}g_{l2}^t - g_{l3}g_{l3}^t$. Pour les mêmes raisons que pour la distribution Bingham, on a : $\frac{\partial E}{\partial v} = \frac{\partial^2 E}{\partial w \partial v} = 0$. Les autres dérivées, prises en $w = 0$ et $v = 0$ s'écrivent :

$$\left. \frac{\partial E}{\partial w_i} \right|_{\substack{v=0 \\ w=0}} = \sum_l \left\langle \left[\frac{\kappa}{2} G_{l1} + \beta_l G_{l\Delta_{23}} R X \right] P_l X^t, \left. \frac{\partial R}{\partial w_i} \right|_{w=0} \right\rangle \quad (7.41)$$

$$\begin{aligned} \left. \frac{\partial^2 E}{\partial w_j \partial w_i} \right|_{\substack{v=0 \\ w=0}} &= \sum_l \left\langle \left[\frac{\kappa}{2} G_{l1} + \beta_l G_{l\Delta_{23}} R_0 X \right] P_l X^t, \left. \frac{\partial^2 R}{\partial w_j \partial w_i} \right|_{w=0} \right\rangle \\ &+ \left\langle \left. \frac{\partial R}{\partial w_j} \right|_{w=0} X^t P_l^{\frac{1}{2}}, \left. \frac{\partial R}{\partial w_i} \right|_{w=0} X^t P_l^{\frac{1}{2}} \right\rangle_{\beta_l G_{l\Delta_{23}}} \end{aligned} \quad (7.42)$$

Résultats

L'étude présentée dans ce chapitre se concentre uniquement sur une modélisation de la variabilité sulcale par des modèles SPAMs (section 4.2.3) parce qu'il s'agit du seul modèle qui a montré des améliorations notables dans le contexte de l'estimation jointe de normalisations et de labélisations. Les autres modélisations proposées (approche Riemannienne sur des modèles Gaussiens 7.6.2) gardent un intérêt essentiellement théorique pour le moment. Leur intérêt pratique devrait s'exprimer en les couplant avec une représentation paramétrique (section B.4.1.2) des modèles SPAMs et ainsi permettre une accélération notable du processus d'optimisation.

8.1 Modèles

Dans cette section, nous explorons 8 modèles différents divisés en trois catégories. Le premier groupe nommé *méthodes standards (image IRM)*, comprend des méthodes composées d'un recalage iconique 7.1 (basé sur l'intensité du signal IRM) suivi de l'estimation d'une modélisation de la variabilité sulcale depuis l'espace commun défini par la méthode de normalisation considérée. Pour représenter les modèles, nous utiliserons les noms suivants en se basant sur le nom de la normalisation appliquée : **SPM-SPAM**, **Baladin-SPAM** et **Talairach-SPAM**.

Dans la seconde catégorie nommée *recalage rigide global*, la labélisation est réalisée de façon couplée au recalage global (équation 7.6) : **basique-globale-SPAM** est basé sur des SPAM estimés à partir de l'espace standard de Talairach (le même que celui utilisé par **Talairach-SPAM** model), **globale-SPAM** et **affine-globale-SPAM** sont basés sur des modèles SPAM raffinés par un recalage global (utilisant la méthode définie par l'équation 7.13). Les modèles **basique-globale-SPAM** et **globale-SPAM** optimisent une transformation rigide globale (1 rotation et 1 translation) alors que la méthode **affine-globale-SPAM** optimise une transformation affine (2 rotations, 1 une matrice diagonale de facteurs d'échelles et 1 translation). Ces trois approches sont basées sur l'approche **Talairach-SPAM** en tant qu'initialisation qui fournit un bon point de départ pour des raffinements ultérieurs.

La dernière catégorie (*recalage rigide local*) regroupe les approches locales non linéaires : **Talairach-locale-SPAM** utilise le modèle **Talairach-SPAM** en tant qu'initialisation alors que **local-SPAM** est basé sur le modèle **globale-SPAM**. Ces deux modèles raffinent les précédents grâce à un ensemble de recalages locaux (une transformation rigide par sillon : équation 7.13) où le processus de labélisation est couplé avec l'optimisation locale des paramètres de recalage (équation 7.8) avec un ensemble de points de référence (permettant de définir les centres des rotations locales et comment appliquer les a priori sur les translations, voir section 7.5.3.1) déterminé à partir de l'espace des sujets.

Tous les modèles étudiés ici se basent sur une représentation à base de SPAM. Ils peuvent être appréciés visuellement par une représentation adaptée permettant de percevoir une fonction de probabilité dense dans \mathbb{R}^3 . Nous avons utilisé 3 isosurfaces emboîtées correspondant chacune à 30, 60 et 80% de la probabilité totale, calculée en intégrant les vraisemblances depuis les plus hautes aux plus basses.

La variabilité de deux structures sulcales particulières (sillon central et fissure pariéto-occipitale)

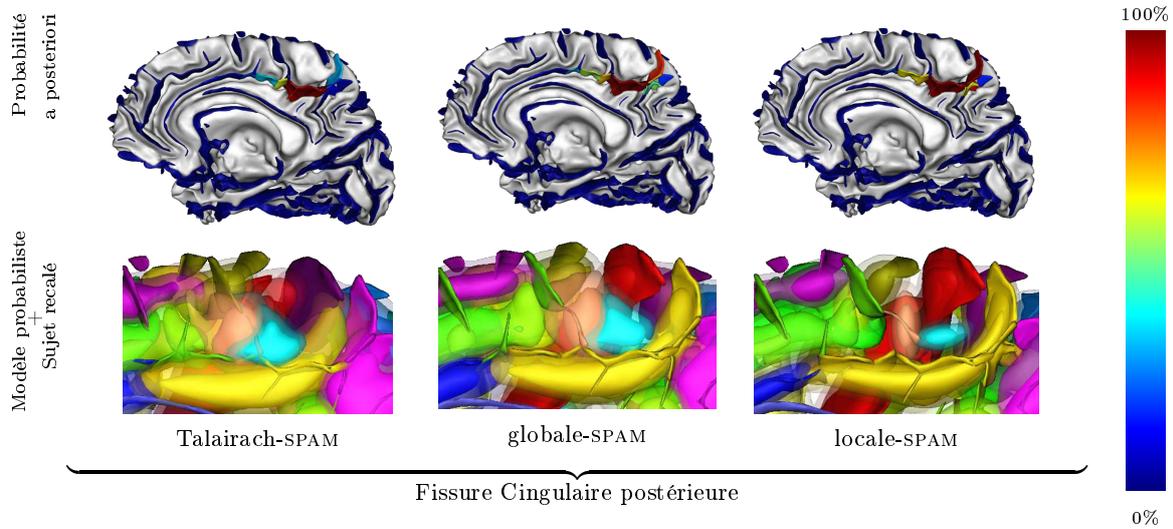


FIG. 8.1 – Effets pratiques de la méthode proposée couplant labélisation et recalage (global et local) sur 2 sillons primaires : la fissure cingulaire postérieure (post.Cing.F) ci-dessus et la fissure Calcarine en figure 8.2. Le phénomène illustré ici a été constaté sur l'ensemble des 62 sujets de notre base de données, mais nous avons choisi le sujet sur lequel les améliorations étaient les plus prononcées sur le sillon considéré. **Haut** : probabilités a posteriori locales du sillon étudié, projetées sur les segments sulcaux correspondants. **Bas** : détails de la zone concernée où l'on constate l'amélioration de l'adéquation entre le modèle et le sujet considéré. Les segments sulcaux sont coloriés selon la labélisation automatique de chacun des modèles étudiés, basé sur le modèle SPAM (de gauche à droite : *Talairach-SPAM*, *globale-SPAM* et *locale-SPAM*), et mélangé avec la représentation visuelle des distributions sulcales. Les deux illustrations mettent en relief l'intérêt de l'approche jointe globale pour la labélisation et de l'approche jointe locale pour les probabilités a posteriori locales. Les modèles considérés sont détaillés en détail au début de la section 8.

est alors comparée entre les différentes approches de normalisation (en phase de précalcul ou couplée avec l'estimation du modèle) en figure 8.3. Le *template* ICBM152 a été utilisé pour les approches Baladin et SPM permettant une comparaison visuelle directe. De la même façon, les 3 autres méthodes sont représentées dans l'espace de Talairach et sont comparables entre elles. La comparaison entre les deux groupes n'est malheureusement pas directe. Leur orientation relative a été réglée manuellement. Parmi les méthodes standards, l'espace de Talairach semble donner les résultats les moins bons. Notre approche globale couplée à la modélisation sulcale donne de meilleurs résultats, eux-mêmes distancés par l'approche locale. Ces résultats se confirment dans leur capacité à inférer une labélisation sulcale (tableau 8.1). Une comparaison plus générale, comprenant l'ensemble des labels, mais limitée à uniquement 3 méthodes : l'approche de référence depuis Talairach, l'approche globale et locale, est réalisée en figure 8.4.

8.2 Résultats et discussion

L'ensemble des résultats à suivre sont estimés par une procédure de validation croisée décrite en section 5.3. Pour une meilleure compréhension, nous invitons le lecteur à se référer au tableau 8.1 pour comparer l'ensemble des mesures d'erreurs globales obtenues. Les taux d'erreurs locaux (sillon par sillon) sont représentés en figure 8.5. La figure 8.3 permet également d'observer en détail l'évolution de la qualité de l'alignement de deux structures sulcales particulières (sillon central et fissure parieto-occipitale) au travers des différentes méthodes.

Tout d'abord, étudions le cas des recalages affines en tant que pré-traitement (voir colonne

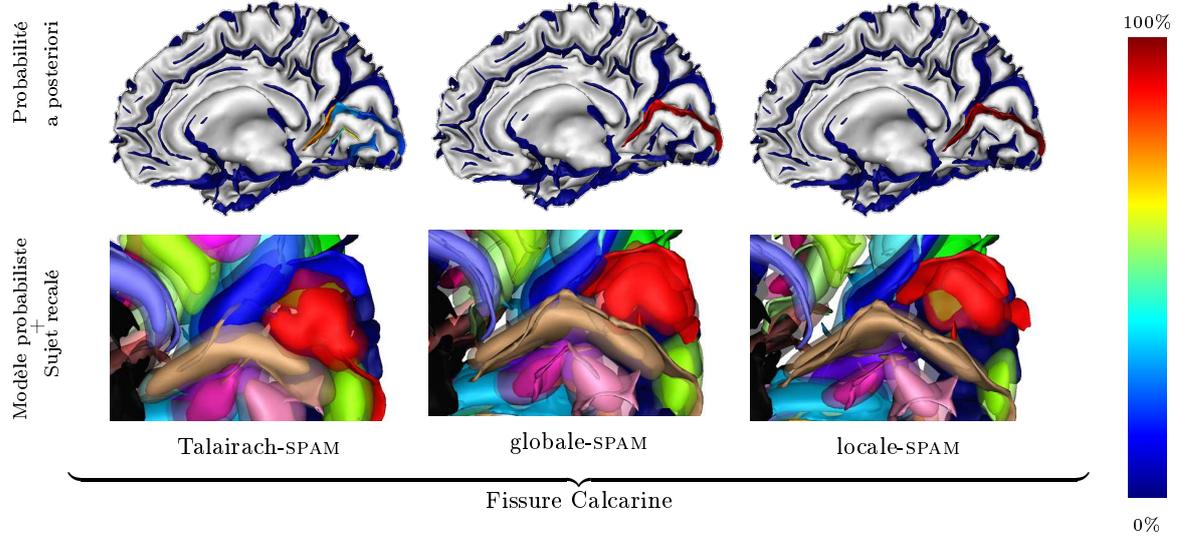


FIG. 8.2 – Même légende que la figure 8.2 mais pour la fissure Calcarine.

méthodes standards (image IRM) du tableau 8.1). Dans ce contexte les qualités des labélisations obtenues sont plus proches pour les trois méthodes testées (Talairach, SPM et Baladin) avec une légère préférence pour l'espace de Talairach où les résultats sont plus équilibrés entre les deux hémisphères. L'avantage des deux autres méthodes étant d'être complètement automatiques alors que dans notre cas, nous rappelons que la normalisation de Talairach est réalisée manuellement à partir de l'identification des commissures antérieure et postérieure (section 7.1). Les labélisations obtenues suite à ces normalisations présentent encore certaines erreurs importantes de même nature, même sur les plus grands sillons. Les erreurs les plus communes et les plus visibles concernent des dédoublements de structures, des déplacements ou des labels mal placés au coeur même de ces structures. La qualité de la segmentation des plissements corticaux en segments sulcaux influe considérablement sur l'apparition de telles erreurs.

La prise en compte d'informations contextuelles permet de limiter ces erreurs dans une certaine mesure que ce soit dans le cadre d'une approche structurale haut-niveau [Rivière 2002, Perrot 2008] (approche «historique», section 2) ou au travers d'une approche plus locale (partie III, chapitre 11). Dans les deux cas, il s'agit de modèles Markoviens de la variabilité sulcale. Leurs performances dépendent de la qualité de l'espace commun à partir duquel est estimé le modèle.

De cette façon, la plupart des erreurs mentionnées plus tôt peuvent s'interpréter comme la conséquence directe d'une normalisation qui s'est mal déroulée, de telle façon que les modèles sulcaux et les sujets considérés sont mal alignés. Ainsi, comparé à la labélisation des sillons corticaux à partir d'un espace commun fixé a priori, notre approche jointe (avec réestimation des paramètres de normalisation) basée sur une transformation globale (rigide ou affine) apporte des améliorations significatives¹ à l'alignement des structures sulcales et par suite, à la qualité de labélisation automatique. Comparé au modèle *Talairach-SPAM*, on obtient une *p-value* de $6.5e^{-8}$ pour les hémisphères gauches et de $6.2e^{-9}$ pour les hémisphères droits. Dans le tableau 8.1, la colonne intitulée *recalage rigide global* montre également l'intérêt de remettre en cause les paramètres de normalisation pendant la phase d'estimation du modèle (*basique-globale-SPAM*) et pendant la labélisation d'un nouveau sujet (*globale-SPAM*). Chacune de ces étapes réduit les incertitudes potentielles sur les paramètres de normalisation, améliore l'alignement des structures et améliore la qualité de la labélisation. La version affine (*affine-globale-SPAM*), n'améliore que très légèrement les résultats par rapport à l'ap-

¹test de Wilcoxon, voir détails en section 5.2

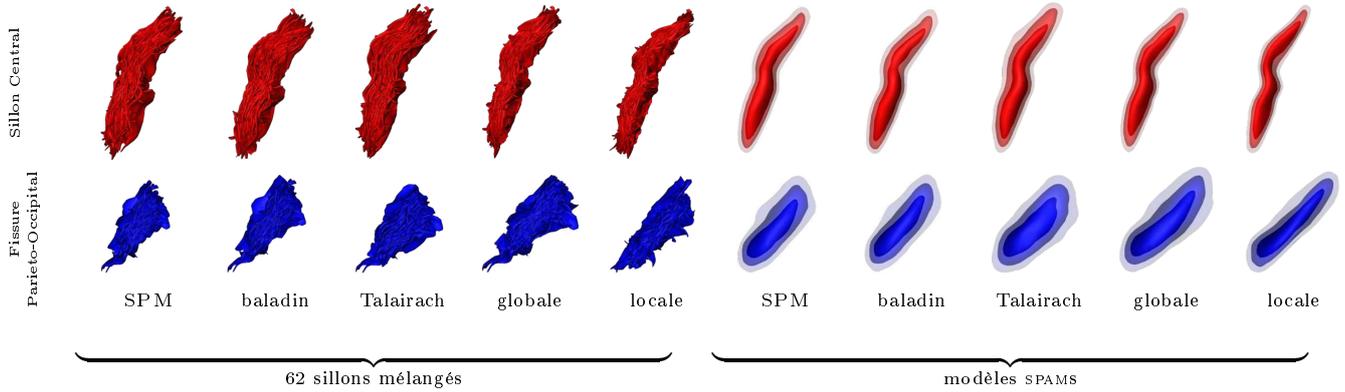


FIG. 8.3 – Comparaison de plusieurs méthodes affines standard et des méthodes de recalage proposées (approche globale et locale) sur deux sillons primaires. **Haut** : sillon central, **Bas** : fissure pariéto-occipitale. **Gauche** : 62 sillons recalés. **Droit** : représentation 3D des modèles SPAMS.

proche rigide, mais est bien plus difficile à optimiser. De plus, des contraintes supplémentaires (a priori sur la rotation définissant les axes d'applications des facteurs d'échelle de la transformation affine) seraient sans doute nécessaires à l'approche affine, afin d'améliorer la robustesse et l'efficacité du processus. Nous avons également mesuré grâce à des mesures de différence d'entropie que chaque distribution SPAM définie avec l'étape de raffinement global (*globale*-SPAM) était plus piquée que celle définie directement depuis Talairach (*Talairach*-SPAM), ce qui indique que la variabilité sulcale a été réduite dans cet espace commun après optimisation.

Alors que la plupart des sillons sont plutôt bien alignés sur cet atlas probabiliste, d'autres peuvent être décalés localement. L'utilisation de recalages locaux à partir d'un espace commun bien choisi permet d'améliorer ou de corriger localement certains ajustements. L'espace de Talairach ne semble pas assez bon pour définir des a priori suffisants (section 7.5) pour contraindre le recalage. En effet, les résultats de labélisation de *Talairach-locale*-SPAM montrent l'amélioration du taux d'erreur si par rapport à ceux obtenus par le modèle *Talairach*-SPAM, mais les dégrade comparé au modèle *globale*-SPAM. Finalement, initier le processus local à partir de l'espace commun optimisé par un recalage rigide global (*globale*-SPAM) permet de constater des améliorations (*locale*-SPAM). En fait, les résultats de labélisation obtenus par l'intégration d'un recalage rigide global ou d'une approche non-linéaire locale sont très proches. Les taux d'erreurs moyens mesurés (voir tableau) sont à la limite de la significativité ($p\text{-value}^1$ de 7.5^{-2} pour les hémisphères gauches et de 8.2^{-2} pour les hémisphères droits). Seulement quelques sillons présentent des améliorations locales significatives (voir les cartes de $p\text{-values}$ corrigées dans les 4 premières lignes de la figure 8.5). Néanmoins, on constate une augmentation nettement significative des probabilités a posteriori locales (voir les 4 dernières lignes de la figure 8.5), ce qui révèle une meilleure adéquation entre chaque sujet labélisé automatiquement et ce modèle de la variabilité sulcale. Pourtant, ces résultats s'expliquent partiellement par la forme plus piquée des modèles SPAMS qui ont également bénéficié d'un raffinement par cette méthode locale de recalage pendant leur estimation ; de cette façon ils donnent des probabilités a posteriori plus contrastées dans le cas d'alignements corrects. Plus concrètement, cet effet est illustré en figure 8.1 sur deux sillons spécifiques sur deux sujets différents.

Finalement, les meilleurs résultats de labélisation sont obtenus avec le modèle *local*-SPAM avec une erreur moyenne globale de 14.22% et 13.48% respectivement pour les hémisphères gauches et droits, ce qui correspond à des taux de bonnes reconnaissances de 85.78% et de 86.52%.

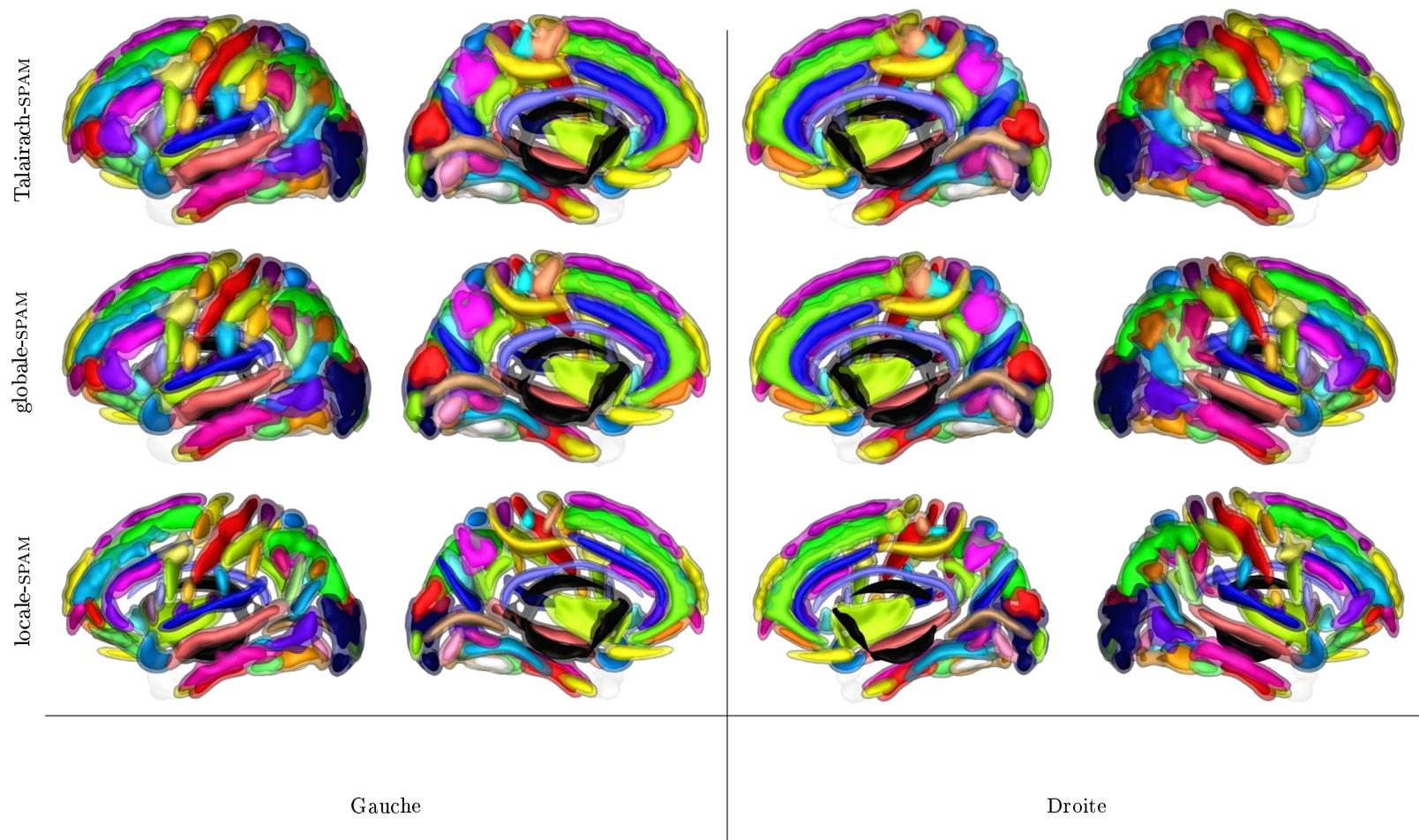


FIG. 8.4 – Vues latérales de modèles de la variabilité sulcale basés sur des SPAMs pour des hémisphères gauches et droits, construits à partir de 62 sujets (voir section A pour plus de détails sur la base de données). Les modèles sont décrits en début de ce chapitre 8.

Base de test :

		méthodes standards (image IRM)			recalage rigide global			recalage rigide local	
		SPM	Baladin	Talairach	basique - globale	globale	affine - globale	Talairach - locale	locale
E_{SI}	Gauche	18.06 (5.31)	17.52 (4.59)	17.55 (5.93)	15.27 (2.74)	14.59 (2.88)	14.31 (2.87)	16.79 (4.35)	14.22 (2.96)
	Droit	16.65 (4.61)	15.82 (3.79)	16.83 (3.77)	14.70 (3.09)	13.97 (2.91)	13.80 (2.62)	15.61 (3.72)	13.48 (3.13)
E_{mass}	Gauche	17.24 (4.95)	16.84 (4.21)	16.64 (3.69)	14.56 (2.54)	14.01 (2.62)	13.75 (2.57)	16.61 (4.17)	14.11 (2.79)
	Droit	15.87 (4.37)	15.17 (3.63)	15.96 (3.66)	13.95 (2.98)	13.37 (2.82)	13.17 (2.52)	15.29 (3.51)	13.30 (2.99)

Base d'apprentissage :

		méthodes standards (image IRM)			recalage rigide global			recalage rigide local	
		SPM	Baladin	Talairach	basique - globale	globale	affine - globale	Talairach - locale	locale
E_{SI}	Gauche	15.09 (5.30)	14.78 (4.34)	14.34 (3.51)	13.11 (2.78)	11.61 (2.38)	11.64 (2.49)	14.71 (4.23)	11.97 (2.30)
	Droit	14.29 (4.58)	13.43 (3.55)	13.64 (3.13)	12.66 (2.68)	11.22 (2.39)	11.66 (2.26)	13.93 (3.66)	11.96 (2.77)
E_{mass}	Gauche	14.44 (4.95)	14.19 (4.10)	13.65 (3.26)	12.55 (2.46)	11.23 (2.27)	11.26 (2.32)	14.63 (4.16)	11.89 (2.05)
	Droit	13.66 (4.34)	12.88 (3.20)	12.87 (2.87)	12.08 (2.57)	10.80 (2.27)	11.19 (2.15)	13.72 (3.46)	11.81 (2.62)

TAB. 8.1 – **Haut** : mesures d'erreurs moyennes de labélisation calculées par *Leave-one-out* (sur 62 sujets) : taux d'erreurs *SI* et *mass* (et l'écart-type associé entre parenthèses) pour chacun des modèles étudiés (une description détaillée est donnée au début de la section 8.2). **Bas** : mêmes mesures mais à partir d'un unique modèle estimé à partir de la base de données complète. L'ensemble des sujets sont labélisés à partir de ce modèle. Entre les résultats de test et d'apprentissage, une différence de près de 3% est observée, ce qui permet de vérifier que les modèles proposés sont plutôt robustes de telle façon que les résultats se généralisent bien à de nouvelles données. On peut remarquer que les résultats obtenus sur la base d'apprentissage sont plus équilibrés entre les deux hémisphères que ceux obtenus sur des données de test, ce qui s'explique par une plus grande variabilité des structures sulcales sur l'hémisphère gauche.

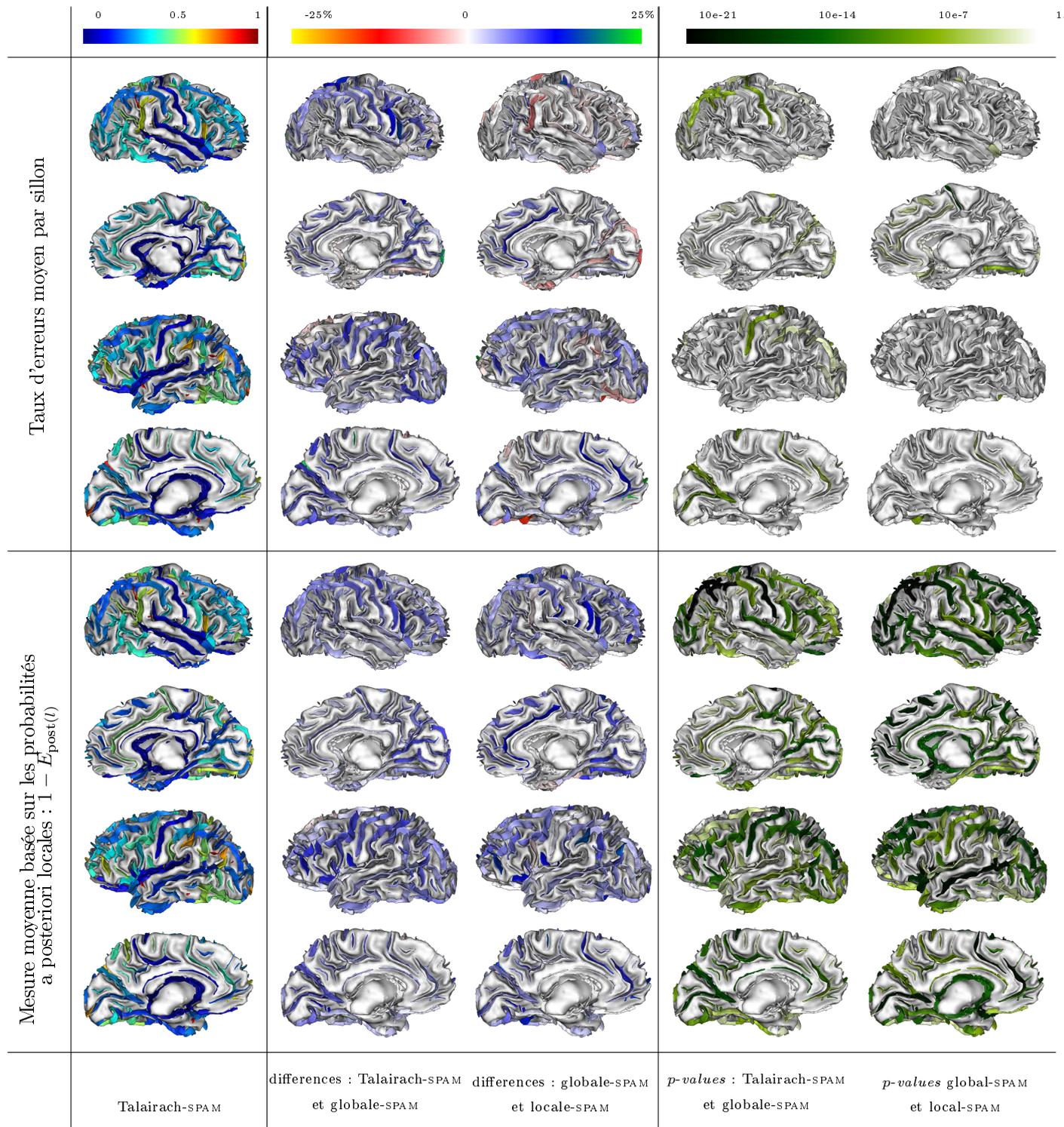


FIG. 8.5 – Améliorations locales (au niveau de chaque sillon) des approches couplant labélisation et recalage. **Haut** : taux d'erreurs moyen par validation croisée (équation 5.2). **Bas** : complémentaire ($1 - P_l$) des probabilités a posteriori moyennes locales P_l (équation 5.5). Tous les résultats sont projetés sur deux hémisphères représentatifs, choisis parce qu'ils présentent la plupart des sillons étudiés. 3 colonnes : **Gauche** : mesures d'erreurs du modèle *Talairach-SPAM*, **Milieu** : améliorations locales (différences entre les erreurs locales) du modèle *globale-SPAM* par rapport au modèle *Talairach-SPAM* (gauche) et *locale-SPAM* versus *globale-SPAM* (droite). **Droite** : projection du logarithme des *p-values* corrigées (voir détails en section 5.2) indiquant la significativité des différences observées dans la colonne précédente. On constate ainsi que la plupart des détériorations ne sont pas significatives. La mesure E_{post} (bas) donne des résultats sensiblement similaires en moyenne comparés aux taux d'erreurs si (haut) pour l'approche globale et des améliorations légères pour l'approche locale. Par contre, les résultats obtenus sont nettement plus significatifs dans le cas des mesures a posteriori (voir mesures de *p-values*). Cette mesure permet donc de caractériser des améliorations plus subtiles.

Conclusion et perspectives

9.1 Conclusion

Les modélisations proposées dans cette partie ont été motivées par l'idée d'extraire et d'exploiter autant que possible des informations de forme des modèles sulcaux (SPAMs principalement) présentés dans la partie précédente. La variabilité de ces formes s'exprime pleinement une fois qu'on a soustrait la part de la variabilité induite par des différences de localisation, d'orientation, voire d'échelle. Chaque objet est alors caractérisé par sa position spatiale dans un espace normalisé. Ces considérations ont été formulées dans un cadre Bayésien où la labélisation sulcale d'un sujet ainsi que la modélisation de la variabilité sulcale sont réalisées conjointement à l'estimation d'une transformation vers un espace normalisé inconnu.

La modélisation proposée est très simple. Elle exploite au maximum la localisation spatiale des structures dans un cadre où la plupart des variables statistiques manipulées sont considérées indépendantes entre elles : les transformations locales, les segments sulcaux, les voxels entre les segments, ainsi que la plupart des a priori. Les résultats observés sont déjà très prometteurs et peuvent être vus comme une base qu'il devrait être facile de dépasser en relevant certaines hypothèses d'indépendance.

Dans cette partie, nous avons proposé plusieurs modèles originaux qui démontrent l'intérêt de proposer une modélisation adaptée, dédiée à un type d'information en particulier (la localisation des segments sulcaux). Les modélisations les plus efficaces reposent sur une représentation de la variabilité sulcale par une mixture de SPAMs, qui fournissent également une compréhension visuelle bien pratique de la position moyenne et de la forme des plissements corticaux. Nous avons proposé 3 façons d'estimer un tel modèle. La première est basée sur une technique de normalisation fixée a priori qui peut être mesurée indifféremment sur des sujets labélisés ou non. Nous avons limité notre étude aux techniques de normalisations affines (Talairach, SPM et Baladin) afin de permettre une comparaison directe avec les modèles que nous avons proposés. Il serait, bien entendu, très intéressant de comparer notre méthode affine locale aux recalages non linéaires standards [Klein 2009].

Deuxièmement, nous avons proposé un raffinement de cette approche en remettant en cause le choix de l'espace normalisé commun. Pour chaque nouveau sujet, une transformation affine est estimée de façon couplée à la labélisation. Cette méthode donne des améliorations significatives comparée à toutes les normalisations affines fixées a priori, que ce soit en terme de qualité de la labélisation sulcale ou du recalage. En effet, l'alignement des structures corticales et sulcales en particulier devient un critère de plus en plus utilisé pour apprécier l'intérêt d'une technique de normalisation, et est évalué généralement par l'intermédiaire d'indices de recouvrement de parcellisations corticales [Klein 2009]. L'approche globale (une seule transformation affine) proposée maximise cet alignement même dans le cas de labels inconnus et surpasse toutes les autres méthodes de recalage affine testées. Ainsi, les capacités de généralisation des modèles SPAMs se comportent comme un indicateur de la qualité de l'espace normalisé à partir duquel le modèle est estimé.

Finalement, nous avons suggéré une extension non-linéaire basée sur un ensemble de recalages rigides locaux (un par sillon). Grâce aux approximations proposées, le modèle est rapide à l'utilisation. Les faiblesses liées à ces simplifications sont compensées en partie par l'utilisation d'a priori

estimés empiriquement à partir d'une base d'apprentissage, alors que dans la littérature l'étendue des déformations possibles est généralement fixée a priori et selon un schéma spatialement stationnaire. Cette approche donne des labélisations assez similaires à celles inférées par l'approche globale précédente, mais les probabilités a posteriori obtenues (à quel point le modèle est confiant dans la labélisation qu'il propose) pour les véritables labels ont été significativement améliorées, ce qui signifie que certaines incertitudes et confusions entre labels ont été réduites.

Il est possible que parmi les meilleures techniques de recalage diffeomorphique de la littérature [Klein 2009], certaines donnent de meilleurs résultats en termes de normalisation que ceux présentés. Mais nous pensons qu'elles pourraient toutes bénéficier d'améliorations notables en s'intégrant au cadre général proposé dans cette partie (couplage normalisation et modélisation de la variabilité anatomique) et ainsi égaler les performances d'approches de normalisation basées sur l'alignement d'amers anatomiques. Pour conclure, la méthode de recalage non-linéaire proposée reste basique mais plutôt fiable puisqu'elle se base sur des contraintes sulcales implicites, sans nécessiter la moindre intervention manuelle; éliminant ainsi cette étape fastidieuse et sujette aux erreurs humaines. Afin de juger pleinement la qualité du recalage proposé, il serait intéressant de reconsidérer des approches standards de détection d'activation ou de *clustering* de faisceaux de fibres après une telle normalisation.

9.2 Perspectives

9.2.1 Méthodes de recalage

Les modèles proposés dans cette partie ont d'abord été conçus avec pour objectif principal l'amélioration du processus de labélisation des sillons corticaux. Jusqu'à présent, ils n'ont pas la prétention de concurrencer les méthodes de normalisation non-linéaire les plus avancées. En effet, l'approche locale proposée est incapable de prendre en compte une situation où un ensemble de sillons est décalé de façon cohérente (exemple d'une telle situation en figure 9.1). Bien entendu, la prochaine étape de ce travail consistera à introduire cette cohérence dans notre modélisation. Plusieurs approches sont possibles, introduisant une complexité graduelle.

La première consiste à lever l'hypothèse d'indépendance entre les a priori des transformations locales. La modélisation PDM (*Point Distribution Model*) utilisée dans les travaux de [Lohmann 2000] suggère l'exploitation de la covariabilité d'un ensemble de points pour modéliser des co-déplacements de structures anatomiques. Même s'il n'est pas présenté en ces termes, le modèle proposé correspond à un modèle Gaussien (avec une matrice de covariance anisotrope) d'un ensemble de points réunis sous la forme d'un grand vecteur concaténant l'ensemble de leurs coordonnées. Dans le cadre de notre approche locale, un tel modèle peut servir à exprimer un a priori sur l'ensemble des translations locales (exprimées par rapport à leur référentiel local respectif). Plutôt que d'utiliser une représentation vectorielle, nous proposons d'utiliser une représentation matricielle qui semble plus naturelle (regroupement de positions 3D) dont la variabilité peut être étudiée avec une généralisation du modèle Gaussien à des données tensorielles [Basser 2003, Basser 2007].

Une seconde méthode viserait à remplacer la méthode de recalage proposée par une autre proposant une déformation volumique cohérente. Les approches diffeomorphes, très populaires cette dernière décennie, répondent à ce critère. Pour s'intégrer dans notre formulation Bayésienne, il faut également que les composantes de recalages locaux soient reliées d'une façon ou d'une autre aux modèles sulcaux. L'approche la plus directe consisterait à utiliser un recalage polyaffine [Arsigny 2005] qui permet d'obtenir une transformation non-linéaire diffeomorphe (le champ de déformation est spatialement continu et lisse). Cette méthode se base sur une collection de recalages affines locaux

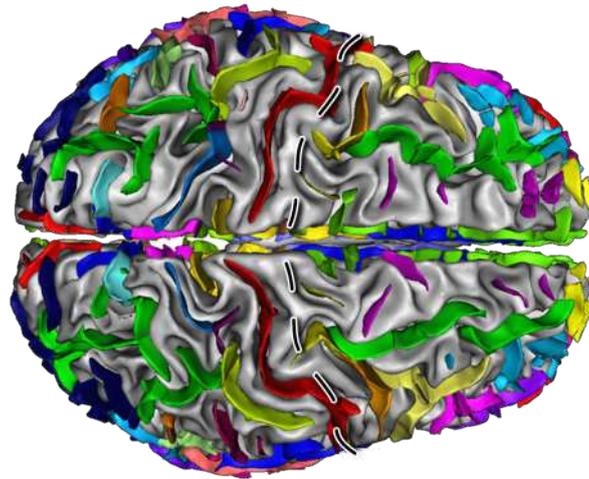


FIG. 9.1 – Ce sujet, représenté ici avec les véritables labels sulcaux identifiés manuellement, illustre parfaitement quelques unes des limites de la modélisation proposée induites par l’hypothèse d’indépendance sur les transformations locales. En fait, le sillon central de ce sujet (couleur rouge au milieu de l’image) et par conséquent son voisinage (pré-central, post-central, partie postérieure des sillons frontaux médians et la partie antérieure du sillon pariétal) est décalé vers l’arrière de près de 2 cm par rapport à l’emplacement moyen représenté par une ligne pointillée. De tels déplacements ne peuvent être traités que par la prise en compte de transformations locales cohérentes et éventuellement par des contraintes supplémentaires entre les deux hémisphères comme le suggère la symétrie du phénomène.

qu’elle interpole de façon cohérente dans l’espace Riemannien correspondant. De façon plus générale, la plupart des méthodes contraintes par des amers anatomiques sont transposables à condition de faire le lien avec notre représentation en segments sulcaux. Nous pensons tout particulièrement aux travaux de nos collaborateurs [Cachier 2001, Auzias 2009] qui utilisent la même représentation et devraient par conséquent être plus faciles à traduire dans notre modèle. Dans tous les cas, une adaptation au cas par cas est nécessaire afin d’intégrer le schéma d’optimisation EM dans le processus de recalage. Il est probable que de telles méthodes de recalage nécessitent des approximations adaptées afin d’éviter des échantillonnages longs et coûteux dans l’espace des déformations.

9.2.2 Remarques diverses

La variabilité sulcale observée sur les plus grands sillons indique qu’un seul point de contrôle par sillon n’est pas suffisant pour modéliser son étendue. En accord avec la théorie qui sous-tend la définition des labels sulcaux utilisés dans le cadre de notre étude de la variabilité corticale, il paraît judicieux d’utiliser les racines sulcales [Régis 2005] (ou les bassins sulcaux [Lohmann 2008]), constituant la partie la plus stable des plissements corticaux, afin d’affiner le jeu de déformations locales autorisées.

Comme l’indique la figure 9.1 des déplacements importants (constatés jusqu’à près de 2 cm) de structures existent même chez le sujet sain. Ces situations restent marginales au sein de la population et ne peuvent être estimées avec précision dans la modélisation actuelle. L’étude des *patterns* sulcaux permettra sans doute d’identifier ces situations et ainsi proposer une modélisation alternative adaptée dans le cadre d’une mixture de modèles sulcaux pour chaque sillon (ou éventuellement par groupe de sillons). L’approche générative proposée peut également servir à générer de nouveaux

cerveaux de test après déformation et ainsi étudier plus finement les capacités de généralisation des situations les plus rares.

Le succès de la modélisation actuelle, malgré l'hypothèse d'indépendance réalisée entre les transformations locales repose sur les contraintes exercées par les a priori $P(\theta_l)$ sur le choix de la déformation limitant ainsi la plupart des incohérences. L'a priori utilisé à l'heure actuelle pour chaque rotation locale reste une modélisation grossière à laquelle on préférera la modélisation Matrix-Von-Mises suggérée en annexe B.3.1.6 ou une modélisation directement sur l'espace Riemannien correspondant [Boisvert 2008]. Dans le cas de techniques de normalisation alternatives, de nouveaux a priori sont à imaginer. Par exemple, de façon standard les champs de déformation sont régularisés par un modèle gaussien isotrope stationnaire. Les résultats observés dans le cadre de notre méthode de recalage rigide par morceau ont démontré l'intérêt de la prise en compte d'a priori locaux spécifiques. Ils permettent la mise en œuvre d'un modèle plus précis qui permet de mieux apprécier la variabilité anatomique.

La modélisation proposée se base essentiellement sur des informations de localisation qui sont plutôt pertinentes sur les principaux plissements corticaux, mais s'avèrent peu fiables pour les structures secondaires comme les branches. Même avec une très bonne méthode de normalisation de telles structures peuvent rester ambiguës. Parfois, seules des informations contextuelles issues des plissements voisins permettent de lever cette difficulté. Cette idée suggère l'utilisation de modèles Markoviens, comme celui présenté en introduction [Rivière 2002, Perrot 2008] de cet ouvrage. Une modélisation Markovienne alternative est présentée dans la partie III à suivre (l'intégration de ce modèle Markovien dans l'approche labélisation-normalisation jointe est discutée en section 13.2.3).

Troisième partie

Approches Markoviennes

Table des Matières

10	Modélisation Markovienne	105
10.1	Présentation	105
10.1.1	Champ de Markov aléatoire	105
10.1.2	Zoologie Markovienne	106
10.2	Apprentissage supervisé	107
10.3	Inférence : optimisation de champs de Markov	109
10.3.1	Risque Bayésien	109
10.3.2	Méthodes d'optimisation	110
11	Sillons corticaux et champ de Markov	115
11.1	Modèle	117
11.1.1	Apprentissage	119
11.1.2	Inférence	119
11.2	Graphe de relations	120
11.2.1	Construction	120
11.2.2	Graphes irréguliers	121
11.3	Potentiels de relations	123
11.3.1	Informations pertinentes	124
11.3.2	Description des potentiels	124
12	Résultats	131
12.1	Approche	131
12.2	Expériences	132
12.2.1	<i>Loopy Belief Propagation</i>	132
12.2.2	Recuit simulé	133
13	Conclusion et perspectives	139
13.1	Conclusion	139
13.2	Perspectives	139
13.2.1	Création du graphe	140
13.2.2	Autres potentiels de relation	142
13.2.3	Couplage : normalisation et modélisation Markovienne	143

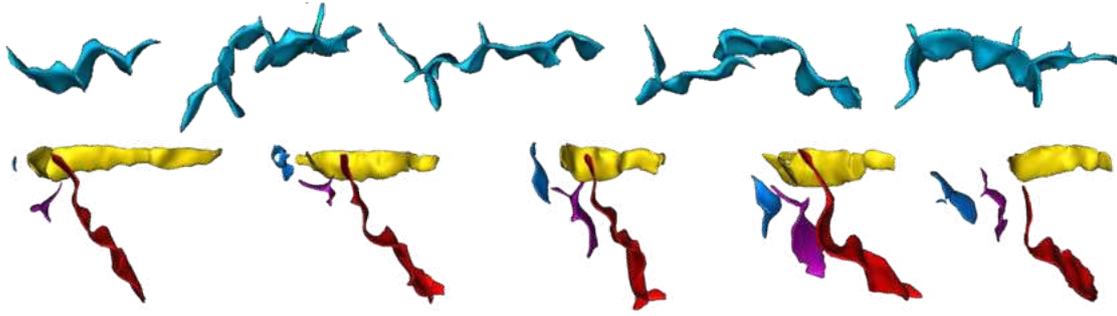


FIG. 9.2 – Intérêt du voisinage dans la labélisation. **Haut** : sillons frontaux intermédiaires de 5 sujets. Ce label regroupe un ensemble de petits plissements s’organisant de façon arborescente. Ce «sillon» n’a pas de structure bien définie, on distingue à peine une tendance dans la direction mais qui reste artificielle puisqu’elle découle de sa définition. En effet, ce sillon est construit comme l’ensemble des plissements entre le sillon frontal supérieur et le sillon frontal inférieur qui suivent la même direction. **Bas** : sillons centraux (rouge), parties supérieures des fissures cingulaires (jaune) et parties supérieures des sillons post-centraux (bleu et mauve) de 5 sujets. Ici, nous observons une configuration très reproductible entre l’extrémité supérieure du sillon central toujours antérieure à l’extrémité postérieure de la fissure cingulaire. Cette configuration se vérifie sur les 62 sujets de notre base de données labélisée manuellement, alors que sur les autres sillons les extrémités sont plutôt variables et se retrouvent indifféremment d’un côté ou de l’autre.

Si l’on se réfère à la démarche de l’expert en anatomie corticale, les sillons ne sont pas identifiés les uns après les autres de façon isolée, mais plutôt en partant des structures les moins variables jusqu’à celles qui le sont plus. De cette façon, chaque structure précédemment identifiée sert de contrainte dans l’identification des suivantes généralement voisines (voir figure 9.2, haut). Et de proche en proche, toutes les structures sont ainsi labélisées, jusqu’aux plus variables (comme le sillon frontal intermédiaire). Malheureusement, la variabilité des *patterns* sulcaux [Sun 2009] est telle qu’il n’existe pas de schéma ordonné unique permettant de réaliser cette tâche, puisque le processus dépend des singularités propres à chaque individu. De plus, ce schéma directeur n’existe que globalement. Localement, des ambiguïtés peuvent exister et amèneraient une telle approche gloutonne dans une mauvaise direction. C’est pourquoi, il est parfois nécessaire de considérer plusieurs structures en même temps afin de lever certaines ambiguïtés. Pour exemple, la stabilité de la configuration présentée au bas de la figure 9.2 permet de garantir l’identification du sillon central dans sa partie supérieure (la plus variable en terme de localisation d’après notre base de données).

Ces remarques montrent l’intérêt de la prise en compte d’informations avoisinantes lors de la labélisation d’une structure sulcale. Ce schéma de labélisation idéal impliquerait une diffusion directionnelle de l’information tant qu’il n’y a pas d’ambiguïté et dans le cas contraire, un échange bidirectionnel d’informations entre les structures en conflit. Cette approche suggèrerait alors un codage du problème sous forme d’un modèle graphique incluant des dépendances statistiques directionnelles ou non selon les situations entre les labels portés par des segments sulcaux voisins. Dans l’état actuel de nos connaissances, il nous semble que seule une analyse a posteriori (une fois les labels identifiés automatiquement) nous permettrait de distinguer les zones difficiles (ambiguës) des autres. De plus, nous ne saurions définir à l’avance le sens à donner aux liens de dépendances à partir de la seule forme des structures, sans même connaître leurs labels. Une solution à ce problème consiste à négliger la direction des dépendances. Si l’on considère uniquement des dépendances localisées (selon un voisinage à définir), alors la modélisation résultante, qui fait l’objet de cette partie, répond au nom de champ de Markov, traditionnellement utilisé en traitement d’images.

Historiquement cette thèse a débuté dans la lignée de travaux réalisés par Denis Rivière [Rivière 2000, Rivière 2002] où une modélisation Markovienne (présentée en introduction 2) permettait de labéliser les sillons corticaux. Cette approche combinait un ensemble d'experts locaux (jouant le rôle de potentiels du champ de Markov) s'exprimant au niveau de chaque sillon et chaque paire de sillons potentiellement voisins. Il s'agit d'une représentation haut niveau du problème où il est possible de caractériser dans le détail la forme ou l'agencement des structures sulcales. Toutefois, l'espace de recherche pour chaque label implique généralement toute la combinatoire engendrée par plus d'une dizaine de segments sulcaux en moyenne.

C'est la raison pour laquelle, nous avons remis en cause jusqu'aux bases de l'approche «historique», au profit d'une modélisation alternative afin d'en dépasser les limites. En effet, l'approche structurelle précédente négligeait l'utilité d'informations localisées, sous-sulcales, qui ne pose aucun problème de combinatoire à l'échelle des segments sulcaux puisqu'ils peuvent tous être traités indépendamment les uns des autres. Nous nous sommes donc évertué à définir des critères discriminants pertinents à cette échelle afin de tirer le meilleur partie de ces informations tout en conservant un modèle relativement simple. Pour tirer le meilleur parti de cette information, nous avons finalement proposé une méthode couplant recalage et labélisation. Cette fois la méthode est hybride : **structurelle** en ce qui concerne l'utilisation des sillons pour contraindre le recalage et **localement iconique** dans la mise en relation des modèles et des segments sulcaux. Cette approche reste faiblement structurelle et nous sommes convaincu que l'exploitation d'information contextuelle est nécessaire. Mais avant d'explorer les niveaux structurels les plus élevés (sillons et relations entre paires de sillons), il existe un niveau intermédiaire qui limite l'impact combinatoire : les relations locales entre segments sulcaux voisins. Il fait l'objet de cette partie.

Nous débuterons cette partie par une description succincte du modèle Markovien en mettant en avant les difficultés qui lui sont propres. Nous replacerons cette approche dans notre contexte d'utilisation et indiquerons comment nous l'avons adapté à la modélisation d'informations locales (entre segments sulcaux voisins) discriminantes.

Modélisation Markovienne

Dans ce chapitre, nous dresserons une revue partielle et succincte des champs de Markov aléatoires [Geman 1984]. Nous nous concentrerons ici principalement sur les aspects explorés pendant cette thèse et quelques méthodes potentiellement intéressantes pour notre champ de recherche. Pour une présentation plus détaillée du sujet, nous invitons le lecteur à consulter les livres suivants [Li 2001] et [Winkler 2006] où cette modélisation est appliquée à l'analyse d'images. Cette description comprend les différentes modélisations Markoviennes standards dans le cadre de l'inférence de variables cachées, l'apprentissage supervisé de ces paramètres et quelques techniques permettant de réaliser cette inférence (c'est à dire, trouver les labels sulcaux inconnus d'un sujet dans notre contexte).

10.1 Présentation

10.1.1 Champ de Markov aléatoire

Intuitivement, un champ de Markov permet de modéliser les valeurs que peuvent prendre un ensemble de variables aléatoires (par exemple les couleurs d'une image) de telle façon que la probabilité de l'état de chaque variable (par exemple, chaque pixel) dépend de ses «voisines». C'est à dire que les dépendances statistiques entre variables sont uniquement locales.

On considère ici un ensemble \mathcal{E} de sites i et leur variables aléatoires associées X_i . L'ensemble $X = \{X_i\}_{i \in \mathcal{E}}$ est un champ de Markov selon le graphe de dépendances \mathcal{G} si :

$$P(X_i | \{X_j; j \neq i, j \in \mathcal{E}\}) = P(X_i | \{X_j; i \sim j, j \in \mathcal{E}\})$$

Que l'on notera plus simplement :

$$P(X_i | X_{\neq i}) = P(X_i | X_{\partial i}) \tag{10.1}$$

où $i \sim j$ signifie que les sites i et j sont voisins au sein du graphe \mathcal{G} , c'est à dire $(i, j) \in \mathcal{G}$. Le graphe reflète ainsi les dépendances statistiques entre les variables considérées.

Le théorème de Hammersley Clifford [Hammersley 1971] stipule qu'un champ de Markov est équivalent à un champ de Gibbs. De cette façon, un champ vérifiant la propriété de Markov (équation 10.1) prend la forme suivante :

$$P(X) = \frac{1}{Z} e^{-\sum_{c \in \mathcal{C}} E_c(X)}, \quad \text{avec } Z = \sum_{X \in \mathcal{X}} e^{-\sum_{c \in \mathcal{C}} E_c(X)} \tag{10.2}$$

où Z est une constante de normalisation, appelée également fonction de partition. Cette expression repose sur un système \mathcal{C} de cliques c (sous-graphes entièrement connectés) issu du graphe \mathcal{G} , où chaque énergie locale $E_c(X)$ est dédiée à l'évaluation des valeurs prises par la restriction de X à la clique c . La modélisation la plus fréquente se base sur des cliques élémentaires (chaque site $i \in \mathcal{E}$ est une clique) et sur les paires de sites en relation. Dans le cadre de notre modélisation de la variabilité sulcale, nous nous sommes restreint à ces systèmes de cliques. L'expression précédente se réécrit

dans ce cas, de la façon suivante :

$$\begin{aligned} P(X) &= \frac{1}{Z} \exp \left[- \sum_{i \in \mathcal{E}} E_i(X) - \sum_{i \sim j \in \mathcal{E}^2} E_{ij}(X) \right] \\ &= \frac{1}{Z} \prod_{i \in \mathcal{E}} \left[e^{-E_i(X)} \right] \prod_{i \sim j \in \mathcal{E}^2} \left[e^{-E_{ij}(X)} \right] = \frac{1}{Z} \prod_{i \in \mathcal{E}} \Phi_i(X_i) \prod_{i \sim j \in \mathcal{E}^2} \Psi_{ij}(X_{ij}) \end{aligned} \quad (10.3)$$

où $\Phi_i(X_i) = e^{-E_i(X)}$ et $\Psi_{ij}(X_{ij}) = e^{-E_{ij}(X)}$.

Nous verrons par la suite que cette formulation présente des avantages pratiques utiles lorsqu'on cherche à faire de l'inférence. Il s'agit d'une façon simple et efficace de modéliser la distribution jointe d'un ensemble de variables en limitant les interactions à un voisinage exprimé par le graphe \mathcal{G} . Le modèle n'impose aucune contrainte sur la forme des énergies locales E_i et E_{ij} qui dépendent fortement du problème considéré.

De façon plus concrète, dans le cadre du modèle présenté dans le chapitre 11, les sites correspondent aux segments sulcaux et les cliques s'expriment au niveau segment et paires de segments voisins ; nous discuterons alors de la Markovianité des données sulcales. Dans le cas du modèle « historique », les sites correspondent également aux segments sulcaux, alors que les cliques s'expriment au niveau des sillons et des relations entre sillons voisins.

10.1.2 Zoologie Markovienne

Les champs de Markov dérivent de modèles de Ising ou de Potts issus de la physique statistique. À l'origine cette modélisation rendait compte de phénomènes de magnétisation où des *spins* voisins s'alignaient de façon cohérente par paquet. Ces modèles se sont généralisés à la modélisation statistique des pixels voisins dans les images. C'est dans ce cadre que ce modèle a trouvé la majorité de ses applications : reconstruction, débruitage ou segmentation principalement. Dans ces applications, on observe des données $D = \{D_i\}_{i \in \mathcal{E}}$ (une image avec des zones abîmées ou bruitées) à partir desquelles on souhaite déduire des données non observées $L = \{L_i\}_{i \in \mathcal{E}}$ (une image reconstruite, débruitée ou des parcelles) sur lesquelles on suppose une cohérence spatiale de l'information modélisée par un champ de Markov. Cette modélisation classique s'appelle **champ de Markov aléatoire caché**¹ (*Hidden Markov Random Field*, HMRF [Zhang 2001]). Cette modélisation suppose la Markovianité de L et l'indépendance conditionnelle des données observées par rapport au champ de Markov : $P(D|L) = \prod_{i \in \mathcal{E}} P(D_i|L)$.

Cette approximation n'est pas valable dans le cas où les données D sont corrélées (structurées) à L fixé, comme par exemple dans le cas des textures puisque la description isolée de chaque pixel ne permet pas de traduire l'agencement des motifs caractérisant une texture. Une solution consiste à considérer également la Markovianité de $P(D|L)$, mais cela n'implique pas forcément la Markovianité de $P(LD)$ ou $P(L|D)$, indispensable à la réalisation de l'inférence ou l'échantillonnage du modèle. Dans la littérature, on retrouve naturellement des modélisations supposant la Markovianité de ces quantités.

Dans le cas de $P(LD)$, il s'agit des **Champs de Markov couplés** [Pieczyński 2000] où les deux types de variables sont étudiés de façon conjointe, de telle façon que $P(L|D)$ et $P(D|L)$ vérifient également la propriété Markovienne. Ce modèle s'est vu augmenté d'un champ auxiliaire U permettant la prise en compte d'un modèle de mélange, permettant la modélisation de la Markovianité de

¹La reconnaissance de sillons peut être réalisée dans ce cadre puisque nous recherchons des labels inconnus L , cohérents spatialement à partir de données anatomiques D connues.

$P(DLU)$. Ce modèle s'appelle **champ de Markov triplet** [Pieczynski 2002] et permet de relâcher la contrainte de la markovianité de la marginale d'intérêt $P(LD)$. Ces modélisations sont surtout utilisées en modélisation et segmentation de textures.

Au contraire les **champs de Markov conditionnels** [Lafferty 2001] (introduction simple et rapide proposée par [Wallach 2004], adaptation au cas $2D$: **champ de markov discriminatif** [Kumar 2006]), suppose directement la Markovianité de $P(L|D)$, c'est à dire la quantité que l'on maximise afin d'inferer les variables inconnues L les plus probables. Il s'agit d'une approche essentiellement discriminative, contrairement aux autres basées sur la Markovianité de $P(L)$, $P(D|L)$ ou $P(LD)$. Ces dernières sont des approches génératives (ou descriptives) permettant de caractériser avec précision la variabilité de l'ensemble des paramètres décrivant les objets étudiés. L'approche discriminative repose sur l'idée qu'il n'est pas nécessaire de modéliser chacun de ces objets avec précision si l'on s'intéresse uniquement à leur identification (à inférer L) ; par exemple, deux modélisations génératives très proches peuvent engendrer des modélisations a posteriori très similaires. Tous les champs de Markov conditionnels ne sont pas simples à manipuler. Une formulation populaire, généralisant la classification basée sur le principe de la régression logistique, s'écrit de la façon suivante :

$$P(L|D, w) = \frac{1}{Z(D)} \exp \left[\sum_{c \in \mathcal{C}} w_c^t f_c(D, L_c) \right] \quad (10.4)$$

où f_c correspond à une fonction permettant de calculer un vecteur descriptif de la configuration L_c de la clique c au regard des données D (sur lesquelles aucune indépendance n'est supposée), et w_c un vecteur de poids à estimer. Cette formulation est moins sensible à l'utilisation de descripteurs riches et redondants [Mccallum 2003] que la modélisation générative, ce qui facilite la modélisation (bien qu'elle soit moins explicative). L'estimation supervisée des poids w_c peut se faire par descente de gradient sur $P(L|D, w)$ (fonction concave, garantissant la convergence) à condition de réaliser une procédure d'inférence très coûteuse à chaque pas.

La modélisation Markovienne proposée dans le chapitre 11 est descriptive et considère la Markovianité de $P(D|L)$. Par contre, le modèle «historique» présenté en introduction 2 est discriminatif et s'apparente fortement à un champ de Markov conditionnel avec des cliques au niveau des sillons et des paires de sillons. Comme dans le cas des champs de markov discriminatif, des classifieurs locaux sont utilisés en guise de potentiel. Par contre, le modèle «historique» ne bénéficie pas d'une estimation rigoureuse des facteurs de pondération entre les différents potentiels.

10.2 Apprentissage supervisé

Dans le cadre de champs de Markov sur des images, de façon courante, les propriétés statistiques du champ (encodées sous forme de paramètres contrôlant l'expression des potentiels) sont fixées a priori ou déduites directement de l'image étudiée. L'estimation des paramètres associée à l'inférence de L est difficile, mais reste possible grâce à la stationnarité des potentiels (c'est à dire qu'ils sont invariants spatialement) et au très grand nombre de sites étudiés. La méthodologie permettant de traiter ces situations ne sera pas présentée dans cette thèse puisque notre application ne rentre pas dans ce cadre.

En effet, la modélisation des liens statistiques entre labels et structures ne peut se déduire à partir d'un seul sujet, par surcroît non labélisé. Dans notre application, nous manipulons une soixantaine de labels pour quelques centaines de sites, l'expression de ces labels est donc très limitée ; de plus, le champ considéré est fortement non stationnaire. C'est dans l'étude de nombreux sujets labélisés que ces liens s'expliquent. Il est donc nécessaire de réaliser un apprentissage supervisé des paramètres du champ.

Plaçons-nous dans un cadre supervisé où L et D sont connus, supposons la Markovianité de $P(LD)$ que l'on notera $X = (L, D)$ pour simplifier le propos. De la même façon, nous supposons que chaque élément de notre base correspond à une réalisation d'un champ de Markov reposant sur une structure de dépendance commune (ce qui n'est pas le cas pour nos données sulcales de telle façon que chaque sujet à sa propre constante de normalisation). Notons alors θ l'ensemble des paramètres contrôlant la forme des potentiels. L'estimation par maximum a posteriori des paramètres à partir d'une base de données d'éléments $\{X_a\}_{a \in \mathcal{A}}$ (où chaque X_a correspond à une réalisation du champ de Markov) consiste à maximiser en θ la quantité suivante :

$$P(\theta|\{X_a\}_{a \in \mathcal{A}}) \propto P(\theta) \prod_{a \in \mathcal{A}} P(X_a|\theta) = P(\theta) \prod_{a \in \mathcal{A}} \left[\frac{1}{Z(\theta)} e^{-E_a(X_a;\theta)} \right] \quad (10.5)$$

où $E_a(X_a;\theta) = \sum_{c \in \mathcal{C}} E_c(X_a;\theta)$. Cette quantité est très difficile à maximiser du fait de la dépendance en θ de la fonction de normalisation $Z(\theta)$, dont l'expression est généralement impossible à calculer puisqu'elle comporte une sommation sur toute les configurations possibles du champ (équation 10.3).

La littérature reste assez restreinte dans le cas supervisé et connaît un essort récent notamment grâce aux champs de Markov conditionnels que l'on retrouve généralement dans un contexte supervisé. On y distingue deux familles d'approximations complémentaires que l'on ne détaillera pas ici. Quel que soit le choix retenu pour maximiser la probabilité a posteriori, la fonction de normalisation $Z(\theta)$ est requise.

La première approche consiste à estimer cette quantité par échantillonnage MCMC (*Monte-Carlo Markov Chain*, [Descombes 1997, Murray 2004] proposent des revues de telles méthodes dans le cadre des modèles graphiques non-orientés), voire même de se passer de la constante de normalisation par le truchement d'une variable auxiliaire [Møller 2004], ou l'intégrer à une méthode de descente de gradient stochastique [Younes 1988].

La seconde approche propose de simplifier l'expression de la vraisemblance $P(X|\theta)$ en négligeant certaines inter-dépendances. L'approximation la plus ancienne et qui reste la plus utilisée, consiste à considérer indépendamment chaque site i associé à son voisinage. Il s'agit de la pseudo-vraisemblance [Besag 1975] dont l'expression est la suivante :

$$P(X|\theta) \approx \prod_{i \in \mathcal{E}} P(X_i|X_{\neq i}; \theta) = \prod_{i \in \mathcal{E}} P(X_i|X_{\partial i}; \theta) = \prod_{i \in \mathcal{E}} \frac{e^{-E_{i\partial i}(X_i;\theta)}}{\sum_x e^{-E_{i\partial i}(X_i=x;\theta)}} \quad (10.6)$$

avec $E_{i\partial i}(X;\theta) = E_i(X|\theta) + \sum_{i \sim j \in \mathcal{E}^2} E_{ij}(X;\theta)$. Ici, $\sum_x e^{-E_{i\partial i}(X_i=x;\theta)}$ joue le rôle de constante de normalisation locale au site i . La pseudo-vraisemblance à l'avantage d'être un estimateur consistant de la vraisemblance [Gidas 1987, Comets 1992]. Plus récemment, une décorrélation du problème par clique a été proposée (*piecewise training* [Sutton 2005]), donnant lieu à l'approximation suivante :

$$P(X|\theta) \approx \left[\prod_{i \in \mathcal{E}} \frac{\Phi_i(X_i|\theta)}{\sum_x \Phi_i(X_i=x|\theta)} \right] \left[\prod_{i \sim j \in \mathcal{E}^2} \frac{\Psi_{ij}(X_{ij}|\theta)}{\sum_{x,y} \Psi_{ij}(X_{ij}=(x,y)|\theta)} \right] \quad (10.7)$$

qui a montré son intérêt dans des contextes où l'information locale à chaque clique est très informative. Cette approche a été récemment combinée avec la pseudo-vraisemblance afin de tirer partie des avantages des deux approches [Sutton 2007]. Dans le même esprit, des généralisations de la pseudo-vraisemblance à des régions ont été proposées [Zhu 2002, Huang 2002], où il faut alors trouver une taille de région réalisant un bon compromis entre qualité de l'approximation (plus la taille est grande, plus on se rapproche de l'expression réelle de la vraisemblance) et difficulté du calcul de la fonction de partition associée à la région (qui augmente avec la taille de la région puisque le nombre de configurations possibles augmente).

Pour les deux modèles Markoviens présentés dans cette thèse (en introduction et plus loin dans cette partie), aucune de ces approches n'a été explorée pour l'instant. Nous nous sommes contenté d'ignorer la constante de normalisation $Z(\theta)$ et de réaliser un apprentissage des paramètres de chaque potentiel de façon indépendante (section 11.1.1).

10.3 Inférence : optimisation de champs de Markov

La phase d'inférence vise à déduire les données inconnues recherchées L (les labels sulcaux dans notre cas) à partir des données D connues (des descripteurs anatomiques dans notre cas). Les dépendances entre les variables, introduites par le champ de Markov, rendent cette tâche difficile ; différentes solutions ont été proposées dans la littérature, nous en détaillerons quelques-unes dans la suite (une revue récente et plus large est proposée dans [Szeliski 2008]).

10.3.1 Risque Bayésien

Selon l'application, l'importance accordée aux erreurs d'inférence de L (se tromper de label sulcal dans notre cas) peut varier. La modélisation Bayésienne permet d'associer une fonction de perte (*loss function*) qui évalue le coût associé à un type d'erreur particulier. Notons l (une labélisation automatique des sillons dans notre cas) une réalisation de la variable L selon notre modèle et t (la labélisation manuelle associée) sa véritable valeur, $\text{loss}(l, t)$ une fonction de perte évaluant les différences entre t et l . La valeur optimale l^* est celle qui minimise le risque Bayésien dont l'expression est la suivante :

$$R(L = l|D) = \mathbb{E}[\text{loss}(l, t)] = \sum_t P(L = t|D)\text{loss}(l, t) \quad (10.8)$$

Les fonctions de coût les plus répandues sont celle du MAP (Maximum A Posteriori) et celle du MPM (*Maximum Posterior Mode* [Marroquin 1987]) :

$$\begin{cases} \text{loss}_{\text{MAP}}(l, t) &= \mathbb{I}_{t \neq l} \\ \text{loss}_{\text{MPM}}(l, t) &= \sum_{i \in \mathcal{E}} \mathbb{I}_{t_i \neq l_i} \end{cases} \quad (10.9)$$

où \mathbb{I}_c est une fonction indicatrice qui vaut 1 si la condition c est vraie et 0 sinon.

La première loss_{MAP} évalue de la même façon toutes les configurations erronées alors que la seconde loss_{MPM} est fonction du nombre d'erreurs locales. En fait, le risque bayésien associé au critère loss_{MAP} s'écrit : $R_{\text{MAP}}(l) = \sum_{t \neq l} P(L = t|D) = 1 - P(L = l|D)$. La minimisation de ce risque se traduit donc par le critère classique du maximum a posteriori :

$$l_{\text{MAP}}^* = \arg \max_l P(L = l|D) \quad (10.10)$$

Dans le cas du critère loss_{MPM} , $R_{\text{MPM}} = \sum_{i \in \mathcal{E}} \sum_t P(L = t|D)\mathbb{I}_{t_i \neq l_i}$ donc :

$$\begin{aligned} l_{\text{MPM}}^* &= \arg \min_l R_{\text{MPM}} = \arg \max_l \sum_i \sum_t P(L = t|D)\mathbb{I}_{t_i = l_i} \\ &= \arg \max_l \sum_i P(L_i = l_i|D) = \left\{ \arg \max_{l_i} P(L_i = l_i|D) \right\}_{i \in \mathcal{E}} \end{aligned} \quad (10.11)$$

Ici, $P(L_i = l_i|D)$ peut s'obtenir par *Belief Propagation* (section 10.3.2.3) ou par MCMC en remarquant que $P(L_i = l_i|D) = \sum_{L_{\neq i}} P(L_i|L_{\neq i}D)P(L_{\neq i}|D) = \mathbb{E}[P(L_i|L_{\neq i}D)]$. Cette approche revient à échantillonner l'espace des configurations possibles pour L par un recuit simulé (section 10.3.2.2) dont le paramètre de température T est fixé arbitrairement à 1 (il est envisageable d'optimiser le choix de T pour favoriser une bonne exploration [Fielding 2000]).

Nous avons largement privilégié l’approche MAP dans cette thèse. Une étude comparative avec le critère MPM a été réalisée dans [Perrot 2008] et présentait des résultats sensiblement similaires. Il serait éventuellement possible de définir une fonction de perte dépendant des labels considérés afin d’en privilégier certains au détriment d’autres.

10.3.2 Méthodes d’optimisation

Dans les méthodes à suivre, nous n’avons pas encore eu l’occasion de tester la méthode du *Graph cut* ou l’approche variationnelle (approximation en champ moyen) dans le cadre de notre application. Elles sont évoquées ici parce qu’elles pourraient présenter un intérêt pour la suite de nos travaux.

10.3.2.1 Méthodes des modes conditionnels itérés

La méthode des modes conditionnels itérés (*Iterated Conditional Modes* : ICM [Besag 1986]) est une approche gloutonne et localement déterministe. Elle consiste à itérer le procédé suivant :

- choisir au hasard un site i du champ considéré.
- tester toutes les valeurs possibles pour x_i et retenir celle qui possède l’énergie la plus basse.

Si le problème considéré est suffisamment simple (fonction localement convexe autour d’une bonne initialisation), elle peut suffire. Par ailleurs, elle s’utilise également après une première méthode d’optimisation stochastique (comme le recuit simulé) afin d’accélérer la convergence finale.

10.3.2.2 Recuit simulé

Le recuit simulé [Kirkpatrick 1983] est une méthode générique de minimisation d’énergie particulièrement bien adaptée aux modélisations Markoviennes puisque les évaluations y sont essentiellement locales et donc rapides à réaliser comme nous pourrions le constater dans cette présentation. Sous un angle probabiliste, le champ de Gibbs associé à l’énergie E s’écrit : $P(X) = \frac{1}{Z} e^{-E(X)}$. Le principe de cette méthode s’appuie sur un paramètre de température T qui contrôle artificiellement la forme de la distribution considérée et permet de la faire tendre vers une loi uniforme. De cette façon, l’exploration est réalisée sur $P_T(X) = \frac{1}{Z_T} e^{-\frac{E(X)}{T}}$ qui possède le même minimum global que $P(X)$ quel que soit T . On notera que $P_\infty(X) = U(X)$ correspond à une loi uniforme alors que $P_0(X = x) = \delta(x - x^*)$ correspond à un Dirac localisé sur le minimum global x^* de l’énergie (s’il est unique). À partir de ce constat, la distribution P_T est d’abord échantillonnée à une température élevée qui permet une exploration assez uniforme de l’espace des configurations. Ensuite la température est régulièrement abaissée ce qui restreint petit à petit la zone de recherche à des modes de l’énergie de plus en plus localisés.

L’échantillonnage peut être réalisé par de nombreuses méthodes différentes. Les approches les plus standards restent l’échantillonnage de Gibbs ou de Métropolis-Hasting.

Échantillonnage de Gibbs. Dans le cas où l’on dispose d’un modèle complet de la distribution des données, il est possible d’utiliser l’échantillonnage de Gibbs. Tour à tour, chaque variable X_i est échantillonnée par rapport aux valeurs des autres variables $X_{\neq i}$ laissées fixes. Dans le cas d’un champ de Gibbs, cela revient à itérer l’échantillonnage de x_i selon la probabilité $P(X_i = x_i | X_{\neq i})$,

avec :

$$\begin{aligned}
P(X_i = x_i | X_{\neq i}) &= \frac{P(X_i = x_i; X_{\neq i})}{\sum_x P(X_i = x; X_{\neq i})} \\
&\propto \Phi_i(x_i) \prod_{i \sim j} \Psi_{ij}(x_i, x_j) \\
&= \exp(-E_{i\partial i}(X)) \\
&\propto \exp(-\Delta E_{i\partial i}(X))
\end{aligned} \tag{10.12}$$

avec $\Delta E_{i\partial i}(X) = E_{i\partial i}(X) - E_{i\partial i}(X_0)$ où X_0 correspond à une configuration ne différant que pour le site i concerné par l'échantillonnage. Ainsi, l'avantage de cette approche est de nécessiter uniquement une évaluation locale des potentiels. Dans le cas où le nombre de configurations locales impliquant un site et ses voisins reste limité, il est envisageable d'en précalculer les potentiels afin d'accélérer le processus d'échantillonnage. Cette approche a pu être suivie dans le cas de la modélisation Markovienne de la variabilité sulcale proposée en section 11.

Échantillonnage de Métropolis-Hastings. Lorsqu'il n'est pas possible d'échantillonner directement une distribution P par l'approche précédente, on peut toujours avoir recours à l'échantillonnage de Métropolis-Hastings. Il repose sur une fonction de transition $T(x, y)$ définie par l'utilisateur. Elle évalue la probabilité de passer d'une configuration x à une configuration y . L'algorithme d'échantillonnage consiste à tirer $\alpha \sim U(0, 1)$ (loi uniforme) et accepter une proposition de changement y depuis une configuration x si la condition suivante est remplie :

$$\frac{P(y)T(x, y)}{P(x)T(y, x)} > \alpha \tag{10.13}$$

L'efficacité de l'algorithme est fonction du choix de T et s'améliore si T suit la distribution à échantillonner, c'est à dire si $T(y, x) \approx P(y)$. Dans le cas symétrique ($T(x, y) = T(y, x)$), le rapport de vraisemblance $P(y)/P(x)$ d'un champ de Gibbs s'exprime de façon assez simple :

$$\frac{P(y)}{P(x)} = e^{E(x) - E(y)} \tag{10.14}$$

et ne dépend pas de la constante de normalisation Z . Comme dans le cas de Gibbs, si les changements sont réduits à un unique site, le rapport de vraisemblance s'exprime localement également.

10.3.2.3 Loopy Belief Propagation

La méthode *Belief Propagation* [Pearl 1982] a été formulée initialement pour calculer de façon efficace des probabilités jointes ou marginales sur modèles graphiques arborescents. Dans le cas de graphes (dépendance avec boucle), cette méthode est appelée *Loopy Belief Propagation* et diffère uniquement par le fait qu'il est nécessaire d'itérer jusqu'à convergence (qui n'est pas garantie) des mesures de «croyances», alors que dans le cas précédent le nombre d'itérations dépend directement de la structure de l'arbre. Cette approche s'est popularisée par son succès pratique [Murphy 1999] (notamment dans le cadre des codes correcteurs d'erreurs). Elle a finalement trouvé des fondements théoriques en physique statistique dans l'approximation de Bethe et de Kikuchi de l'énergie libre du graphe [Yedidia 2001], qui a permis le développement d'approches plus générales où les approximations sont régionales et non locales (*Generalized Belief Propagation* [Yedidia 2005]).

Il s'agit d'une méthode itérative, locale, où chaque site i propage des messages m_{ij} à chacun de ses voisins $j \in \partial i$ dans le graphe, fonctions de leurs potentiels de clique respectifs. Dans le cas d'un

champ de Markov X vérifiant l'équation 10.3, le schéma itératif suit la procédure suivante :

$$\left\{ \begin{array}{l} m_{ij}(x_j) \leftarrow \propto \sum_{x_i} \Psi_{ij}(x_i, x_j) \Phi_i(x_i) \prod_{\substack{k \in \partial i \\ k \neq i}} m_{ki}(x_i) \\ b_i(x_i) \leftarrow \propto \Phi_i(x_i) \prod_{k \in \partial i} m_{ki}(x_i) \\ b_{ij}(x_i, x_j) \leftarrow \propto \Psi_{ij}(x_i, x_j) \Phi_i(x_i) \Phi_j(x_j) \prod_{\substack{k \in \partial i \\ k \neq i}} m_{ki}(x_i) \prod_{\substack{k \in \partial j \\ k \neq j}} m_{kj}(x_j) \end{array} \right. \quad (10.15)$$

Le processus part de messages $m_{ij}(x_j)$ uniformes pour toutes les paires de sites (i, j) et de propositions x_j . La procédure peut se formuler sous forme matricielle pour un traitement efficace. Finalement, cette méthode approche les probabilités marginales suivantes : $P(X_i = x_i) \approx b_i(x_i)$ et $P(X_i = x_i, X_j = x_j) \approx b_{ij}(x_i, x_j)$ à partir desquelles il est possible d'inférer une réalisation optimale du champ au sens du critère MPM (section 10.3.1) par exemple.

Cette approche fournit une solution approchée dont l'intérêt principal est sa rapidité d'exécution. Cette méthode est séduisante parce qu'elle mime dans une certaine mesure la démarche suivie par l'expert en anatomie lors d'une labélisation manuelle. En effet, la solution est construite dans les deux cas de proche en proche à partir des zones les plus fiables.

10.3.2.4 Graph cut

L'approche du *Graph cut* [Boykov 2001] provient du domaine de la théorie des graphes. Sous certaines conditions sur l'expression des potentiels de clique utilisés [Kolmogorov 2002], cette méthode s'avère relativement efficace sur certains champs de Markov cachés dont l'optimisation est difficile (nombreux minima locaux) puisqu'elle permet des mouvements complexes (de nombreux labels peuvent être modifiés en une seule fois contre un seul dans les autres méthodes) dans l'espace des configurations. Il s'agit d'une approximation dès que le nombre de labels est supérieur à 2, mais qui converge généralement en un nombre d'itérations assez limité.

L'approche se base sur deux types de mouvement dans l'espace des configurations. Le premier, α - β *swap* consiste à réaliser des échanges de labels entre les sites portant le label α ou le label β . Le second, α *expansion* consiste à étendre le nombre de sites portant le label α . La méthode consiste finalement à itérer pour chaque label ou paires de labels ces deux types de mouvement, en choisissant à chaque fois les modifications de labels minimisant l'énergie à optimiser. Le mouvement optimal est calculable de façon très efficace en utilisant un algorithme de calcul de coupe minimale sur un graphe.

Cette approche semble inutile dans le cas du modèle Markovien proposé dans cette partie (relations entre segments voisins) où l'information reste très localisée. Dans le cas de l'approche «historique» où les relations sont entre les structures sulcales, on peut espérer une accélération du processus d'optimisation puisque l'énergie considérée comporte de nombreux minima locaux.

10.3.2.5 Approche variationnelle

Les techniques variationnelles permettent d'approcher un modèle complexe $P(X)$ (un champ de Markov dans notre cas), par un modèle simplifié Q dont l'expression est estimée en minimisant la divergence de Kullback-Leibler entre Q et P . De cette façon, l'inférence ou l'estimation de paramètres se retrouvent simplifiées sur le modèle Q .

Le modèle Q le plus simple consiste à «casser» tous les liens Markovien du modèle P , tel que $Q(X) = \prod_{i \in \mathcal{E}} q_i$, et donne lieu à l'approximation standard en champ moyen (*mean field* [Chandler 1987, Parisi 1988]). Cette approximation s'appuie sur l'hypothèse suivante : à l'équilibre, chaque variable X_i ne dépend plus que de l'espérance $\mathbb{E}[X_k]$ de chacun de ses voisins $k \in X_{\partial i}$ dont on néglige les fluctuations et s'écrit par conséquent :

$$P(X) \approx \prod_{i \in \mathcal{E}} P(X_i | \mathbb{E}[X_{\partial i}]) \quad (10.16)$$

Cette approximation nécessite l'estimation d'espérances qui se calculent par un processus itératif, qui s'exprime dans le cas de labels discret de la façon suivante : $\mathbb{E}[X_i]^{(n+1)} = \arg \max_{x_i} P(X_i = x_i | \mathbb{E}[X_{\partial i}]^{(n)})$ à partir d'une initialisation quelconque $\mathbb{E}[X_{\partial i}]^{(0)}$.

Cette formulation fournit une approximation directe des marginales : $P(X_i) \approx P(X_i | \mathbb{E}[X_{\partial i}])$ que l'on retrouve dans l'apprentissage semi-supervisé (section 4.1.2) ou l'approche jointe recalage et labélisation (section 7.3 et section 13.2.3 pour l'intégration du champ de Markov).

Sillons corticaux et champ de Markov

La littérature sur les champs de Markov est conséquente, mais se cantonne essentiellement au domaine du traitement d'images. Une grande majorité des applications rencontrées partagent une partie de leur modélisation. L'unité élémentaire à labéliser est généralement le pixel ou le voxel, dont la taille ne varie pas à travers l'image. Généralement, la structure du graphe de relations du champ de Markov découle naturellement du maillage régulier qui décrit l'image, où seule l'amplitude du voisinage peut différer. La plupart du temps, on considère que les propriétés du champ sont homogènes sur toute l'image, ce qui signifie que les potentiels de cliques utilisés ne dépendent que du type de clique mais pas de l'emplacement spatial de la clique dans l'image. Les champs de Markov sont essentiellement utilisés pour imposer des contraintes de régularité (au sens des potentiels du champ) ce qui en fait un outil de choix pour régulariser ou débruiter une image ou un champ dérivé de l'image qu'il soit connu ou inconnu (dans le cas de la segmentation d'images). Une modélisation Markovienne très populaire à cette problématique, consiste à utiliser un modèle d'Ising (2 classes) ou de Potts (plus de 2 classes) qui permet de forcer dans une certaine mesure 2 pixels voisins à partager la même classe. Il s'agit d'une modélisation provenant du domaine de la physique statistique et correspond à la modélisation de l'orientation d'un champ de spins.

En imagerie médicale, l'utilisation classique (au sens du paragraphe précédent) du modèle de champ de Markov est courante. Par contre, dans le cadre de la modélisation de la variabilité sulcale il reste encore très peu utilisé. Nous détaillerons les quelques travaux qui à notre connaissance profitent d'une telle modélisation.

Dans [Fischl 2004], le problème de labélisation de structure est posé sur toute la surface corticale, où au préalable des a priori quant à la localisation des structures ou à leurs propriétés de courbure ont été estimés suite à une projection sphérique d'un ensemble de données. L'inférence des labels est réalisée au niveau de chaque triangle du maillage cortical en tenant compte d'un a priori Markovien non stationnaire basé sur les co-occurrences de labels voisins spatialement. La force de cette modélisation Markovienne réside principalement dans sa non-stationnarité qui lui permet de tirer partie des spécificités locales, dont la qualité dépend fortement de la qualité de la projection sphérique. La structure du graphe et le système de voisinage quant à eux découlent naturellement du support utilisé : le maillage cortical. L'auteur utilise la méthode *icm* (détaillée en section 10.3.2.1) pour réaliser l'inférence des labels. Dans cet exemple, on ne s'éloigne pas trop de la modélisation Markovienne classique utilisée couramment en traitement d'images et la modélisation Markovienne ne permet pas de faire beaucoup plus que des ajustements relativement localisés.

Dans les travaux de [Shi 2007, Shi 2009], il s'agit d'extraire les lignes sucales de huit sillons majeurs. À partir d'une squelettisation de la surface corticale [Shi 2008], une collection d'extrémités potentielles est extraite ainsi qu'un ensemble de courbes candidates les reliant. Cette approche permet de réduire considérablement l'espace de recherche. La restriction de l'étude à 8 grands sillons permet à l'auteur de fixer a priori une structure de relation Markovienne entre les différents sillons (bas de la figure 11.1). L'auteur se restreint ici à un arbre de dépendance ce qui lui fournit un cadre optimal d'utilisation d'inférence par la méthode *Belief Propagation* (description en section 10.3.2.3). De façon similaire, à l'approche suivie par le modèle «historique» [Rivière 2002], les potentiels du champ de Markov sont dérivés de classifieurs locaux. Il s'agit en l'occurrence d'une version

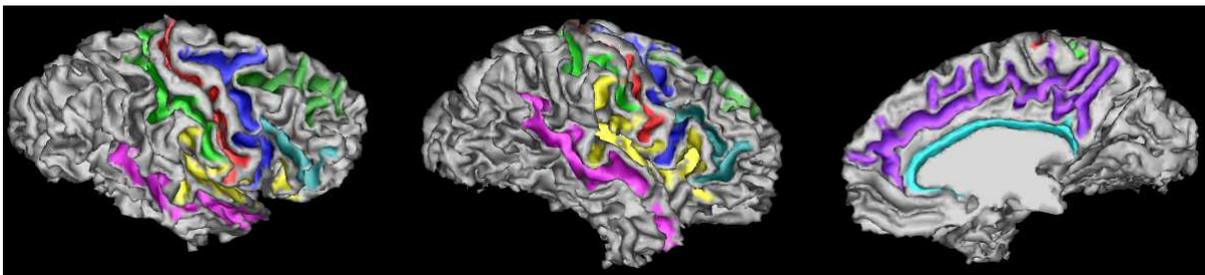
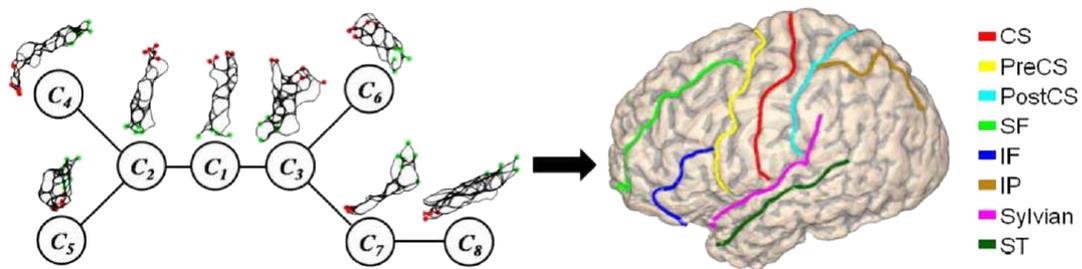
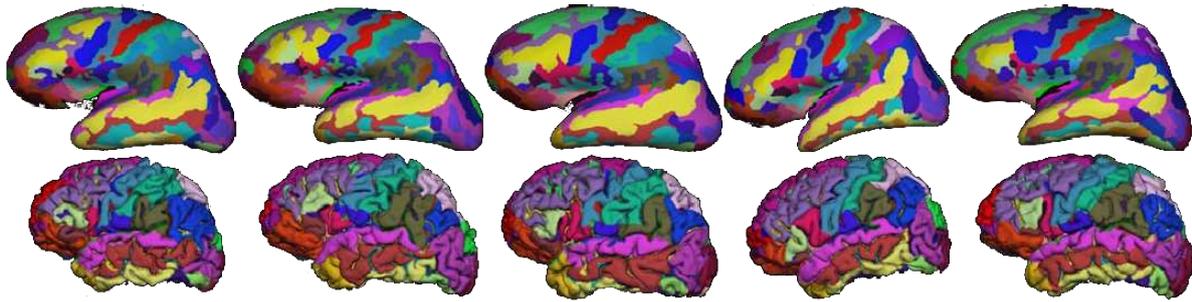


FIG. 11.1 – Modélisation Markovienne et identification des sillons corticaux. **Haut** : labélisation des structures sulcales et gyrales sur la surface contrainte par un champ de Markov sur les labels [Fischl 2004]. **Centre** : modèle graphique sur 8 sillons majeurs proposé dans [Shi 2009]. L'espace d'états sur lequel sont optimisées les lignes sulcales est composé de l'ensemble des combinaisons possibles d'un nombre limité de lignes candidates. **Bas** : labélisation de *patches* surfaciques sulcaux par la mise en correspondance structurelle d'un cortex avec un graphe-modèle suivant une modélisation Markovienne [Yang 2009].

approchée de la régression logistique basée sur la combinaison de classifieurs faibles (*weak classifiers*) par la méthode AdaBoost. Une description multi-échelles, dérivée de la transformée en *wavelet* de Haar des données sulcales, est utilisée pour définir un ensemble de descripteurs qui serviront à la discrimination dans le modèle proposé. Pour un label donné, le classifieur vise à distinguer les lignes sulcales issues des données d'apprentissage de lignes candidates (jouant le rôle de contre-exemples, deuxième point commun avec l'approche «historique») suffisamment dissemblables des véritables données. Le même principe est suivi pour les paires de labels en relation, où cette fois les formes jointes de paires de structures sont à distinguer des formes jointes de paires de lignes candidates. Cette méthode exploite des informations haut niveau à l'échelle des sillons et des paires de sillons, mais semble avoir fortement canalisé l'espace des solutions possibles en ne considérant non pas un espace d'états composé de labels, mais de sillons candidats. La difficulté principale de cette approche étant de garantir que la véritable configuration se trouve parmi les candidates tout en limitant le plus possible le nombre de propositions.

Dans [Yang 2007, Yang 2009], les auteurs ont proposé une méthode où chaque sujet est représenté par un graphe décrivant ses structures corticales. Elle vise à mettre en correspondance ces structures avec un «graphe-modèle» décrivant des propriétés à l'échelle d'un groupe de sujets. Bien que les auteurs présentent leurs travaux sous la forme d'une méthode de *graph matching*, il s'agit également d'une formulation Markovienne des liens statistiques entre données et labels, où les potentiels s'expriment à l'échelle des structures haut niveau (sillons et relations entre paires de sillons). Leur approche s'appuie sur une représentation sous forme de *patches* surfaciques des plissements corticaux. Une projection sphérique des surfaces corticales permet de définir une paramétrisation commune au travers des individus et ainsi y définir des statistiques sur la localisation des labels (sorte de SPAMs surfaciques, sans lissage ici). Les sillons sont décrits par leur forme (moments invariants [Mangin 2003b, Sun 2007]), leur orientation (axes principaux). Dans cette approche 2 structures sont en relation si les parcelles surfaciques correspondantes se touchent (après dilatation géodésique). Le voisinage est alors caractérisé par la fréquence de mise en relation de 2 structures. Leur modélisation est descriptive puisqu'elle suppose une modélisation Gaussienne de chaque descripteur. Cette approche partage ainsi certains préceptes à la base du modèle «historique» [Rivière 2002].

Dans notre situation, la labélisation est réalisée sur un ensemble de segments sulcaux. Il n'y a donc pas de transposition directe des approches suivies par les travaux présentés dans les paragraphes précédents. L'approche proposée par [Shi 2009] est séduisante et nous envisageons de l'explorer dans de futurs travaux pour la prise en compte d'informations haut niveau. Mais dans un premier temps, nous souhaitons limiter notre étude aux informations locales pour simplifier le modèle et l'espace des configurations à explorer. Ce qui devrait nous fournir des indices intéressants pour la suite.

La description de notre modélisation débutera par un descriptif général du modèle, suivi par la construction du graphe de relations, support du champ de Markov, qui indique l'influence du choix du label d'un segment sulcal sur le reste de la labélisation. Nous détaillerons quelques propositions originales de potentiels pour les cliques de relations. Cette partie n'apporte aucune solution révolutionnaire mais tente simplement de combler certaines lacunes du modèle précédent [Rivière 2002].

11.1 Modèle

Nous reprenons ici notre modélisation précédente introduite en partie I. Dans celle-ci une indépendance statistique était supposée entre chaque structure à labéliser. En réalité cette supposition est fautive et on peut lui imputer certaines erreurs systématiques constatées sur les labélisations automatiques. En particulier, les branches des sillons constituent des rameaux secondaires dont la forme, l'emplacement ou la direction ne sont pas réellement caractéristiques d'un label particulier. Même si elles sont fortement reliées à une structure principale bien identifiée, cette information reste

ignorée par le modèle qui suppose l'indépendance des segments entre eux.

L'interdépendance est avérée et s'exprime de différentes manières sur les données corticales étudiées :

(1) La cohérence spatiale des plissements limite le nombre de labels voisins les uns des autres. La fréquence d'apparition des labels portés par des paires de segments voisins pourrait être contrôlée par une modélisation Markovienne de $P(L|E)$ ¹. Le graphe de voisinage (contenu dans la variable E) apporte une première réponse à la modélisation de ces dépendances entre structures, mais n'exploite pas les informations précieuses fournies par les descripteurs D .

(2) En réalité, la forme, l'emplacement et l'orientation de chaque structure sont contraints par les autres (principalement par celles qui l'avoisinent). Ce constat se vérifie à l'échelle des segments sulcaux, même si c'est dans une moindre mesure. Cette information s'exprime sur l'ensemble des données D elles-mêmes. Cela implique que certains descripteurs sont corrélés entre deux segments voisins : comme la localisation, la profondeur ou l'ouverture des plissements et éventuellement la direction pour les sillons aux formes les plus simples. Pour ce type de descripteurs, $P(D)$ serait Markovien.

(3) Cette information peut être affinée si elle est conditionnée spatialement ou encore mieux si elle est conditionnée par les labels (qui eux-mêmes ont un support spatial restreint), dans ce cas : $P(D|L)$ est Markovien.

Toutes les interdépendances constatées ci-dessus existent, mais elles ne sont pas toutes exploitables.

Dans le cas (1), $P(L|E)$ nous pensons que l'information fournie par la structure du graphe (codée dans la variable E) ne suffit pas à caractériser les co-occurrences de labels. Il ne faut pas oublier que dans notre cas, les graphes sont très différents d'un sujet à l'autre et leur découpage en segments sulcaux reste plutôt sensible au bruit d'acquisition. De ce fait, un apprentissage des co-occurrences devrait être contraint à l'extrême afin d'éviter autant que possible le risque de sur-apprentissage. À l'issue d'un tel processus peu d'informations discriminantes risquent d'être encore présentes.

Le cas (2), « $P(D)$ Markovien» ne nous intéresse pas directement. En effet il n'est pas relié aux labels et n'intervient donc pas dans le processus de labélisation. Par contre, si l'on suppose $P(D|L)$ Markovien (cas (3)), alors $P(D) = \sum_L P(D|L)P(L)$ l'est également sous certaines conditions ($P(L)$ Markovien ou $P(L) = \prod_i P(L_i)$ par exemple). Cette formulation permet d'utiliser des données D pertinentes, discriminantes et idéalement peu sensibles à l'effet de la segmentation et ainsi contraindre la labélisation des segments en relation.

Finalement la modélisation retenue s'exprime sous la forme d'un champ de Markov conditionnel aux labels et s'écrit formellement de la façon suivante :

$$\begin{aligned}
 P(\text{Labélisation}|\text{Données}) &= P(L|D; \mathcal{M}) = \frac{P(L|\mathcal{M})P(D|L; \mathcal{M})}{P(D|\mathcal{M})} \\
 &= \frac{P(L|\mathcal{M})}{Z(\mathcal{M})} \prod_{i \in \mathcal{E}} P(D_i|L; \mathcal{M}) \prod_{i \sim j \in \mathcal{E}} P(D_{ij}|L; \mathcal{M}) \quad (11.1) \\
 &\propto P(L|\mathcal{M}) \prod_{i \in \mathcal{E}} P(D_i|L_i; \mathcal{M}) \prod_{i \sim j \in \mathcal{E}} P(D_{ij}|L_i L_j; \mathcal{M})
 \end{aligned}$$

avec la constante de normalisation $Z(\mathcal{M}) = \sum_L P(L|\mathcal{M}) \prod_{i \in \mathcal{E}} P(D_i|L; \mathcal{M}) \prod_{i \sim j \in \mathcal{E}} P(D_{ij}|L; \mathcal{M})$ qui n'intervient pas dans la phase d'inférence. Cette fois $D = \{D_i, D_{ij}; (i, j) \in \mathcal{E}^2\}$. C'est à dire une

¹Ici on reprend la notation introduite en section 4.3 où E désigne certaines informations directement liées à la segmentation des données corticales en segments sulcaux, à savoir la structure même du modèle : les segments et les relations entre segments à la base du graphe, support du champ de Markov.

description D_i de chaque segment i et une description D_{ij} d'informations partagées entre les segments i et j , de façon à ce que ces informations soient indépendantes entre elles conditionnellement à la labélisation L . Ici $i \sim j$ fait référence à l'ensemble des paires de segments en relation dans le graphe.

Dans cette formulation, les composantes $P(D_i|L_i; \mathcal{M})$ correspondent idéalement aux modèles génératifs de segments. L'ensemble des modèles locaux présentés dans le cadre de la modélisation indépendante (partie I) peuvent prendre place ici. Dans la suite, nous focaliserons notre étude sur le modèle SPAM qui a donné lieu aux meilleurs résultats avec la modélisation indépendante.

11.1.1 Apprentissage

Les paramètres du modèle \mathcal{M} devraient être estimés idéalement de façon couplée puisqu'ils sont reliés entre eux par la constante de normalisation $Z(\mathcal{M})$ qui en dépend. Puisque la structure du graphe dépend du sujet considéré, la constante de normalisation en dépend également. Chacune de ces constantes de normalisation est très difficile à calculer puisqu'elle implique une sommation sur l'ensemble des labélisations possibles. Leurs valeurs dépendent des paramètres du modèle de telle façon qu'il faudrait les recalculer à chaque proposition de paramètres fournie par une méthode d'optimisation quelconque. Pour l'instant, nous avons suivi une approximation grossière souvent utilisée mais dont il est difficile de mesurer l'impact. Elle consiste à négliger la dépendance entre constante de normalisation et paramètres du modèle. Ceci permet un découplage des paramètres du modèle : les potentiels (qu'ils soient au niveau des segments ou des relations entre segments) sont alors estimés indépendamment les uns des autres. Il est probable que cette approximation soit à l'origine de certains déséquilibres constatés entre les différents potentiels (nous reviendrons sur ce sujet en section 11.2.2).

11.1.2 Inférence

Dans le cas d'une modélisation Markovienne au niveau des sillons (non traité dans cette partie, voir modèle «historique» en introduction 2), le nombre de propositions de sillons possibles à partir des segments candidats à un label donné peut aller de quelques centaines à quelques milliers de possibilités. Pour une paire de labels, plusieurs millions de propositions sont possibles, chacune à évaluer pour chaque paire de labels possibles (presque 4000), soit au total plusieurs dizaines de Gigaoctets de données. Dans une telle situation, le calcul des descripteurs et des potentiels associés ne peut se faire qu'au fur et à mesure des propositions.

Dans notre modélisation, les liens Markoviens se situent à l'échelle des segments et présentent un avantage notable par rapport à des liens au niveau des sillons. En effet, ils limitent considérablement la combinatoire associant labels et structures en relation. Pour chaque paire de segments en relation (il y en a environ un millier en général), les potentiels associés à chaque paire de labels peuvent être directement précalculés à partir des descripteurs de relations. Au total, seulement quelques millions d'informations doivent être stockées en mémoire, c'est à dire quelques Mégaoctets. On peut encore réduire cette quantité d'informations en filtrant spatialement les labels possibles (en se basant sur le modèle SPAM présenté en partie I), réduisant généralement le nombre de labels en concurrence par segment à une dizaine environ.

Cette étape de précalcul apporte un gain net dans le processus d'optimisation, il a été testé avec succès dans le cas de l'optimisation des labels par maximum a posteriori basé sur un recuit simulé (section 10.3.2.2) avec échantillonnage de Gibbs, par MPM (section 10.3.1) ou par *Loopy Belief Propagation* (section 10.3.2.3). Le modèle Markovien ainsi proposé permet d'obtenir en moins de 10 minutes la labélisation correspondant au maximum global des critères proposés, contre plusieurs heures auparavant.

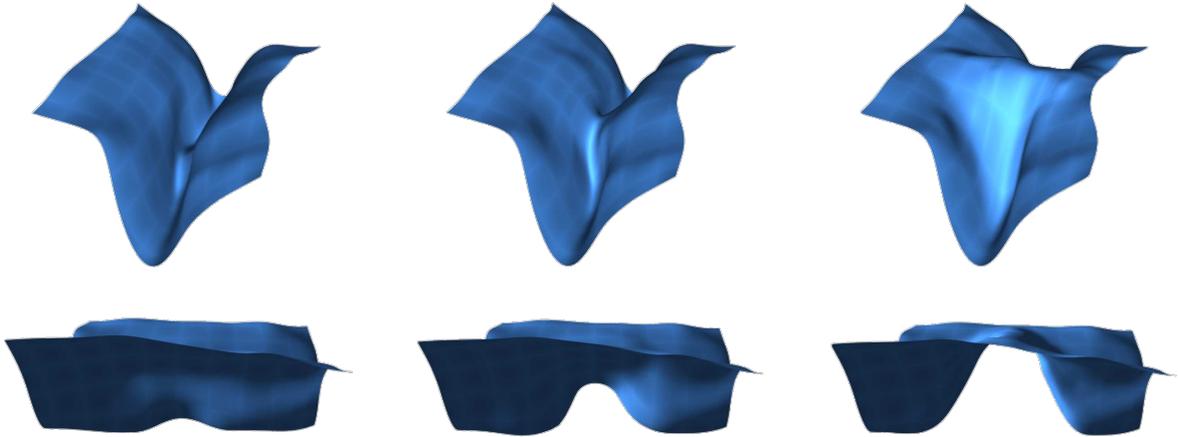


FIG. 11.2 – Exemple synthétique de la continuité entre *pli de passage* (gyrus enfoui) à gauche et *relation corticale* (gyrus séparant 2 structures sulcales) à droite.

11.2 Graphe de relations

Au delà de la modélisation Markovienne que nous suggérons dans cette partie, l'étude de la variabilité des sillons corticaux a bénéficié par le passé d'approches exploitant une représentation de la topographie corticale sous forme de graphe [Goualher 1998]. Dans [Vivodtzev 2005], les auteurs proposent un découpage du cortex en parcelles surfaciques mettant en évidence des plissements élémentaires connectés entre eux sous la forme d'une topologie reliant les structures voisines. Des techniques ad hoc de mise en correspondance de graphes (*graph matching*) permettent finalement de reporter les labels d'un sujet à un autre.

Comme nous l'avons indiqué plus tôt, dans les cas classiques d'utilisation de champ de Markov en traitement d'images, le graphe associé découle directement de la structure de l'image et d'un système de voisinage par plus proche voisin. De ce fait, le graphe est régulier et homogène sur toute l'image hormis éventuellement les bords. Le même principe est applicable à d'autres données structurées comme une surface maillée régulière.

11.2.1 Construction

Dans un modèle Markovien, le graphe de relations reflète les dépendances statistiques entre variables. Dans l'idéal, il s'agirait de répondre à la question suivante : quelles structures se sont influencées mutuellement au cours du développement cérébral. Cette question, fort intéressante au demeurant, est un domaine de recherche à part entière qui est loin d'avoir décrypté tous les mécanismes sous-jacents qui permettraient d'y apporter une réponse. Nous supposons donc simplement une influence locale que nous allons maintenant définir. Le jeu de relations que nous allons décrire correspond à celui utilisé dans l'approche «historique» [Rivière 2002].

Si on se réfère au processus d'extraction des segments sulcaux présenté précédemment (section 3.2), un jeu de relations découle directement de l'étape de sursegmentation appliquée aux structures corticales. En effet, cette étape vise à garantir autant que possible que chaque segment sulcal correspond au plus à un label. Par conséquent, elle induit de nombreux surdécoupages artificiels. Selon le critère utilisé pour le découpage, on distinguera deux types de relations. Le découpage en segments élémentaires topologiquement simples induit des relations nommées *jonctions topologiques* entre les morceaux correspondants. Le découpage dans la hauteur d'un pli au niveau d'un gyrus enfoui au fond d'un sillon sera nommé *pli de passage*.

Pour deux plissements se faisant face, on constate en pratique (sur notre base de données) que toutes les configurations existent (illustration de ces situations en figure 11.2) allant de la situation où les plissements fusionnent, engendrent un pli de passage ou sont séparés nettement par un gyrus. Le premier cas est à l'origine de la plupart des problèmes de segmentation constatés en pratique puisqu'aucun indice a priori (sans les labels) ne permet de détecter le problème. Le dernier cas par contre, n'induit aucune relation entre les deux structures par le procédé indiqué, alors que la situation semble plutôt proche du deuxième cas. Ainsi, il a semblé nécessaire de mettre en relation tous les segments séparés par un gyrus. Malheureusement, les structures gyrales sont plutôt complexes et bruitées par les nombreuses branches qui viennent en perturber la caractérisation. Quand bien même ces structures seraient clairement identifiées, dans une situation où deux sillons sont parallèles, faudrait-il mettre en relation tous les segments sulcaux les composant ? Cela aurait pour conséquence directe d'alourdir le modèle et sa manipulation. Au niveau local (au sens spatial du terme), la dépendance statistique s'exprime de façon visible comme l'influence mécanique du processus de plissement de chaque sillon sur ses voisins.

Pour répondre à ces besoins une procédure assez simple a été suivie, permettant de ne conserver qu'un nombre limité de relations. Ce dernier type de relations, sera nommé *relation corticale*. Il s'agit d'une approche standard permettant de déduire simplement un graphe de voisinage local à partir d'un ensemble de données organisées spatialement. Classiquement cette adjacence est déduite d'un diagramme de Voronoï appliqué à un ensemble de points. Ici, la procédure est appliquée à partir des segments sulcaux (définis par un ensemble de voxels dans le volume IRM du sujet), géodésiquement à l'enveloppe du cerveau [Mangin 1995d].

11.2.2 Graphes irréguliers

Comme nous l'avons indiqué précédemment, dans les cas d'utilisation classique des champs de Markov en traitement d'images, le modèle graphique associé s'appuie naturellement sur la structure de voisinage des pixels ou des voxels et prend alors la forme d'une grille régulière. Dans un tel graphe, chaque noeud est relié à un nombre de voisins constant (sauf éventuellement sur les bords de l'image). De cette façon, l'importance relative des potentiels dédiés aux arêtes ou aux noeuds reste homogène sur tout le graphe.

Dans la littérature, on retrouve principalement deux situations dans lesquelles la structure d'un champ de Markov est irrégulière. La première vise à simplifier une représentation structurée (image [Turtinen 2006, Nwogu 2008] ou maillage par exemple) sous la forme d'un ensemble de régions connectées à leurs voisines. La seconde correspond à des modélisations dédiées à des applications spécifiques. L'inférence ne souffre pas de l'irrégularité du graphe tant que les potentiels respectent l'importance relative à accorder à chaque noeud du graphe.

Notre cas d'utilisation est proche de la première situation pour laquelle la grande majorité des études ne considère pas les difficultés éventuelles liées à l'irrégularité du graphe. Les seules références dont nous avons connaissance, se limitent à la modélisation d'interactions spatiales par des modèles autorégressifs (CAR ou SAR), il a été montré que l'irrégularité du système de voisinage privilégiait artificiellement certains noeuds du graphe [Wall 2004]. Dans [Cla 1992], l'auteur suggère de pondérer chaque lien par le nombre de sites le reliant (et dans le cas de CAR la contrainte de symétrie impose également une telle pondération sur les variances conditionnelles locales). Dans notre cas, le graphe résulte d'une segmentation en segments dont les tailles et l'agencement sont inhomogènes, expliquant ainsi l'irrégularité des systèmes de voisinage (voir figure 11.3). Nous allons maintenant illustrer les difficultés induites par cette irrégularité pendant un processus d'inférence d'une labélisation par une procédure de recuit simulé par échantillonnage de Gibbs. Comme nous l'avons indiqué précédemment (section 10.3.2.2), pour un segment donné i , un label local l lui est rattaché proportionnellement à $e^{-\frac{E_i \vartheta_i(l)}{T}}$ (où T est la variable de température du recuit simulé)

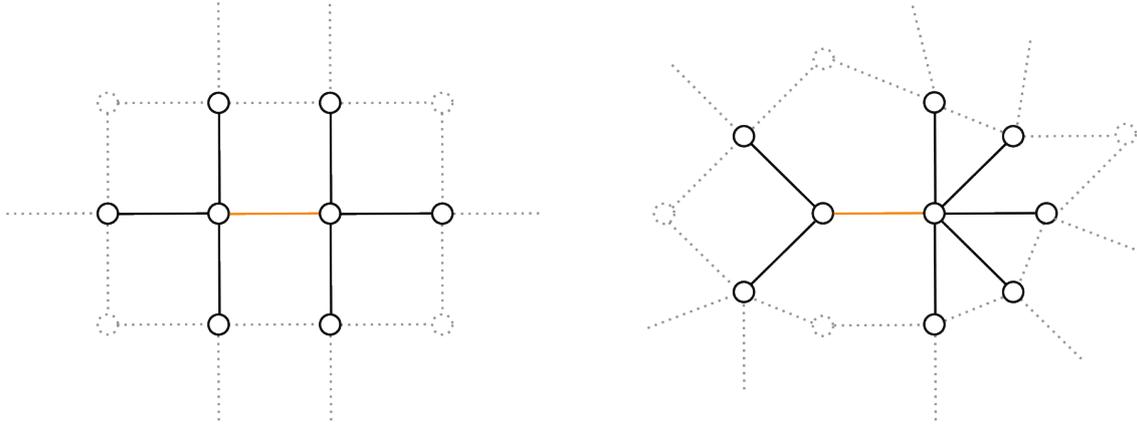


FIG. 11.3 – Régularité topologique d'un modèle graphique. **Gauche** : grille régulière standard basée sur un 4-voisinage des pixels d'une image. **Droite** : graphe irrégulier. Ici les 2 noeuds reliés à l'arête orange sont reliés à 3 et 6 voisins, créant un déséquilibre.

avec :

$$E_{i\partial i}(l) = E_i(L_i = l) + \sum_{j \sim i} E_{ij}(L_i = l, L_j = l_j) \quad (11.2)$$

avec E_i et E_{ij} les potentiels locaux du segment i et de la paire de segments (i, j) en relation dans le graphe. Ainsi, plus le nombre de relations est grand, plus il pèse dans la définition de cette énergie et donc dans la détermination du label tiré. L'influence du potentiel E_i peut alors devenir négligeable. Dans notre application, on constate en pratique que plus le nombre de relations est grand, plus il y a de chance qu'ils rentrent en contradiction, ce qui favorisera un tirage uniforme des labels en compétition, alors que l'information portée par E_i seule pourrait être fortement discriminante. Si l'on interprète le graphe de relations comme la discrétisation de liens continus sous-jacents (dépendance statistiques entre les voxels du cortex ou les nœuds du maillage corticale), il semblerait naturel que chaque énergie locale soit pondérée à hauteur de la part d'information qu'elle représente. Cette remarque suggère une reformulation de l'énergie locale de la façon suivante :

$$E_{i\partial i}(l) = w_i E_i(L_i = l) + \sum_{j \sim i} w_{ij} E_{ij}(L_i = l, L_j = l_j) \quad (11.3)$$

où nous suggérons $w_i = \sum_{j \sim i} w_{ij}$ pour équilibrer² l'importance de chaque noeud du graphe par rapport aux arêtes auxquelles il est rattaché. Dans cette formulation, il semble naturel de poser la contrainte $w_{ij} = w_{ji}$ puisque l'énergie E_{ij} correspondante n'est pas liée à une direction particulière. Cette formulation équilibre chaque voisinage local, mais déséquilibre l'importance relative des potentiels rattachés aux segments qui s'exprime par la pondération w_i . Posons maintenant $\bar{E}_{i\partial i}(l) = \frac{E_{i\partial i}(l)}{w_i}$, une version normalisée de notre énergie où l'importance moyenne d'un noeud ou de l'ensemble des relations auxquelles il est rattaché vaut 1. L'échantillonnage de Gibbs précédent s'exprime alors comme un tirage proportionnellement à $e^{-\frac{E_{i\partial i}(l)}{T}} = e^{-\frac{\bar{E}_{i\partial i}(l)}{w_i}}$. Le facteur de pondération w_i revient à considérer chaque segment exploré par le recuit à des températures différentes

²Comme nous l'avons indiqué plus tôt, l'importance relative à accorder à chaque modèle (leur entropie relative en quelque sorte) découle normalement de l'estimation couplée des paramètres du modèle \mathcal{M} . Nous avons choisi de réaliser leur apprentissage de façon indépendante afin de simplifier le processus. Par conséquent nous ne disposons pas de cette pondération naturelle entre les différents modèles.

impliquant une exploration inhomogène de l'espace des labélisations. C'est pourquoi nous suggérons d'utiliser directement $\bar{E}_{i\partial i}(l)$ à la place de $E_{i\partial i}(l)$.

Concernant le choix des pondérations, aucune réponse évidente ne s'est imposée de fait à partir de la méthode de création du graphe utilisée. Pour apporter une réponse nette et unique à cette question, il faudrait sans doute revenir à l'expression originale continue des relations et suivre une approche intégrative afin de définir directement les potentiels de cliques pendant le processus d'extraction du graphe et non a posteriori comme c'est le cas actuellement. Dans l'attente d'une telle étude, nous proposons néanmoins plusieurs façons de pondérer ces liens :

- L'approche faisant le moins d'a priori consiste à se baser uniquement sur le nombre de relations, ce qui revient à poser $w_{ij} = 1$ et par conséquent w_i correspond au nombre d'arêtes en contact avec le nœud i du graphe.
- On peut estimer que l'importance relative d'une relation par rapport à une autre dépend de la taille des sillons impliqués, ainsi : $w_{ij} = s_i s_j$ où s_i correspond à la taille du segment i (après normalisation, la composante j disparaît de l'énergie $\bar{E}_{i\partial i}$).
- De façon à exploiter directement une information reliée à la création du graphe, nous proposons d'utiliser l'aire de la surface de contact entre les cellules de Voronoï des deux segments impliqués dans une *relation corticale* et la longueur de la zone de contact entre deux segments se touchant dans le cas d'une *jonction topologique* ou d'un *pli de passage*.

11.3 Potentiels de relations

Dans l'expression de la modélisation Markovienne proposée, un potentiel est rattaché à chaque paire de structures en relation dans le graphe (le support du champ de Markov). Ces potentiels traduisent l'adéquation de données structurales observées sur un sujet d'étude particulier avec la variabilité constatée à l'échelle d'un groupe d'individus constituant une base d'apprentissage. À ce stade, les travaux visant l'étude de la variabilité anatomique, et notamment ceux impliquant la co-variabilité de structures (comment plusieurs structures se déforment, se déplacent de façon conjointe), permettent la définition de tels potentiels.

Les modèles de distribution de points (*Points Distribution Model*, PDM) demeurent l'outil le plus répandu pour étudier la covariabilité d'un ensemble de structures appariées sur un groupe de sujets. Ainsi, [Lohmann 2000] met en correspondance des «bassins sulcaux», [Tao 2001] des lignes sulcales en se basant sur la méthode du procrustes, alors que [Caunce 1998, Caunce 2001] réalise un appariement automatique d'amers anatomiques en utilisant une variante de la méthode ICP (décrite en annexe B.1) incorporant des informations structurales.

L'étude de la co-variabilité de lignes sucales a été proposée par [Royackkers 1999] au travers de la position relative de points homologues (extrémités, centre, point médian), de l'orientation relative (produit scalaire entre 2 vecteurs indiquant la direction générale de chaque structure), de connectivité (position et abscisse curviligne du point de contact, directions locales au niveau de la zone de contact). Dans [Corouge 2002], l'auteur définit un repère propre à chaque structure (directions principales) et un repère médian pour chaque paire de structures, à partir duquel sont évaluées les positions relatives des centres de gravité et les orientations relatives des repères. Plus récemment, [Fillard 2007b] a proposé d'étudier les corrélations entre structures, par une analyse de corrélation canonique (*Canonical Correlation Analysis*, CCA) de la matrice de covariance totale (inter et intra sillons).

11.3.1 Informations pertinentes

Dans la littérature, la variabilité anatomique est le plus souvent étudiée après appariement de structures identifiées : les points sont mis en correspondance par paires, les lignes et les surfaces sont paramétrisées. Notre choix de représentation, sous la forme de segments sulcaux, prend en compte toute la complexité des données sulcales (branches, interruptions). Elle ne permet pas toujours de proposer une telle paramétrisation (notamment dans le cas de sillons arborescents, c'est à dire composé d'un amas de branches comme les sillons frontaux). De plus, sur un nouveau sujet, à labels inconnus, on ne peut pas estimer une telle paramétrisation. Il serait alors nécessaire de déterminer une paramétrisation spécifique pour chaque proposition de labélisation, ce qui pourrait s'avérer extrêmement coûteux.

Dans le contexte de cette thèse, nous avons limité notre modèle Markovien à l'échelle de relations entre segments sulcaux voisins, ce qui nous permet (pour l'instant) de conserver une énergie relativement simple à optimiser. Ainsi, dans notre cas, il nous faut étudier la covariabilité structurelle de paires de labels à une échelle sous-sulcale. Cette façon de procéder est plutôt originale, mais il ne faut pas en espérer une description détaillée de l'organisation des sillons corticaux. En effet, la plupart des mesures décrites plus tôt s'expriment à l'échelle de paires de sillons et s'avèrent peu pertinentes ou très bruitées à notre niveau d'étude. Il nous faut donc nous restreindre à des mesures grossières mais robustes.

Les branches de sillons sont principalement à l'origine de ces difficultés. Dans leur grande majorité, elles peuvent s'interpréter comme un bruit venant s'ajouter au processus de plissement général, perturbant le profil idéal des lignes sulcales. Mais dans le cas des sillons arborescents, l'agencement des branches masque complètement le profil initial s'il existe. Le sillon frontal intermédiaire est le paroxysme de cette situation et ne peut être identifié qu'une fois ses voisins caractérisés (que ce soit manuellement ou automatiquement). Ainsi, leur position est très variable le long de la structure principale, d'orientation souvent orthogonale à celle-ci et perturbe toute tentative de mise en correspondance locale de points. Sans modélisation explicite des branches (ajout de sous-labels pour les différencier dans le modèle et capter des propriétés qui leurs sont spécifiques), il est nécessaire de proposer des mesures qui y sont insensibles.

Déjà problématique à l'échelle des segments, la segmentation induit une variabilité d'autant plus élevée sur la structure de graphe reliant les segments voisins. La mise en relation de deux segments résulte de leur connexion suite à une dilatation volumique (détails en section 11.2.1). Ainsi, pour 2 sillons parallèles se faisant face, une inhomogénéité de segmentation de part et d'autre peut engendrer une relation entre deux segments de tailles relativement différentes. Dans une telle situation, il semblerait que la relation entre les deux structures ne soit que partielle et qu'il faudrait dans l'idéal identifier quelles parties des deux segments sont en relation afin de limiter la définition des descripteurs de relation à cette zone. Nous ne disposons pas pour le moment d'une telle information, les mesures descriptives des relations locales devront donc s'affranchir de cette difficulté autant que possible.

En gardant à l'esprit les restrictions évoquées plus tôt, seules deux catégories d'informations semblent garder une signification à l'échelle des relations entre segments : la distance et la direction entre les segments. Nous définirons dans la section suivante comment ces informations sont intégrées pour former un potentiel de relation.

11.3.2 Description des potentiels

Les potentiels proposés dans la suite ne sont sans doute pas encore optimaux et posent encore bon nombre de problèmes de robustesse auxquels il faudra remédier par la suite. À cet effet, nous

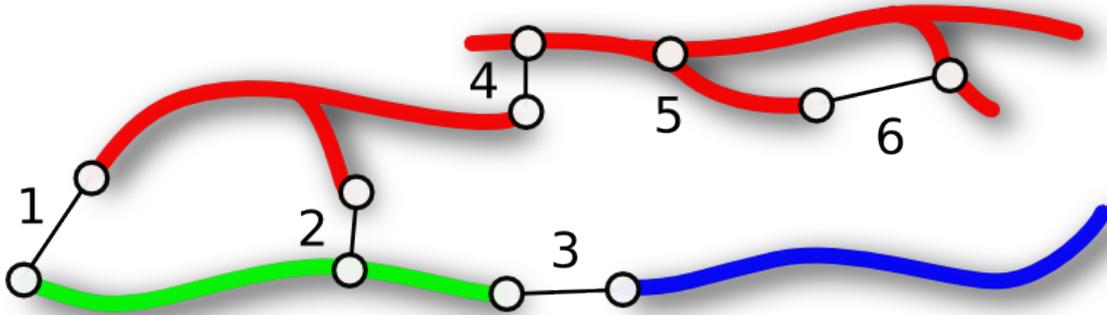


FIG. 11.4 – Schématisation du calcul de distances minimales entre quelques segments en relation : inter-labels (1-3) et intra-label (4-6). Pour les *jonctions topologiques* (5) ou les *plis de passage* cette distance est nulle. Elle approxime la largeur locale d'un gyrus séparant deux sillons (1,3,4) : cette information est sensible à la variabilité de localisation des extrémités. Les branches perturbent également l'estimation de cette information (2,6).

agrémenterons leur description de pistes pour dépasser les limitations actuelles.

La formulation Markovienne proposée indique qu'un potentiel de relation doit être décliné pour chaque paire de labels susceptibles d'être en relation. En pratique de nombreuses relations s'expriment plutôt rarement empêchant une bonne estimation des potentiels correspondants. Ces situations sont identifiées et traitées séparément. Pour les détecter deux seuils assez drastiques sont utilisés afin de ne conserver que les relations les plus communes et les plus faciles à estimer : il faut qu'elles soient présentes chez plus de la moitié des sujets (plus de 31 sujets pour notre base) et qu'il en existe au moins une centaine d'occurrence sur l'ensemble des graphes. En pratique, toutes les configurations possibles, même les plus rares devront être évaluées, il est donc nécessaire de définir un potentiel par défaut pour les situations ne passant pas les seuils définis précédemment. Pour chaque label, il existe des potentiels intra-label (impliquant deux fois le même label) et inter-labels (impliquant des labels différents). Il convient de les distinguer puisqu'ils caractérisent généralement des situations assez différentes. Plutôt que de poser ce modèle intra par défaut (respectivement inter) a priori, il est estimé de façon empirique à partir de l'ensemble des mesures intra-label (respectivement inter) rejetées par le seuil (dont le nombre d'instances est insuffisant). De cette façon, la répartition de l'ensemble des événements rares contribue à caractériser ces modèles par défaut.

11.3.2.1 Distance

L'introduction d'informations de distance dans notre modélisation des relations visait à corriger des erreurs fréquentes mais impliquant généralement des petites structures, particulièrement des branches ou des segments aux extrémités (généralement en queue de distribution des modèles de localisation et moins bien modélisés). L'idée sous-jacente est que deux segments sulcaux qui se touchent (5) ont des chances d'avoir le même label. Bien entendu les choses ne sont pas si simples. La figure 11.4 présente la plupart des situations que l'on peut rencontrer, que l'on désignera dans le texte par leur numéro. Par exemple, si un sillon présente plusieurs composantes connexes (par exemple le sillon temporal supérieur), certaines paires de segments sont séparées par un gyrus (4). De plus, dans le cas de sillons arborescents (6), deux segments (correspondant alors chacun à une branche) portant le même label peuvent être en relation sans pour autant se toucher. Dans le cas de relations inter-labels, on retrouve également toutes ces situations. On observe couramment une

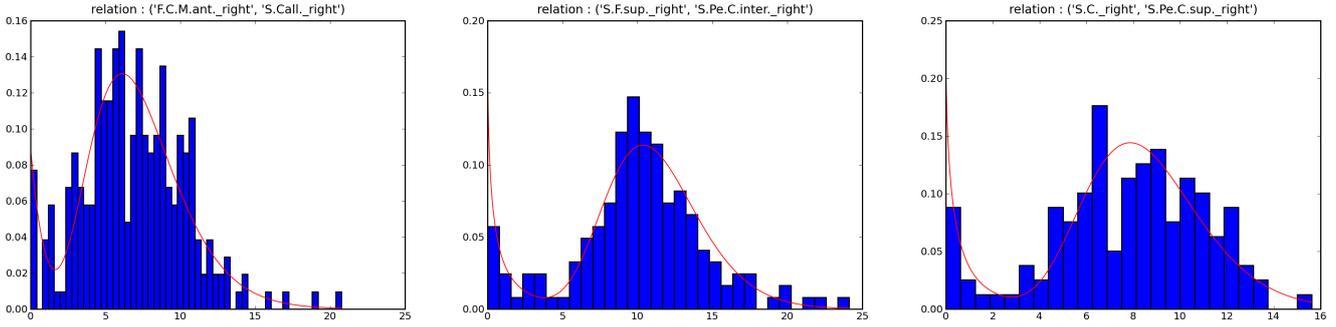


FIG. 11.5 – Pour 3 paires de labels : histogramme (en bleu) des distances minimales superposé à la vraisemblance du modèle Gamma-Exponentielle (courbe rouge) estimée à partir de ces données. À Gauche : fissure cingulaire et sillon calleux. Au centre : sillon frontal supérieur et sillon précentral intermédiaire. À droite : sillon central et sillon précentral supérieur.

connexion entre le sillon frontal supérieur et le sillon précentral. Ainsi, selon le sujet, les différentes configurations présentées en figure 11.2 sont observées (où les sillons se rencontrent orthogonalement et non de front comme dans l'exemple), allant du simple contact à une séparation nette par un gyrus **(3)**. Une autre situation courante est le cas de deux structures parallèles se faisant face **(1)**. Ici, la largeur du gyrus les séparant correspond à l'idée intuitive qu'on se fait de la distance entre de telles structures. Les branches **(2)** perturbent néanmoins cette mesure (comme le cas **(6)** précédent).

Il est probable qu'une analyse surfacique de ces quantités et partant des fonds de sillons, permettrait de mieux gérer le cas des branches et sans doute réduire également la variabilité des extrémités de sillons. En effet, seules les parties les plus enfouies de ces structures semblent réellement pertinentes pour notre étude.

Dans l'attente d'une telle analyse, nous proposons d'utiliser simplement la distance minimale (distance entre les deux points les plus proches) entre chaque paire de segments en relation :

$$\text{dist}(S_i, S_j) = \arg \min_{\substack{x_i \in S_i \\ x_j \in S_j}} \|x_i - x_j\|^2 \quad (11.4)$$

où S_i et S_j correspondent à deux segments en relation, chacun composé d'un ensemble de voxels. Cette mesure n'est pas pleinement satisfaisante puisqu'elle est assez sensible à de légères variations le long des structures étudiées. Malgré ces difficultés, on constate en pratique que cette mesure représente des informations intéressantes. En effet, l'histogramme de cette mesure sur nos 62 sujets suit un profil spécifique à chaque paire de labels qui peut être utile pour les discriminer. Autre point intéressant et facilitant notre modélisation, ces profils dérivent tous d'une base commune (voir figure 11.5 pour quelques exemples). Il s'agit d'un mélange de deux distributions. L'importance relative de chaque modèle, leur forme et leur localisation varient alors d'une paire de labels à l'autre.

Le premier mode, situé en zéro, correspond aux *jonctions topologiques* et aux *plis de passage*. La variabilité des situations coupant un sillon par un gyrus (présentées en figure 11.2) induit une variabilité prenant une forme proche d'une distribution exponentielle. Afin de contrôler la vitesse d'aplatissement du premier modèle, nous avons utilisé une généralisation de la loi exponentielle. Il s'agit en réalité d'une famille restreinte de distributions extraite d'une généralisation de la loi Gamma [Stacy 1962]. Sa vraisemblance sur le paramètre x (représentant dans notre cas la distance

minimale entre deux segments) s'exprime de la façon suivante :

$$P_e(x|\mu\alpha) = \mu \exp\left(-\left(x\mu\alpha\Gamma(\alpha+1)\right)^{\frac{1}{\alpha}}\right) \quad (11.5)$$

avec α et μ deux nombres positifs, désignant respectivement la forme (vitesse d'appâtissement de la distribution) et la valeur de la densité au point $x = 0$.

Le second mode avoisine le centimètre, correspond aux *relations corticales* et est approché par une loi Gamma classique :

$$P_\Gamma(x|k\theta) = x^{k-1} \frac{\exp\left(-\frac{x}{\theta}\right)}{\Gamma(k)\theta^k} \quad (11.6)$$

avec k un paramètre de forme et θ un paramètre d'échelle.

Le modèle complet proposé (mixture Gamma-Exponentielle généralisée) s'écrit donc :

$$P_\Psi(x|\pi_e\pi_\Gamma\mu\alpha k\theta) = \pi_e P_e(x|\mu\alpha) + \pi_\Gamma P_\Gamma(x|k\theta) \quad (11.7)$$

avec π_e et π_Γ les poids des différentes composantes du modèle vérifiant $\pi_e + \pi_\Gamma = 1$. Un a priori de Dirichlet est utilisé sur ces poids afin de forcer artificiellement la présence des deux classes et prévenir certains problèmes numériques. L'estimation de l'ensemble des paramètres $\pi_e, \pi_\Gamma, \mu, \alpha, k$ et θ est réalisée par l'algorithme EM (*Expectation-Maximization* [Dempster 1977]) présenté plus tôt dans un autre contexte en section 4.1.2. Le procédé est itératif et raffine une proposition initiale. On se base sur le fait que la distribution exponentielle est forcément à gauche de la distribution Gamma. Initialement, les deux modes (qui n'existent pas toujours) sont séparés par une méthode ad hoc visant à garantir que le nombre de points différents de zéro dans la première classe est suffisant de manière à permettre l'estimation de la loi exponentielle. Le procédé est lancé 5 fois de suite avec des initialisations différentes ; la solution maximisant la vraisemblance est finalement retenue. La figure 11.5 semble indiquer que notre modélisation permet de bien représenter les données. En pratique, ce constat se vérifie pour l'ensemble des paires de labels traitées.

Le cas des données inter et intra labels rejetées par notre seuillage (voir section précédente) sont distingués pour constituer deux modèles par défaut dans les situations où trop peu de données sont disponibles pour espérer une bonne estimation. En pratique, on constate une vraisemblance 4 fois plus élevée en 0 pour le modèle intra (les contacts entre segments y sont plus fréquents) et le deuxième mode se situe proche du centimètre pour le modèle inter et du demi-centimètre pour le modèle intra avec une variance plus grande pour le second.

11.3.2.2 Direction

Les structures sulcales suivent un schéma d'organisation assez stable à l'échelle macroscopique. Par exemple, le sillon central est toujours situé entre le sillon précentral et le sillon postcentral ; il est également au dessus de la vallée sylvienne. De proche en proche, l'ensemble des structures s'agencent ainsi les unes par rapport aux autres. Si un sillon est absent (une petite dizaine de sillons sont plutôt rares) chez un sujet leur organisation générale n'en est pas pour autant perturbée. Cette information est portée d'une part par la distance entre les structures (que nous avons traitée dans la section précédente) et d'autre part par leurs directions relatives (c'est à dire un vecteur 3D de norme 1).

Ces données directionnelles sont plutôt précises à l'échelle d'une paire de sillons, mais beaucoup moins au niveau des segments où l'information est perturbée par des spécificités locales et leur variabilité (branches, extrémités). Par exemple, dans le cas du sillon temporal supérieur, l'orientation locale varie d'un sujet à l'autre si le sillon est d'un seul tenant ou séparé en plusieurs composantes. Cette orientation locale peut alors fortement différer de l'orientation globale du sillon.

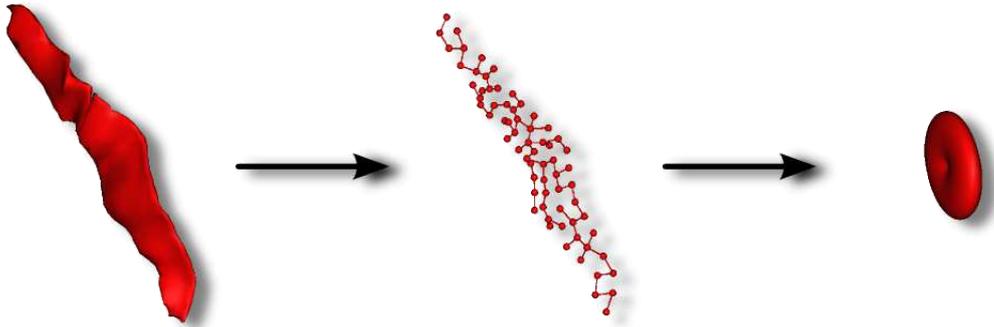


FIG. 11.6 – Modélisation des directions des relations intra-label. Estimation d’une distribution Bingham (à droite) à partir des directions induites par un arbre de recouvrement minimal sur le résultat d’un *k-means*. Cette opération est répétée pour chaque composante connexe. L’exemple porte ici sur un sillon central.

Chaque sillon étant composé de plusieurs segments, notre modélisation doit permettre de modéliser la variabilité directionnelle des segments qu’elle soit inter ou intra labels. Les deux cas sont assez différents et nécessitent des traitements adaptés.

Avant toute chose, il convient de distinguer la phase d’inférence de labels sur un sujet aux labels inconnus et la phase d’estimation de la variabilité qui est réalisée à partir de notre base de données de 62 sujets labélisés.

Dans le premier cas, l’information directionnelle a été réduite à son expression la plus simple, à savoir, la direction entre les centres de gravité des deux segments impliqués dans une relation. La simplicité avec laquelle l’information directionnelle est caractérisée participe à la rapidité du processus d’inférence. Cette mesure ne permet de caractériser les directions locales que grossièrement, en particulier dans le cas de segments de grande taille. À cette échelle, il semble difficile de proposer mieux. Pour aller plus loin, un appariement des voxels composant les structures en relation serait nécessaire afin de prendre en compte l’intégralité de la variabilité locale des directions. Pendant la phase d’apprentissage, l’information directionnelle issue des centres de gravité s’avère être trop grossière pour permettre une estimation correcte de la variabilité. En effet, ces données sont artificiellement restreintes aux relations induites par la segmentation des sillons en segments sulcaux élémentaires (voir section 3.2). L’approche que nous proposons est résumée visuellement en figure 11.6. Elle consiste à échantillonner l’espace des segmentations possibles pour chacun des sujets considérés. Nous avons pu constater que ces segmentations ont parfois lieu dans le sens de la longueur du plissement ou de biais contrairement au cas le plus fréquent où la coupure a lieu dans le sens de la profondeur. Dans le cas des relations intra-label, nous estimons qu’aucune information directionnelle pertinente n’est portée dans le sens de la profondeur des sillons. Finalement l’échantillonnage des directions est réalisé de la façon suivante. Pour un label donné, chaque sillon de la base d’apprentissage est simplifié en utilisant la méthode des *k-means*. Un arbre de recouvrement minimal (*Minimum Spanning Tree*) est calculé à partir du résultat précédent. L’ensemble des directions locales engendrées par les liens ainsi obtenus est utilisé pour caractériser leur variabilité. Finalement, on modélise cette variabilité par une distribution Bingham (voir annexe B.3.1.5) qui généralise la distribution Gaussienne multivariée à des données axiales (vecteur 3D de norme 1 dont le signe importe peu).

Dans le cas des relations inter-labels, on constate localement une grande variabilité dans l’expression des directions induites par les centres de gravité des segments en relation. En particulier, les branches, la sous-segmentation des segments sulcaux et la mise en relation arbitraire de structures

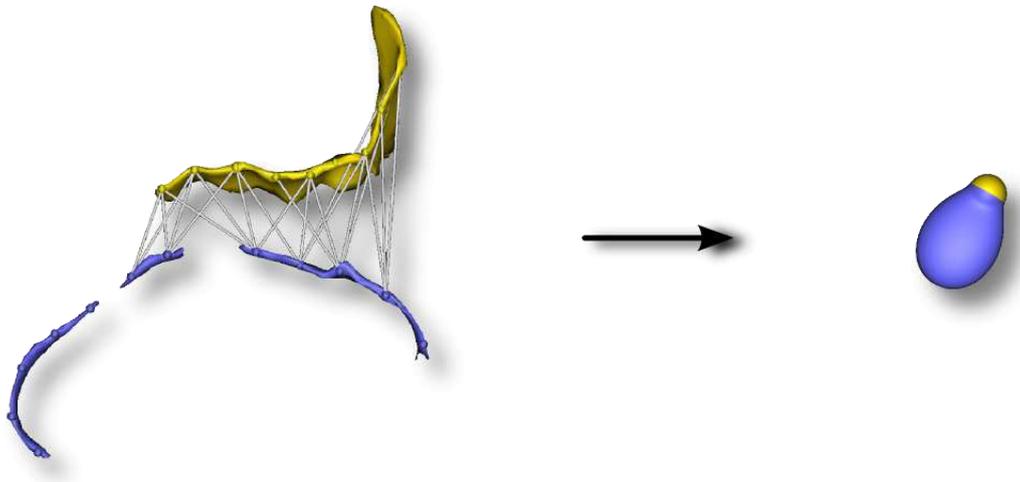


FIG. 11.7 – Modélisation des directions des relations inter-labels. Estimation d'une distribution Kent (à droite) à partir des directions induites par le produit cartésien du résultat de la quantification vectorielle (obtenue par la méthode du *k-means*) des deux sillons considérés (seul un sous-échantillonnage est représenté pour des raisons de visibilité). Les sillons représentés sont la fissure cingulaire (en jaune) et le sillon calleux (en violet).

éloignées explique principalement cette variabilité. Toutefois une constante demeure, il semblerait qu'il soit toujours possible de déterminer un demi-plan séparateur permettant de caractériser grossièrement la direction privilégiée d'une relation entre deux labels. Dans bien des cas il est possible d'être plus précis. Nous suivons finalement la procédure suivante, résumée en figure 11.7. Pour une paire de labels donnée, chacune des structures (indépendamment pour chaque label et pour chaque sujet) est simplifiée par la méthode des *k-means*. Pour chaque sujet, l'information directionnelle à apprendre résulte alors du produit cartésien des deux nuages de points obtenus à l'étape précédente. Finalement, la variabilité de ces directions est modélisée par une distribution Kent (voir annexe B.3.1.3). Cette approche permet d'aplatir artificiellement la distribution estimée afin d'être plus robuste.

Cette fois encore, un modèle par défaut est proposé afin de modéliser des situations rares ou non représentées et être ainsi en mesure de les évaluer. Les lois uniformes relatives aux données directionnelles (pour le cas inter-sillon) et axiales (pour le cas intra sillon) ont été utilisées.

Dans ce chapitre nous décrivons la stratégie adoptée pour étudier et raffiner la modélisation Markovienne proposée et les résultats obtenus sur une sélection limitée de potentiels avec et sans potentiels locaux au niveau segment.

12.1 Approche

Les potentiels de relation décrits plus tôt en section 11.3.2 n'ont pas été imaginés directement sous la forme que nous avons présentée. Nous avons effectué un processus de va et vient entre *design* de potentiels et évaluation de leur capacité à discriminer les labels sulcaux. Plus précisément, pour un type de potentiel donné, on estime les paramètres spécifiques à chacun des labels de façon indépendante. Ensuite, la phase d'évaluation vise à identifier les faiblesses des potentiels proposés ; elle est réalisée en labélisant automatiquement des sujets tests (non utilisés durant la phase d'apprentissage). Afin de répondre à cette question il convient de trouver des situations où l'étiquetage manuel diffère de l'étiquetage automatique (ce qui est relativement facile à déterminer). En pratique, il existe de multiples origines possibles à une mauvaise labélisation, chacune correspondant à des faiblesses de notre modélisation Markovienne :

- **sous-segmentation**. Chaque potentiel de relation correspond à la log-vraisemblance d'un descripteur caractéristique de l'agencement d'une paire de segments sulcaux. L'étape de segmentation est imparfaite ; parfois, un segment sulcal correspond alors à la fusion entre deux structures dont les labels devraient être différents. Comme nous l'avons évoqué plus tôt, une solution est de rendre ce descripteur peu sensible aux problèmes de segmentation. Notre potentiel basé sur des informations de distance y est peu sensible. Dans le cas de potentiels basés sur des informations directionnelles, nous avons fait le choix d'un modèle peu informatif mais très robuste.
- **voisinage**. Si une ou plusieurs structures voisines d'un segment portent un label erroné, leurs erreurs peuvent impacter la labélisation du segment considéré au travers des potentiels de relations. Cette situation correspond généralement à un problème d'optimisation. Dans ce cas, il est alors nécessaire de suivre l'évolution locale des potentiels associés à un segment et à son voisinage pendant le processus de labélisation, afin de cerner l'origine du problème d'optimisation. Notre outil de labélisation permet d'accéder à de telles informations de suivi et d'analyse du processus d'étiquetage.
- **mélange**. Comme nous l'avons indiqué plus tôt, l'estimation des paramètres des différents potentiels (relations et segments) devrait être réalisée de façon conjointe. Nous avons choisi de découpler leur apprentissage ce qui pose un problème de calibration de l'importance relative de chaque potentiel. Il est malheureusement difficile d'évaluer l'impact pratique de cette approximation. De plus, lorsqu'on optimise le label d'un segment particulier, on se doit d'évaluer le potentiel de segment qui lui est rattaché, additionné à l'ensemble des potentiels de relations (équation 11.2). On a alors pu constater que chaque relation n'avait pas forcément la même importance (effet pris en compte par l'équation 11.3) dans ce mélange de potentiels. Afin de faciliter l'interprétation, chaque proposition de potentiel a été évaluée indépendamment des autres.

- **relations locales peu informatives.** Afin de s'affranchir des problèmes de pondération entre potentiels de segments et de relations, il est tentant de considérer une modélisation sans potentiels de segments. Malheureusement cette approche s'avère limitée (voir section suivante) puisque les relations ont besoin d'avoir des points d'accroche (portés par les potentiels de segments) afin d'être bien définies. Dans le cas du potentiel de relation directionnel, plusieurs structures en alignement avec le segment considéré présenteront des directions et donc des potentiels très proches. Dans le cas de la distance minimale, la plupart des sillons présentent des profils très proches, sans potentiels de segments le modèle se révèle beaucoup moins informatif (voir section suivante).

En pratique, il n'est pas rare qu'une combinaison de facteurs explique un mauvais étiquetage, rendant l'interprétation difficile. On observe généralement dans ce genre de cas des contradictions entre le potentiel de segment et les potentiels de relations (c'est dans ce genre de situation qu'une pondération adéquate entre les différents potentiels à le plus d'importance) qui induisent des probabilités a posteriori $P(L_i = l | D_i)$ plus faibles pour le véritable label l . De façon pratique, notre attention est focalisée sur les segments dont la probabilité a posteriori est très faible. Ce sont de très bons indicateurs de la qualité du modèle, puisque dans le cas où une telle situation extrême est corrigée (en proposant un potentiel mieux défini par exemple), elle aura des répercussions directes sur des situations moins marquées. De cette façon seul un nombre limité de configurations est nécessaire pour faire progresser notre modélisation, accélérant ainsi le processus de va et vient¹.

L'approche que nous venons de suggérer vise à découpler le plus possible les difficultés auxquelles on peut s'attendre et à se focaliser sur quelques configurations problématiques individuelles restreintes mais révélatrices de problèmes globaux. Pendant l'étude de notre modélisation Markovienne, nous avons suivi cette démarche, ce qui nous a permis d'apporter des améliorations notables en une durée assez courte. Nous n'en sommes encore qu'à un stade exploratoire ; il est fort à parier que le modèle proposé peut encore être amélioré à l'échelle des relations entre segments. Ce processus est donc loin d'être terminé. Il gagnerait sans doute à être automatisé davantage dans l'avenir ; l'identification des situations pertinentes reste encore assez fastidieuse pour l'instant.

12.2 Expériences

12.2.1 *Loopy Belief Propagation*

Une étude préliminaire a été menée sur une première version, plutôt basique, de la modélisation Markovienne proposée ici. Elle se basait sur notre ancienne base de données comprenant 26 sujets seulement (utilisée dans [Rivière 2002, Perrot 2008]). Nous ne reporterons aucun résultat chiffré pour cette étude ; leur intérêt est très limité puisque les potentiels de cliques ont évolué depuis et les bases de données utilisées sont différentes, rendant les comparaisons très difficiles.

Dans ce cadre, l'optimisation des labels a été réalisée par la méthode *Loopy Belief Propagation* (section 10.3.2.3). Il s'agit d'une méthode itérative qui permet de calculer les probabilités a posteriori locales aux segments $P(L_i | D_i)$ ou aux paires de segments en relation : $P(L_i L_j | D_{ij})$, de façon exacte dans le cas où les variables sont liées entre elles par une structure arborescente de dépendances et de façon approchée lorsqu'il s'agit d'un graphe (sans garantie de convergence d'ailleurs). Ces estimations locales sont utilisées ensuite pour déterminer une labélisation selon le critère Bayésien MPM 10.3.1. Nous avons pu constater en pratique sur plusieurs potentiels différents, une extrême sensibilité de la méthode à l'irrégularité du graphe (section 11.2.2). En effet, cette méthode procède à une diffusion de l'information de proche en proche à travers le graphe. Dans le cas d'un graphe irrégulier, la proportion d'information diffusée augmente avec le nombre de relations d'un segment,

¹Bien entendu, cette dernière remarque n'est pas propre à la modélisation Markovienne et peut s'appliquer à d'autres modélisations proposées dans cet ouvrage

provoquant des déséquilibres induisant au final un étalement excessif de certains labels sur les segments. Des pondérations similaires à celles proposées dans l'équation 11.3 seraient sans doute utiles au processus. À notre connaissance, aucun résultat théorique connu ne permet d'identifier les conditions nécessaires à la convergence de la méthode. En pratique, les divergences constatées semblent s'expliquer principalement par la diffusion d'informations contradictoires lorsque plusieurs relations d'un même segment sont en désaccord (phénomène sujet-dépendant). Une solution consiste à donner plus d'importance aux potentiels de segments qu'aux potentiels de relation. De cette façon, le modèle de relations implique des ajustements très localisés uniquement. Il semblerait qu'il existe une valeur critique de ce facteur de pondération, propre à chaque sujet, à partir de laquelle le processus converge. Pour l'instant, nous avons procédé à des ajustements manuels afin de déterminer une valeur unique satisfaisante pour tous les sujets.

En conclusion, cette méthode s'est révélée assez décevante dans notre cas d'utilisation : trop sensible, imprévisible, difficile à calibrer pour obtenir finalement, lorsqu'il y a convergence, seulement une approximation de mesures locales. Nous avons indiqué plus tôt une modélisation proposée dans [Shi 2009] où la méthode *Belief Propagation* (détaillée au début de la section 11) est utilisée sur un arbre modélisant des dépendances statistiques au niveau de paires de sillons voisins. Dans ce cas, la structure de dépendance peut être fixée une fois pour toutes (indépendante des sujets), le nombre de liens contrôlé et les potentiels à cette échelle sont moins sensibles au bruit local et devraient engendrer moins de contradictions. Il serait sans doute intéressant de revenir sur cette approche dans ce cadre.

12.2.2 Recuit simulé

L'étude suivante a été menée sur les potentiels de relation présentés plus tôt (section 11.3.2). Elle a pu bénéficier de la base de données de 62 sujets décrite en annexe A. Fort des limitations constatées lors de l'utilisation précédente de la méthode *Loopy Belief Propagation*, nous lui avons préférée une optimisation par un recuit simulé 10.3.2.2 où il est possible d'approcher la solution optimale autant que souhaitée, à condition d'itérer la méthode suffisamment longtemps. De cette façon, une évaluation précise de l'efficacité des potentiels est désormais possible. L'ensemble des résultats présentés dans cette section suit le schéma de validation croisée décrit en section 5.3. Dans cette étude, nous avons étudié plusieurs facteurs, orthogonaux en termes d'informations représentées, mais interagissant au travers de la modélisation proposée : le choix de l'a priori $P(L)$, de la pondération adaptative entre chaque potentiel de segment et les potentiels de relations associés, et le choix des potentiels de relation. Afin de ne pas alourdir inutilement cette partie, seule une sélection des résultats est compilée sous forme chiffrée (il s'agit des meilleurs résultats obtenus en pratique), le reste est généralisé dans le corps du texte. Dans les résultats à venir, la meilleure modélisation locale à l'échelle des segments est utilisée afin de fournir la meilleure base possible aux potentiels de relation. Nous avons donc utilisé le modèle SPAM avec prise en compte de l'incertitude spatiale vue en partie II (nous nous sommes restreints à la normalisation rigide globale pour des raisons techniques). En pratique, des résultats comparables (avec un écart constant de 3% d'erreur en faveur du modèle avec estimation de la normalisation) sont obtenus avec le modèle de SPAM défini dans Talairach (partie I), nous ne les présenterons donc pas davantage.

12.2.2.1 Étude préliminaire : a priori et pondérations

Concernant le choix de l'a priori $P(L)$, l'étude s'est limitée à quatre possibilités (le nom utilisé dans le tableau est indiqué en gras) : aucun a priori (**Aucun**), la fréquence des labels sur les segments (**fréq. segments** : $P(L|E)$, section 4.3.1), la fréquence des labels sur les voxels (**fréq. tailles** : $P(T|L; E)$, même section), modèle de Dirichlet généralisé (**Dirichlet G. tailles** : $P(T|L; E)$, section 4.3.2). Comme nous l'avons vu plutôt tôt, l'expression de la distribution de Dirichlet vérifie une

vraisemblance nulle dès que la fréquence d'apparition d'un des labels est nulle, imposant de fait la présence de tous les labels. Afin d'éviter ce phénomène, dans nos essais nous avons artificiellement remplacé chaque fréquence $f_l = s_l / \sum_{l \in \mathcal{L}} s_l$ d'apparition du label l par $(s_l + S) / (\sum_{l \in \mathcal{L}} s_l + S)$ avec s_l la taille du sillon et S une valeur permettant de contrôler dans une mesure limitée et de façon indirecte la fréquence d'apparition des labels (opération réalisée à l'estimation du modèle et pour l'inférence des labels).

L'irrégularité du graphe de relations semble suggérer la pondération locale des différents potentiels impliqués dans chaque proposition de labels au cours de l'échantillonnage de Gibbs. On se propose de comparer l'absence de pondération (**Aucun** : équation 11.2) aux différentes suggestions exposées plus tôt dans ce manuscrit (équation 11.3) : normalisation selon le **nombre de connexions**, d'après la **longueur des segments** en relation, ou par l'**aire de contact** des zones de Voronoï des segments. Deux façons d'utiliser ces pondérations ont été présentées plus tôt avec ou sans normalisation des potentiels de segments entre eux (utilisation de $\bar{E}_{i\partial i}(l)$ ou $E_{i\partial i}(l)$ respectivement). En pratique, les résultats obtenus sont extrêmement similaires, nous indiquerons donc uniquement ceux obtenus par la méthode normalisée.

Les essais concernant le choix de l'a priori $P(L)$ ont été réalisés en fixant arbitrairement la méthode de pondération des potentiels. Nous avons choisi celle basée sur le nombre de connexions. De la même façon, les tests sur le choix des pondérations ont été réalisés en fixant l'a priori à la fréquence des labels sur les segments. ces deux premières séries de résultats sont résumées dans le tableau ci-dessous (taux d'erreurs globaux moyen SI et écarts-types associés entre parenthèse) :

	A priori $P(L)$				Pondération des relations			
	Aucun	Dirichlet G. tailles	fréq. tailles	fréq. segments	Nombre de connexions	Longueurs segments	Aire de contact	Aucun
Gauche	14.52 (2.89)	14.08 (2.64)	14.36 (2.89)	13.87 (2.48)	13.92 (2.67)	13.89 (2.55)	16.85 (3.63)	
Droite	13.39 (2.59)	13.45 (2.56)	13.16 (2.83)	12.93 (2.56)	12.37 (2.53)	13.14 (2.47)	15.81 (2.91)	

Ces résultats indiquent, qu'en l'état, les modèles les plus simples restent les plus efficaces. Dans le cas de l'a priori $P(L)$, la fréquence des labels est théoriquement optimale dès qu'on suppose l'indépendance statistique des labels sur les segments (section 4.3.1). Au contraire, la modélisation de Dirichlet généralisée proposée prend en compte cette dépendance, mais reste limitée pour prendre en compte l'absence de sillons. Nous lui préférerons dans l'avenir la distribution de Pólya multivariée (équation 4.33) qui permet de modéliser l'absence de labels de façon appropriée.

Les résultats obtenus démontrent également l'utilité de pondérer les potentiels de relations et de segments afin de compenser les déséquilibres induits par un graphe de dépendance irrégulier (section 11.2.2). L'intérêt de pondérations entre les relations n'apparaît pas de façon flagrante. Pour la suite, nous avons donc conservé le modèle le plus simple proposant une normalisation du nombre de connexions.

12.2.2.2 Potentiels de relation

Nous présentons maintenant l'étude comparative du modèle originel sans relation (**SPAM**) et d'un ensemble de propositions de potentiels de relation, exploitant des informations de distance (**mindist**, section 11.3.2.1), de direction : basées sur la différence entre les centres de gravité des segments (**g-dir**, section 11.3.2.2) ou la modélisation basée sur un arbre de recouvrement minimal en intra-label et le produit cartésien de toutes les directions élémentaires entre paires de voxel en inter-labels (**a-dir**, même section). Les capacités de généralisation de chaque potentiel ont été testées individuellement. Comme nous l'avons évoqué précédemment, ces deux types d'information sont complémentaires. Ainsi, une combinaison de deux potentiels (**mindist** + **a-dir**) a été explorée sans qu'aucune précaution particulière n'ait été prise pour pondérer les contributions respectives des informations de distance et de direction, de telle façon que localement chaque potentiel de segment

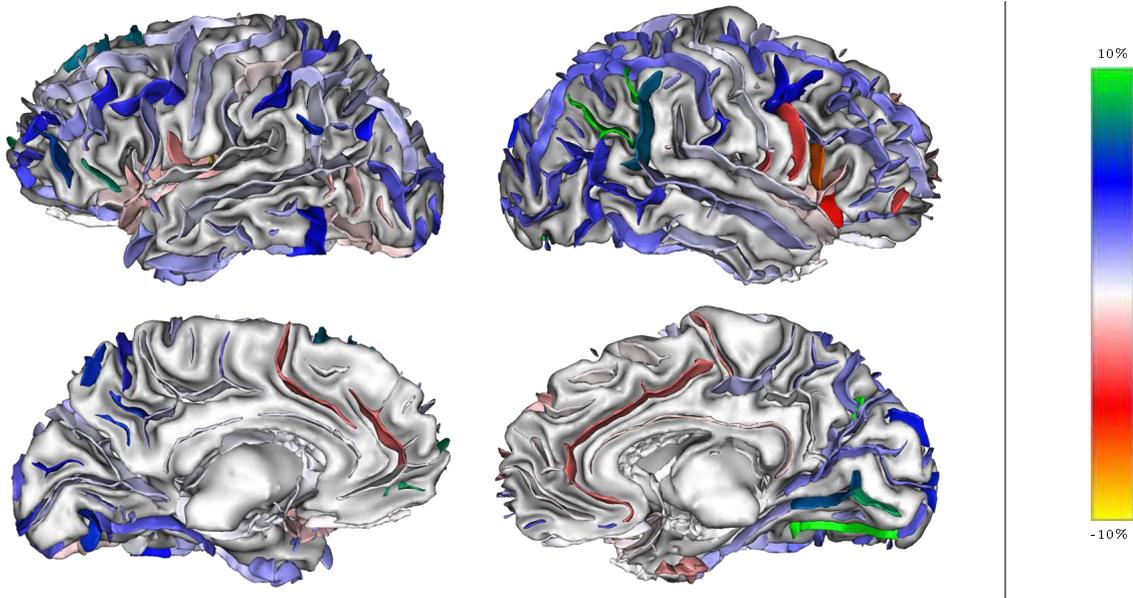


FIG. 12.1 – Comparaison du modèle SPAM avec prise en compte de l’incertitude spatiale (normalisation rigide, partie II) avec et sans relations. Les améliorations apportées par le modèle Markovien sont indiquées en bleu et vert alors que les dégradations le sont en rouge et jaune.

a autant d’importance que l’ensemble des potentiels de relations auquel il est rattaché et ceci pour chaque type d’information considérée (distance et direction).

Comme nous l’avons indiqué plus tôt, les potentiels de segment utilisés ici prennent en compte l’incertitude spatiale. Dans le cas des potentiels basés sur des directions, cette information n’a pas été prise en compte, les directions ont donc été calculées dans l’espace de Talairach.

Les résultats² (taux d’erreurs globaux moyen SI et $mass$ en pourcentage et leur écart-type entre parenthèse) de ces essais sont présentés dans le tableau suivant :

		SPAM	SPAM + mindist	SPAM + g-dir	SPAM + a-dir	SPAM + mindist a-dir
E_{SI}	Gauche	14.59 (2.88)	13.87 (2.48)	15.20 (2.83)	15.72 (3.17)	15.25 (2.94)
	Droite	13.97 (2.91)	12.93 (2.56)	15.59 (3.44)	15.13 (3.04)	14.31 (3.00)
E_{mass}	Gauche	14.01 (2.62)	13.27 (2.25)	14.22 (2.53)	14.85 (2.82)	14.37 (2.64)
	Droite	13.37 (2.82)	12.24 (2.43)	14.31 (3.13)	14.19 (2.92)	13.45 (2.89)

Ce tableau indique que l’utilisation de potentiels basés sur la distance minimale entre les segments permet d’améliorer la qualité des labélisations obtenues par rapport au modèle de SPAM sans liens Markoviens. Les résultats obtenus sont significativement meilleurs avec une p -values³ de $4.2e^{-4}$ pour l’hémisphère gauche et de $1.0e^{-6}$ pour l’hémisphère droit. Les différences locales entre les deux modèles sont soulignées par les images de la figure 12.1. Les améliorations sont réparties de façon assez homogène sur le cerveau. Les détériorations semblent localisées à la zone de Broca et au sillon

²Comme pour tous les résultats de cette section, les potentiels de SPAM sont définis avec prise en compte de l’incertitude spatiale.

³test de Wilcoxon, voir détails en section 5.2

Cingulaire. Dans le premier cas, il s'agit d'une zone où les relations sont nombreuses et complexes (liens au niveau de l'enveloppe corticale et dans des structures enfouies comme l'insula). Dans le deuxième cas, le sillon cingulaire prend souvent la forme d'un ensemble de grands plissements reliés à de nombreux petits sillons en frontal, aux formes assez variables. Ce modèle présente des taux d'erreurs extrêmement faibles (les meilleurs sur l'ensemble des modèles testés pendant cette thèse) sur le sillon central : 2.89% pour l'hémisphère gauche et 2.68% pour l'hémisphère droit, où une quarantaine de sujets présente une reconnaissance parfaite du sillon central, une vingtaine un taux d'erreur inférieur à 10% et environ 5 sujets posent encore des difficultés réelles.

Concernant les potentiels basés sur des informations directionnelles, les deux modèles proposés (*a-dir* et *g-dir*) présentent des résultats globaux comparables en moyenne; on constate pourtant de nombreuses différences locales dans la forme des modèles directionnels de Kent et de Bingham (voir figure 12.2). L'approche *g-dir* est moins robuste que *a-dir* dans le cas de sillons parallèles ou orthogonaux l'un à l'autre. Par contre si les deux sillons considérés sont dans le prolongement l'un de l'autre, la situation s'inverse puisque le fait de considérer l'ensemble des paires de directions élémentaires entre voxels (approche *a-dir*) a tendance à sur-pondérer de façon artificielle l'importance de la direction générale des deux sillons (ce qui n'est pas le cas de l'approche *g-dir*). Le compromis réalisé entre robustesse (distributions peu piquées) et quantité d'informations discriminantes ne permet pas de contribuer à l'amélioration des résultats.

Malgré les conclusions précédentes, individuellement les potentiels de relations seuls portent des informations permettant d'obtenir des résultats comparables à ceux observés au début de cette thèse avec l'approche «historique». Dans cette expérience, les potentiels de segments ont été enlevés, de telle façon que les labels n'aient plus d'attache spatiale spécifique. Ces observations sont résumées dans le tableau suivant, où le taux d'erreur global σ est indiqué pour les 3 potentiels présentés précédemment :

	mindist	g-dir	a-dir
Gauche	32.08 (4.95)	24.65 (3.80)	27.13 (4.05)
Droite	30.08 (3.74)	42.28 (3.59)	25.97 (4.35)

Il est plutôt intéressant de constater que les résultats obtenus ne sont pas en accord avec et sans l'utilisation de potentiels de segments. Cela confirme notre intuition selon laquelle l'information portée par la distance minimale serait moins caractéristique de labels spécifiques que les informations directionnelles (confirmée par les résultats obtenus avec le modèle *a-dir*). Le modèle *g-dir* utilisé seul donne des résultats assez déséquilibrés entre les deux hémisphères. À droite, les résultats sont systématiquement moins bons et de façon très nette. Dans les deux cas, les grandes structures sont généralement bien identifiées. Par contre, à droite, certains labels ont tendance à s'étaler spatialement empêchant l'expression d'autres labels. L'origine du problème n'est pas encore clairement identifiée.

12.2.2.3 Remarques complémentaires

La méthode d'optimisation utilisée pour inférer les labélisations est stochastique et ne garantit d'obtenir l'optimum que dans des conditions limites impossibles à vérifier en pratique (décroissance de température logarithmique, équilibre thermodynamique, ...). Néanmoins, nos essais nous ont permis de confirmer que le choix d'une approche locale (modélisation des segments et des liens Markoviens locaux entre segments) simplifiait l'expression de l'énergie optimisée. En effet, nos essais ont permis de constater qu'une simple approche gloutonne comme l'ICM permettait d'atteindre des résultats très proches (voire exactement similaires dans bien des situations) à ceux obtenus avec un recuit simulé et un échantillonnage de Gibbs. Dans ces conditions, la méthode approchée *Loopy*

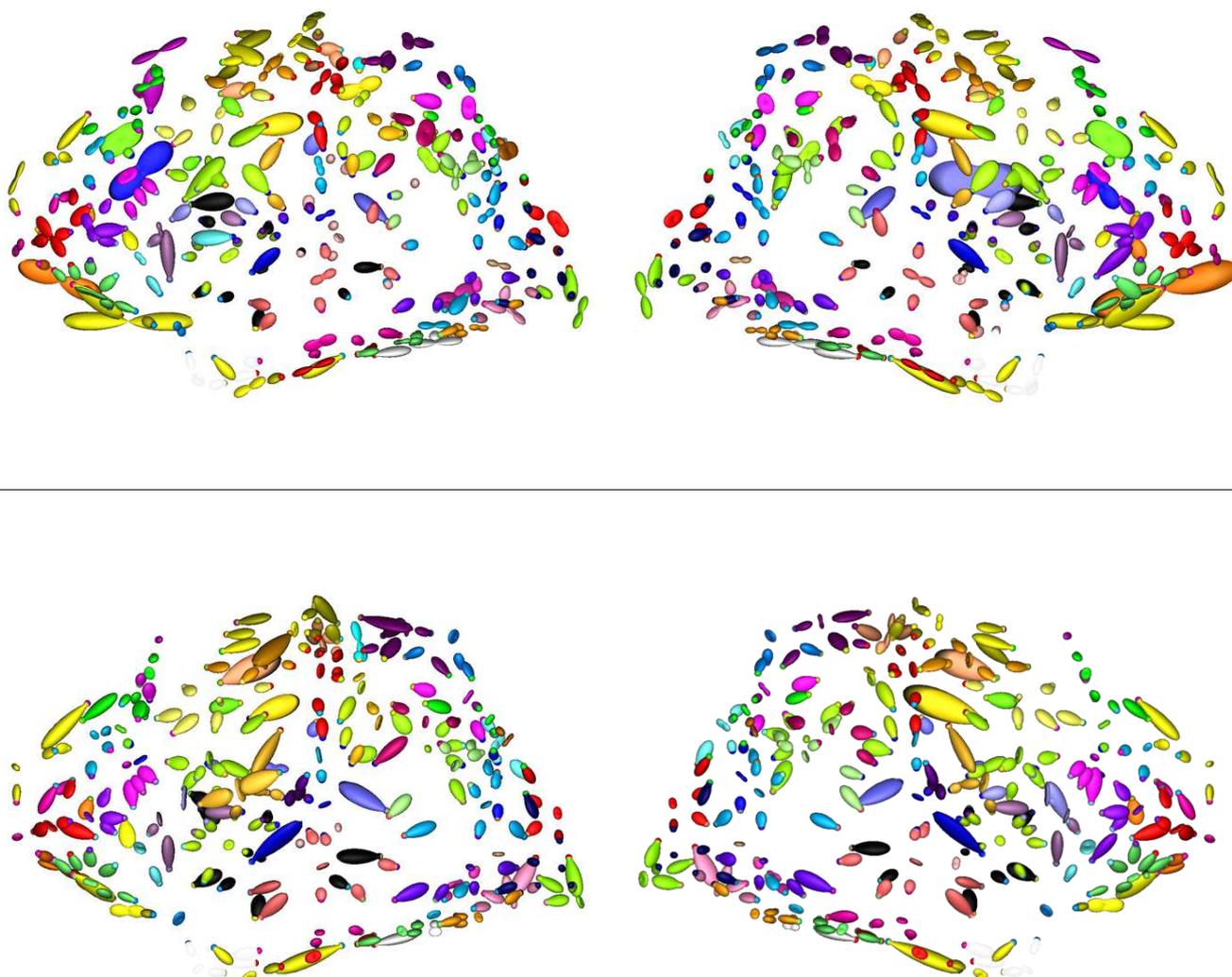


FIG. 12.2 – Représentation visuelle des potentiels de relation sur des données directionnelles (modèles hémisphères gauche et droit) : distribution Kent pour les relations inter-labels et distribution Bingham pour les relations intra-label. Les couleurs correspondent aux labels en relation. Les distributions sont placées au niveau du centre de gravité inter-sujets du label concerné pour les potentiels intra-label et d'un point médian pour les potentiels inter-labels. Dans le cas des directions inter-labels, le sens du vecteur est très important, ainsi pour chaque relation un ordre arbitraire a été fixé entre les labels. L'asymétrie des distributions Kent reflète cet ordre arbitraire. **Haut** : basée sur la différence entre centre de gravité ($g-dir$). **Bas** : basée sur la représentation sous forme d'arbre de recouvrement minimal en intra-label et sur le produit cartésien des voxels des structures en relation pendant la phase d'apprentissage, et sur la différence entre les centres de gravité pendant la phase d'inférence ($a-dir$).

Belief Propagation (étudiée lors d'une étude préliminaire en section 12.2.1) ne présente aucun intérêt puisqu'un recuit simulé assez rapide suffit à l'obtention de la labélisation optimum.

Conclusion et perspectives

13.1 Conclusion

L'hypothèse d'indépendance des labels entre les segments sulcaux a permis de proposer un modèle basique mais déjà efficace (partie I), suivi de raffinements remettant en cause l'étape de normalisation des données, partie II). Cette hypothèse entraîne toutefois de nombreuses petites erreurs de labélisation au niveau des branches de sillons (il s'agit généralement de petites structures comptant peu dans les erreurs mesurées).

La modélisation Markovienne proposée dans cette partie répond à ces limitations et complète l'approche locale décrite en partie I. À l'instar de l'information locale portée par les segments sulcaux (positions des voxels ou orientation locale), les relations entre segments voisins portent des informations descriptives de l'agencement local des sillons (distances et directions entre les structures). Nos analyses ont démontré l'efficacité et l'intérêt de l'approche. Une amélioration significative des résultats a été observée lors de l'utilisation d'un potentiel de relations basé sur une modélisation de la distance minimale entre segments pour chaque paire de labels. La méthode s'est avérée extrêmement rapide et fiable en pratique (voir annexe C.1), ce qui la rend pleinement utilisable dans des études impliquant de nombreux sujets. On peut sans mal envisager l'intégrer dans un modèle plus complexe. Ces atouts tiennent principalement à l'attention privilégiée portée aux informations localisées dans le modèle qui donnent une forme assez régulière à la fonction optimisée pendant la phase d'inférence.

Cette étude préliminaire a permis de démontrer une efficacité globale de l'approche Markovienne proposée. Localement, par contre, les résultats sont inhomogènes. En l'état, cette méthode ne semble donc pas assez fiable pour être utilisée sur des applications à visée neuroscientifique. Maintenant que la «preuve de concept» est entérinée, il sera sans doute nécessaire de reconsidérer jusqu'à la structure même du graphe de relations Markoviennes pour rendre l'approche plus robuste et moins sensible à la segmentation des plissements corticaux.

13.2 Perspectives

Cette étude a été l'occasion de dresser un bilan des possibilités et des limitations d'une modélisation Markovienne des labels portés par des segments sulcaux voisins. Les raffinements possibles sont assez nombreux, des choix stratégiques seront forcément nécessaires. L'approche locale proposée présente de nombreux atouts mais elle est intrinsèquement limitée et ne peut prétendre gérer des situations où des sillons entiers sont fortement déformés ou déplacés (observé en pratique sur tous les modèles présentés dans cette thèse). La prise en compte d'informations haut niveau sous la forme de potentiels à l'échelle des sillons et de paires de sillons est donc inévitable dans de futurs travaux. Par contre, cela ne remet nullement en cause l'intérêt de l'approche actuelle qui permet de fournir une première labélisation relativement proche de la labélisation optimale sur bien des sujets.

De ce fait, les perspectives envisagées s'orientent vers **(1)** un transfert méthodologique (méthodes d'optimisation, modèles directionnels, ...) de l'échelle des segments sulcaux aux sillons, **(2)** l'amélioration (création du graphe de relations, nouveaux potentiels de relation) du modèle Markovien présenté afin d'alléger autant que possible les modélisations structurelles plus haut niveau (sillons

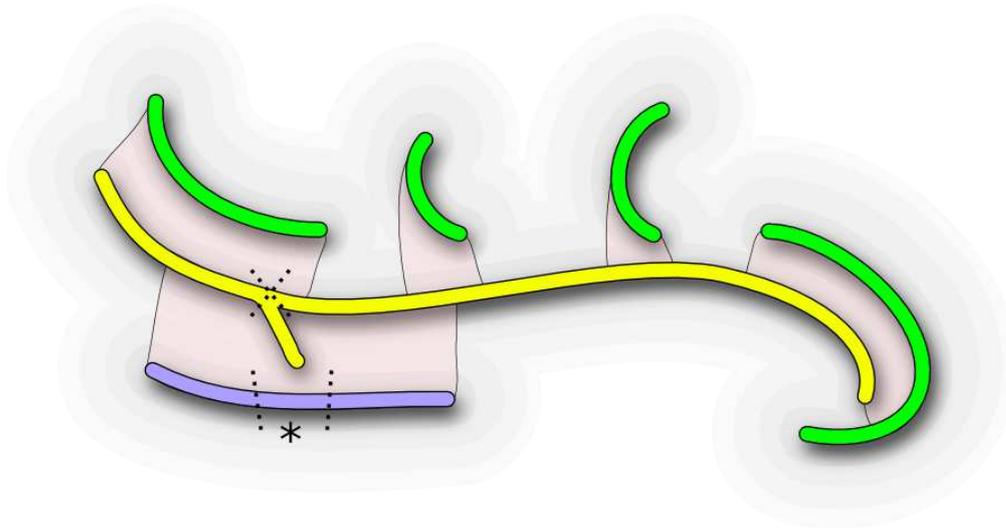


FIG. 13.1 – Segmentation et relations Markoviennes : exemples de mises en relation idéales représentées par les zones rosâtres. Chaque couleur représente un sillon spécifique. Les lignes en pointillés indiquent des séparations induites par la phase de segmentation. L'étoile indique une situation possible de sur-segmentation, où un segment présente des relations avec les segments du sillon de couleur jaune qui dépendent de sa longueur. S'il est trop petit, il ne sera en relation qu'avec la branche jaune (rendant difficile l'estimation d'une mesure de distance qui ait du sens), sinon il sera également en relation avec la partie principale du sillon jaune. En présence de sous-segmentation (sillon jaune), une structure peut se retrouver en relation avec de nombreux sillons interagissant avec elle sur des zones bien distinctes. Dans ce cas, une mise en correspondance très localisée semble nécessaire.

et paires de sillons) ce qui devrait simplifier la phase d'inférence des labels, **(3)** la prise en compte des incertitudes de normalisation dans le cas du modèle Markovien, c'est à dire le couplage complet des modèles des parties II et III et non un simple enchaînement comme nous l'avons proposé jusqu'à présent. Le premier point permettra de réconcilier pleinement l'esprit des modélisations proposées dans cette thèse et l'approche Markovienne «historique» présentée en introduction.

13.2.1 Création du graphe

La qualité du modèle proposé dépend fortement de la qualité du graphe de relations qui indique s'il y a ou non une dépendance statistique entre deux segments sulcaux. Le processus de création de ce graphe, basé sur une dilatation des structures locales (section 11.2.1) rend le résultat sensible à la position des structures et à la segmentation en plissements élémentaires (le problème est illustré en figure 13.2). Par exemple, si un champ de déplacement très lisse était appliqué pour déformer les données, de nombreuses relations corticales disparaîtraient alors que de nouvelles apparaîtraient avec le processus de création actuel. La transposition d'une réponse standard à ce type de problème serait de mettre en relation toutes les paires de segments et de pondérer l'importance de chaque relation de façon inversement proportionnelle à cette distance. Une telle approche ne serait alors pas Markovienne. Cette solution ne convient pas à tous les choix de mesures et de potentiels, en particulier celles proposées dans cette thèse.

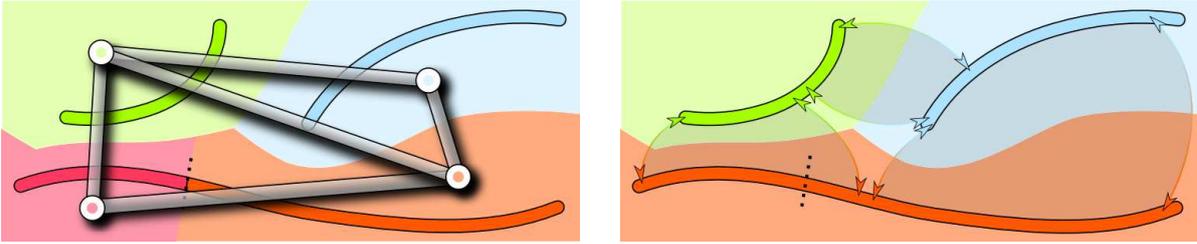


FIG. 13.2 – Mise en relation de segments voisins. **Gauche** : méthode utilisée dans cette thèse : les segments dont les zones de voronoï sont en contact sont mis en relation. Selon l'emplacement de la segmentation représentée en pointillés (entre le segment rose et le segment orange), les segments vert et orange sont en relation (comme dans l'exemple) ou les segments rose et bleu. **Droite** : mise en relation partielle (les structures en relation sont indiquées par les flèches et les zones grisées) des segments indépendamment de la segmentation (description dans le corps du texte).

En effet, dans le cadre de mesures locales telles que celles qui nous intéressent, on constate en pratique que les capacités de description de l'agencement local des structures sont sous-tendues par l'homogénéité de la segmentation (en termes de taille des éléments en relation) induite par un problème de sous-segmentation ou de sur-segmentation manifeste ou par de simples différences de tailles entre les structures. Dans ces situations, l'utilisation de liens Markoviens dans le modèle a tendance à dégrader la qualité de la labélisation.

Comme l'illustre la figure 13.1, si les segments en relation sont trop petits, il est difficile d'en extraire des informations utiles à la labélisation. Les structures environnantes sont alors assez peu prises en compte alors qu'elles comportent parfois des indices précieux à leur mise en relation. L'identification des branches secondaires permettrait sans doute d'améliorer le processus.

Au contraire, un segment trop grand risque d'être en relation avec de nombreux segments assez éloignés les uns des autres. Dans cette situation, il semble nécessaire de restreindre l'expression de chaque relation à une sous-partie des segments, où l'influence est directe et avérée. Une telle mise en relation locale est schématisée en figure 13.1. Elle semble permettre un appariement des voxels des segments se faisant face et par conséquent la caractérisation de mesures de distance et de direction moyenne plus robustes que celles proposées dans cette thèse. Le choix du processus de mise en correspondance dépend sans doute du choix de potentiels de relations.

Dans le cas des potentiels locaux proposés, nous envisageons une procédure en trois temps : **(1)** chaque plissement (composé éventuellement de plusieurs segments sulcaux) est dilaté géodésiquement à la matière grise jusqu'à ce que les zones ainsi définies rentrent en contact, **(2)** au niveau des interfaces de connexion, les structures en relation sont identifiées, **(3)** l'information est ensuite diffusée (par un processus d'advection [Weiskopf 2004]) depuis l'interface vers les structures le long du gradient de la carte de distance aux structures. De cette façon, les structures sont mises en relation localement indépendamment de la segmentation sulcale.

Une telle mise en relation partielle de paires de segments semble difficilement se généraliser aux solutions proposées en introduction de cette section. De plus, la méthode décrite dans le paragraphe précédent semble beaucoup plus sensible aux variations de segmentation et de positionnement des structures comme l'illustre la figure 13.2.

13.2.2 Autres potentiels de relation

13.2.2.1 Mesures locales

À l'échelle de relations entre segments, les informations pertinentes et caractéristiques des configurations sulcales sont assez restreintes. Les deux critères retenus jusqu'à présent (distance minimale et direction) fournissent des indices précieux permettant de reconstituer des sillons (en intra-label) à partir des segments sulcaux les composant.

Dans le cas de sillons très réguliers, une information peut compléter ce descriptif. Il s'agit du produit scalaire entre les directions locales des zones en relation, qui permet de caractériser si deux segments sont dans le prolongement l'un de l'autre (ou parallèles) ou orthogonaux. Cette information devrait également permettre de caractériser l'agencement local de certains sillons à condition que l'orientation locale des structures soit plutôt stable.

La mesure de distance proposée permet de caractériser la présence d'un gyrus entre deux structures. Toutefois, dans le cas où la distance mesurée est nulle, cela indique juste que les deux structures sont en contact. Le contact peut n'être qu'en surface indiquant alors la présence d'un gyrus. Pour ces configurations plusieurs solutions sont envisageables : mesurer la «force de connexion» entre les deux segments en contact (le pourcentage de la longueur du contact par rapport à la profondeur du sillon), ou réaliser les mesures de distance directement depuis le fond du sillon (actuellement notre représentation des fonds de sillon n'est pas suffisamment fiable pour cette tâche).

Il serait tentant de généraliser le concept de SPAM, qui a fait ses preuves pour caractériser la variabilité en localisation des structures sulcales, aux relations entre ces structures. Un tel modèle permettrait de mesurer la variabilité couplée de deux sillons et donc d'évaluer la probabilité que deux segments agencés d'une façon particulière portent une paire de labels donnée. L'estimateur à noyaux (équation B.96) utilisé pour le calcul des SPAMs (équation 4.19) se généralise facilement à une paire de voxels x et y , appartenant chacune à deux labels différents (dans le cas inter-labels) notés L_x et L_y :

$$P(xy|L_x L_y) \propto \sum_{u \in \Omega} \sum_{v \in \Omega} K_\sigma(x, u) K_\sigma(y, v) f(u, v) \quad (13.1)$$

où Ω correspond au domaine de définition des labels (à restreindre éventuellement à l'aide d'une boîte englobante) et $f(u, v)$ à la fréquence d'apparition conjointe, mesurée sur une base d'apprentissage, d'un voxel de label L_x en position u et d'un voxel de label L_y en position v . Reste désormais à définir une façon de résumer cette mesure sur les voxels d'un couple de segments en relation (en appariant les voxels par paire par exemple). Cette expression s'exprime dans un espace à 6 dimensions rendant son estimation et sa manipulation difficile (stockage en mémoire). L'utilisation d'une modélisation paramétrique (annexe B.4.1.2) sera sans doute nécessaire pour lever cette difficulté. Une telle mesure permet de définir également des probabilités conditionnelles $P(L_i|L_j D_i D_j)$ qui devraient trouver leur utilité dans le cadre de l'étude des *patterns* sulcaux et dans l'aide à la labélisation manuelle.

13.2.2.2 Mesures longues distances

Les potentiels de relation évoqués jusqu'à présent reposent sur un graphe de voisinage où les liens sont locaux et géodésiques à la surface corticale. L'influence locale réciproque d'un sillon sur les plissements voisins ne fait aucun doute. Il semblerait que certaines structures distantes soient également corrélées [Fillard 2007b]. Il s'agit peut-être de traces du processus de développement cortical qui subit une série de plissements successifs par vagues [Welker 1990, Dubois 2008], impliquant successivement les sillons primaires, secondaires et tertiaires, de telle façon que des sillons distants connaissent un développement simultané. Le développement cérébral est sous-tendu par la structuration des faisceaux de fibres de la matière blanche. Ces structures sous-corticales participent sans doute dans une certaine mesure à l'organisation des plissements corticaux. L'activité fonctionnelle

du cerveau s'appuie sur de tels réseaux fonctionnels, à travers lesquels des zones anatomiques corticales distantes communiquent. L'imagerie de diffusion permet grâce aux méthodes de *tracking* de fibres d'estimer la densité de fibres reliant deux zones corticales. Cette mesure de densité, la forme des fibres les reliant ou le label (si on dispose d'un atlas) du faisceau sont autant d'indicateurs potentiellement spécifiques à une paire de sillons donnée.

Cette fois encore, ce type d'information peut être considéré localement, au niveau de paires de segments distants ou à l'échelle de paires de sillons en fonction de la fiabilité locale de l'information. Il est possible qu'il soit également nécessaire de considérer la connectivité au niveau des structures gyrales environnantes où les densités de fibres seraient plus fortes qu'ailleurs.

13.2.3 Couplage : normalisation et modélisation Markovienne

Dans la partie II nous avons présenté une méthode permettant de remettre en cause l'étape de normalisation spatiale des données considérées en exploitant nos connaissances sur la variabilité sulcale pendant le processus de reconnaissance automatique des labels. Cette approche a démontré son intérêt dans le cadre d'une modélisation à base de SPAMs où les labels des segments sont considérés indépendants les uns des autres. D'autre part, dans cette partie, l'approche Markovienne a permis une amélioration de la cohérence locale des labélisations proposées. Ces deux approches corrigent des effets assez différents et peuvent se compléter. À cet effet, l'approche Markovienne finalement retenue se base sur la normalisation obtenue par notre modélisation jointe (normalisation et labélisation, partie II) et réalise l'inférence par le modèle Markovien dans cet espace. Dans cette approche, le modèle Markovien, pourtant plus efficace que le modèle indépendant n'est pas exploité lors de l'étape de normalisation. Une meilleure intégration des deux approches permettrait sans doute d'améliorer les résultats.

La méthode de normalisation et de labélisation jointe s'appuie sur une optimisation itérative conduite par l'algorithme EM. L'approche présentée en équation 7.4 se généralise facilement au modèle Markovien présenté en équation 11.1. L'algorithme consiste à trouver les paramètres de normalisation θ qui maximisent l'expression suivante (étape M) :

$$\begin{aligned}
Q\left(\theta \mid \theta^{(n)}\right) &= \sum_{L=l} P(L|D; \theta^{(n)}) \log [P(DL; \theta)] \\
&= \sum_{L=l} P(L|D; \theta^{(n)}) \log \left[P(L) \prod_{i \in \mathcal{E}} P(D_i|L_i) \prod_{i \sim j \in \mathcal{E}} P(D_{ij}|L_i L_j) \right] \\
&= \sum_{L=l} P(L|D; \theta^{(n)}) \log [P(L; \theta)] + \sum_{i \in \mathcal{E}} \sum_{L_i=l_i} P(L_i = l_i | D_i; \theta^{(n)}) \log [P(D_i | L_i = l_i; \theta)] \\
&\quad + \sum_{i \sim j \in \mathcal{E}} \sum_{L_i=l_i} \sum_{L_j=l_j} P(L_i = l_i \ L_j = l_j | D_{ij}; \theta^{(n)}) \log [P(D_{ij} | L_i = l_i \ L_j = l_j; \theta)]
\end{aligned} \tag{13.2}$$

Cette expression se décompose en trois parties impliquant l'a priori, les modèles locaux à l'échelle des segments et les modèles de relations. L'étape d'estimation de l'algorithme EM consiste à évaluer les probabilités a posteriori : $P(L_i = l_i | D_i; \theta^{(n)})$ pour tous les segments et tous les labels, et $P(L_i = l_i \ L_j = l_j | D_{ij}; \theta^{(n)})$ pour toutes les paires de segment et de labels. Dans le cadre d'une modélisation Markovienne, l'estimation de ces mesures peut être réalisée par MCMC (*Monte-Carlo Markov Chain* [Neal 1993]) en suivant la stratégie suivie par la méthode MPM (section 10.3.1) ou de façon approchée par la méthode *Loopy Belief Propagation* (section 10.3.2.3) ou en utilisant une approximation de type champ moyen (section 10.3.2.5).

Conclusion générale

Conclusion générale

14.1 Résumé des contributions

La contribution majeure de cette thèse est la mise au point d'une nouvelle méthode d'identification automatique des sillons corticaux. Elle améliore considérablement les résultats précédents avec un gain de plus de 15% par rapport au modèle «historique» présenté en introduction de cette thèse. Désormais, le taux de reconnaissance des structures atteint des résultats supérieurs à 86% en moyenne et plus de 97% pour le sillon central. De telles performances n'avaient été observées dans la littérature que sur une dizaine de structures tout au plus (hormis quelques travaux), alors que nous avons considéré ici près de 62 plissements corticaux. Ces résultats promettent de nombreuses applications que nous évoquerons dans la section suivante.

Pour commencer, le modèle «historique» mis au point précédemment par Denis Rivière [Rivière 2002] a été étendu. Des machines à vecteurs de support (SVM) ont été utilisées à la place de réseaux neuronaux dans la définition des experts locaux chargés de l'identification de sillons spécifiques. Ce changement a permis d'améliorer la robustesse de l'apprentissage grâce à l'optimisation par validation croisée des paramètres du modèle. Une étape de réduction de dimensions a été introduite dans chaque modèle local afin de limiter le surapprentissage induit par un nombre de descripteurs trop élevé face au nombre restreint de sujets.

Sur le plan méthodologique, nous avons proposé une formulation Bayésienne du problème de labélisation automatique, afin de faire face à l'étendue de la variabilité corticale et à l'ambiguïté intrinsèque qui la caractérise. Nous avons privilégié une approche descriptive (au travers de modèles génératifs), plus proche de la représentation mentale de l'expert en anatomie, ce qui facilite la validation du modèle et l'incorporation d'a priori. Nous nous sommes focalisé sur des informations locales (au niveau des segments sulcaux à labéliser et leur voisinage) permettant ainsi la mise en œuvre de méthodes d'optimisation efficaces, facilitant ainsi leur validation ou leur utilisation pratique.

Trois modèles principaux ont été proposés, répartis selon les trois grandes parties centrales de ce manuscrit : tout d'abord, une approche basique exploitant des informations de localisation des structures sulcales (partie I) à l'origine des deux modèles suivants :

- le problème d'identification des sillons corticaux a été reformulé comme un problème de normalisation spatiale contraint par des structures sulcales dont les labels sont inconnus (partie II). Il s'agit d'une formulation inédite dans ce cadre, qui permet d'imaginer de futurs raffinements par l'intégration de méthodes de normalisation plus avancées dans ce modèle. Dès à présent, elle a permis de développer une méthode de normalisation affine qui surpasse les approches affines standards testées, ainsi qu'une méthode affine par morceaux (une par label sulcal) à partir de laquelle une normalisation non-linéaire peut être définie.
- une approche Markovienne en a été dérivée afin de tenir compte de la direction et de la distance entre segments sulcaux voisins (partie III). Elle a permis de corriger des erreurs locales impliquant principalement des embranchements secondaires. Elle constitue un premier pas vers des informations plus haut niveau impliquant les sillons ou leur agencement.

Au cours de cette thèse, 62 sujets (hémisphères gauches et droits) ont été labélisés manuellement selon les recommandations du neuroanatomiste Jean Régis [Régis 2005]. Cette nouvelle base de données a déjà permis d'affiner notre modélisation de la variabilité sulcale et permettra sans doute d'améliorer notre compréhension de l'anatomie corticale qui reste encore assez limitée dans les zones les plus variables.

14.2 Applications

L'ensemble des méthodes proposées pendant cette thèse ont été intégrées au sein de la plateforme logicielle Brainvisa [Rivière 2009]. Les performances annoncées en terme de qualité d'identification des structures sulcales et vitesse de traitement, facilitent leur utilisation, favorisant ainsi le développement de nouvelles applications (nous retiendrons¹ ici uniquement celles s'appuyant sur nos travaux [Rivière 2002, Perrot 2009]).

L'intérêt pour les structures corticales tend à se généraliser mais il s'agit d'une préoccupation encore récente motivée principalement par son intérêt dans la prise en compte de la variabilité inter-individuelle dans les études de groupe. Notre méthode définit automatiquement de nombreux amers anatomiques sur l'ensemble du cerveau. Ces points de repère peuvent être utilisés afin de contraindre la normalisation d'un groupe de sujets : [Cachier 2001], plus récemment en utilisant une approche à base de courants [Auzias 2009], ou par son intégration dans le processus d'identification automatique et la prise en compte des incertitudes de labélisation [Perrot 2009]. En s'appuyant sur l'alignement préférentiel des plissements selon une organisation en méridiens et parallèles, il est possible de définir un système de coordonnées surfacique basé sur les sillons s'exprimant sous la forme d'une projection sphérique de la surface corticale [Clouchoux 2005].

Ces plissements permettent également de définir une segmentation de la surface corticale en parcelles s'appuyant sur les gyri, depuis la projection sphérique évoquée plus tôt [Clouchoux 2006] ou à partir de diagrammes de Voronoï géodésiques s'appuyant sur les sillons [Cachia 2003b]. Plus directement, chaque modèle proposé fournit un atlas anatomique probabiliste de 62 structures sulcales : depuis un espace normalisé standard ou raffiné par co-recalage selon notre approche. De tels atlas sont communément utilisés en VBM (*Voxel Based Morphometry*) ou en analyse fonctionnelle afin d'identifier les structures impliquées dans le processus étudié. Une version déterministe de notre atlas normalisé a été impliqué dans une étude fonctionnelle visant à réduire l'incertitude spatiale dans le cadre de détection d'activations fonctionnelles [Keller 2009].

Les études standards en morphométrie cérébrale se basent généralement sur une étude voxel à voxel (VBM) dont l'efficacité reste limitée par la variabilité inter-individuelle. Notre laboratoire a mis au point une approche alternative basée sur une description morphométrique (longueur, profondeur, direction, topologie, épaisseur corticale,...) des plissements corticaux [Mangin 2003a] qui a permis la mise en place d'outils de classification dédiés [Duchesnay 2007]. Cette approche descriptive se poursuit dorénavant par la mise en évidence de *patterns* sulcaux, c'est à dire un ensemble de profils locaux particuliers de structures sulcales ou d'agencements de plis partagés par un groupes de sujets [Sun 2009]. Cette approche présente un intérêt certain, particulièrement dans le cas de maladies neurodégénératives et de certains troubles psychiatriques qui pourraient être liés à un développement anormal du cerveau. Elle s'est illustrée récemment par l'identification de plissements corticaux anormaux chez des schizophrènes aux hallucinations auditives résistantes [Cachia 2008] et dans le cadre du diagnostic précoce de la maladie d'Alzheimer [Fouquet 2009]. Puisque l'identification automatique des structures sulcales n'est pas parfaite (et ne le sera sans doute jamais complètement compte tenu de l'extrême variabilité inter-individuelle), nous suggérons

¹La liste de travaux cités est non-exhaustive.

de prendre en compte les probabilités a posteriori de labélisation dans les calculs statistiques menés ultérieurement.

L'intérêt des sillons n'est pas cantonné à la recherche, qu'elle soit clinique ou méthodologique. Certains plissements corticaux primaires délimitent des aires fonctionnelles qu'il est souhaitable de repérer pendant la phase préopératoire précédant certaines interventions chirurgicales : traitement de tumeurs cérébrales ou stimulation corticale par exemple. Dans le cadre du projet Miniara (pôle de compétitivité Medicen), impliquant notamment la plateforme de recherche Neurospin et l'entreprise Dosisoft², certains travaux de cette thèse seront mis à profit pour détecter automatiquement les aires sensorimotrices à partir du sillon central alors que les approches suivies en routine clinique demandent généralement l'intervention d'un expert en neuroanatomie [Uwano 2008].

14.3 Perspectives

Au cours du manuscrit de nombreuses perspectives ont été proposées (à la fin de chacune des grandes parties) ; nous en rappellerons ici les grandes lignes. D'autres perspectives plus générales sont également proposées.

Du point de vue de la formalisation, un manque important est à noter : l'absence de modèle génératif bien défini pour les segments sulcaux. Dans cette approche, les modèles SPAMs modélisent un processus ponctuel (la probabilité qu'un sillon passe par un voxel donné d'un espace normalisé) à partir duquel une vraisemblance ad hoc est calculée au niveau d'un segment entier. En pratique, les labélisations erronées liées à cette approximation semblent assez rares (l'ampleur du phénomène dépend sûrement des paramètres de segmentation des sillons en segments) et pourraient généralement être corrigées par l'apport d'informations additionnelles. Néanmoins, il s'agit d'un point méthodologique à éclaircir.

Concernant la représentation de l'information sulcale, une alternative populaire vise à considérer une approche surfacique. À ce détail près, il serait sans doute possible de «rejouer» l'intégralité des méthodes proposées à partir d'une définition surfacique du concept de SPAM. Une telle approche faciliterait la manipulation des données et accélérerait certainement de nombreux calculs puisque le nombre de données à analyser est bien moins élevé que dans le cas volumique. Dans ce cadre, il est possible de définir facilement une représentation multi-échelle de l'information, ce qui permettrait sans doute d'estimer des mesures locales plus robustes (par exemple l'information de courbure). Dans le cadre d'une modélisation Markovienne, la description du voisinage pourrait également impliquer des mesures géodésiques à cette surface dont les calculs seraient alors facilités par cette représentation. La motivation principale de cette approche repose sur les limitations connues de la méthode d'extraction des sillons. Elle éprouve des difficultés certaines lorsque les structures sulcales sont peu marquées, notamment au niveau de l'Insula, alors que les plissements y sont plutôt stables.

Qu'il s'agisse de cerveaux dont les plissements ont été altérés par une pathologie ou une particularité individuelle chez un sujet sain, il est nécessaire de tenir compte des configurations corticales atypiques. Actuellement leur gestion est très limitée. En effet, dans ces situations les vraisemblances mesurées des différentes composantes du modèle peuvent être en contradiction avec la réalité anatomique. Seule leur mise en compétition permet actuellement de résoudre certaines situations problématiques. Malheureusement, parfois l'ensemble des composantes est mis en défaut, d'autres solutions sont donc à envisager.

La première consiste naturellement à introduire des informations de plus haut niveau (niveau sillon ou paire de sillons voisins), ou plutôt à les ré-introduire puisqu'il s'agit de l'approche suivie

²<http://www.dosisoft.com>

par les travaux de Denis Rivière [Rivière 2002] dans le cadre du modèle «historique», et ainsi compléter les informations locales (niveau segment et paire de segments voisins) utilisées actuellement. Concernant l'agencement des sillons et la cohérence spatiale des labels voisins, nous avons montré dans cette thèse deux façons de le modéliser : par l'intermédiaire d'un champ de Markov ou par la modélisation de déformations induites par une normalisation. Ces approches sont complémentaires et gagneraient à être utilisées de façon conjointe.

Une autre façon de prendre en considération les configurations atypiques consiste à utiliser un jeu de labels supplémentaires constituant un modèle de mélange pour chaque label où y seraient distinguées les situations standards des autres. En conséquence, autant de modèles SPAMs et de modèles locaux Markoviens que nécessaire seraient estimés. Cette approche permettrait également de différencier la modélisation des sillons présentant plusieurs modes dans la population (les situations très rares seraient alors associées à une classe de rejet).

Les performances des modèles proposées sont assez élevées et sans doute assez proches de la variabilité inter-expert dans certaines zones du cerveau. Pour poursuivre ce travail et mesurer effectivement des améliorations significatives, un nouvel agrandissement de la base de données sera nécessaire. Puisqu'il s'agit d'une opération longue et difficile, nous suggérons deux approches alternatives beaucoup plus faciles à mettre en œuvre.

Une solution envisagée et exposée partiellement dans ce manuscrit est l'utilisation d'une approche semi-supervisée. Il s'agit d'associer un ensemble de sujets déjà labélisés manuellement à un grand nombre de sujets non-labélisés. Si le modèle utilisé est suffisant, cette approche permettrait d'estimer un modèle représentatif de l'ensemble des sujets considérés et identifier leurs labels de façon conjointe.

Une autre approche consisterait à limiter la labélisation manuelle d'une grande base de données à quelques sillons majeurs facilement identifiables. Cette opération est envisageable rapidement pour plusieurs centaines de cerveaux. Une étude poussée des caractéristiques de ces structures serait alors conduite afin d'améliorer spécifiquement la qualité de leur identification automatique, plutôt qu'une stratégie de modélisation indépendante des spécificités locales (le choix du modèle et des informations représentées et non les paramètres du dit modèle). Notre choix se portera naturellement sur les structures les plus utilisées en pratique, que ce soit pour repérer des zones fonctionnelles ou réduire la variabilité inter-individuelle.

L'identification des structures sulcales n'est pas une fin en soi, mais un moyen sur lequel s'appuient d'autres études. De plus, l'amélioration des performances de reconnaissance est intrinsèquement limitée par nos connaissances en anatomie corticale (représentées essentiellement par le choix des labels sulcaux). C'est pourquoi il est nécessaire de considérer l'intérêt du processus dans un cadre plus global afin d'améliorer notre compréhension du cerveau qu'il s'agisse de son anatomie ou de son fonctionnement.

L'étude de la formation et du développement cérébral [Lefèvre 2009] devrait permettre de mieux comprendre l'origine des plissements, information qui profitera directement à leur modélisation chez l'adulte. Une modélisation longitudinale de la variabilité sulcale pourrait également être envisagée dans ce cadre.

L'apparition récente de données à haut-champ (IRM 7T à Neurospin) nous invite à identifier les différentes couches de la matière grise, et ainsi distinguer des zones cytoarchitectoniques homogènes, connues sous le nom d'aires de Brodmann.

L'ensemble des connaissances accumulées et modélisées sur les sillons pourraient profiter à la compréhension d'informations issues d'autres modalités comme les données fonctionnelles et de diffusion. De telles informations pourraient également bénéficier à la qualité de l'identification des structures sulcales. En effet, certaines zones fonctionnelles sont associées à des structures anatomiques bien définies comme les aires motrices primaires. L'IRM de diffusion permet de détecter

les fibres de la matière blanche qui sont impliquées dans la formation des plissements corticaux [Essen 1997] et participent à l'échange d'information entre aires fonctionnelles distantes en passant par certaines structures sous-corticales. À terme, il serait profitable de prendre en compte l'ensemble de ces inter-connexions au sein d'un unique modèle permettant l'identification couplée de noyaux, de faisceaux de fibres et de structure sulcales.

Quatrième partie

Annexes

Table des Matières

Notations	155
A Base de données de sillons	157
A.1 Motivations	157
A.2 Élaboration d'une base de données sulcales	158
A.3 Outils	159
A.4 Planches anatomiques	160
B Méthodes diverses	169
B.1 Iterative Closest Point (ICP)	169
B.1.1 Association	170
B.1.2 Estimation	171
B.2 <i>k-means</i> Riemannien de données directionnelles et axiales	173
B.2.1 <i>k-means</i> classique	173
B.2.2 Approche Riemannienne	174
B.3 Quelques distributions de probabilités	177
B.3.1 Distributions circulaires, directionnelles, axiales, rotations	178
B.3.2 Distribution sur un simplexe	190
B.4 Alternatives et extensions au modèle SPAM	193
B.4.1 Approximations	193
B.4.2 Extensions	199
B.4.3 Alternatives	201
C Développements informatiques	205
C.1 Introduction	205
C.2 Technologies utilisées	205
C.3 Commandes et fonctionnalités	206
C.3.1 Mesures d'erreurs	207
C.3.2 Visualisation	207
C.3.3 Apprentissage de modèles et labélisation automatique	208
C.3.4 Parallélisation de tâches	208
Bibliographie	211

Notations

Nous regroupons ici les notations principales utilisées tout au long du manuscrit :

\mathcal{A}	Base d'apprentissage.
\mathcal{T}	Base de Test.
\mathcal{B}	Union de la base d'apprentissage et de la base de Test.
\mathcal{L}	Ensemble des labels possibles pour désigner des structures sulcales.
\mathcal{E}	Ensemble des segments sulcaux élémentaires d'un sujet donné.
\mathcal{E}_a	<i>idem</i> pour le sujet a de la base d'apprentissage \mathcal{A} .
\mathcal{E}_t	<i>idem</i> pour le sujet t de la base de test \mathcal{T} .
\mathcal{E}_l	<i>idem</i> mais limité au sillon l .
\mathcal{M}	Modèle de la variabilité sulcale dont la formulation diffère selon le contexte. Contient toutes les informations permettant de réaliser la labélisation d'un sujet (qu'il appartienne à la base d'apprentissage ou de test).
\mathcal{M}_s	restriction de \mathcal{M} aux paramètres contrôlant la partie générative du modèle sulcale.
\mathcal{M}_l	restriction de \mathcal{M}_s aux paramètres liés au label l .
\mathcal{M}_p	restriction de \mathcal{M} aux paramètres contrôlant l'a priori sur la labélisation

a	Élément de la base d'apprentissage \mathcal{A} .
t	Élément de la base de test \mathcal{T} .
b	Élément de l'union de la base d'apprentissage et de la base de test \mathcal{B} .
l	information dédiée au label l appartenant à \mathcal{L} .
i, j	segments sulcaux élémentaires appartenant à un sujet donné pris dans l'ensemble \mathcal{E} .

$L_i = l_i$	Label porté par le segment i sur un sujet donné.
$L = l$	} Labélisation complète d'un sujet donné, avec $l \in \mathcal{L}$ et $l_i \in \mathcal{L}$.
$\{L_i = l_i\}_{i \in \mathcal{E}}$	
L_a	Labélisation complète du sujet a de la base d'apprentissage \mathcal{A} .
L_t	Labélisation complète du sujet t de la base de test \mathcal{T} .
D	Ensemble des données structurales, morphométriques, positionnelles dont on dispose sans que la connaissance des labels des sillons soit nécessaire.
D_a	<i>idem</i> mais concernant le sujet a de la base d'apprentissage \mathcal{A} .
D_t	<i>idem</i> mais concernant le sujet t de la base de test \mathcal{T} .
D_i	<i>idem</i> mais limité au segment i .
D_{ij}	<i>idem</i> mais limité au couple de segments i et j , et éventuellement à la zone d'interactions entre ces deux segments.

$\Theta = \theta$	Ensemble des paramètres de normalisation sans précisé le type de normalisation.
$\Theta_a = \theta_a$	<i>idem</i> mais concernant le sujet a de la base d'apprentissage \mathcal{A} .
$\Theta_t = \theta_t$	<i>idem</i> mais concernant le sujet t de la base de test \mathcal{T} .
$\Theta_g = \theta_g$	Ensemble des paramètres régissant une normalisation globale (rigide ou affine).
$\Theta_s = \{\theta_{s,l}\}_{l \in \mathcal{L}}$	Ensemble des paramètres régissant une normalisation rigide par morceaux (avec une transformation rigide par label l).

Base de données de sillons

Le cerveau humain présente de nombreux plissements tortueux aux formes labyrinthiques (où les sillons jouent le rôle des parois alors que les gyri constituent les zones de passage) qui varient fortement d'un individu à l'autre. Seul les plus grands et les plus profonds plis font consensus parmi les experts en anatomie sulcale. Il s'agit des sillons primaires, les premiers à apparaître dès les premières phases du développement cérébral. À ce stade, ces plissements élémentaires [Dubois 2008] (ou proto-sillons), c'est à dire les plissements à l'origine des futurs sillons possèdent une localisation inter-individus plutôt stable. Au delà, concernant les plissements secondaires ou tertiaires dont l'apparition se fait plus tard, il n'existe pas de vérité de terrain. Il s'agit encore d'un domaine de recherche actif où de nombreuses théories [Régis 2005, Lohmann 2000, Ono 1990] se côtoient. Elles partagent de grandes lignes directrices communes concernant les structures anatomiques les plus stables, mais peuvent diverger lorsqu'il s'agit de structures plus variables. Nous sommes persuadés que l'étude du développement cérébral, et en particulier un suivi longitudinal anténatal permettrait de mieux comprendre l'origine de la variabilité cérébrale et par suite, de mieux la modéliser. La comparaison conjointe des plissements corticaux aux fibres de la matière blanche ou à des activations fonctionnelles spécifiques pourrait lever certaines incertitudes dans la nomenclature sulcale et participer à unifier les théories à ce sujet.

A.1 Motivations

À partir de l'analyse des travaux que nous avons mené dans [Perrot 2008] nous avons mis en avant que la modélisation de la variabilité sulcale proposée permettait d'évaluer a posteriori la fiabilité du choix d'une nomenclature sulcale particulière. Nous avons utilisé à ce propos, un modèle exploitant des informations de localisation grâce à des SPAMS (voir section 4.2.3) pour analyser la qualité de la labélisation manuelle de notre base d'apprentissage contenant à l'époque 26 sujets. Cela nous a permis de détecter certaines erreurs ou labels manquants. Bien entendu, une telle base de données ne pouvait prétendre être une bonne représentation de l'étendue de la variabilité inter-individuelle sulcale. En fait, cette étude nous a permis d'acquérir une bonne compréhension concernant la plupart des grands plissements, mais seulement une vision limitée avec de nombreuses questions sans réponses pour les sillons les plus variables. À cet égard, la nomenclature sulcale qui avait été définie jusqu'alors a été retravaillée et raffinée (certaines structures trop variables ont été regroupées, d'autres ont été découpées en plusieurs composantes lorsque leur description était trop grossière) en se basant cette fois sur l'étude de 62 sujets (sur les hémisphère gauches et droits contre uniquement les droits précédemment). Les résultats ont été à la hauteur du temps passé à élaborer cette base de données puisque la labélisation a gagné en cohérence au travers l'ensemble des sujets et entre les deux hémisphères (il y avait près de 5% d'écart de qualité d'identification des structures sulcales entre les deux hémisphères auparavant contre 1% dorénavant pour cette nouvelle base), même sur les plus petits plissements. Néanmoins, certaines incertitudes, plutôt limitées et localisées, demeurent et nécessiteraient l'étude d'un nombre encore plus important de sujets. Mais la taille de la nouvelle base de données proposée au cours de ce travail de thèse a permis, et permettra encore dans l'avenir, de suggérer des améliorations pour la modélisation sulcale et par conséquent pour la labélisation automatique.

A.2 Élaboration d'une base de données sulcales

Ces 62 sujets (y compris les 26 sujets considérés précédemment) ont été totalement labélisés ou relabélisés par mes soins avec l'aide d'Alan Tucholka, pendant de nombreuses séances d'une heure qui se sont étalées dans le temps sur plusieurs mois au début de la troisième année de thèse. Par ses mécanismes d'apprentissage évolués, le cerveau humain est extrêmement doué pour trouver de la cohérence dans un jeu de données même si seul le hasard a fait apparaître ces structures. On peut alors tirer de fausses conclusions assez facilement si l'on ne considère pas assez de sujets en même temps. Pour cette raison, les cerveaux ont été comparés les uns avec les autres plusieurs fois par paquets de 10 sujets environ (il s'agit d'une sorte de procédure de validation croisée pour rendre plus robuste la labélisation manuelle). Pour mener ce travail à bout, nous avons piloté Anatomist, le logiciel de visualisation de données médicales 2D/3D de la plateforme logicielle Brainvisa [Rivière 2009]. Ainsi, chaque choix de label a été discuté longuement pour rendre les labélisations plus consistantes entre elles. Les hémisphères gauches et droits ont été labélisés individuellement : l'ensemble des hémisphères droits suivis par les gauches. Constat intéressant, les modèles associés présentent de fortes symétries (voir figure 8.4). Ainsi, nous espérons avoir évité certains biais potentiels. Le processus de labélisation nous a permis de mettre en avant certains résultats connus concernant des asymétries entre les deux hémisphères : les aires de Broca et de Wernicke de l'hémisphère gauche (la plupart du temps), impliquées dans le traitement du langage, sont plus variables que les zones correspondantes sur l'autre hémisphère. De plus, les sillons de la partie inférieure du cortex (fissure collatérale, sillon lingual, sillon occipito-temporal...) semblent être moins variables et donc plus faciles à reconnaître sur les hémisphères gauches, alors même que la segmentation des sillons a quelques difficultés dans cette zone (indifféremment de l'hémisphère).

La nomenclature des sillons choisie et les principes de labélisation utilisés pour construire cette base de données sont basés sur la théorie des racines sulcales proposée par Jean Régis [Régis 2005]. Ces racines sulcales correspondent à la position des premiers signes de plissement de chaque sillon pendant la croissance du cortex. Elles pourraient être liées au développement des fibres de la matière blanche. Nous pensons que ces plissements élémentaires sont relativement fiables pour mettre en correspondance des cerveaux. Au delà des facteurs génétiques¹, il est probable que la majorité de la complexité apparente des configurations corticales résulte de la diversité des stimuli extérieurs auxquels il est confronté pendant le processus de développement cérébral. Ce dernier serait alors conditionné par la structuration en partie chaotique, aux échelles microscopiques, de réseaux neuro-naux complexes possédant de nombreuses boucles de rétroactions qui donnent lieu à des différences importantes aux échelles macroscopiques. À cause des limitations de notre compréhension actuelle de l'anatomie et des limitations du processus de segmentation, l'ensemble des labels que nous avons utilisés a été limité aux racines sulcales les plus fiables (par exemple, 5 labels ont été utilisés pour modéliser le sillon pré-central) et à des groupes de plissements voisins pour les zones les plus variables et les moins comprises. Par exemple, un unique label a été utilisé pour les plis corticaux du lobe occipital où aucune structure n'a pu réellement être mise en avant. De cette façon, nous avons défini un ensemble de 63 labels pour les hémisphères gauches et de 62 pour les droits. Ce label supplémentaire à gauche correspond à une différence anatomique nette ; il est situé entre la branche postérieure de la vallée sylvienne et le bas du sillon post-central. Les racines sulcales possèdent de fortes connexions avec le concept des bassins sulcaux [Lohmann 2000] qui comprennent les parties les plus profondes des sillons corticaux, sauf ceux qui ont été poussés vers l'extérieur lors du développement de zones avoisinantes.

La construction de cette base de données sulcales repose sur la réunion de plusieurs bases de données hétérogènes (notre ancienne base de données de 26 sujets, une base de données dédiées aux données de diffusion [Poupon 2006], une base de données de jumeaux dédiées à des études

¹Les cerveaux de vrais jumeaux sont moins dissemblables que deux cerveaux pris au hasard dans la population, mais les différences restent très importantes [Biondi 1998].

fonctionnelles et quelques sujets provenant de la base ICBM). Chaque base de données provient d'un site spécifique : SHFJ (Orsay), CHU La Pitié Salpêtrière (Paris), CHU La Timone (Marseille), McGill (Montréal, Canada), possédant chacun un scanner 1.5T avec ses caractéristiques propres (résolutions spatiales différentes par exemple). La plupart des sujets sont des hommes droitiers, entre 25 et 35 ans. Le projet scientifique associé à l'élaboration de chacune de ces bases de données n'était pas forcément lié à l'étude des sillons, mais toutes contenaient les volumes IRM T1 requis. De plus, pour une bonne moitié de la base, des acquisitions en IRM de diffusion et en IRM fonctionnelle (protocole du *Localizer*) sont disponibles, ce qui permettra sans nul doute, de futures études comparatives entre ces différentes modalités. Il est probable que la nature hétérogène des données a pu contribuer dans une certaine mesure, difficile à quantifier, à améliorer la robustesse des modèles sulcaux estimés.

A.3 Outils

Toute la phase d'élaboration de la base de données (y compris l'extension, le nettoyage, la mise en correspondance, les comparaisons, ...) a été accomplie avec l'aide du logiciel Anatomist [Rivière 2009]. Nous avons développé pour l'occasion une méthode de labélisation manuelle ergonomique, dont le but est de rendre l'utilisateur très efficace dans sa démarche en l'aidant à se concentrer au maximum sur la tâche de réflexion (le choix des labels) plutôt que de perdre du temps à appliquer ses choix sur les structures. Cette approche permet de partir de rien ou d'une première labélisation (obtenue par exemple par une des méthodes automatiques proposées dans cette thèse : approche plus rapide mais plus risquée puisqu'en cas de doute, on risque de favoriser le choix de label proposé par le modèle). Cette méthode est basée uniquement sur 3 actions : la sélection des structures à labéliser, la copie et le collage de labels d'une structure à l'autre (s'inspirant des opérations de copier/coller des traitements de texte). En partant d'une labélisation pré-établie et avec un entraînement intensif et régulier un cerveau entier peut être labélisé manuellement en moins d'une heure (contre 2 à 3 fois plus auparavant).

Finalement, nous avons dû créer un outil permettant de détecter et retirer les sujets-doublons, ceux apparaissant dans plusieurs des bases que nous avons regroupées (quelques sujets ayant subi plusieurs acquisitions à plusieurs années d'intervalle). À cet effet, un outil de recalage classique mais plutôt basique, appelé ICP [Besl 1992] (Iterative Closest Points, détaillé en section B.1) a été utilisé pour recaler par une transformation rigide chaque paire de sujets pris deux à deux. L'énergie résiduelle du recalage, autrement dit la distance entre chaque paire de sujets recalés a été utilisée dans ce contexte comme un indice de similarité entre les formes (dans [Sun 2009] cette mesure est utilisée pour faire du *clustering* de plissements corticaux), qui peut de manière équivalente s'appliquer aux segments de sillons, aux interfaces gris-blanc ou blanc-LCR. Nous avons accordé une attention toute particulière à cette étape, sachant que nous comptions utiliser des méthodes de validation par *leave-one-out* (validation croisée en laissant un seul sujet de côté à chaque étape) ; sans ça, les résultats obtenus auraient alors été fortement biaisés par ce sur-apprentissage.

L'ensemble des méthodes de labélisation automatique présenté dans ce manuscrit est limité par la qualité de la segmentation en segments sulcaux. Dans certains cas particuliers, lorsque deux sillons différents se retrouvent en parfait alignement, la segmentation ne permet pas de détecter le passage d'une structure à l'autre à partir de l'analyse du fond de sillon ou de sa courbure. Par conséquent, un seul label pourra être attribué au plissement. Dans ces configurations, il serait nécessaire de découper a posteriori les plissements posant problème. Nous disposons désormais d'un outil qui permet de réaliser une telle découpe à partir de la sélection manuelle d'un simple point. Ultérieurement, cette étape sera automatisée et intégrée directement au processus de labélisation automatique.

A.4 Planches anatomiques

Nous avons la volonté de présenter sous une forme visuelle la base de données d'informations sulcales qui a été créée et utilisée pendant ce travail de thèse. D'une part, parce qu'il s'agit d'un investissement en temps important et d'autre part car ce travail est à l'origine des avancées majeures que nous avons proposé par la suite en matière de reconnaissance automatique des sillons. De plus, cela peut permettre au lecteur intéressé de se faire une idée plus précise de la variabilité sulcale et pourquoi pas servir de support à des cours d'anatomie corticale (notre base de données devrait être mise à la disposition de la communauté scientifique dans un futur proche).

Pour des raisons de place et de visibilité, nous avons choisi de limiter nos planches anatomiques à 25 sujets (environ 40% de la base). Nous proposons alors pour chaque hémisphère 4 planches, une par vue (face externe, face interne, vue du dessus, vue du dessous). Les hémisphères gauches et droits se font face pour faciliter les comparaisons.

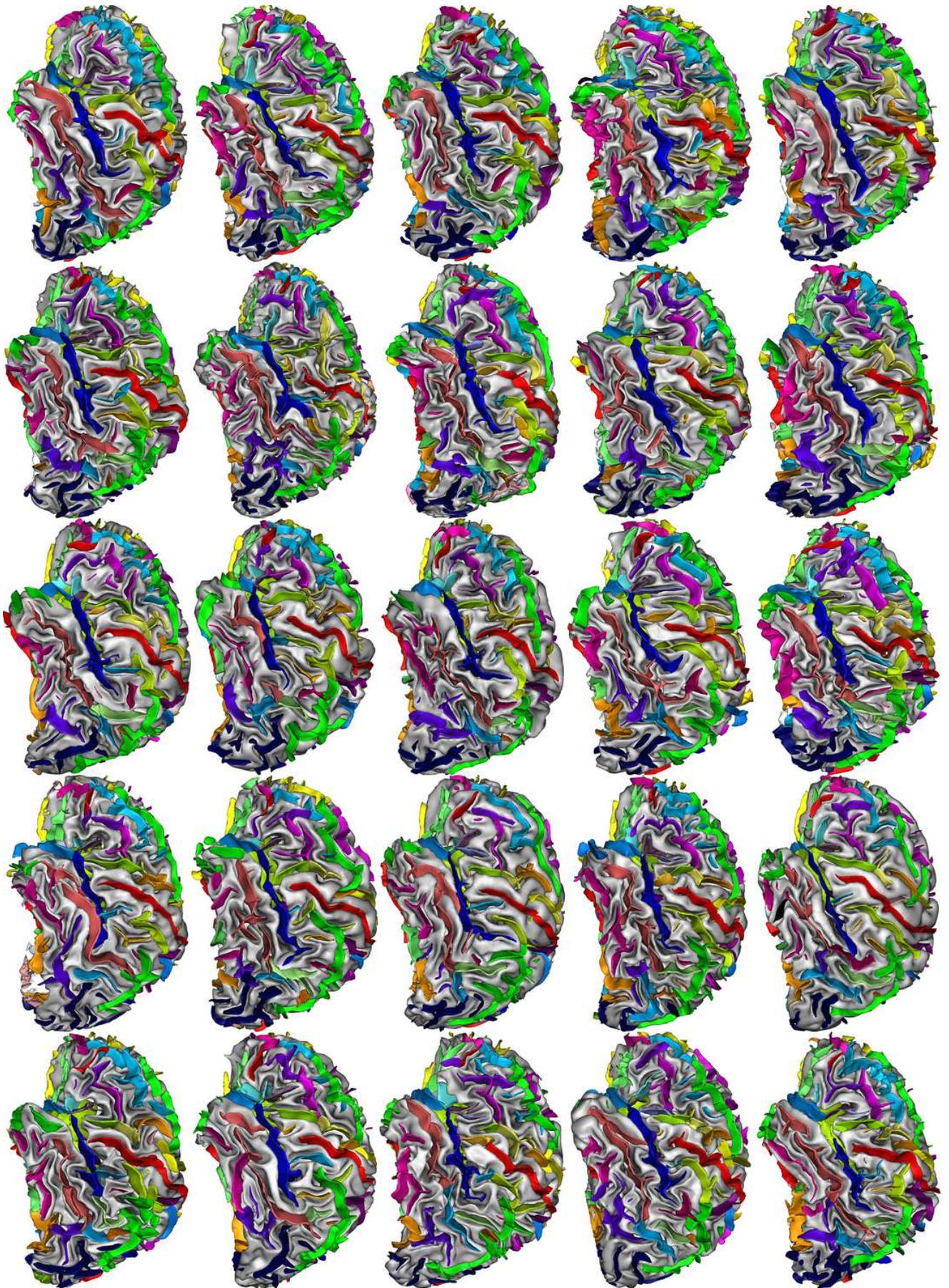


FIG. A.1 – Hémisphères gauches - Face externe

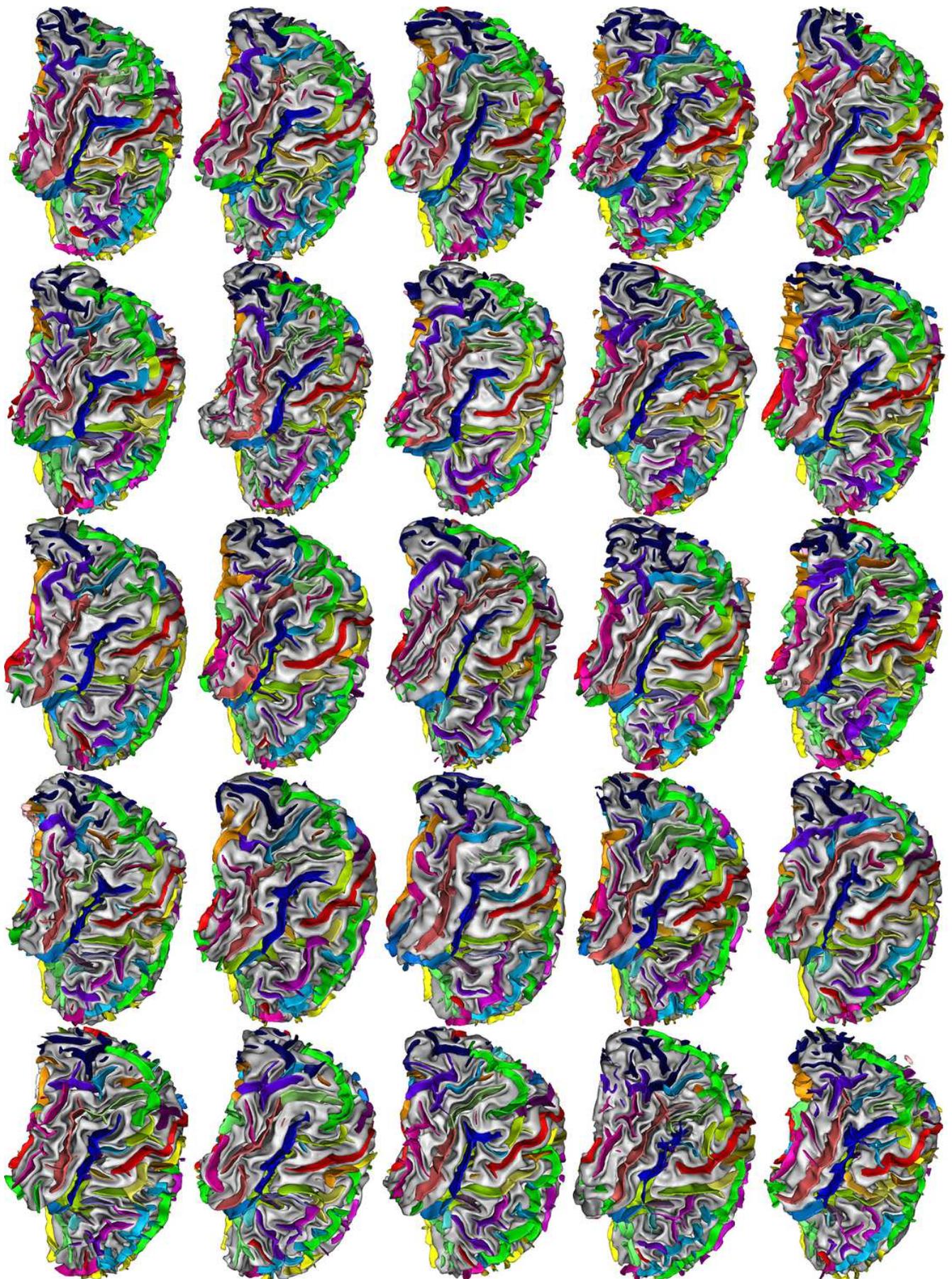


FIG. A.2 – Hémisphères droits - Face externe

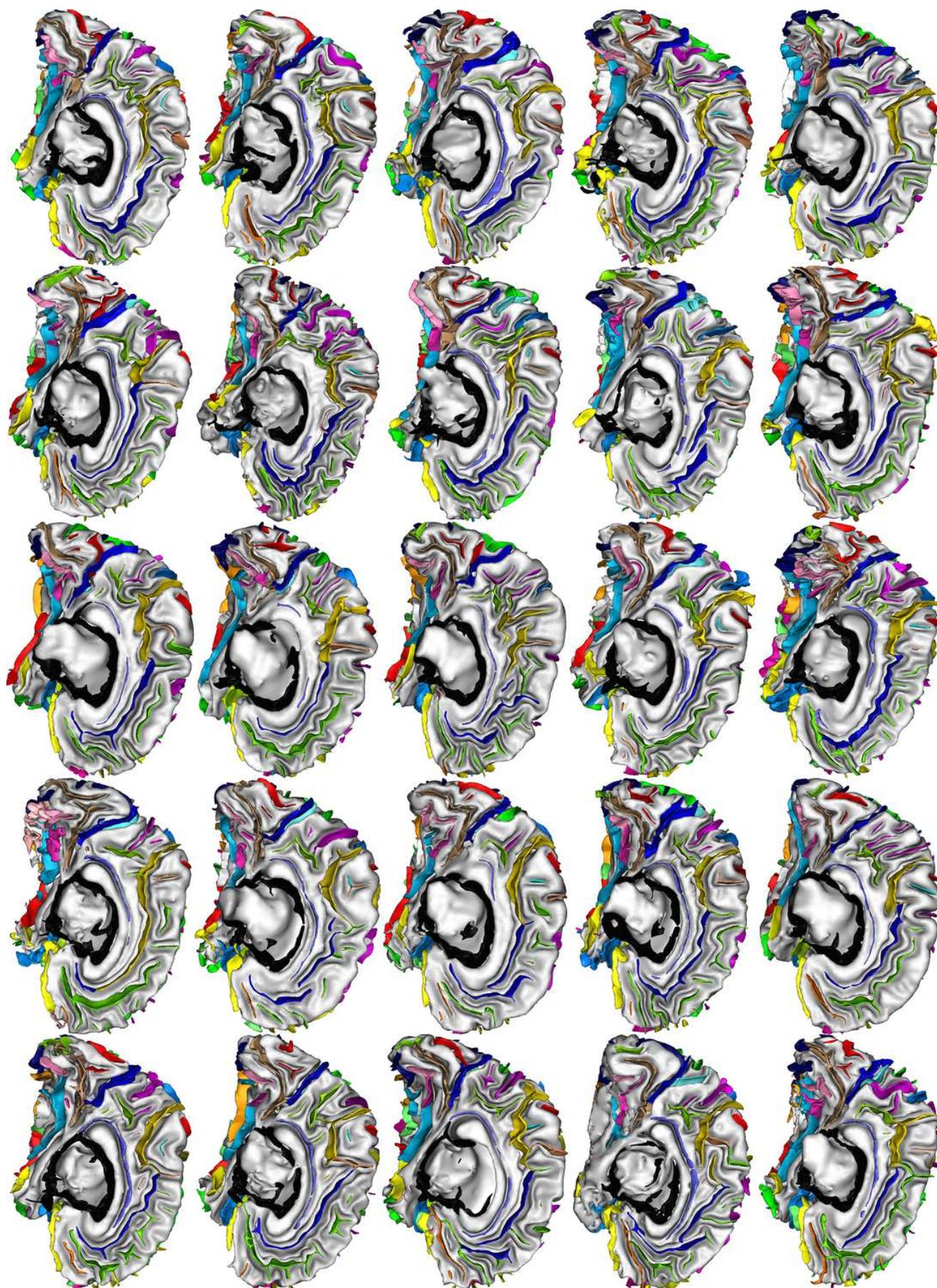


FIG. A.3 – Hémisphères gauches - Face interne

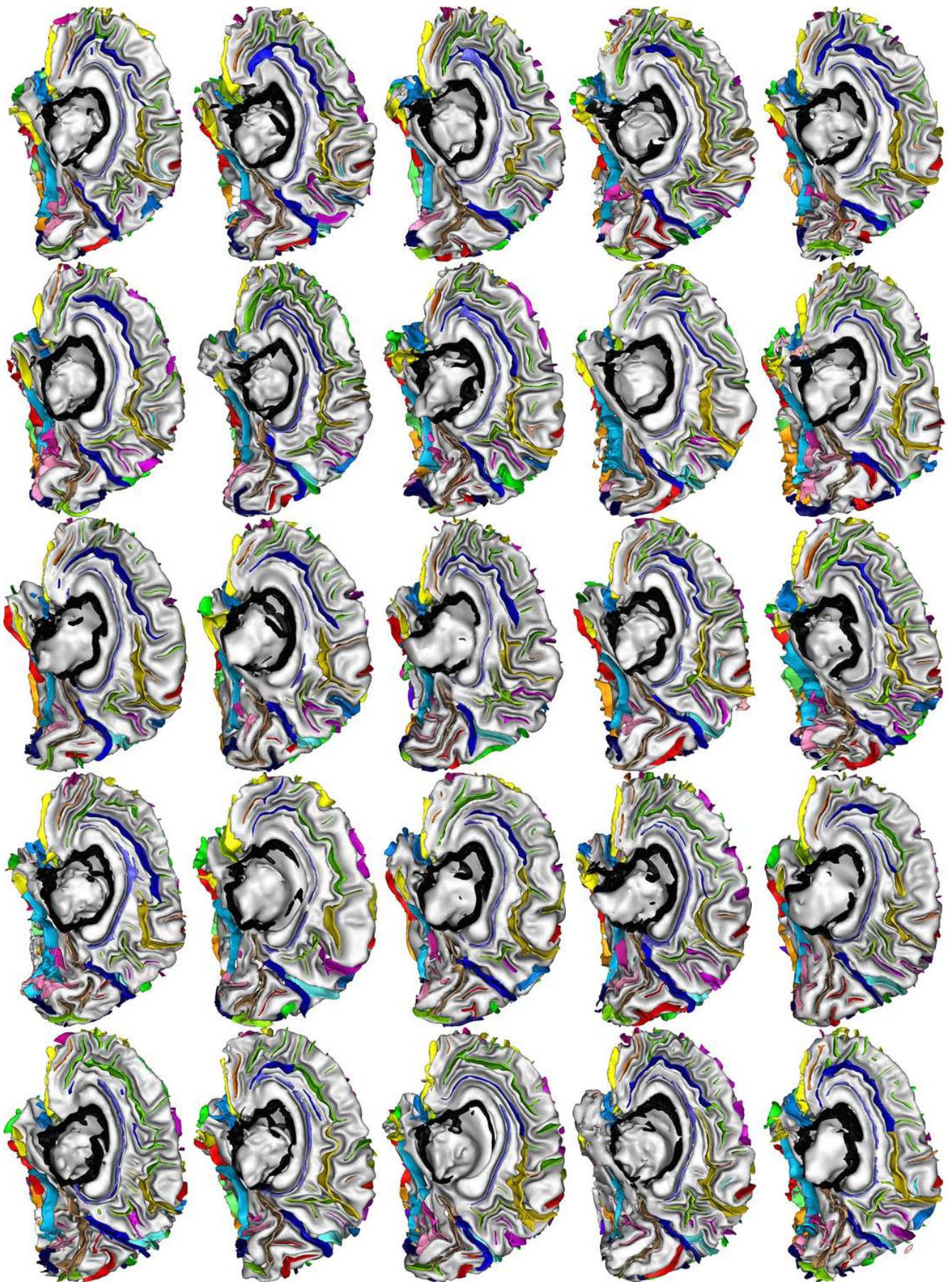


FIG. A.4 – Hémisphères droits - Face interne

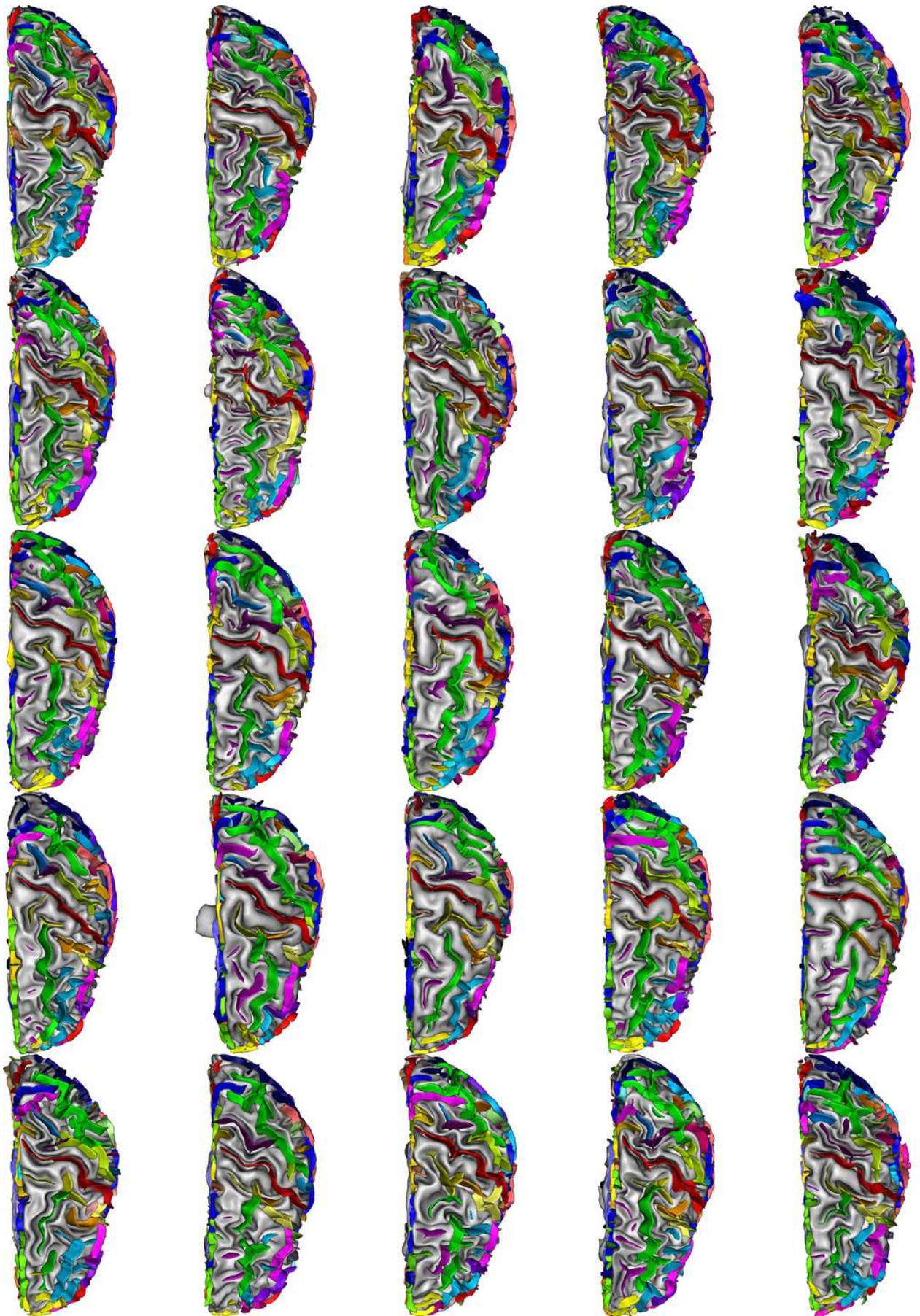


FIG. A.5 – Hémisphères gauches - Vue du dessus

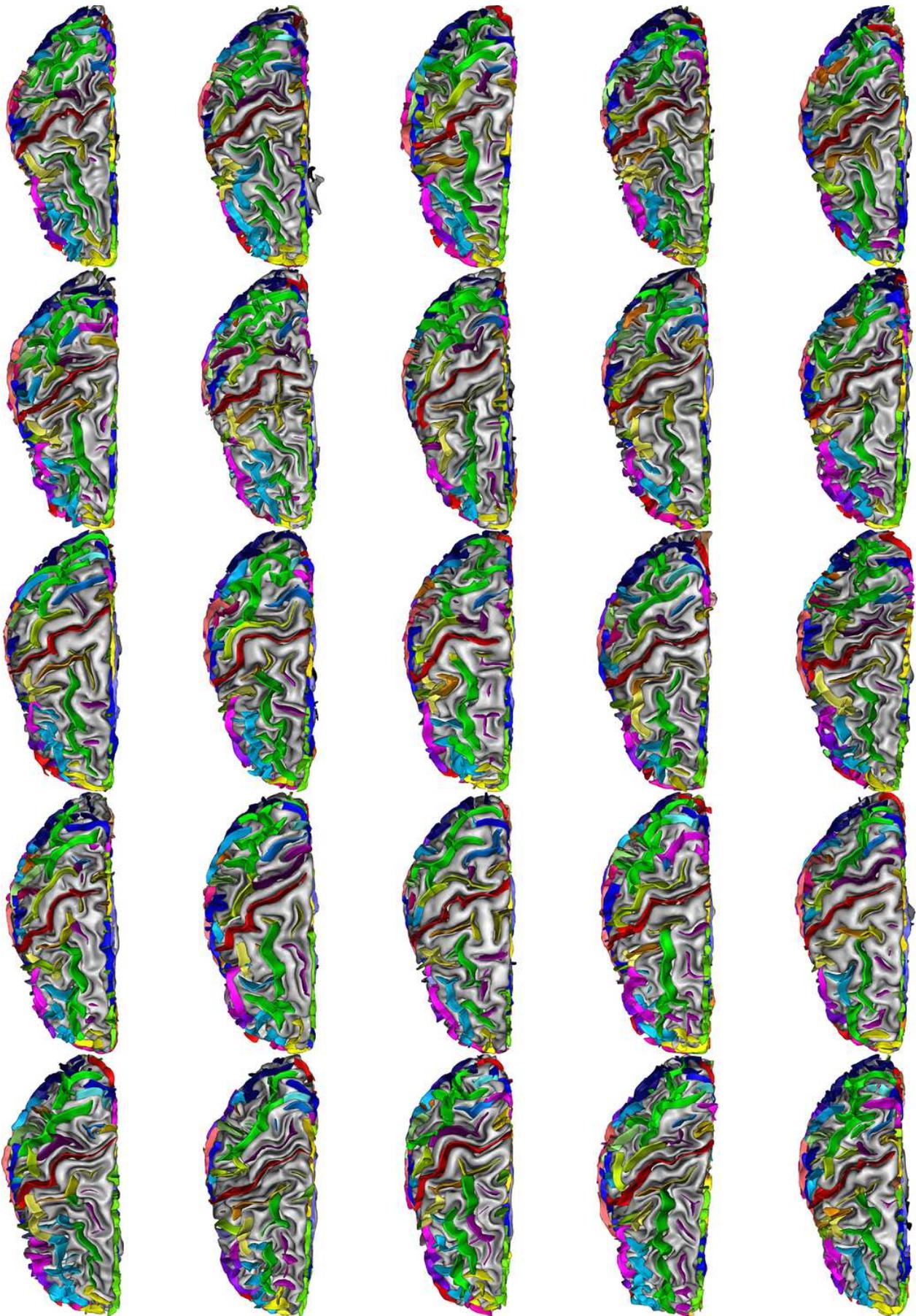


FIG. A.6 – Hémisphères droits - Vue du dessus

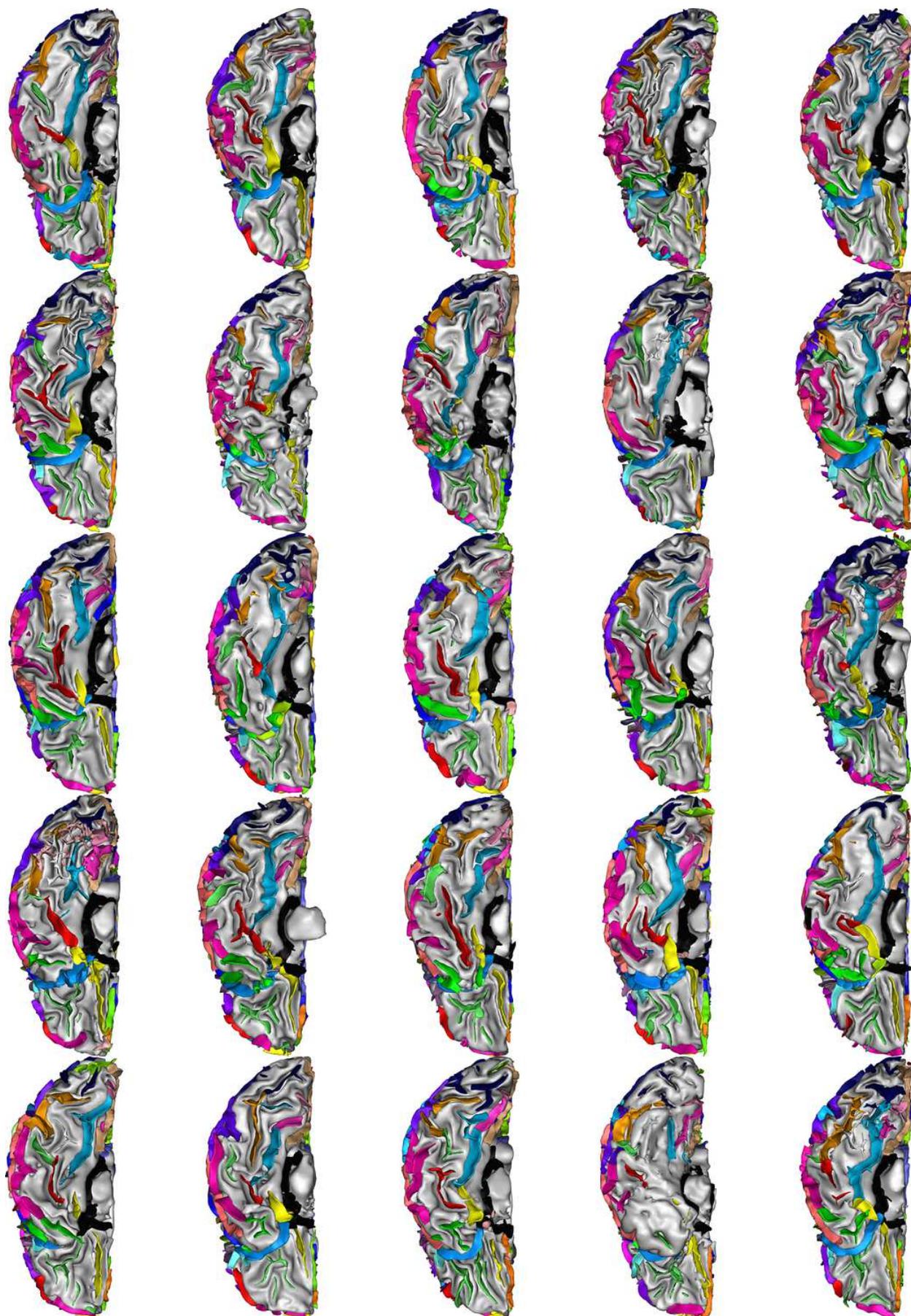


FIG. A.7 – Hémisphères gauches - Vue du dessous



FIG. A.8 – Hémisphères droits - Vue du dessous

Méthodes diverses

Cette annexe regroupe diverses méthodes utilisées dans le cœur du manuscrit en les détaillant et les replaçant si possible dans leur contexte d'utilisation. On y trouvera : la description de l'algorithme ICP, une méthode de clustering Riemannienne de données directionnelles ou axiales, des précisions sur un ensemble de distributions de probabilités : notamment des estimateurs de leurs paramètres (par maximum de vraisemblance ou méthode des moments). Et pour finir, quelques idées de modélisations alternatives pour remplacer les SPAMS.

B.1 Iterative Closest Point (ICP)

La méthode Iterative Closest Point (ICP [Besl 1992]) trouve sa place ici pour plusieurs raisons. La normalisation de données tient une place importante dans cette thèse. La méthode ICP fait partie des méthodes de recalage les plus simples, elle a donc participé à l'initiation de la prise en compte de la normalisation au sein du processus de labélisation ou lors de l'estimation des modèles de variabilité sulcale. Dans le cadre de la normalisation, nous avons aussi évoqué des méthodes dérivées du Procrustes [Gower 1975] qui possèdent certaines connexions avec la méthode ICP (l'étape d'estimation de la transformation). Dernier point, cette méthode a trouvé son utilité et a prouvé son efficacité par sa rapidité et sa capacité de détection de doublons au sein d'une base de données de sujets (voir A), c'est à dire plusieurs acquisitions d'un même sujet. La distance moyenne entre chaque paire de sujets est évaluée au sens de l'ICP. En seuillant assez bas sur cette valeur, on met en valeur les données correspondant à un même sujet. Dans ce cadre, la méthode ICP trouve son utilité puisque deux acquisitions d'un même sujet (même séparées de plusieurs années d'intervalles) donnent des données relativement proches. Dans un cadre plus général d'appariement de plus de deux objets, cette méthode nécessite un objet destination commun souvent représenté par un objet moyen courant mis à jour à chaque itération de la méthode. Il est très difficile de définir la moyenne de structures sulcales, à moins d'avoir une représentation paramétrique adaptée ; nous avons donc rejeté cette méthode pour ce type d'utilisation.

L'idée de cette méthode est extrêmement simple, c'est la raison pour laquelle il existe de très nombreuses variantes, chacune s'ajustant à des besoins spécifiques en raffinant telle ou telle étape. La méthode de base a été introduite par [Besl 1992] et permet l'alignement rigide entre 2 structures : X (l'objet source) que l'on déplacera au cours du processus et Y (l'objet destination) qui restera statique. Chaque objet est représenté par une matrice où chaque vecteur ligne désigne un point 3D caractérisant l'objet. Le nombre de points entre les 2 formes peut être différent. En interne, l'algorithme rendra ce nombre égale afin de faciliter l'estimation des paramètres de recalage. Pour y parvenir, des points sont mis en association à chaque itération de la méthode (voir étape 2.). À l'étape n , ces points constituent la matrice X_n , un sous-ensemble de la matrice X (avec éventuellement des points qui se répètent), de même taille que la matrice Y . On considérera alors que 2 vecteurs lignes possédant le même numéro de ligne dans ces matrices sont des points à mettre en correspondance. Si les objets possèdent énormément de points, il peut être judicieux de les sous-échantillonner afin de gagner en temps de calcul (temps directement relié au carré du nombre de points) à condition d'en garder suffisamment pour conserver leurs traits caractéristiques. L'algorithme consiste à optimiser

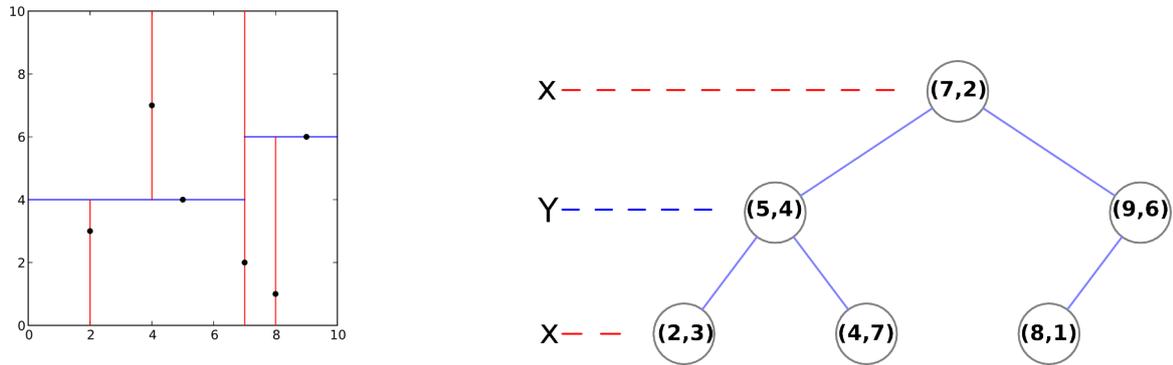


FIG. B.1 – KD-tree construit à partir de l'ensemble des points noirs selon la règle de la médiane (voir texte). Chaque dimension est représentée par une couleur donnée (rouge et bleu). Gauche : vision géométrique du découpage des données. Droite : vision hiérarchique de ce même découpage. (source des images : *wikipedia*).

une matrice de rotation R et un vecteur de translation t afin d'aligner au mieux les 2 formes. Il se présente comme suit :

1. **Initialisation** : $R = I$ et $t = \vec{0}$. On considère Y un objet centré en $\vec{0}$. On pose le numéro d'itération initial : $n = 0$.
2. **Association** : Chaque point de Y est associé à un point de X . Le critère le plus classique est de se baser sur le plus proche voisin au sens de la distance euclidienne. La liste des points de X sélectionnés par ce processus forme la matrice X_n dont les vecteurs lignes sont notés x_i .
3. **Estimation** : On cherche la transformation (R, t, s) (R : matrice de rotation, t : vecteur de translation, s facteur d'échelle) optimisant un critère évaluant la qualité de l'alignement des deux objets. Souvent, on utilise à cet effet la minimisation de la distance quadratique entre les points associés : $\sum_i \|sRx_i + t - y_i\|^2$. Ce qui peut se réécrire sous forme matricielle : $\|sRX_n + T - Y\|^2$ avec $T = \mathbb{1}t^t$ la matrice répétant le vecteur ligne t^t autant de fois qu'il y a de points dans Y . Ce critère moyenné par le nombre de points définit une distance entre les 2 objets considérés qui permet de mesurer la qualité du recalage.
4. **Itérer** : $n \leftarrow n + 1$. Retour au 2., tant qu'un critère d'arrêt (basé sur l'évolution de la distance de l'ICP d'une itération à l'autre) n'est pas vérifié.

Nous allons maintenant détailler l'étape d'association des points ainsi que celle d'estimation des transformations.

B.1.1 Association

La phase d'appariement des deux structures est basée dans notre cas sur la recherche pour chaque point du premier objet de son plus proche voisin dans le deuxième objet. Afin d'accélérer cette recherche nous avons utilisé un algorithme de recherche standard basé sur un KD-tree. Le principe général de cette approche est de structurer l'ensemble des données de façon hiérarchique (au sein d'un arbre) afin de diriger la recherche du plus proche voisin de manière très efficace (avec une complexité de l'ordre de $\mathcal{O}(\log(n))$ où n est le nombre de points de la structure cible). Il existe de multiples façons de construire un tel arbre à partir d'un ensemble de données (dans notre cas, l'intégralité des points du maillage cible considéré). L'approche standard (qui reste basique mais ne

fait pas trop d'hypothèses sur la répartition des données) consiste à itérer le procédé suivant (voir la figure B.1 pour un support visuel) :

- Choisir une dimension de séparation (l'approche classique est d'itérer sur les dimensions entre chaque niveau du *KD-tree* : voir figure B.1).
- Projeter les données le long de cette dimension.
- Sélectionner le point médian après projection.
- Séparer l'espace courant en deux demi-espaces selon le plan orthogonal à l'axe choisi et passant par le point médian sélectionné.
- Recommencer sur chacun de ces demi-espaces en se limitant aux points de la zone correspondante.

Cet algorithme basique, mais déjà efficace, peut connaître des raffinements selon les besoins (recherche de plus proches voisins, intersection d'une droite avec des objets, détection de collisions, ...). Dans l'approche décrite ici, les plans de découpe correspondent à des plans orthogonaux aux axes canoniques. Les tests de bord permettant de guider la recherche pour savoir dans quelle boîte se trouve le point d'intérêt sont alors extrêmement simples et rapides. Si on enlève cette contrainte, cela permet d'avoir des plans de séparation mieux adaptés aux données contre des tests plus coûteux pour l'exploration du *KD-tree*. Il y a donc des compromis à faire, dépendant de la répartition des données afin de profiter au mieux d'une telle structure.

B.1.2 Estimation

Les paramètres de recalage sont estimés en cherchant à minimiser l'erreur quadratique entre les points associés. Chaque point du premier objet est relié à un unique point du deuxième objet. Le problème que l'on cherche à résoudre se formule sous la forme de la minimisation de l'expression suivante :

$$\begin{aligned}
 E &= \sum_{i=1}^n \|sRx_i + t - y_i\|^2 \\
 \text{On pose : } &\begin{cases} \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \\ \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \end{cases} \\
 &= \sum_{i=1}^n \left\| \underbrace{sR(x_i - \bar{x})}_{x'_i} - \underbrace{(y_i - \bar{y})}_{y'_i} + \underbrace{(sR\bar{x} - \bar{y} + t)}_{t'} \right\|^2 \\
 &= \sum_{i=1}^n \|(sRx'_i - y'_i) + t'\|^2 \\
 &= \sum_{i=1}^n \|sRx'_i - y'_i\|^2 + \sum_{i=1}^n \|t'\|^2 + 2 \underbrace{\sum_{i=1}^n Rx'_i - y'_i \cdot t'}_{=0} \\
 &= \sum_{i=1}^n \|sRx'_i - y'_i\|^2 + n \|t'\|^2
 \end{aligned} \tag{B.1}$$

Cette réécriture permet d'optimiser de façon séparée la composante de rotation et de translation, à condition que le facteur d'échelle soit connu. L'annulation de la dérivée partielle en t donne : $t = \bar{y} - sR\bar{x}$. Ainsi, le deuxième terme de l'énergie est nul :

$$E = \sum_{i=1}^n \|sRx'_i - y'_i\|^2 \tag{B.2}$$

Le facteur d'échelle optimal peut être déterminé par simple annulation de la dérivée de l'énergie :

$$\begin{aligned}\frac{\partial E}{\partial s} &= \sum_{i=1}^n \frac{\partial \|sRx'_i\|^2}{\partial s} - 2 \frac{\partial (sRx'_i) \cdot (y'_i)}{\partial s} \\ &= 2s \sum_{i=1}^n \|Rx'_i\|^2 - 2 \sum_{i=1}^n (Rx'_i) \cdot y'_i = 0\end{aligned}\quad (\text{B.3})$$

donc finalement : $s = \frac{\sum_{i=1}^n (Rx'_i) \cdot y'_i}{\sum_{i=1}^n \|x'_i\|^2}$ puisque $\|Rx'_i\|^2 = \|x'_i\|^2$ (invariance par rotation). Il serait souhaitable que les deux objets jouent un rôle symétrique sur le résultat de la méthode. Afin que le facteur d'échelle ne dépende pas de l'ordre arbitraire entre X et Y , il faudrait donc que le problème E soit équivalent au problème $E2$ suivant :

$$E_2 = \sum_{i=1}^n \left\| \frac{1}{s} R^t y_i - \frac{1}{s} R^t t - x_i \right\|^2 = \sum_{i=1}^n \left\| \frac{1}{s} R^t y'_i - x'_i \right\|^2 \quad (\text{B.4})$$

dont le facteur d'échelle s optimal vaut : $s = \frac{\sum_{i=1}^n \|y'_i\|^2}{\sum_{i=1}^n (R^t y'_i) \cdot x'_i}$, ce qui viole l'hypothèse de symétrie du facteur d'échelle entre les deux sens possibles pour réaliser le recalage. Une solution simple à ce problème, proposée par [Horn 1987], est de combiner les 2 sens (equations B.2 et B.4) au sein d'une même expression afin d'avoir une expression symétrique :

$$\begin{aligned}E_3 &= \sum_{i=1}^n \left\| \sqrt{s} Rx'_i - \frac{1}{\sqrt{s}} y'_i \right\|^2 \\ &= \sum_{i=1}^n s \|Rx'_i\|^2 - 2(Rx'_i) \cdot y'_i + \frac{1}{s} \|y'_i\|^2\end{aligned}\quad (\text{B.5})$$

ce qui donne finalement : $s = \frac{\sum_{i=1}^n \|y'_i\|^2}{\sum_{i=1}^n \|x'_i\|^2}$, expression qui ne dépend que des données (contrairement aux précédentes qui dépendaient également du paramètre R).

Nous pouvons maintenant nous consacrer à l'optimisation du paramètre de rotation R . En se basant sur la représentation sous forme de quaternion de R (une généralisation de l'espace des nombres complexes à un espace vectoriel de dimension 4, où les quaternions correspondant à des rotations appartiennent à l'hypersphère unitaire), l'énergie à optimiser s'écrit alors (voir [Horn 1987] pour les détails) :

$$\begin{cases} q^* = \arg \min_q q^t A q \\ \text{sous contrainte : } \|q\|^2 = 1 \end{cases} \quad (\text{B.6})$$

où A est défini ci-dessous :

$$A = \begin{pmatrix} S_{xx} + S_{yy} + S_{zz} & D_{yz} & D_{zx} & D_{xy} \\ D_{yz} & S_{xx} - S_{yy} - S_{zz} & A_{xy} & A_{zx} \\ D_{zx} & A_{xy} & -S_{xx} + S_{yy} - S_{zz} & A_{yz} \\ D_{xy} & A_{zx} & A_{yz} & -S_{xx} - S_{yy} + S_{zz} \end{pmatrix} \quad (\text{B.7})$$

avec :

$$S = X^t Y = \begin{pmatrix} S_{xx} & S_{xy} & S_{xz} \\ S_{yx} & S_{yy} & S_{yz} \\ S_{zx} & S_{zy} & S_{zz} \end{pmatrix} \quad \begin{array}{l} D_{yz} = S_{yz} - S_{zy} \\ D_{zx} = S_{zx} - S_{xz} \\ D_{xy} = S_{xy} - S_{yx} \end{array} \quad \left| \quad \begin{array}{l} A_{xy} = S_{xy} + S_{yx} \\ A_{zx} = S_{zx} + S_{xz} \\ A_{yz} = S_{yz} + S_{zy} \end{array} \right.$$

Il s'agit alors d'un problème de minimisation classique, vérifiant $Aq = \lambda q$ avec λ un Lagrangien permettant d'introduire la contrainte dans l'énergie considérée tel que le problème à minimiser

devienne : $q^t A q + \lambda(1 - \|q\|^2)$. Ainsi, q est le vecteur propre de A associé à la plus grande valeur propre.

Finalement, la matrice de rotation R se calcule de la façon suivante :

$$Q = q^t q$$

$$R = \begin{pmatrix} Q_{00} + Q_{11} - Q_{22} - Q_{33} & 2(Q_{12} - Q_{03}) & 2(Q_{13} + Q_{02}) \\ 2(Q_{12} + Q_{03}) & Q_{00} - Q_{11} + Q_{22} - Q_{33} & 2(Q_{23} - Q_{01}) \\ 2(Q_{13} - Q_{02}) & 2(Q_{23} + Q_{01}) & Q_{00} - Q_{11} - Q_{22} + Q_{33} \end{pmatrix}$$

B.2 *k-means* Riemannien de données directionnelles et axiales

La méthode du *k-means* est une méthode très populaire pour réaliser le *clustering* (ou parcellisation) de données vectorielles ou la quantification vectorielle (*vector quantization* : sous-échantillonnage de données), à la fois par son côté intuitif, sa rapidité ou son efficacité dans bien des situations. Elle offre aussi une visualisation simplifiée, résumant de grande quantité de données. Toutefois, cette approche reste limitée pour des données non-vectorielles. Nous avons exploré de telles données au cours de cette thèse, notamment les données directionnelles (vecteurs de norme 1) et axiales (vecteurs de norme 1 où le sens du vecteur n'a pas de signification). Un *k-means* classique sur de telles données ne prendrait pas en compte la courbure de la sphere sur laquelle vivent ces données, ce qui conduirait inévitablement à des erreurs d'appréciation. C'est pourquoi, nous avons souhaité généraliser le principe de cette méthode à ce type de données. Tout d'abord nous allons présenter la méthode classique du *k-means* avant d'en proposer une généralisation.

Dans notre cas, cette version particulière du *k-means* permet de réaliser une quantification vectorielle de nombreuses données directionnelles ou axiales afin de vérifier visuellement la pertinence du choix des distributions pour modéliser leur variabilité en utilisant les *clusters* ainsi défini comme un estimateur de densité empirique.

B.2.1 *k-means* classique

La méthode du *k-means* [MacQueen 1967, Forgy 1965] est une approche non-supervisée qui vise à parcelliser un ensemble de données $\{x; x \in \mathcal{X}\}$ en k *clusters* où l'on notera \mathcal{C} l'ensemble des *clusters* et c désignera un de ces *clusters*. Chaque *cluster* est représenté par un centre $\bar{x}_c = \frac{1}{|c|} \sum_{x \in c} x$ avec $|c|$ le nombre de points du *cluster* (qui peut varier au cours de l'algorithme). On considère qu'une bonne parcellisation se résume à des parcelles concentrées chacune autour de leur point central respectif \bar{x}_c . Ce problème se formule alors sous la forme de la minimisation de l'énergie suivante :

$$E = \frac{1}{2} \sum_{c \in \mathcal{C}} \sum_{x \in c} \|x - \bar{x}_c\|^2 \quad (\text{B.8})$$

L'optimisation de cette énergie peut être réalisée par une méthode de type descente de gradient [Bengio 1995] qui permet la mise à jour des centres $\bar{x} = \{\bar{x}_c; c \in \mathcal{C}\}$ (simultanée ou successive) :

$$\bar{x}^{(n+1)} \leftarrow \bar{x}^{(n)} - \varepsilon^{(n)} \nabla^{(n)} E \quad \text{avec } \varepsilon^{(n)} \text{ un petit nombre réel.} \quad (\text{B.9})$$

avec $\nabla^{(n)} E = \left(\frac{\partial E}{\partial \bar{x}_c^{(n)}}, c \in \mathcal{C} \right)^t$, où $\frac{\partial E}{\partial \bar{x}_c^{(n)}} = |c^{(n)}| \left(\bar{x}_c^{(n)} - \frac{1}{|c^{(n)}|} \sum_{x \in c^{(n)}} x \right) = |c^{(n)}| \left(\bar{x}_c^{(n)} - \mathbb{E}(x|x \in c^{(n)}) \right)$. De manière plus générale la méthode de Newton-Raphson peut être utilisée. Elle permet d'approcher localement l'énergie par une fonction quadratique en se basant sur la matrice Hessienne (matrice des dérivées secondes) de l'énergie, ce qui permet d'orienter plus efficacement la recherche du minimum :

$$\bar{x}^{(n+1)} \leftarrow \bar{x}^{(n)} - H^{-1} \nabla^{(n)} E \quad \text{avec } H \text{ la matrice Hessienne.} \quad (\text{B.10})$$

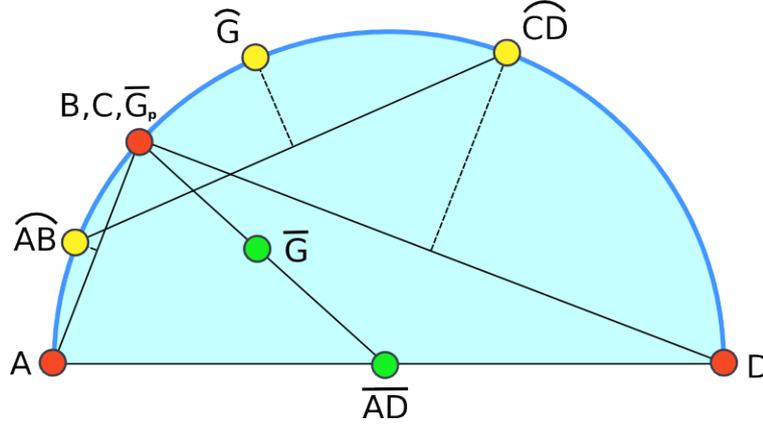


FIG. B.2 – En rouge : A, B, C, D : 4 points (B et C sont confondus pour simplifier l'exemple) du cercle dont on cherche à calculer la moyenne sur le cercle, en vert : les moyennes euclidiennes des points indiqués, en jaune : les moyennes intrinsèques, \bar{G} est la moyenne euclidienne des 4 points, \bar{G}_p correspond à la moyenne projetée sur le cercle, alors que \widehat{G} correspond à la véritable moyenne intrinsèque.

où les composantes de H se calculent ainsi :

$$\frac{\partial^2 E}{\partial \bar{x}_{c_i}^{(n)} \partial \bar{x}_{c_j}^{(n)}} = |c_j^{(n)}| \frac{\partial \left(\bar{x}_{c_i}^{(n)} - \mathbb{E}(x | x \in c_j^{(n)}) \right)}{\partial \bar{x}_{c_i}^{(n)}} = \begin{cases} 0 & \text{si } c_i^{(n)} \neq c_j^{(n)} \\ |c_i^{(n)}| & \text{sinon.} \end{cases} \quad (\text{B.11})$$

La matrice Hessienne est donc diagonale, pour un *cluster* c donné, cette approche se traduit donc de manière pratique par la règle simplifiée suivante :

$$\bar{x}^{(n+1)} \leftarrow \mathbb{E}(x | x \in c^{(n)}) \quad (\text{B.12})$$

qui correspond à la méthode standard du *k-means*. Que l'on peut alors exprimer de manière pratique par le schéma suivant :

1. **Initialisation** : les centres $\bar{x}_c^{(0)}$ peuvent être tirés au hasard parmi les données étudiées, ou basés sur une estimation de la matrice de variance des *clusters* [Art 1982, Tarsitano 2003]. La sensibilité de l'algorithme à l'initialisation est d'autant plus grande que la dimensionnalité des données est grande [Bradley 1998] (comme toute méthode de descente de gradient qui converge vers un minimum local de l'énergie). Il est donc courant de relancer plusieurs fois l'algorithme avec des initialisations différentes et de conserver finalement la meilleure solution.
2. **Association** : chaque point est associé au *cluster* le plus proche selon la métrique choisie (distance euclidienne dans l'approche classique).
3. **Mise à jour** : le centre \bar{x}_c de chaque *cluster* c est réévalué et correspond à la moyenne des points appartenant à ce *cluster* (c'est à dire la règle énoncée dans l'équation B.12).
4. **Itérer** : retour à l'étape 2. jusqu'à stabilisation des *clusters*.

B.2.2 Approche Riemannienne

Nous avons évoqué que certaines données vivant dans des espaces courbés n'étaient pas traitées de manière optimale par l'approche classique du *k-means* où la métrique considérée est euclidienne,

que ce soit pour la mise à jour des centres des *clusters* ou pour la phase d'association des points aux *clusters* basée encore une fois sur la distance euclidienne. En effet, si on considère des données directionnelles $x \in \mathcal{X}$ (vecteurs de norme 1 : $\|x\|^2 = 1$), la moyenne classique de telles données $\mathbb{E}(x) = \frac{1}{n} \sum x$ à toutes les chances de ne pas être de norme 1. De plus, la projection de cette moyenne sur l'espace des données de norme 1 (obtenue en normalisant cette quantité) qui vaut $\frac{\mathbb{E}(x)}{\|\mathbb{E}(x)\|}$ ne correspond pas forcément à la moyenne idéale ou naturelle de telles données (voir figure B.2).

C'est ici, que l'approche Riemannienne [Klingenberg 1982, Do Carmo 1992] rentre en jeu. Elle dérive de la géométrie différentielle qui permet de définir une métrique (produit scalaire) intrinsèque, adapté à la courbure de l'espace considéré. La distance entre deux points d'une variété correspond alors à la longueur de la géodésique (un chemin continu inclus dans l'espace Riemannien considéré) la plus courte. L'approche Riemannienne consiste à se ramener localement au cas euclidien classique en approchant l'espace courbe par un espace tangent (voir figure 7.4). Dans le cas général, l'approximation n'est vraie que localement, des méthodes itératives sont alors nécessaires afin de progresser par petits pas (dans la mesure de la validité de l'approximation dans l'espace tangent) le long de la variété considérée [Edelman 1998, Pennec 1999, Pennec 2006].

Dans le cas plutôt simple de données vivant sur la sphère, la distance naturelle entre 2 points correspond à la distance géodésique de l'arc reliant ces 2 points. Pour deux vecteurs de norme 1 : x_i et x_j , cette distance est égale à l'angle entre ces deux vecteurs (puisque le périmètre d'un cercle de rayon 1 vaut 2π). Finalement, cette distance s'exprime de la façon suivante :

$$\text{dist}_{\text{dir}}(x_i, x_j) = \arccos(x_i^t x_j) \tag{B.13}$$

Dans le cas de données axiales, le sens des vecteurs étudiés n'a pas d'importance. Ainsi, dans le plan constitué par les 2 vecteurs considérés les vecteurs sont définis à une rotation π près. La distance géodésique correspond alors au minimum entre $\text{dist}_{\text{dir}}(x_i, x_j)$ et $\text{dist}_{\text{dir}}(-x_i, x_j)$. Ce qui de manière pratique se traduit de la façon suivante :

$$\text{dist}_{\text{axe}}(x_i, x_j) = \begin{cases} \theta & \text{si } \theta \leq \frac{\pi}{2} \\ \pi - \theta & \text{sinon} \end{cases} \text{ avec } \theta = \arccos(x_i^t x_j) \tag{B.14}$$

Ces formulations peuvent être utilisées en lieu et place de la distance euclidienne dans l'algorithme du *k-means* pendant la phase d'association des points au *cluster* le plus proche. Concernant la phase de mise à jour des centres des *clusters* (étape 3, équation B.12), elle revient à calculer la moyenne intrinsèque (encore appelée moyenne de Fréchet [Fréchet 1948]), le long de la variété considérée. Afin de mieux comprendre cette notion et trouver un moyen naturel de l'estimer, il faut revenir à la définition de la moyenne dans le cas euclidien qui peut s'exprimer de différentes manières :

$$\mathbb{E}(x) = \frac{1}{n} \sum_{i=1}^n x_i = \arg \min_{\mu} \sum_{i=1}^n \|x_i - \mu\|^2 = \arg \min_{\mu} \sum_{i=1}^n \text{dist}(x_i, \mu) \tag{B.15}$$

En effet, on peut vérifier (puisque le point μ^* minimisant cette expression vérifie : $\frac{\partial}{\partial \mu} \sum_{i=1}^n \|x_i - \mu\|^2 \Big|_{\mu=\mu^*} = n\mu - \sum_{i=1}^n x_i = 0$) que le point qui minimise la somme des distances quadratiques à un ensemble de points, correspond à leur moyenne. On définit alors la moyenne intrinsèque au sens Riemannien selon ce principe en remplaçant la distance euclidienne par les distances définies précédemment (équation B.13 et B.14). Dans le cas général, il n'existe pas de formulation explicite de la moyenne, il faut alors avoir recours à des méthodes itératives [Pennec 2006, Fletcher 2007] afin de l'estimer.

L'algorithme de calcul de la moyenne intrinsèque part d'une initialisation $\mu^{(0)}$ bien choisie. Pour des raisons de performance, un bon choix est de partir du calcul d'une moyenne euclidienne classique

projetée sur la surface de la sphère. L'approche Riemannienne permet alors de raffiner le calcul et de trouver la véritable moyenne par le procédé de mise à jour suivant (correspondant à une descente de gradient le long de la variété Riemannienne de l'énergie B.15 avec les distances intrinsèques précisées plus tôt) :

$$\left\{ \begin{array}{ll} \bar{T} = \frac{1}{n} \sum_{i=1}^n \text{Log}_{\mu^{(n)}}(x_i) & \text{moyenne dans l'espace tangent } \mathcal{T}. \\ \mu^{(n+1)} \leftarrow \text{Exp}_{\mu^{(n)}}(\tau \bar{T}) & \text{avec } 0 < \tau \ll 1, \text{ le pas de la descente de gradient.} \end{array} \right. \quad (\text{B.16})$$

où $\text{Exp}_{\mu^{(n)}}(\cdot)$ est la carte exponentielle ($\mathcal{T} \rightarrow \zeta$) et $\text{Log}_{\mu^{(n)}}(\cdot)$ la carte logarithmique ($\zeta \rightarrow \mathcal{T}$) permettant de naviguer entre l'espace tangent \mathcal{T} et l'espace courbe ζ . Attention à ne pas confondre ces notations avec $\exp(\cdot)$ et $\log(\cdot)$ les fonctions logarithmique et exponentielle généralisées aux matrices.

Cet algorithme converge vers la moyenne intrinsèque plutôt rapidement puisque l'énergie considérée est convexe dans le cas des données directionnelles (pour des données axiales, l'énergie optimisée est convexe sur les 2 demi-sphères centrées respectivement sur μ et $-\mu$ les 2 expressions équivalentes de la solution recherchée). Un bon critère d'arrêt consiste à regarder si la norme $\|\bar{T}\|$ devient plus petite qu'un seuil donné correspondant à la précision souhaitée. En effet, \bar{T} correspond au déplacement vectoriel dans l'espace tangent, qui tend vers 0 avec l'augmentation des itérations n .

Nous proposons ici une approche non-standard, basée sur une équivalence entre les données directionnelles (ou axiales dont on précisera les différences plus loin) et un sous-ensemble des matrices de rotation afin de réutiliser les cartes exponentielles et logarithmiques de cet espace que nous avons déjà développé par ailleurs. Chaque donnée étudiée x est alors représentée par la rotation $R_{u_0 \rightarrow x}^w$ qui déplace un vecteur arbitraire u_0 en x par une rotation autour de l'axe w . Puisque nous cherchons à représenter un vecteur unitaire par une matrice de rotation, tous les w appartenant au disque défini par les vecteurs $u_0 \times x$ et $\frac{u_0 \times x}{2}$ forment une classe d'équivalence représentant le vecteur x . Au sein de cette classe d'équivalence tous les $R_{u_0 \rightarrow x}^w$ jouent le même rôle, on peut alors choisir un vecteur w pratique pour nos calculs. Notre choix s'est porté sur $w = u_0 \times x$, l'angle de rotation associé est alors : $\widehat{u_0 \times x} = \arccos(x^t u_0)$. Pour simplifier les notations on notera alors : $R_{u_0 \rightarrow x} = R_{u_0 \rightarrow x}^{u_0 \times x}$.

En notant $M = R_{u_0 \rightarrow \mu}$, les cartes exponentielle et logarithmique d'une matrice de rotation R_{x, u_0} s'expriment ainsi :

$$\left\{ \begin{array}{l} \text{Log}_{M^{(n)}}(R_{x, u_0}) = \log \left(R_{u_0 \rightarrow \mu^{(n)}}^t R_{u_0 \rightarrow x} \right) = \log \left(R_{\mu^{(n)} \rightarrow x}^{w^{(n)}} \right) \\ \quad \text{où } w^{(n)} = u_0 \times \left(\frac{\widehat{u_0 \times x}}{\sin(\widehat{u_0 \times x})} x - \frac{\widehat{u_0 \mu}}{\sin(\widehat{u_0 \mu})} \mu \right) \\ \quad \text{En utilisant la formule de Rodrigues (voir section B.56) :} \\ = J(w^{(n)}) \quad \text{avec } J(w^{(n)}) = \begin{pmatrix} 0 & -w_z & w_y \\ w_z & 0 & -w_x \\ -w_y & w_x & 0 \end{pmatrix} \\ \quad \text{où } \left\{ \begin{array}{l} w^{(n)} \text{ correspond à l'axe de rotation} \\ \text{et } \|w^{(n)}\| \text{ à l'angle de rotation} \end{array} \right. \\ \text{Exp}_{M^{(n)}}(J(w^{(n)})) = M^{(n)} \exp(J(w^{(n)})) = M^{(n)} R_{\mu^{(n)} \rightarrow \mu^{(n+1)}}^{w^{(n)}} \end{array} \right. \quad (\text{B.17})$$

Si on revient à la moyenne exprimée sous la forme du vecteur unitaire $\mu^{(n)}$, la règle de mise à jour se résume à :

$$\mu^{(n+1)} \leftarrow R_{\mu^{(n)} \rightarrow \mu^{(n+1)}}^{w^{(n)}} \mu^{(n)} \quad (\text{B.18})$$

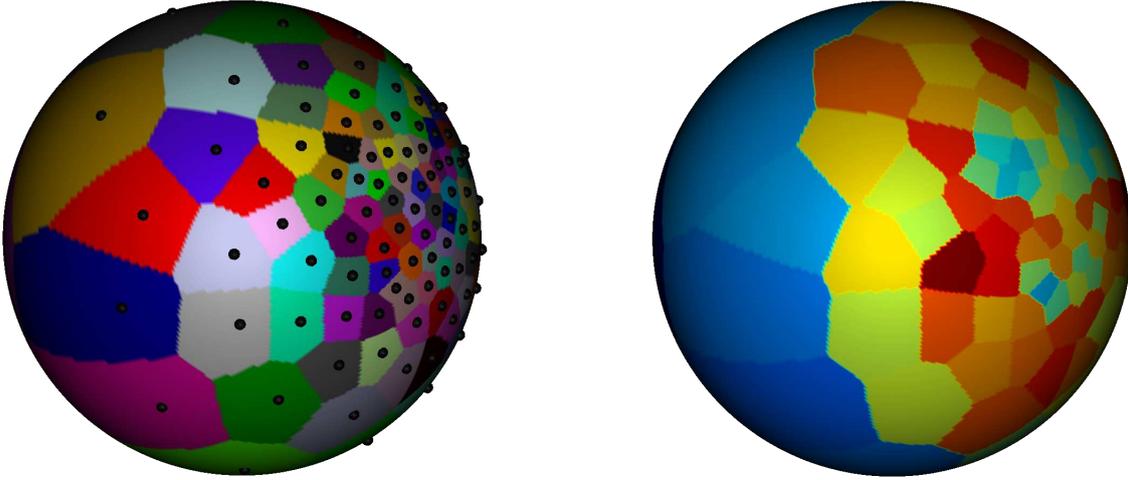


FIG. B.3 – k -means sur la sphère (données directionnelles normées). Gauche : *clusters* et leurs centres. Droite : la densité de points correspondant à chaque *cluster* est représenté.

Compte tenu de la classe d'équivalence mentionnée plus tôt, on choisit d'utiliser $R_{\mu^{(n)} \rightarrow x}$ à la place de $R_{\mu^{(n)} \rightarrow x}^{w^{(n)}}$, afin de faciliter les calculs, ce qui donne alors :

$$w^{(n)} = \frac{1}{n} \sum_{i=1}^n \log(R_{\mu \rightarrow x_i}^{\mu \times x_i}) = \frac{1}{n} \mu \times \left(\sum_{i=1}^n \frac{\theta_i}{\sin(\theta_i)} x_i \right) \quad (\text{B.19})$$

avec $\theta_i = \widehat{\mu x_i}$, puisque $\|\mu \times x_i\| = \sin(\theta_i)$ et les rotations $R_{\mu \rightarrow x_i}^{\mu \times x_i}$ correspondent respectivement à une rotation d'angle θ_i autour de l'axe $\mu \times x_i$.

Pour le cas de données axiales, les mêmes calculs peuvent être dérivés. Il suffit alors d'appliquer la même règle de définition des angles θ_i que dans le calcul de la distance dist_{axe} (équation B.14).

Pour information, la méthode développée ici permet de *clusteriser* 100000 données en 100 *clusters* en environ 10 secondes. Ces chiffres correspondent au nombre de vecteurs que nous avons rencontré dans la modélisation de la variabilité de la direction de grands sillons comme le sillon central (environ 1600 directions locales (une par voxel) \times 62 sujets). Une centaine de *clusters* permet de représenter visuellement (voir figure B.3) la densité des données et juger de la pertinence du choix de modèles Von Mises Fisher, Kent, Watson, Bingham vis-à-vis des données étudiées.

B.3 Quelques distributions de probabilités

Cette section détaille quelques distributions de probabilités d'intérêt rencontrées dans le coeur de l'ouvrage, divisées en 2 catégories principales : 1) les distributions directionnelles ou circulaires, 2) les distributions sur un simplexe. Leur formulation est accompagnée d'estimateurs de leurs paramètres par maximum de vraisemblance ou par la méthode des moments. On s'intéressera notamment aux distributions suivantes : Von Mises, Von Mises Fisher, Kent, Bingham, Dirichlet, Dirichlet généralisée. Pour chaque distribution, on rappellera brièvement dans quel contexte elle a été rencontrée au cours de cet ouvrage. Cette retranscription a été principalement motivée par la difficulté rencontrée pour trouver ces d'informations dans la littérature et leur offrir une nouvelle visibilité (même si elle reste encore limitée). De plus, la plupart des distributions présentées ci-après n'ont pas d'implémentation standard ou de représentations visuelles adaptées dans le cas de la distribution Matrix Von

Mises. Leur manipulation pratique s'est effectuée en Python dans notre cas. Il serait donc tout à fait envisageable de prévoir leur inclusion dans le module scientifique standard Python : SciPy.

B.3.1 Distributions circulaires, directionnelles, axiales, rotations

Dans cette section, nous avons regroupé dans une même catégorie des distributions sur des données circulaires (Von Mises), directionnelles (Von Mises Fisher, Kent), axiales (Watson, Bingham) et sur des données appartenant à la variété de Stiefel (Matrix Von Mises Fisher, Matrix Bingham) comprenant notamment l'espace des matrices de rotation. Toutes ces données et les distributions correspondantes ont une particularité commune : elle vivent dans des espaces courbés et «circulaires» (rebouclant sur eux mêmes). Par exemple, dans le cas de données angulaires, elles sont définies à 2π près. Le modèle le plus connu et sans doute le plus simple pour estimer des densités de probabilité est le modèle Gaussien. Toute une branche des statistiques (voir [Jupp 1989, Jammalamadaka 2001, Mardia 2000] pour des revues détaillées sur ces thèmes) s'est évertuée à généraliser cette notion aux données circulaires au sens large. Il existe d'autres distributions généralisant celles présentées ici ou proposant des approximations alternatives, nous n'avons cité ici que celles qui ont été les plus explorées dans la littérature et où il y avait suffisamment de résultats pour permettre leur exploitation. La plupart des modèles présentés dans cette section correspondent à la projection d'une distribution Gaussienne classique estimée sur les données brutes vers la variété Riemannienne correspondante. Il s'agit donc d'approximations puisque la Gaussienne n'est pas directement estimée sur la variété Riemannienne. Mais elles ont toutes l'avantage d'avoir des expressions relativement simples (si on omet la constante de normalisation qui peut être difficile à calculer) prenant la forme de modèles exponentiels qui sont connus pour leurs bonnes propriétés.

Les applications potentielles de telles distributions sont très nombreuses, mais elles restent plutôt limitées à ce jour et assez peu présentent dans la littérature neuroscientifique et se cantonne au domaine de l'IRM de diffusion avec la modélisation de la fonction de densité de probabilité de l'orientation des fibres [Seunarine 2007] ou du processus de diffusion [Rathi 2009] afin de réaliser la tractographie des fibres de la matière blanche [Zhang 2009]. On trouve des applications dans les domaines suivants : géologie [Watson 1968], paleomagnétisme [Jupp 1987], vectorcardiographie [Downs 1972, Khatri 1977, Prentice 1986, Wood 1993], astrophysique [Mardia 1975, Jupp 1995], cristallographie [Krieger Lassen 1994], analyse de forme [Mardia 1999] et biomécanique [León 2006, Rivest 2008].

Contexte d'utilisation Dans le cadre de l'étude de la variabilité de la forme des sillons, nous avons défini (que ce soit lors de travaux précédents ou au cours de cette thèse) des descripteurs morphométriques spécifiques caractérisant chaque sillon ou des relations entre paires de sillons. L'idée générale des méthodes de labélisation des sillons présentées tout au long de ce manuscrit consiste en 1) la détermination d'un modèle de la variabilité sulcale appris sur une base de données labélisées manuellement et 2) à exploiter ensuite ces modèles afin d'inférer une labélisation à partir de nouvelles données. Certaines des données susceptibles de décrire l'organisation et la forme des sillons sont justement de type *circulaire*. Afin d'exploiter au mieux ces données, il est nécessaire d'utiliser des distributions dédiées évitant ainsi des biais grossiers et incontrôlés.

La forme de la plupart des sillons est caractérisée notamment par 3 axes principaux : l'orientation du sillon (dans le sens de sa longueur) définie à un signe près, la normale moyenne (tangentielle à l'enveloppe du cerveau) définie à un signe près, et pour finir un vecteur normal à l'enveloppe du cerveau et pointant vers l'extérieur. Les deux axes peuvent ainsi être estimés par des distributions de Bingham alors que la troisième information de nature directionnelle peut se représenter par des distribution de Kent ou de Von Mises Fisher. Seule la normale à l'enveloppe a été étudiée pour l'instant. Afin d'exploiter au mieux l'ensemble de ces 3 axes et tenir compte des contraintes qui les lient (l'ensemble des 3 axes forme une matrice appartenant à la variété de Stiefel), il faut utiliser une distribution dédiée. C'est pourquoi, nous avons développé dans cette section une description

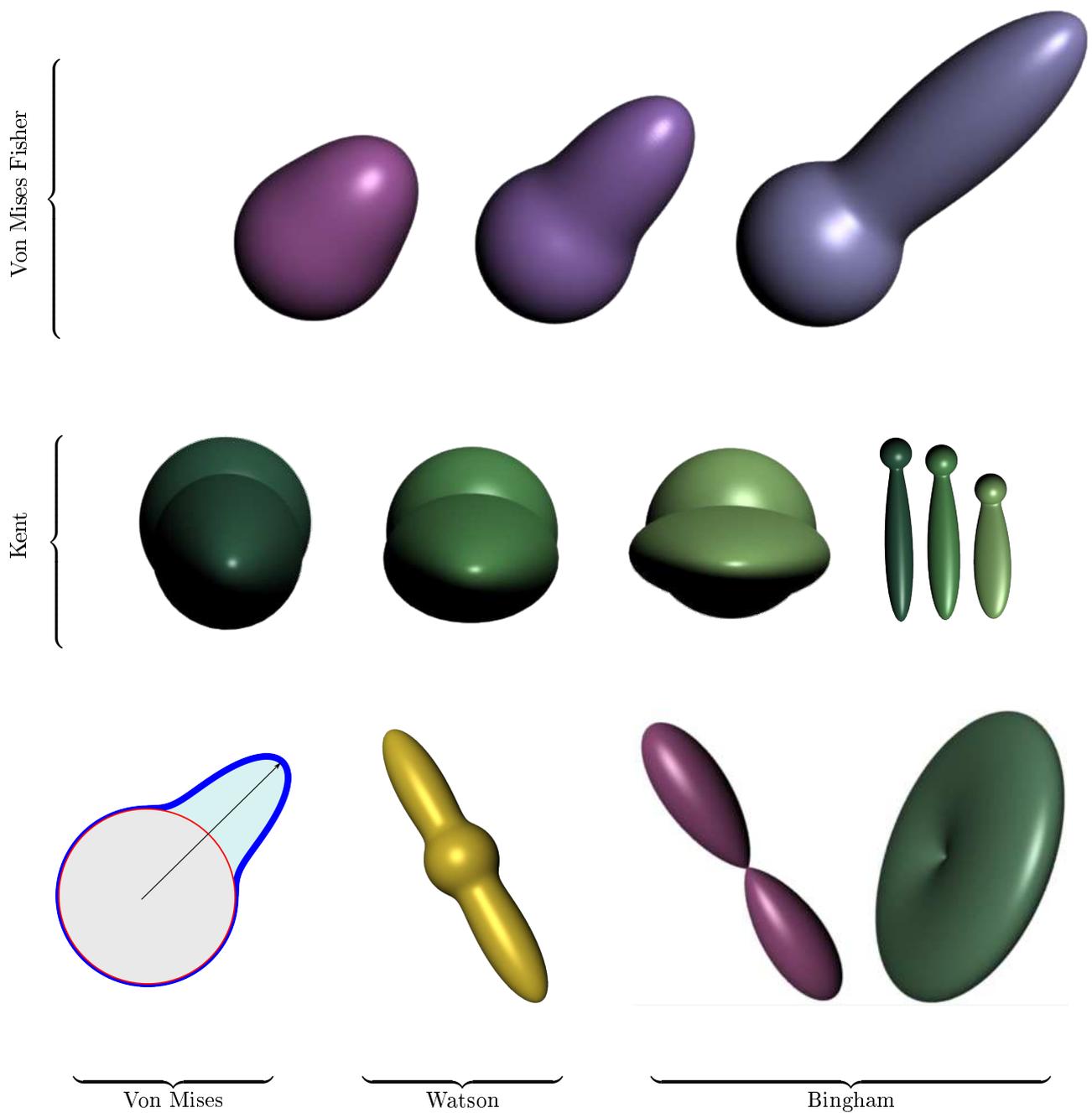


FIG. B.4 – Représentation visuelle 2D et 3D de distributions directionnelles et axiales. Haut et Milieu : distributions sur des directions (vecteurs normés). Haut : Von Mises Fisher en faisant varier de gauche à droite le paramètre κ qui prend les valeurs : 5, 10, 20. Milieu : Distribution de Kent à κ fixé à une valeur de 60 et en faisant varier β de gauche à droite : 0, 10, 20. Bas : de gauche à droite : distribution Von Mises sur des angles, distribution Watson puis Bingham pour traiter des données axiales.

des distributions Matrix Von Mises Fisher et Matrix Bingham. Au départ, cette dernière a d'ailleurs été considérée à tort comme une distribution permettant de modéliser un ensemble d'axes liés par la contrainte de Stiefel. Cette dernière distribution impose en plus des contraintes de symétrie qui sont de trop dans notre étude et nuiraient à la précision de la modélisation. Dans l'état actuel de nos connaissances, une solution serait de considérer un mélange de 2 distributions Matrix Von Mises Fisher de paramètres respectifs $F_1 = PDQ$ et $F_2 = F_1S$ où S est une matrice diagonale comportant des 1 lorsque la dimension considérée est directionnelle (comme notre composante normale à l'enveloppe cérébrale) et des -1 lorsque celle-ci est axiale (comme les 2 autres composantes considérées), on obtiendrait alors la vraisemblance suivante :

$$P(W|F_1F_2S) \propto \frac{1}{2} [\exp(\text{Tr}(F_1)) + \exp(\text{Tr}(F_1S))] \quad (\text{B.20})$$

Ces distributions (Von Mises Fisher, Kent) ont été également envisagées dans le cadre de l'étude de la covariabilité de sillons voisins, notamment pour caractériser la direction (signée) entre 2 sillons. Les log-vraisemblances associées jouent alors le rôle de potentiel dans un champ de Markov.

Finalement, une dernière utilisation est la modélisation d'a priori sur l'espace des matrices de rotation afin de contraindre une méthode de recalage. Dans un premier temps nous avons envisagé une solution basée sur la paramétrisation du vecteur rotation associé à un découplage des composantes de direction et d'angle de rotation. La première, axiale, a été modélisée par une distribution de Bingham, alors que la deuxième, angulaire a été modélisée grâce à une distribution de Von Mises.

B.3.1.1 Von Mises

La distribution de Von Mises [Mardia 2000] (voir figure B.4) permet de modéliser une dispersion de type Gaussien pour des données angulaires (ce qui correspond à une restriction de la distribution Von Mises Fisher à des données de dimension 2).

$$p(\alpha|\kappa\mu) = \frac{\exp(\kappa \cos(\alpha - \mu))}{2\pi I_0(\kappa)} \quad (\text{B.21})$$

où $I_0(\kappa)$ est la fonction de Bessel modifiée d'ordre 0, κ contrôle la concentration autour de la moyenne et μ correspond à l'angle moyen.

À partir d'une base de données d'apprentissage $\mathbb{A} = \{\alpha_i; i \in 1 : n\}$, on peut estimer les paramètres κ et μ par maximum de vraisemblance, on doit donc résoudre :

$$\kappa^*, \mu^* = \arg \max_{\kappa, \mu} \prod_{i=1}^n p(\alpha_i|\kappa\mu) \quad (\text{B.22})$$

La résolution de ce problème est un cas particulier de l'estimation par maximum de vraisemblance de la distribution Von Mises Fisher, dont on pourrait se contenter. Mais il existe toutefois des approximations permettant d'accélérer les estimations dans le cas restreint de Von Mises sur l'espace \mathbb{S}^2 que nous allons présenter.

D'après les calculs formulés plus loin dans le cas général des distributions de Von Mises Fisher, on obtient le vecteur 2D correspondant à l'angle moyen : $\frac{\bar{w}}{\|\bar{w}\|}$, où $\bar{w} = \sum_{i=1}^n w_i$ avec w_i la représentation vectorielle de chaque angle α_i de la base de données considérée telle que : $w_i = \begin{pmatrix} \cos(\alpha_i) \\ \sin(\alpha_i) \end{pmatrix}$, donc :

$$\mu = \frac{\bar{w}}{\|\bar{w}\|} = \arctan \left(\frac{\sum_{i=1}^n \sin(\alpha_i)}{\sum_{i=1}^n \cos(\alpha_i)} \right) \quad (\text{B.23})$$

Concernant le paramètre κ , cela revient à résoudre le problème suivant :

$$A_2(\kappa) = \frac{I_1(\kappa)}{I_0(\kappa)} = \|\bar{w}\| = \bar{w} \cdot \mu = \frac{1}{n} \sum_{i=1}^n \cos(\alpha_i - \mu) = \mathcal{W} \quad (\text{B.24})$$

Des approximations [Hill 1977, Hussin 2008] basées sur des développements de Taylor ont été proposées permettant une estimation rapide et assez précise du paramètre κ :

$$\left\{ \begin{array}{ll} 0 \leq \kappa < 0.53 & \Rightarrow \kappa \approx 2\mathcal{W} + \mathcal{W}^3 + (5\mathcal{W}^5)/6 \\ 0.53 \leq \kappa < 0.85 & \Rightarrow \kappa \approx -0.4 + 1.39\mathcal{W} + \frac{0.43}{1 - \mathcal{W}} \\ \kappa \geq 0.85 & \Rightarrow \kappa \approx \frac{1}{\mathcal{W}^3 - 4\mathcal{W}^2 + 3\mathcal{W}} \end{array} \right. \quad (\text{B.25})$$

Des corrections de biais ont été suggérées [Best 1981] lorsque la taille n de la base de données est réduite. le paramètre κ sans biais est noté κ_b est s'obtient ainsi :

$$\left\{ \begin{array}{ll} \kappa < 2 & \Rightarrow \kappa_b = \kappa - \frac{2}{n\kappa} \\ \kappa \geq 2 & \Rightarrow \kappa_b = \kappa \frac{(n-1)^3}{n(n^2+1)} \end{array} \right. \quad (\text{B.26})$$

B.3.1.2 Von Mises Fisher

La distribution de Von Mises Fisher (appelée aussi tout simplement Fisher ou Langevin, voir [Mardia 2000] pour plus de détails et [Banerjee 2005] pour son utilisation dans un contexte Bayésien) généralise la distribution Von Mises aux données directionnelles w de dimension p supérieure à 3, c'est à dire appartenant à l'hypersphère \mathcal{S}^{p-1} de dimension $p-1$. Sa vraisemblance s'exprime de la façon suivante :

$$p(w|\kappa\mu) = \frac{\exp(\kappa w \cdot \mu)}{C_p(\kappa)} \quad (\text{B.27})$$

avec $\kappa \geq 0$ le facteur de dispersion du modèle, μ la direction moyenne et la constante de normalisation $C_p(\kappa) = \frac{\kappa^{p/2-1}}{(2\pi)^{p/2} I_{p/2-1}(\kappa)}$ où I_k est la fonction de Bessel modifiée de première sorte et d'ordre k .

On va maintenant formuler l'estimation des paramètres κ et μ par maximum de vraisemblance à partir d'une base de données représentée par une matrice $W = [w_1, \dots, w_n]^t$ où chaque ligne représente un vecteur w_i de dimension p et de norme 1. On cherche donc à résoudre :

$$\kappa^*, \mu^* = \arg \max_{\kappa, \mu} \prod_{i=1}^n p(\alpha_i|\kappa\mu) = \arg \max_{\kappa, \mu} n \log(C_p(\kappa)) + \kappa \mu \cdot \bar{w} \quad (\text{B.28})$$

avec $\bar{w} = \frac{1}{n} \sum_{i=1}^n w_i$. On introduit alors la contrainte $\|\mu\|^2 = 1$ grâce au Lagrangien λ , on cherche alors à minimiser le problème modifié suivant :

$$L = n \log(C_p(\kappa)) + \kappa \mu \cdot \bar{w} + \lambda(1 - \|\mu\|^2) \quad (\text{B.29})$$

En dérivant L par rapport à μ , λ et κ , l'annulation des dérivées permet d'obtenir :

$$\left\{ \begin{array}{l} \mu = \frac{n\kappa}{2\lambda} \bar{w} \\ \frac{C'_p(\kappa)}{C_p(\kappa)} = -\mu \cdot \bar{w} \end{array} \right. \quad (\text{B.30})$$

ce qui donne finalement :

$$\left\{ \begin{array}{l} \lambda = \frac{n\kappa}{2} \|\bar{w}\| \quad \text{puisque } \|\mu\|^2 = 1. \\ \mu = \frac{\bar{w}}{\|\bar{w}\|} \quad \text{la moyenne empirique normalisée.} \\ \frac{C'_p(\kappa)}{C_p(\kappa)} = -\|\bar{w}\| \end{array} \right. \quad (\text{B.31})$$

Or la dérivée de la constante de normalisation par rapport à κ s'écrit :

$$\frac{C'_p(\kappa)}{C_p(\kappa)} = C_p(\kappa) \left(\frac{p/2 - 1}{\kappa} - \frac{I'_{p/2-1}(\kappa)}{I_{p/2-1}(\kappa)} \right) \quad (\text{B.32})$$

et I_k suit la règle de récurrence suivante : $\kappa I_{s+1}(\kappa) = \kappa I'_s(\kappa) - s I_s(\kappa)$, donc finalement :

$$-\frac{C'_p(\kappa)}{C_p(\kappa)} = \frac{I_{p/2}(\kappa)}{I_{p/2-1}(\kappa)} = A_p(\kappa) = \|\bar{w}\| \quad (\text{B.33})$$

Or A_p^{-1} n'a pas d'expression analytique ou de bonnes approximations dans le cas général, il faut donc avoir recours à des stratégies numériques. Des approximations asymptotiques ont été proposées [Banerjee 2005] ou d'autres plus précises [Tanabe 2007]. Nous suivons ici une méthode itérative basée sur la méthode de Newton-Raphson suggérée par [Banerjee 2005] en partant de la solution approchée initiale κ_0 :

$$\kappa_0 = \frac{\|\bar{w}\| p - \|\bar{w}\|^3}{1 - \|\bar{w}\|^2} \quad (\text{B.34})$$

La méthode de Newton-Raphson permet de trouver les zéros d'une fonction. Nous cherchons donc à résoudre le problème suivant : $f(\kappa) = A_p(\kappa) - \|\bar{w}\| = 0$. L'algorithme itératif s'exprime comme suit :

$$\kappa_{n+1} = \kappa_n - \frac{f(\kappa_n)}{f'(\kappa_n)} \quad (\text{B.35})$$

$$= \kappa_n - \frac{A_p(\kappa) - \|\bar{w}\|}{A'_p(\kappa)} \quad (\text{B.36})$$

$$(\text{B.37})$$

avec $A'_p(\kappa) = 1 - A_p(\kappa) \left(A_p(\kappa) + \frac{p-1}{\kappa} \right)$.

B.3.1.3 Kent

La distribution de Kent [Kent 1982] (aussi connue sous le nom de distribution Fisher-Bingham FB5) est restreinte aux données directionnelles normées de dimension 3, c'est à dire appartenant à la sphère unitaire $\mathcal{S}^2 \in \mathbb{R}^3$. Dans ce cadre restreint, elle généralise la distribution de Von Mises Fisher en permettant de contrôler pleinement la structure de covariance, elle généralise donc la notion de Gaussienne bivariée à ce type de données. Sa vraisemblance s'exprime comme suit :

$$p(w|\kappa\beta\gamma_1\gamma_2\gamma_3) = \frac{\exp(\kappa w \cdot \gamma_1 + \beta((w \cdot \gamma_2)^2 - (w \cdot \gamma_3)^2))}{\exp(-\kappa)\sqrt{\kappa^2 - 4\beta^2}} \quad (\text{B.38})$$

où γ_1 correspond à la direction moyenne, γ_2 et γ_3 aux axes de plus grande et plus petite dispersion respectivement, le long de la variété considérée. Les vecteurs colonnes $\gamma_1, \gamma_2, \gamma_3$ forment une matrice orthogonale notée Γ . Le paramètre κ contrôle la dispersion de la distribution alors que le paramètre β l'anisotropie de la distribution. Les paramètres vérifient : $\kappa > 0$ et $0 \leq 2\beta < \kappa$.

La méthode du maximum de vraisemblance ne semble pas permettre la formulation explicite des paramètres optimaux. Nous avons alors suivi la méthode des moments suggérée par [Kent 1982] consistant à trouver les paramètres alignant au mieux les moments empiriques avec ceux du modèle. Il indique que les estimations fournies par la méthode se rapprochent de celles que pourraient fournir le maximum de vraisemblance à condition que $2\beta/\kappa$ soit suffisamment petit ou que κ soit suffisamment grand. Ce qui correspond aux cas qui nous intéressent en pratique (les modèles les plus piqués donnant lieu à une bonne discrimination par la suite).

Nous rappelons maintenant sans les détailler les grandes étapes de l'estimation des paramètres de la distribution Kent. Pour une base d'apprentissage de taille n composée de vecteurs normés de dimension p et notés w_i , on notera $\bar{w} = \frac{1}{n} \sum_{i=1}^n w_i$ leur moyenne et $S = \frac{1}{n} \sum_{i=1}^n w_i w_i^t$ leur matrice de dispersion autour de 0.

La matrice Γ^t permet de passer du repère naturel des données au repère canonique où $(1, 0, 0)$ correspond au vecteur moyen, $(0, 1, 0)$ et $(0, 0, 1)$ aux 2 axes principaux de dispersion. Ainsi, Γ vérifie $\Gamma \bar{w} = (\|\bar{w}\|, 0, 0)^t$. Pour trouver Γ , l'opération est réalisée en 2 temps pour simplifier l'écriture et la résolution du processus.

Tout d'abord, on choisit une matrice de rotation H qui permet d'orienter \bar{w} le long de l'axe $(1, 0, 0)^t$, le reste de la transformation est noté K . On a alors $\Gamma = HK$. En notant (ϕ, θ) les coordonnées polaires de \bar{w} telles que $\phi = \arctan\left(\frac{\bar{w}_z}{\bar{w}_y}\right)$ et $\theta = \arccos\left(\frac{\bar{w}_x}{\|\bar{w}\|}\right)$ en notant $\bar{w}_x, \bar{w}_y, \bar{w}_z$ les coordonnées spatiales de \bar{w} . Cette matrice s'écrit donc :

$$H = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) \cos(\phi) & \cos(\theta) \cos(\phi) & -\sin(\phi) \\ \sin(\theta) \sin(\phi) & \cos(\theta) \sin(\phi) & \cos(\phi) \end{bmatrix} \quad (\text{B.39})$$

Dans ce référentiel la matrice de dispersion s'écrit :

$$B = H^t S H = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} \quad (\text{B.40})$$

Puisque la moyenne est alignée sur le premier axe canonique, il suffit alors de diagonaliser la sous-matrice 2×2 des composantes restantes de B afin de calculer K en déterminant l'angle de rotation autour du vecteur $(1, 0, 0)$:

$$K = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\Psi) & -\sin(\Psi) \\ 0 & \sin(\Psi) & \cos(\Psi) \end{bmatrix} \quad (\text{B.41})$$

En se basant sur le calcul des vecteurs propres de $B_{2:3,2:3}$, cet angle se détermine de la façon suivante : $\Psi = \frac{1}{2} \arctan\left(2 \frac{b_{23}}{b_{22} - b_{33}}\right)$.

Finalement, en notant $r_1 = \|\bar{w}\|$ qui mesure l'importance donnée à la mesure moyenne et $r_2 = T_{22} - T_{33}$ où $T = \Gamma^t S \Gamma$ correspond à la matrice de dispersion exprimée dans le repère canonique (après réalignement). Pour κ suffisamment grand on obtient finalement :

$$\begin{cases} \kappa &= \frac{1}{2 - 2r_1 - r_2} + \frac{1}{2 - 2r_1 + r_2} \\ \beta &= \frac{1}{2 - 2r_1 - r_2} - \frac{1}{2 - 2r_1 + r_2} \end{cases} \quad (\text{B.42})$$

B.3.1.4 Watson

La distribution de Watson [Watson 1965] (voir la figure B.4) correspond à l'approximation d'une Gaussienne isotrope à des données axiales. Son utilisation est marginale dans le contexte des neurosciences, toutefois elle a été utilisée comme modèle élémentaire pour des études de diffusion dans

[Schwartzman 2005]. La vraisemblance d'un axe w est donnée ci-après :

$$p(w|\kappa\mu) = P(-w|\kappa\mu) = \frac{\exp(\kappa(w \cdot \mu)^2)}{C(\kappa)} \quad (\text{B.43})$$

où κ est un paramètre d'échelle contrôlant la dispersion de la distribution et μ correspond à la direction de l'axe moyen. La constante de normalisation vaut : $C(\kappa) = \int_0^1 e^{\kappa t^2}$. On fera ici l'approximation suivante : $C(\kappa) \approx \frac{k}{\pi \exp(\kappa)}$, valide uniquement pour des distributions suffisamment concentrées autour de leur moyenne (pour de grandes valeurs de κ).

À partir d'une base de données $\{w_i; i = 1 : n\}$ composée de données axiales normées définies à un signe près, on se propose d'estimer les paramètres κ et μ . La log-vraisemblance à maximiser s'écrit :

$$L = \kappa \sum_{i=1}^n (\mu \cdot w_i)^2 + n \log C(\kappa) \quad (\text{B.44})$$

$$= n (\kappa \mu^t S \mu + \log C(\kappa)) \quad (\text{B.45})$$

en notant S la matrice de dispersion autour de 0 telle que $S = \frac{1}{n} \sum_{i=1}^n w_i w_i^t$. L'estimateur μ^* par maximum de vraisemblance est donc solution du problème suivant :

$$\begin{cases} \mu^* = \arg \max_{\mu} \mu^t S \mu \\ \text{sous contrainte } \mu^t \mu = 1 \end{cases} \Rightarrow \begin{cases} \text{En introduisant le Lagrangien } \lambda : \\ \mu^* = \arg \max_{\mu} \mu^t S \mu + \frac{\lambda}{2} (1 - \mu^t \mu) \end{cases} \quad (\text{B.46})$$

En annulant la dérivée, on obtient : $(S - \lambda Id)\mu = 0$, donc μ^* est un vecteur propre de S , or la quantité qu'on cherche à maximiser vaut $\mu^t S \mu = \lambda$, donc il s'agit du vecteur propre associé à la plus grande valeur propre notée λ^* .

Finalement, en dérivant B.45 par rapport à κ , on obtient : $\lambda^* = \mu^{*t} S \mu^* = -\frac{C'(\kappa)}{C(\kappa)} \approx 1 - \frac{1}{\kappa}$ en se basant sur l'expression approchée $C(\kappa) \approx \frac{k}{\pi \exp(\kappa)}$ signalée plus tôt.

B.3.1.5 Bingham

La distribution de Bingham [Bingham 1974] (voir la figure B.4) correspond à l'approximation d'une Gaussienne bivariée appliquée à des données axiales. Elle raffine la distribution de Watson en permettant un contrôle complet de la structure de covariance. Cette distribution est également bimodale puisqu'elle traite des données directionnelles. Elle est très populaire dans l'étude du paléomagnétisme [Onstott 1980]. Son expression est la suivante :

$$P(w|A) = P(-w|A) = \frac{\exp(\|w\|_A^2)}{{}_1F_1(0.5, 1.5, Z)} \quad (\text{B.47})$$

où $\|w\|_A^2 = w^t A w$ est le produit scalaire canonique défini par la matrice symétrique définie positive A et $A = M Z M^t$ (M orthogonale et Z diagonale) où ${}_1F_1$ est la fonction conflente hypergéométrique d'argument matriciel Z .

On considère maintenant une base de données d'apprentissage $\{w_i; i = 1 : n\}$ de taille n , où chaque w_i est un vecteur normé de dimension p . L'estimation par maximum de vraisemblance de la matrice A n'a pas de formulation analytique connue. Nous nous proposons de suivre la méthode proposée par [Onstott 1980, Tanaka 1999]. En notant, $S = \frac{1}{n} \sum_{i=1}^n w_i w_i^t$ la matrice de dispersion des données étudiées, l'estimateur par maximum de vraisemblance de M correspond à la matrice

des vecteurs propres de S ordonnés de la plus petite à la plus grande valeur propre. Dans le cas particulier d'une direction w indéterminée (que nous noterons : $w=?$), il est toute de même possible de déterminer sa contribution à la matrice de dispersion en intégrant cette mesure sur l'ensemble des directions possibles :

$$C(w = ?) = \oint_{S(2)} ww^t dw = \frac{4}{3}\pi Id \quad (\text{B.48})$$

Concernant les facteurs de dispersion codés par la matrice Z , seule leur estimation numérique est possible. Nous avons utilisé la méthode de Powell [Powell 1964] afin de maximiser l'expression de la log-vraisemblance. Elle vaut :

$$L = WMZM^tW^t - \log({}_1F_1(0.5, 1.5, Z)) \quad (\text{B.49})$$

où W^t est la matrice des vecteurs colonnes w_i . La fonction ${}_1F_1(0.5, 1.5, Z)$ est plutôt difficile à calculer, nous avons donc utilisé une implémentation efficace fournie par [Koev 2006].

B.3.1.6 Matrix Von Mises Fisher

La distribution de Von Mises Fisher décrite plus tôt permet de modéliser une dispersion Gaussienne autour d'une direction privilégiée pour des vecteurs normés de dimension p . Pour une collection d'objets rigides $\{i = 1 : n\}$, chacun décrit par plusieurs directions d'intérêt dont les angles sont fixes entre eux et notées sous forme matricielle : $W_i = [w_{i,1}, \dots, w_{i,q}]$ avec $q \leq p$ (c'est à dire q directions dans p dimensions). Pour fixer les idées, on peut donner comme exemple en 3D, l'étude de la modélisation d'un objet 2D en forme de L caractérisée par 2 directions naturelles que constituent chacune des branches du L . On peut alors modéliser chaque composante directionnelle de façon indépendante (toutes les $\{w_{i,1}; i = 1 : n\}$ entre elles, ... , toutes les $\{w_{i,q}; i = 1 : n\}$ entre elles) par une distribution de Von Mises Fisher. Pour répondre à cette limitation [Downs 1972] a proposé une généralisation de ce modèle traitant conjointement de telles données : il s'agit du modèle Matrix Von Mises Fisher (aussi appelé Matrix Langevin), dont l'étude a été détaillée par la suite dans [Khatri 1977, Jupp 1979, Chikuse 2003]. Un modèle alternatif basé sur la décomposition de Cayley, a été proposé dans [León 2006]. De telles données W vérifient $W^tW = C$ où C est une matrice symétrique définie positive contrôlant les angles entre les différentes directions. L'ensemble des matrices vérifiant cette condition forment une C -variété de Stiefel. Nous nous restreindrons au cas $W^tW = Id$ (la variété de Stiefel) puisqu'il suffit d'appliquer la transformation linéaire suivante $W' = WC^{-\frac{1}{2}}$ sur la donnée initiale W afin de se ramener au cas de l'identité. L'expression de la vraisemblance d'un tel objet W s'écrit alors :

$$p(W|F) = \frac{\exp(\text{Tr}(FW^t))}{{}_0F_1\left(\frac{p}{2}, \frac{D^2}{4}\right)} \quad \text{avec} \quad \begin{cases} F = PDQ & \text{la décomposition en valeur singulière (SVD)} \\ \text{tel que} & \begin{cases} P \text{ et } Q \text{ soient des matrices de rotation} \\ D \text{ une matrice diagonale} \end{cases} \end{cases} \quad (\text{B.50})$$

Pour donner une interprétation plus fine à cette paramétrisation, nous allons déterminer quelle valeur de W maximise cette probabilité (le mode principal de cette densité de probabilité). Pour se faire, il faut introduire la contrainte $W^tW = I$ dans l'optimisation, ce qui revient à maximiser l'énergie E suivante :

$$E = FW^t + \text{Tr}\Lambda(W^tW - I) \quad (\text{B.51})$$

avec la matrice Lagrangienne Λ définie telle que $\Lambda^t = \Lambda$ pour satisfaire la contrainte. La dérivée de cette expression vaut $\frac{\partial E}{\partial W} = F + 2W\Lambda$, en utilisant la contrainte d'orthogonalité, on obtient :

$W^t F = F^t W$, et ainsi :

$$W^* = F^{-t}(F^t F)^{\frac{1}{2}} \quad (\text{B.52})$$

$$= (PD^{-1}Q)(Q^t D^2 Q)^{\frac{1}{2}} \quad (\text{B.53})$$

$$= PQ \quad (\text{B.54})$$

Ainsi, $W^* = PQ$ correspond à une configuration la plus probable pour cette distribution qui est aussi la configuration moyenne [Downs 1972]. On a alors $F = PDQ = (PDP^t)W^*$.

Afin de poursuivre notre compréhension des paramètres de cette distribution, appliquons un changement de repère sur les données considérées afin d'aligner la configuration moyenne W^* sur la matrice identité (ou identité tronquée si $q < p$). Dans ce contexte, l'expression de la vraisemblance est proportionnelle à $\exp(\text{Tr}(PDP^t W^t))$. Ce qui suggère finalement que P définit les axes de plus grande dispersion de la distribution, alors que D correspond à l'étendue de cette dispersion.

Étude des fonctions marginales : La distribution Matrix Von Mises Fisher étudiée agit sur des données de grandes dimensions. Nous l'étudierons ici en nous limitant aux cas des matrices de rotation 3×3 . Il s'agit de données paramétrisées de façon minimale par 4 paramètres. Il n'est donc pas possible de visualiser directement cette distribution. L'étude des fonctions de densité marginales permet de réduire le nombre de dimensions étudiées en intégrant la vraisemblance sur une ou plusieurs variables. Selon la paramétrisation choisie pour représenter la rotation et ce que l'on souhaite mettre en valeur, il existe plusieurs marginalisations possibles. Pour simplifier les explications, nous supposons que $W^* = I$ et donc que $F = PDP^t$ sans nuire à la généralité puisque cette supposition ne fait qu'aligner la moyenne de la distribution avec les axes canoniques par une simple rotation. Nous proposons ici de décomposer chaque rotation $W = [w_x, w_y, w_z]$ selon les 3 vecteurs colonnes qui la compose. Ces vecteurs représentent les coordonnées du repère canonique après transformation par la rotation W . Nous allons donc pouvoir étudier chacun de ces axes séparément. Les calculs concernant l'étude du premier axe w_x va être détaillé ci-après, par un jeu d'écriture en remplaçant x par y et y par z et z par x , on étudie W_y ou x par z , y par x et z par y revient à considérer W_z . La densité de probabilité marginale s'écrit alors :

$$P(w_x|F) = \int_0^{2\pi} \exp \text{Tr}(FW_\theta^t) d\theta \quad \text{avec } W_\theta = [w_x, w_y^\theta, w_z^\theta] \quad (\text{B.55})$$

où w_y^θ est défini en faisant tourner un vecteur w_y^0 de référence (tel que $w_x \cdot w_y^0 = 0$) d'un angle θ autour du vecteur w_x . On définit alors $w_z^\theta = w_x \times w_y^\theta$ afin que W_θ représente un repère évoluant autour de l'axe w_x . La matrice de rotation autour du vecteur w_x d'un angle θ s'écrit :

$$R_x^\theta = I + \sin(\theta)J(w_x) + (1 - \cos(\theta))J^2(w_x) \quad \text{où } J(w_x) = \begin{pmatrix} 0 & -w_{xz} & w_{xy} \\ w_{xz} & 0 & -w_{xx} \\ -w_{xy} & w_{xx} & 0 \end{pmatrix} \quad (\text{B.56})$$

en utilisant la formule de Rodrigues. Finalement, on obtient : $w_y^\theta = R_x^\theta w_y^0 = w_y^0 \cos(\theta) + w_z^0 \sin(\theta)$, en notant $w_z^0 = w_x \times w_y^0$ et par suite : $w_z^\theta = w_z^0 \cos(\theta) - w_y^0 \sin(\theta)$. Pour simplifier les notations, on notera : $w_y = w_y^0$ et $w_z = w_z^0$.

Ainsi, en notant $F = [f_x, f_y, f_z]$ (décomposition en vecteurs colonnes), la densité marginale étudiée s'écrit $P(w_x|F) = \exp(w_x^t f_x) \int_0^{2\pi} \exp[\cos(\theta)(w_y^t f_y + w_z^t f_z) + \sin(\theta)(w_z^t f_y - w_y^t f_z)] d\theta$. Or,

les intégrales de ce type vérifient :

$$\begin{aligned}
\int_0^{2\pi} \exp[\cos(\theta)a + \sin(\theta)b] d\theta &= \int_0^{2\pi} \exp\left[\sqrt{a^2 + b^2}(\cos(\theta)\cos(t) + \sin(\theta)\sin(t))\right] d\theta \\
&= \int_0^{2\pi} \exp\left[\sqrt{a^2 + b^2}\cos(\theta - t)\right] d\theta \\
&= \int_0^{2\pi} \exp\left[\sqrt{a^2 + b^2}\cos(\theta)\right] d\theta \\
&= 2\pi I_0(\sqrt{a^2 + b^2})
\end{aligned}$$

où $I_0(\kappa)$ est la fonction de Bessel modifiée d'ordre 0, déjà rencontrée en tant que constante de normalisation des distributions Von Mises. Ici $a^2 + b^2 = (w_y^t f_y + w_z^t f_z)^2 + (w_z^t f_y - w_y^t f_z)^2$. Nous allons maintenant réécrire cette expression de façon à enlever toute référence au choix arbitraire des vecteurs de référence w_y et w_z :

$$\begin{aligned}
a^2 + b^2 &= (e_y^t F w_y + e_z^t F w_z)^2 + (e_y^t F w_z - e_z^t F w_y)^2 \quad \text{avec } I = [e_x, e_y, e_z] \\
&= w_y^t F (e_y e_y^t + e_z e_z^t) F w_y + w_z^t F (e_y e_y^t + e_z e_z^t) F w_z + 2 [w_y^t F (e_y e_z^t - e_z e_y^t) F w_z] \\
&= \text{Tr} [F(I - e_x e_x^t) F (I - w_x w_x^t)] + 2 \text{Tr} [F e_y e_z^t F (w_z w_y^t - w_y w_z^t)]
\end{aligned}$$

$$\text{puisque } WW^t = I = \sum_{i \in \{x, y, z\}} w_i w_i^t \text{ et } I = \sum_{i \in \{x, y, z\}} e_i e_i^t$$

$$= \text{Tr} [F(I - e_x e_x^t) F] - \|w_x\|_{F(I - e_x e_x^t) F}^2 - 2 \text{Tr} [F e_y e_z^t F S J(w_x) S]$$

avec $S = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ et $J(w_x)$ comme défini dans l'équation B.56.

On se concentre maintenant sur la deuxième partie de l'équation :

$$\begin{aligned}
\text{Tr} [F e_y e_z^t F S J(w_x) S] &= \text{Tr} [e_z^t F S J(w_x) S F e_y] \\
&= \text{Tr} [e_z^t F S J^t (S F e_y) w_x]
\end{aligned}$$

puisque $J(x) = J^t(-x)$ et $J(x)y = -J(y)x$

$$\begin{aligned}
&= \text{Tr} [(J(S F e_y) S F e_z)^t w_x] \\
&= -\text{Tr} [(S J(F e_y) F e_z)^t w_x] \\
&= -\text{Tr} [(S J(P D P^t e_y) P D P^t e_z)^t w_x] \\
&= -\text{Tr} [(S J P (D P^t e_y) D P^t e_z)^t w_x] \\
&= -\text{Tr} [(S |D| P D^{-1} P^t J(e_y) e_z)^t w_x] \quad \text{avec } |D| \text{ le déterminant de } D \\
&= -|D| \text{Tr} [(S F^{-1} e_x)^t w_x] \\
&= -d_y d_z e_x^t F^{-1} S w_x
\end{aligned}$$

en notant $[d_x, d_y, d_z]$ les éléments diagonaux de D .

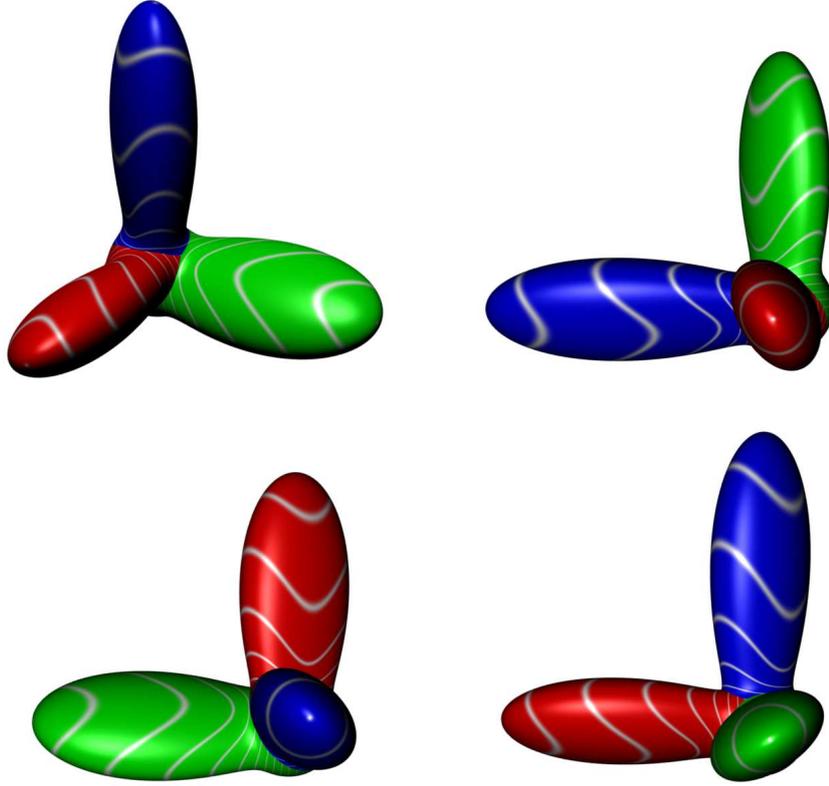


FIG. B.5 – Représentation jointe des 3 fonctions de densité de probabilités marginales de la distribution Matrix Von Mises Fisher sous 4 angles différents. Chaque couleur correspond à une des marginales. Les lignes de niveau de couleur blanche représentent chacune l'ensemble des vecteurs unitaires 3D équidistants du mode de la marginale correspondante. Les lignes sont alors réparties uniformément autour de ce même mode. Les valeurs utilisées ici sont $P = Q^t = \frac{\sqrt{3}}{3}J(x)$ avec $x = [1, 0.8, 0.7]$ et la matrice diagonale $D = \text{diag}([2, 5, 20])$.

Finalement, on obtient la relation suivante :

$$p(w_x|F) = 2\pi \exp(w_x^t f_x) I_0 \left(\sqrt{\text{Tr} [F(I - e_x e_x^t)F] - \|w_x\|_{F(I - e_x e_x^t)F}^2 + 2d_y d_z e_x^t F^{-1} S w_x} \right) \quad (\text{B.57})$$

où $\exp(w_x^t f_x)$ ressemble à une distribution de Von Mises Fisher classique à quelques détails près puisque f_x n'est pas de norme 1.

Ces formules peuvent maintenant être mises à profit pour faciliter la représentation visuelle d'une telle distribution de probabilités. Dans beaucoup d'applications sur données réelles, il est raisonnable de penser que l'ensemble des rotations que l'on cherche à étudier se concentre autour d'une rotation moyenne avec une variabilité limitée qui rend improbable des rotations de l'ordre de $\pi/2$ (quel que soit le vecteur rotation choisi). Sous cette condition, les 3 densités marginales $p(w_x|F)$, $p(w_y|F)$ et $p(w_z|F)$ ont des supports (zones où la vraisemblance est sensiblement différente de zéro) spatiaux distincts, on peut donc les représenter ensemble, ce qu'on peut observer dans la figure B.5.

Estimation des paramètres : À partir d'une base de données $\{W_i; i = 1 : n\}$, il est possible d'estimer par maximum de vraisemblance une partie des paramètres; les détails calculatoires ne seront pas précisés ici (voir [Jupp 1979] pour une démonstration complète). On considère alors

l'expression de la moyenne empirique $\bar{W} = \frac{1}{n} \sum_{i=1}^n W_i$. Sa décomposition en valeur singulière $\bar{W} = \bar{P}\bar{D}\bar{Q}$ telle que \bar{P} et \bar{Q} soient des matrices de rotation et \bar{D} une matrice diagonale. Cette expression est fortement liée aux estimateurs P^* et Q^* du maximum de vraisemblance puisque :

$$\begin{cases} P^* &= \bar{P} \\ Q^* &= \bar{Q} \\ \bar{D}_{i,i} &= \frac{\partial}{\partial D_{i,i}} {}_0F_1\left(\frac{p}{2}, \frac{D^2}{4}\right) \end{cases} \quad (\text{B.58})$$

où $D_{i,i}$ et $\bar{D}_{i,i}$ désignent respectivement le i ème élément de la diagonale des matrices D et \bar{D} .

Cette fois encore, il n'est pas possible de calculer explicitement l'estimateur de vraisemblance des composantes de dispersion. La méthode de Powell [Powell 1964] associée à l'estimation de la constante de normalisation ${}_0F_1(\cdot)$ [Koev 2006] permet d'estimer D^* .

[Chang 1992] et [Wood 1993] ont proposé des approximations permettant de simplifier et d'accélérer l'estimation de ces paramètres. [Wood 1993] propose dans le cas restreint de $SO(3)$ (ensemble des matrices de rotation 3×3) d'exploiter l'équivalence entre la distribution de Bingham 4D sur l'ensemble des quaternions unitaires (voir aussi [Prentice 1986]) et la distribution Matrix Von Mises sur les rotations. Il propose également une approche basée sur le calcul d'une pseudo vraisemblance évitant l'intégration numérique de la constante de normalisation.

B.3.1.7 Matrix Bingham

La distribution Matrix Bingham [Jupp 1979] est à la distribution Bingham ce que la distribution Matrix Von Mises Fisher est à la distribution Von Mises Fisher. En effet, elle permet de modéliser la répartition d'objets X présentant plusieurs axes de symétries (par exemple, en 3D, de tels objets seraient assimilables à un axe, un cylindre ou une sphere selon le nombre d'axes présentant des symétries). Cette distribution permet de traiter des objets appartenant à la variété de Stiefel de dimension p (tel que $XX^t = I_q$ avec I_q la matrice identité de dimension q et $X^tX \neq I_p$ si $p \neq q$). Sa formulation comporte donc de nombreuses symétries qui contraignent fortement sa forme :

$$p(W|F) = \frac{\exp(\text{Tr}(WFW^t))}{{}_1F_1\left(\frac{q}{2}, \frac{p}{2}, D^2\right)} \quad \text{avec} \quad \begin{cases} F = PDP \\ \text{tel que} \quad \begin{cases} P \text{ est une matrice de rotation} \\ D \text{ une matrice diagonale} \end{cases} \end{cases} \quad (\text{B.59})$$

Cette distribution vérifie : $P(WS|F) = P(-W|F) = P(W|F)$ ou S est une matrice diagonale prenant comme valeurs 1 ou -1 (correspondant à l'inversion d'un ou plusieurs axes de W). Elle est donc antipodale avec 2^q modes. De plus, $P(RW|F) = P(W|F)$, quelquesoit la matrice de rotation R considérée.

Pour comprendre plus en détail le comportement de cette distribution, nous allons considérer 3 cas différents sur des données 3D. Si le nombre d'axes considérés est égal au nombre de dimensions $p = q$, alors $P(W|F) \propto \exp(\text{Tr}(W))$ correspond à la distribution uniforme, ce qui est logique puisque les données sont alors assimilables à des sphères. Les deux autres cas sont duaux ($p = 1$ et $p = 2$). En effet, si on écrit : $W^t = [w_x, w_y]$ (les deux axes considérés) alors $p(W|F) \propto \exp(\text{Tr}(FW^tW)) = \exp(\text{Tr}(F(I - w_z w_z^t))) = \exp(\text{Tr}(F) - \text{Tr}(w_z^t W w_z))$ avec $w_z = w_x \times w_y$. L'expression $\text{Tr}(w_z^t F w_z)$ correspond à une distribution de Bingham classique. On remarque également, que la distribution est entièrement déterminée par w_z et donc invariante selon w_y et w_x . Les fonctions marginales de cette distribution se présentent donc comme une Bingham pour w_z et une Bingham uniforme dans le plan porté par w_x et w_y .

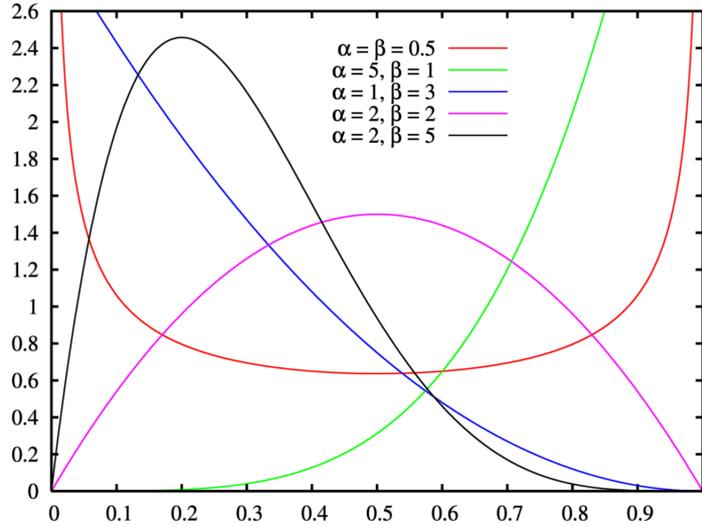


FIG. B.6 – Distribution Beta pour différentes valeurs des paramètres (source de l'image : *wikipedia*). Abscisse : valeur de x_1 . Ordonnée : densité de probabilité au point x_1 .

B.3.2 Distribution sur un simplexe

On considère par définition, qu'un vecteur réel $x = \{x_k; k \in 1 : p\}$ de dimension p appartient au simplexe de dimension p si la somme de ses éléments est égale à 1 : $\sum_{k=1}^p x_k = 1$. Ainsi défini, chaque x_k est compris entre 0 et 1. Un tel vecteur peut ainsi représenter le partitionnement, en termes de proportion, d'une quantité d'intérêt. La suite de cette partie étudie une famille de distributions consacrées à ce type de donnée. Dans le cadre de la modélisation Bayésienne de modèles de mélange, ces distributions permettent d'incorporer un a priori sur la répartition des poids régissant le mélange.

Au cours de ce manuscrit cette distribution a été utilisée à diverses occasions dans ce contexte : pour contraindre les a priori d'un modèle de mélange (section [?]), en tant qu'a priori sur l'ensemble des labels $P(L)$ (section 4.3.2) dont les paramètres sont alors estimés à partir d'une base d'apprentissage.

Trois distributions de complexité croissante (en nombre de paramètres et capacité de modélisation) sont considérées : Beta, Dirichlet et Dirichlet généralisée. Chacune est définie uniquement sur un simplexe et nulle aux bords. En effet, dès qu'une des variables vaut soit 0, soit 1, la densité de probabilité associée est nulle, ce qui en pratique, empêche la dégénérescence d'un modèle de mélange puisqu'aucune composante ne peut disparaître.

B.3.2.1 Beta

La distribution Beta correspond à la distribution de Dirichlet restreinte au 2-simplexe, c'est à dire 2 variables x_1, x_2 sommant à 1, ou exprimé autrement à une seule variable x_1 comprise entre 0 et 1.

Sa densité de probabilité dépend de 2 paramètres α et β et s'exprime comme suit :

$$p(x_1|\alpha, \beta) = \frac{x_1^{\alpha-1} x_2^{\beta-1}}{B(\alpha, \beta)} = \frac{x_1^{\alpha-1} (1-x_1)^{\beta-1}}{B(\alpha, \beta)} \quad (\text{B.60})$$

où la fonction beta $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$. Les paramètres α et β contrôlent à la fois la forme (plus ou moins piquée) et l'asymétrie de la distribution (voir figure B.6).

B.3.2.2 Dirichlet

La distribution de Dirichlet (pour plus de précision se référer à [Minka 2003], voir la figure B.7 pour une appréciation visuelle) généralise la distribution Beta aux simplexes de dimension p quelconque. Sa densité de probabilité en tout vecteur x du simplexe s'écrit :

$$p(x|\alpha) = \frac{1}{B(\alpha)} \prod_{k=1}^p x_k^{\alpha_k - 1} \quad (\text{B.61})$$

où la fonction beta multinomiale $B(\alpha) = \frac{\prod_{k=1}^p \Gamma(\alpha_k)}{\Gamma(\sum_{k=1}^p \alpha_k)}$. Le vecteur α est de dimension p et permet de contrôler les moments d'ordre 1 de la distribution. Par contre les moments d'ordre 2 ne sont pas libres car directement reliés aux moments d'ordre 1. En effet, cette distribution vérifie $\mathbb{E}[x_k|\alpha] = \frac{\alpha_k}{\alpha_0}$ et $Var[x_k|\alpha] = \frac{\alpha_k(\alpha_0 - \alpha_k)}{\alpha_0^2(\alpha_0 + 1)}$, où $\alpha_0 = \sum_{k=1}^p \alpha_k$.

À partir d'une base de données suffisamment grande, l'estimation des paramètres de cette distribution peut être réalisée en alignant les deux premiers moments (moyenne et variance) du modèle : $\mathbb{E}[x|\alpha]$ et $Var[x_k|\alpha]$ sur les moments empiriques calculés sur la base de données étudiée. On les notera $\vec{m} = \{m_k, k = 1 : p\}$ sa moyenne empirique et $\vec{v} = \{v_k, k = 1 : p\}$ sa variance empirique. Ainsi, le vecteur de paramètres α est défini à un facteur c près puisque $\mathbb{E}[x|\alpha] = \frac{\alpha}{\alpha_0}$. On estime alors c à partir des moments d'ordre 2, en minimisant l'erreur quadratique moyenne (*mean square error* ou MSE) entre variance empirique et variance du modèle, noté ici E . En notant $m_0 = \sum_{k=1}^p m_k$, on obtient :

$$c^* = \arg \min_c E \quad (\text{B.62})$$

$$= \arg \min_c \sum_{k=1}^p (Var[x_k|\alpha] - v_k)^2 \quad (\text{B.63})$$

$$= \arg \min_c \left\| \frac{\vec{f}}{m_0 c + 1} - \vec{v} \right\|^2 \quad \text{avec } \vec{f} = \frac{m_k(m_0 - m_k)}{m_0^2} \quad (\text{B.64})$$

$$(\text{B.65})$$

L'annulation de la dérivée en c donne :

$$\frac{\partial E}{\partial c} = 2(m_0 c + 1)^{-2} \left[f \cdot \vec{v} - (m_0 c + 1)^{-1} \|f\|^2 \right] = 0 \quad (\text{B.66})$$

Et finalement :

$$c = \frac{1}{m_0} \left[\frac{\|f\|^2}{f \cdot \vec{v}} - 1 \right] = \frac{1}{m_0} \left[\frac{\sum_{k=1}^p m_k^2 (m_0 - m_k)^2}{m_0^2 \sum_{k=1}^p m_k (m_0 - m_k) v_k} - 1 \right] \quad (\text{B.67})$$

Finalement, le paramètre α recherché vaut tout simplement : $\alpha = c \vec{m}$.

B.3.2.3 Dirichlet généralisé

Pour plus de détails, le lecteur peut se référer à [Wong 1998, Wong 2005] pour son étude comparative entre le modèle Dirichlet et le modèle Dirichlet généralisé dans le cadre d'analyse Bayésienne (pour une comparaison visuelle, voir la figure B.7).

Cette distribution permet de contrôler de façon plus fine la variance de chaque variable. Avec la distribution de Dirichlet, les 2 premiers moments sont liés. Ce qui signifie que 2 variables de même moyenne ont la même variance. Dans le cas où la mesure d'intérêt correspond aux fréquences d'apparition des labels de sillons sur un cortex, cela signifierait que si leur fréquence est la même en moyenne alors leur variabilité aussi, ce qui s'avère souvent faux en pratique.

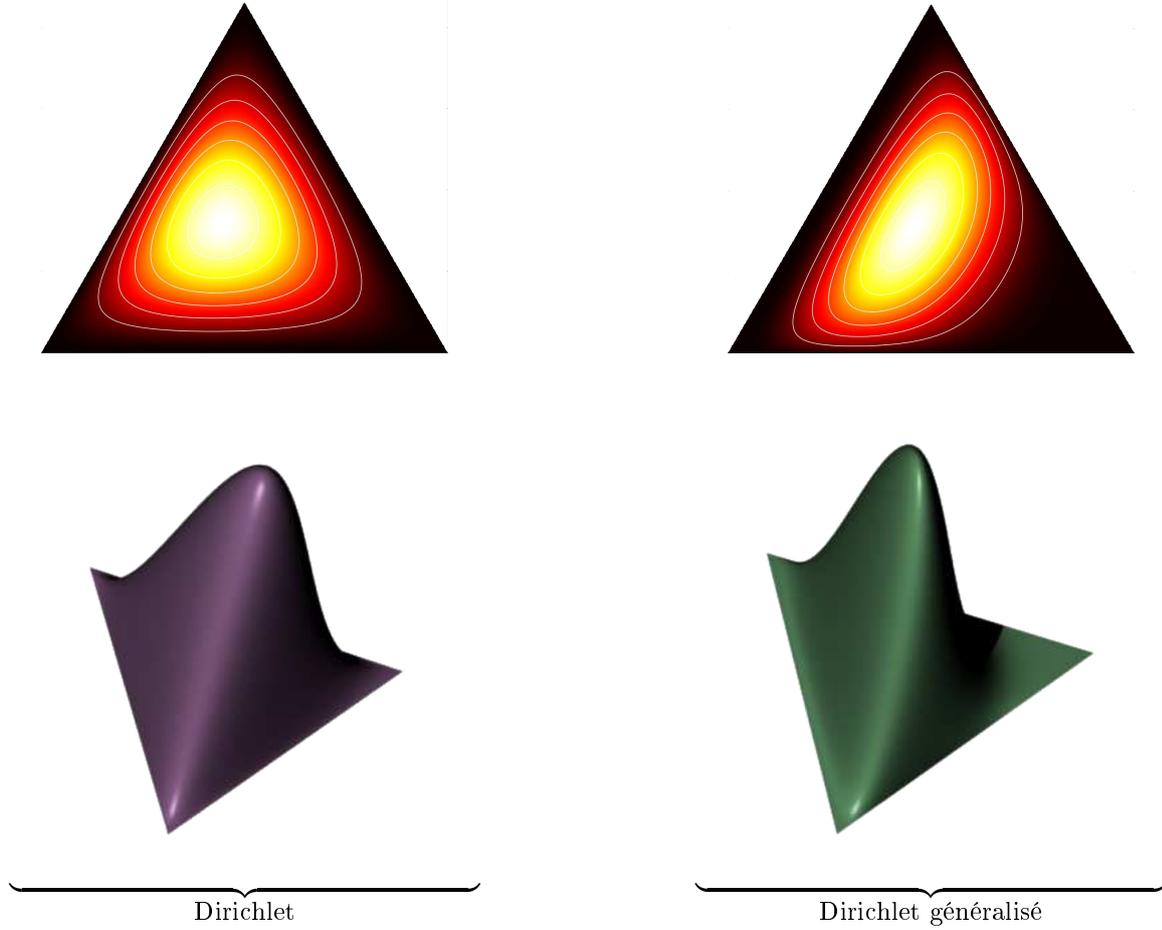


FIG. B.7 – Représentation des distributions de Dirichlet et Dirichlet généralisée pour 3 variables. Haut : carte de vraisemblances et lignes de niveau associées. Bas : représentation 3D où la vraisemblance est représentée par la hauteur. Ici, le vecteur moyenne correspond à 0.29, 0.355, 0.355. Le modèle de droite permet plus de flexibilité et permet de contrôler de façon indépendante la variance de chacune des mesures considérées.

Cette distribution dépend de deux jeux de paramètres : $\alpha = \{\alpha_k; k = 1 : p\}$ et $\beta = \{\beta_k; k = 1 : p\}$ et s'exprime comme suit pour un vecteur x de dimension p dont les composantes somment à 1 :

$$p(x|\alpha\beta) = \prod_{k=1}^p \frac{x_k^{\alpha_k-1} (1-x_1-\dots-x_k)^{\gamma_k}}{B(\alpha_k, \beta_k)} \quad (\text{B.68})$$

avec $\gamma_k = \beta_k - \alpha_{k+1} - \beta_{k+1}$ et $B(\alpha_k, \beta_k) = \frac{\Gamma(\alpha_k)\Gamma(\beta_k)}{\Gamma(\alpha_k+\beta_k)}$.

Wong [Wong 2005] a montré les résultats suivants concernant l'expression des 2 premiers moments marginalisés à une seule des variables d'intérêt x_k . Son espérance $\mathbb{E}[x_k|\alpha\beta]$ et sa variance

$Var[x_k|\alpha\beta]$ s'exprime comme suit, pour k différent de p :

$$\left\{ \begin{array}{l} \mathbb{E}[x_k|\alpha\beta] = \frac{\alpha_k}{\alpha_k + \beta_k} \prod_{i=1}^{k-1} \frac{\beta_i}{\alpha_i + \beta_i} \\ Var[x_k|\alpha\beta] = \frac{\alpha_k(\alpha_k + 1)}{(\alpha_k + \beta_k)(\alpha_k + \beta_k + 1)} \prod_{i=1}^{k-1} \frac{\beta_i(\beta_i + 1)}{(\alpha_i + \beta_i)(\alpha_i + \beta_i + 1)} - \mathbb{E}[x_k|\alpha\beta]^2 \end{array} \right. \quad (\text{B.69})$$

et pour $k = p$:

$$\left\{ \begin{array}{l} \mathbb{E}[x_p|\alpha\beta] = \prod_{i=1}^p \frac{\beta_i}{\alpha_i + \beta_i} \\ Var[x_p|\alpha\beta] = \prod_{i=1}^p \frac{\beta_i(\beta_i + 1)}{(\alpha_i + \beta_i)(\alpha_i + \beta_i + 1)} - \mathbb{E}[x_p|\alpha\beta]^2 \end{array} \right. \quad (\text{B.70})$$

Comme dans le cas de la distribution de Dirichlet nous allons chercher à estimer les paramètres vectoriels α et β à partir d'une base de données par la méthode des moments, en égalant les moments empiriques et ceux du modèle. De la même façon, on notera $\vec{m} = \{m_k, k = 1 : p\}$ sa moyenne empirique et $\vec{v} = \{v_k, k = 1 : p\}$ sa variance empirique.

Pour $k = 1$, l'expression des paramètres α_1 et β_1 ne dépend pas des autres paramètres : $\mathbb{E}[x_1|\alpha\beta] = \mathbb{E}[x_1|\alpha_1\beta_1] = \frac{\alpha_1}{\alpha_1 + \beta_1}$ et $Var[x_1|\alpha\beta] = Var[x_1|\alpha_1\beta_1] = \frac{\alpha_1(\alpha_1 + 1)}{(\alpha_1 + \beta_1)(\alpha_1 + \beta_1 + 1)}$. On obtient alors :

$$\left\{ \begin{array}{l} \alpha_1 = \frac{m_1}{v_1}(m_1 - m_1^2 - v_1) \\ \beta_1 = \frac{1 - m_1}{v_1}(m_1 - m_1^2 - v_1) \end{array} \right. \quad (\text{B.71})$$

Les autres valeurs sont alors estimées en cascade. On notera $\Pi_k^{\mathbb{E}}(\alpha, \beta) = \Pi_k^{\mathbb{E}} = \prod_{i=1}^{k-1} \frac{\beta_i}{\alpha_i + \beta_i}$ et $\Pi_k^{Var}(\alpha, \beta) = \Pi_k^{Var} = \prod_{i=1}^{k-1} \frac{\beta_i(\beta_i + 1)}{(\alpha_i + \beta_i)(\alpha_i + \beta_i + 1)}$ les produits apparaissant dans l'expression des 2 premiers moments dépendant uniquement des paramètres déjà estimés. On obtient finalement :

$$\left\{ \begin{array}{l} \alpha_k = \frac{m_k (m_k (\Pi_k^{Var} - m_k) - v_k)}{(\Pi_k^{\mathbb{E}} - \Pi_k^{Var}) m_k^2 + \Pi_k^{\mathbb{E}} v_k} \\ \beta_k = \frac{(\Pi_k^{\mathbb{E}} - m_k) (m_k (\Pi_k^{Var} - m_k) - v_k)}{(\Pi_k^{\mathbb{E}} - \Pi_k^{Var}) m_k^2 + \Pi_k^{\mathbb{E}} v_k} \end{array} \right. \quad (\text{B.72})$$

B.4 Alternatives et extensions au modèle SPAM

Cette section regroupe quelques modèles explorés rapidement. Il s'agit de bons candidats de substitution aux modèles SPAMs visant une estimation plus robuste ou une représentation moins gourmande en mémoire. Trois thématiques différentes sont abordées : approximation du modèle SPAM, extension de sa formulation et modélisation alternative.

B.4.1 Approximations

Actuellement le modèle SPAM est représenté par un volume 3D, où chaque voxel code la vraisemblance locale qu'une structure portant un label donné passe par ce voxel. Comme nous l'avons vu plus tôt, cette représentation est sensible à la grille d'échantillonnage, l'échelle de représentation, ainsi que le choix de l'espace commun où elle est définie. En effet, la méthode couplant normalisation

et labélisation que nous avons présentée plus tôt, induit de nombreuses transformations au cours de l'algorithme. Nous avons choisi pour accélérer le processus de négliger les effets d'échantillonnage en arrondissant chaque coordonnée après transformation, mais cela nuit en réalité à la précision de la méthode. Une représentation continue et définie en tout point de l'espace serait donc souhaitable. Une simple interpolation des mesures le long de la trame discrète pourrait suffire, mais un deuxième problème se pose, cette fois plutôt handicapant. Les volumes 3D (malgré le fait qu'ils soient restreints à des boîtes englobantes resserrées autour du cœur de la distribution) prennent en général beaucoup de place en mémoire : 1.5Mo en moyenne, pour un total de près de 100Mo par hémisphère. Nous avons évoqué plus tôt que nous souhaitions prendre en compte la variabilité du paramètre σ (la largeur du noyau Gaussien utilisé pour estimer les SPAMs). Cette analyse nécessiterait sans doute une représentation multi-échelles des SPAMs, où chaque échelle correspond à une valeur de σ . Avec la représentation actuelle cette étude ne peut être menée dans de bonnes conditions à cause de la quantité mémoire nécessaire. En réponse, nous proposons donc dans les sections à suivre quelques premières idées pour représenter un modèle de SPAM comme une fonction continue spatialement et prenant une place réduite en mémoire : le modèle RBF et le modèle de mixture de Gaussiennes.

B.4.1.1 Radial Basis Function : RBF

Les fonctions à base radiale (RBF [Powell 1992]) sont plutôt populaires pour mettre au point des méthodes d'interpolation ou d'approximation de fonction. Il s'agit d'une fonction qui a la propriété d'être constante le long de lignes de niveaux concentriques autour d'un point de référence. Plusieurs de ces fonctions peuvent alors être combinées pour donner une représentation intuitive $h_\pi(x)$ d'une fonction cible $f(x)$ (la fonction à approcher ou interpoler) basée sur un ensemble de points de contrôle μ_k et une fonction radiale Φ telle que :

$$\begin{aligned} h_\pi(x) &= \sum_k \pi_k \Phi(\|x - \mu_k\|^2) \\ &= \Phi(\cdot, x)^t \pi \end{aligned} \tag{B.73}$$

avec π l'ensemble des poids π_k , et $\Phi(\cdot, x) = \{\Phi(\|x - \mu_k\|^2); k\}$. Il existe de nombreux choix pour la fonction Φ , une formulation populaire étant celle du noyau Gaussien d'écart-type σ tel que $\Phi(r) \propto \exp(-\frac{r}{2\sigma^2})$.

Dans le cas de l'interpolation, on souhaite que la fonction $h_\pi(x)$ vérifie $h_\pi(x) = y$ pour un ensemble de couple (x, y) , une valeur y rattachée à chaque point de contrôle x . S'il n'y a pas assez de points, l'estimation des points de contrôle n'est pas possible (sans contraintes additionnelles). S'il y en a trop, il s'agit alors d'approximation, comme nous le verrons par la suite. Dans cette situation, le problème peut s'exprimer sous la forme matricielle suivante :

$$\vec{y} = \Phi \pi \rightarrow \pi = \Phi^{-1} \vec{y} \tag{B.74}$$

avec \vec{y} l'ensemble des valeurs à interpoler et $\Phi = \{\Phi(\cdot, x); x\}$ défini sur l'ensemble des points x où l'interpolation est réalisée. Dans le cas des modèles de SPAM, nous avons choisi ces points comme étant le résultat d'un k -means sur une distribution. C'est à dire que nous avons réalisé un k -means sur l'ensemble des points de la boîte englobante du label concerné et pondéré chaque point par sa vraisemblance dans le modèle SPAM; ce qui revient à multiplier le nombre d'occurrences de chaque position à hauteur de cette même vraisemblance. Cette opération vise à favoriser un plus grand nombre de points de contrôle dans les zones où la vraisemblance est plus élevée.

Dans le cas de l'approximation, on suppose cette fois, que l'on souhaite minimiser une erreur d'approximation entre la fonction cible $f(x)$ et la fonction d'approximation $h_\pi(x)$, en l'occurrence

selon les moindres carrés :

$$\begin{aligned} E &= \sum_x \|h_\pi(x) - f(x)\|^2 \\ &= \|\Phi\pi - \vec{y}\|^2 \end{aligned} \quad (\text{B.75})$$

dont la solution est $\pi = (\Phi^t\Phi)^{-1}\Phi^t\vec{y}$, où nous avons étendu la définition de la matrice Φ à l'ensemble des points x participant à l'approximation (en l'occurrence, l'ensemble des voxels de la boîte englobante dans le cas des SPAMs).

Finalement, cette formulation peut ne pas être complètement satisfaisante car elle n'offre pas la garantie que la fonction ainsi approchée soit toujours une distribution vérifiant : $\int_x h_\pi(x)dx = 1$. Dans le cas d'une fonction radiale Gaussienne, il est possible de reconsidérer la fonction $h_\pi(x)$ comme une mixture de Gaussiennes dont le centre μ_k et l'écart-type σ de chaque Gaussienne est prédéterminé et où la valeur des a priori π_k reste inconnue. De cette façon :

$$h_\pi(x) = \sum_k \pi_k \Phi(\|x - \mu_k\|^2) = \sum_k P(z = k)P(x|z = k; \mu_k\sigma) \quad (\text{B.76})$$

où z est une variable indicatrice de la Gaussienne concernée dans la mixture. Dans ce cadre, $\pi_k = P(z = k)$. Il faut donc ajouter la contrainte $\sum_k \pi_k = 1$ à l'optimisation précédente (équation B.77) pour obtenir une distribution de probabilité aux propriétés désirées. En introduisant le lagrangien λ , l'énergie à optimiser devient :

$$E = \frac{1}{2}\|\Phi\pi - \vec{y}\|^2 + \lambda \left(1 - \sum_k \pi_k\right) \quad (\text{B.77})$$

où $\frac{\partial E}{\partial \pi} = \Phi^t\Phi\pi - \Phi^t\vec{y} - \lambda\mathbb{1} = 0$ avec $\mathbb{1}$ un vecteur de la même taille que π mais ne contenant que des 1. On a alors :

$$\pi = (\Phi^t\Phi)^{-1}(\Phi^t\vec{y} + \lambda\mathbb{1}) \quad (\text{B.78})$$

$$\Rightarrow 1 = \mathbb{1}^t (\Phi^t\Phi)^{-1} (\Phi^t\vec{y} + \lambda\mathbb{1}) \quad (\text{B.79})$$

$$\Rightarrow \lambda = \frac{1 - \mathbb{1}^t (\Phi^t\Phi)^{-1} \Phi^t\vec{y}}{\mathbb{1}^t (\Phi^t\Phi)^{-1} \mathbb{1}} \quad (\text{B.80})$$

En combinant les lignes B.78 et B.80, on obtient le résultat désiré. Le cas de la fonction radiale Gaussienne cumulé à la contrainte d'intégration suggère d'approcher le modèle SPAM par une mixture de Gaussiennes où cette fois l'ensemble des paramètres (y compris centres et matrice de variance-covariance) est à déterminer. L'avantage du modèle RBF est que l'estimation des paramètres est directe ce qui permet une estimation rapide contrairement au modèle à suivre.

B.4.1.2 Modèle de mixture de Gaussiennes

La mixture de Gaussiennes (*Gaussian Mixture Model* ou encore GMM) est une distribution de probabilités composée d'un ensemble de Gaussiennes. Considérons un point x issu d'une de ces Gaussiennes désignée par la variable z , la vraisemblance de ce point s'exprime alors :

$$\begin{aligned} P_{\text{gmm}}(x) &= \sum_k P(z = k)P(x|z = k; \mu_k\Sigma_k) \\ &= \frac{1}{(2\pi)^{\frac{d}{2}}} \sum_k \pi_k \frac{\exp(-\frac{1}{2}(x - \mu_k)^t \Sigma_k^{-1} (x - \mu_k))}{\sqrt{|\Sigma_k|}} \end{aligned} \quad (\text{B.81})$$

où $\pi_k = P(z = k)$ correspond à l'a priori qu'un point soit issu de la Gaussienne i , $P(x|z = k; \mu_k\Sigma_k)$ est la vraisemblance en x de cette même Gaussienne, avec μ_k sa moyenne et Σ_k sa matrice de variance-covariance.

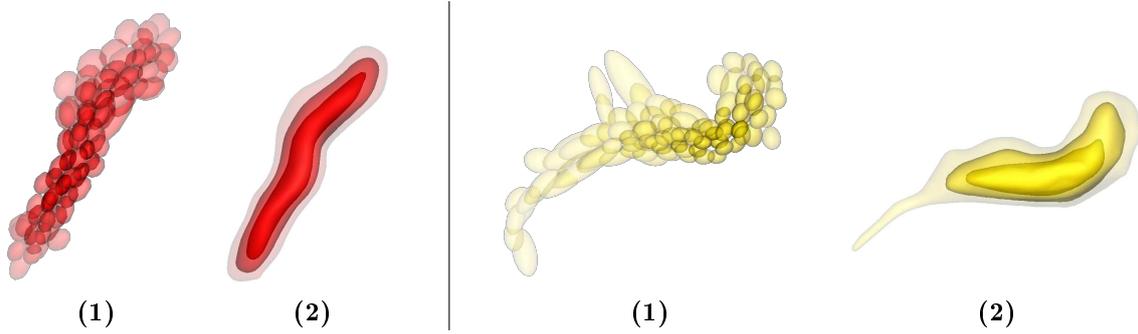


FIG. B.8 – Approximation de SPAMs par une mixture de Gaussiennes sur le sillon central (à gauche) et la partie postérieure de la fissure cingulaire (à droite). Les modèles visualisés correspondent aux GMM sous 2 représentations : **(1)** l'une où chaque Gaussienne du mélange est représentée séparément par une forme ovoïdale délimitant 60% de la distribution et **(2)** la même représentation que celle utilisée pour les SPAMs.

Afin de remplacer notre représentation actuelle de la variabilité sulcale cette distribution peut être utilisée de 2 façons. Une première possibilité consisterait à estimer ce modèle à partir des positions des voxels d'un ensemble de sillons portant le même label. Cette approche comporte un inconvénient majeur par rapport à l'estimation réalisée par le modèle de SPAM : il n'y a pas de lissage spatial de l'information, qui rappelons-le assure la robustesse du procédé. Une autre façon consiste à approcher un modèle de SPAM classique par un GMM. C'est à dire trouver l'ensemble des paramètres de cette distribution qui minimise la distance entre ces deux distributions. Nous avons choisi de minimiser la divergence de Kullback-Leibler entre $P_{\text{gmm}}(x)$ et $P_{\text{spam}}(x)$. Pour un label donné (que nous ne précisons pas dans les notations ici), cette mesure s'exprime de la façon suivante :

$$\begin{aligned} KL(P_{\text{spam}}(x), P_{\text{gmm}}(x)) &= \sum_{x \in \Omega} P_{\text{spam}}(x) \log \frac{P_{\text{spam}}(x)}{P_{\text{gmm}}(x)} \\ &= - \sum_{x \in \Omega} P_{\text{spam}}(x) \log P_{\text{gmm}}(x) + C \end{aligned} \quad (\text{B.82})$$

avec Ω le support de la distribution et C une constante ne dépendant pas des paramètres du modèle à optimiser. La minimisation de cette expression correspond également à une estimation des paramètres par maximum de vraisemblance en pondérant la contribution de chaque voxel de l'espace par la vraisemblance du modèle SPAM en ce point.

L'estimation des paramètres de la mixture de Gaussiennes est réalisée par l'algorithme EM [Dempster 1977] déjà décrit à plusieurs reprises dans ce manuscrit. Il s'agit d'une méthode itérative qui consiste 1) à partir d'une solution initiale des paramètres (dans notre cas, on peut utiliser le résultat d'une méthode à base de RBF proposée précédemment), 2) à calculer les probabilités a posteriori locales (en chaque voxel) d'appartenir à telle ou telle Gaussienne (étape E : Expectation) et 3) à estimer les paramètres du modèle en pondérant les contributions de chaque point par les probabilités a posteriori obtenue précédemment (étape M : Maximization). On notera \mathcal{M} l'ensemble des paramètres du modèle GMM à estimer. L'étape E consiste finalement à calculer :

$$p_{ik}^{(n)} = P(z = k|x; \mathcal{M}^{(n)}) = \frac{P(x|z = k; \mathcal{M}^{(n)}) P(z = k; \mathcal{M}^{(n)})}{\sum_k P(x|z = k; \mathcal{M}^{(n)}) P(z = k; \mathcal{M}^{(n)})} \quad (\text{B.83})$$

Pour l'étape M , il faut trouver les paramètres $\mathcal{M}^{(n+1)}$ qui maximisent la quantité suivante :

$$\mathbb{E}_Z \left[\log P \left(\mathcal{M}^{(n+1)} \right) + \sum_{x_i \in \Omega} w_i \log P \left(x_i z_i | \mathcal{M}^{(n+1)} \right) \middle| \{x_i; i\} \mathcal{M}^{(n)} \right] \quad (\text{B.84})$$

où chaque x_i correspond à un voxel du modèle de SPAM que l'on cherche à approcher. La notation w_i représente la vraisemblance $P_{\text{spam}}(x_i)$ au point x_i . Avec les notations introduites plus tôt, cela revient à maximiser l'énergie suivante :

$$\begin{cases} E = \log P \left(\mathcal{M}^{(n+1)} \right) + \sum_{x_i \in \Omega} \sum_k w_i p_{ik}^{(n)} \log \left[P \left(x_i | z_i = k; \mathcal{M}^{(n+1)} \right) \pi_k \right] \\ \text{s.t.} \quad \sum_k \pi_k = 1 \end{cases} \quad (\text{B.85})$$

Sans la spécification d'a priori approprié pour $P(\mathcal{M}^{(n+1)})$, l'estimation peut s'avérer instable numériquement dans notre contexte d'utilisation. En effet, les poids w_i sont plutôt mal définis loin du coeur de la distribution et peuvent entraîner des imprécisions numériques empêchant une bonne convergence de l'algorithme EM. De plus, dans certains cas, une ou plusieurs composantes du modèle peuvent être amenées à disparaître si leur a priori π_k tend vers 0. Dernier point, sans contrainte sur la structure de covariance des Gaussiennes, l'estimation peut également engendrer des Gaussiennes dégénérées où certaines valeurs propres tendent vers 0. Afin de prévenir tous ces cas pathologiques nous avons ajouté un ensemble d'a priori standard :

$$\begin{cases} \pi \sim \text{Dir}(\alpha) & \text{Dirichlet} \\ \mu_k \sim \mathcal{N}(m_k, \tau_k \Sigma_k^{-1}) & \text{Gaussienne} \\ \Sigma_k^{-1} \sim \mathcal{W}(M_k, \rho) & \text{Wishart} \end{cases} \quad (\text{B.86})$$

dont les expressions sont les suivantes :

$$\begin{cases} P(\pi) \propto \prod_k \pi_k^{\alpha-1} = \left[\prod_k \pi_k \right]^{\alpha-1} \\ P(\mu_k) \propto \exp \left(-\frac{1}{2} \tau_k \|m_k - \mu_k\|_{\Sigma_k^{-1}}^2 \right) \\ P(\Sigma_k^{-1}) \propto |\Sigma_k^{-1}|^{\frac{\rho-d-1}{2}} \exp \left(-\frac{1}{2} \text{Tr} (M_k \Sigma_k^{-1}) \right) \end{cases} \quad (\text{B.87})$$

avec d le nombre de dimensions. Ici, α , τ_k , m_k , ρ et M_k sont des paramètres contrôlant la forme des a priori, dont les valeurs ont été fixées empiriquement après quelques essais. En introduisant la contrainte grâce au lagrangien λ et en omettant les références aux numéros d'itération (n) et ($n+1$), l'énergie à maximiser devient :

$$\begin{aligned} E &= \sum_{x_i \in \Omega} \sum_k w_i p_{ik} \left[-\frac{1}{2} (x_i - \mu_k)^t \Sigma_k^{-1} (x_i - \mu_k) - \frac{1}{2} \log |\Sigma_k| + \log \pi_k \right] \\ &+ \sum_k \left[(\alpha - 1) \log \pi_k - \frac{\tau_k}{2} (m_k - \mu_k)^t \Sigma_k^{-1} (m_k - \mu_k) - \frac{1}{2} \log |\tau_k^{-1} \Sigma_k| \right. \\ &\left. - \frac{1}{2} ((1 + d - \rho) \log |\Sigma_k^{-1}| + \text{Tr} (M_k \Sigma_k^{-1})) \right] + \lambda \left(1 - \sum_k \pi_k \right) + C(\mathcal{M}) \end{aligned} \quad (\text{B.88})$$

où $C(\mathcal{M})$ ne dépend pas des paramètres du modèle. La distribution de Dirichlet empêche qu'une des composantes soit démesurément sous ou sur-représentée dans le mélange de Gaussiennes. L'a priori

sur les centres permet de stabiliser et d'accélérer le processus de convergence puisqu'il empêche les centres de trop se déplacer par rapport à l'initialisation, la tolérance restant conditionnée par la matrice de covariance de chacune des Gaussiennes. Finalement, l'a priori sur la matrice de covariance est choisi de manière à être centré sur une matrice de covariance isotrope telle que $M_k = \nu_k Id$.

L'ensemble des paramètres du modèle $(\{\pi_k \mu_k, \Sigma_k; k\})$ minimisant cette expression est calculée simplement par annulation de la dérivée. Pour l'a priori, on a :

$$\frac{\partial E}{\partial \pi_k} = \frac{1}{\pi_k} \left((\alpha - 1) + \sum_{x_i} w_i p_{ik} \right) - \lambda = 0 \quad (\text{B.89})$$

Dans le cas où l'ensemble du support Ω du modèle de SPAM est utilisé, on a $\sum_k \sum_{x_i} w_i p_{ik} = \sum_{x_i} w_i \sum_k p_{ik} = \sum_{x_i} w_i = 1$. En pratique, il peut être intéressant de ne considérer qu'un sous-échantillonnage régulier de ce support. Dans cette situation, on a seulement $\sum_k \sum_{x_i} w_i p_{ik} = \sum_{x_i} w_i$ dans le cas général. En introduisant la contrainte sur π , on obtient alors :

$$\pi_k = \frac{(\alpha - 1) + \sum_{x_i} w_i p_{ik}}{|z|(\alpha - 1) + \sum_{x_i} w_i} \quad (\text{B.90})$$

en notant $|z|$ le nombre de Gaussiennes du modèle GMM. Pour les centres des Gaussiennes, on a :

$$\frac{\partial E}{\partial \mu_k} = \Sigma_k^{-1} \left[\tau_k (m_k - \mu_k) + \sum_{x_i} w_i p_{ik} (x_i - \mu_k) \right] = 0 \quad (\text{B.91})$$

ce qui donne finalement :

$$\mu_k = \frac{\tau_k m_k + \sum_{x_i} w_i p_{ik} x_i}{\tau_k + \sum_{x_i} w_i p_{ik}} \quad (\text{B.92})$$

Pour les matrices de covariance, on a :

$$\begin{aligned} \frac{\partial E}{\partial \Sigma_k} &= \frac{1}{2} \sum_{x_i} w_i p_{ik} [(\Sigma_k^{-1} (x_i - \mu_k)(x_i - \mu_k)^t \Sigma_k^{-1} - \Sigma_k^{-1}) - \Sigma_k^{-1}] \\ &\quad + \frac{1}{2} [\tau_k \Sigma_k^{-1} (m_k - \mu_k)(m_k - \mu_k)^t \Sigma_k^{-1} - \Sigma_k^{-1} - (\rho - d - 1) \Sigma_k^{-1} + \Sigma_k^{-1} M_k \Sigma_k^{-1}] \\ &= \frac{1}{2} \left[\Sigma_k^{-1} \left(\tau_k (m_k - \mu_k)(m_k - \mu_k)^t + M_k + \sum_{x_i} w_i p_{ik} (x_i - \mu_k)(x_i - \mu_k)^t \right) \right. \\ &\quad \left. - \left(\rho - d + \sum_{x_i} w_i p_{ik} \right) \right] \Sigma_k^{-1} = 0 \end{aligned} \quad (\text{B.93})$$

On obtient alors :

$$\Sigma_k = \frac{\tau_k (m_k - \mu_k)(m_k - \mu_k)^t + M_k + \sum_{x_i} w_i p_{ik} (x_i - \mu_k)(x_i - \mu_k)^t}{\rho - d + \sum_{x_i} w_i p_{ik}} \quad (\text{B.94})$$

À partir de ces formules, l'approximation d'un modèle SPAM par un modèle GMM est réalisable, mais à condition de fixer le nombre de Gaussiennes auparavant. Ce modèle étant très populaire pour modéliser des probabilités, il existe des critères simplifiés (dérivés de la sélection de modèles) permettant de définir le nombre de Gaussiennes optimal pour un problème donné. Un des critères les plus populaires est le critère BIC (*Bayesian Information Criterion*) défini comme suit :

$$\text{BIC} = -2 \sum_{x_i} w_i \log P_{\text{gmm}}(x_i) + f \log(n) \quad (\text{B.95})$$

où f est le nombre de degré de liberté du modèle. n correspond classiquement au nombre de points considérés pour l'estimation. Dans notre cas, puisque nous cherchons à rapprocher 2 distributions l'une de l'autre le nombre de points utilisés pour l'estimation n'a pas vraiment d'importance. Ce critère ne semble donc pas adapté à notre cas d'utilisation. Le nombre de Gaussiennes considérées reflète en fait la précision désirée dans l'approximation. On pourrait alors éventuellement chercher le nombre minimal de Gaussiennes nécessaire pour atteindre une précision donnée au sens d'une erreur quadratique sur Ω par exemple.

B.4.2 Extensions

Le modèle SPAM modélise la probabilité qu'un voxel donné fasse partie d'un sillon portant le label l . Cette densité de probabilité est estimée par un estimateur à noyau (Gaussienne isotrope d'écart-type $\sigma = 2$ dans notre cas). Rappelons l'expression de cet estimateur à la position spatiale x :

$$\begin{aligned} P(x|L=l) &= \frac{1}{n} \sum_{y \in \mathcal{Y}} K_{\sigma}(x, y) \\ &= \frac{1}{n\sigma^3(2\pi)^{\frac{3}{2}}} \sum_{y \in \mathcal{Y}} \exp\left(-\frac{\|x-y\|^2}{2\sigma^2}\right) \end{aligned} \quad (\text{B.96})$$

avec \mathcal{Y} l'ensemble des positions des voxels composant chacun des sillons portant le label l . Cet estimateur de densité considère que le bruit présent dans les données mesurées est uniforme spatialement et son amplitude est caractérisée par l'écart-type σ .

En réalité la précision d'estimation à laquelle on peut prétendre dépend du nombre de points présents dans cette zone, c'est à dire de la densité empirique de points. Dans notre cas, les sujets «*outliers*» entraînent une densité inhomogène dans les zones éloignées du coeur de la distribution : ici l'incertitude spatiale est grande, il faudrait donc utiliser une grande valeur de σ . Au contraire, dans les zones où beaucoup de sujets se superposent, l'incertitude spatiale est plus réduite, une valeur beaucoup plus petite de σ peut alors être utilisée.

Une première réponse à ce problème est donnée par l'estimateur des k -plus proches voisins, où le paramètre k remplace alors le paramètre σ . Il évalue en chaque point x la distance $d_k(x)$ de ce point à son k ième plus proche voisin dans l'ensemble \mathcal{Y} . La densité en ce point est alors approchée par :

$$P(x|L=l) = \frac{k}{nV_k(x)} = \frac{3k}{4\pi n d_k(x)^3} \quad (\text{B.97})$$

où $V_k(x)$ correspond au volume de la sphère centrée en x et de rayon $d_k(x)$. Cette formulation s'interprète de la façon suivante. Si on réalise le tirage de n points selon $P(x|L=l)$ on s'attend à en avoir environ $nV_k(x)P(x|L=l)$ dans le volume $V_k(x)$. Par définition, ce volume contient effectivement k points, ce qui explique la formulation précédente.

Une seconde solution consiste à combiner les deux approches (méthode à noyaux et plus proches voisins). Elle porte le nom d'estimateur des k -plus proches voisins généralisé. La distance $d_k(x)$ entre chaque point x et son k ième plus proche voisin est utilisée pour régler de manière adaptative le paramètre de précision σ . On pose alors $\sigma(x) = d_k(x)$, ce qui donne finalement :

$$\begin{aligned} P(x|L=l) &= \frac{1}{n} \sum_{y \in \mathcal{Y}} K_{\sigma(x)}(x, y) \\ &= \frac{1}{n d_k(x)^3 (2\pi)^{\frac{3}{2}}} \sum_{y \in \mathcal{Y}} \exp\left(-\frac{\|x-y\|^2}{2d_k(x)^2}\right) \end{aligned} \quad (\text{B.98})$$

Une illustration de ces 3 méthodes est proposée en figure B.4.2. Il s'agit de l'estimation de la variabilité de la localisation du sillon central chez 62 sujets.

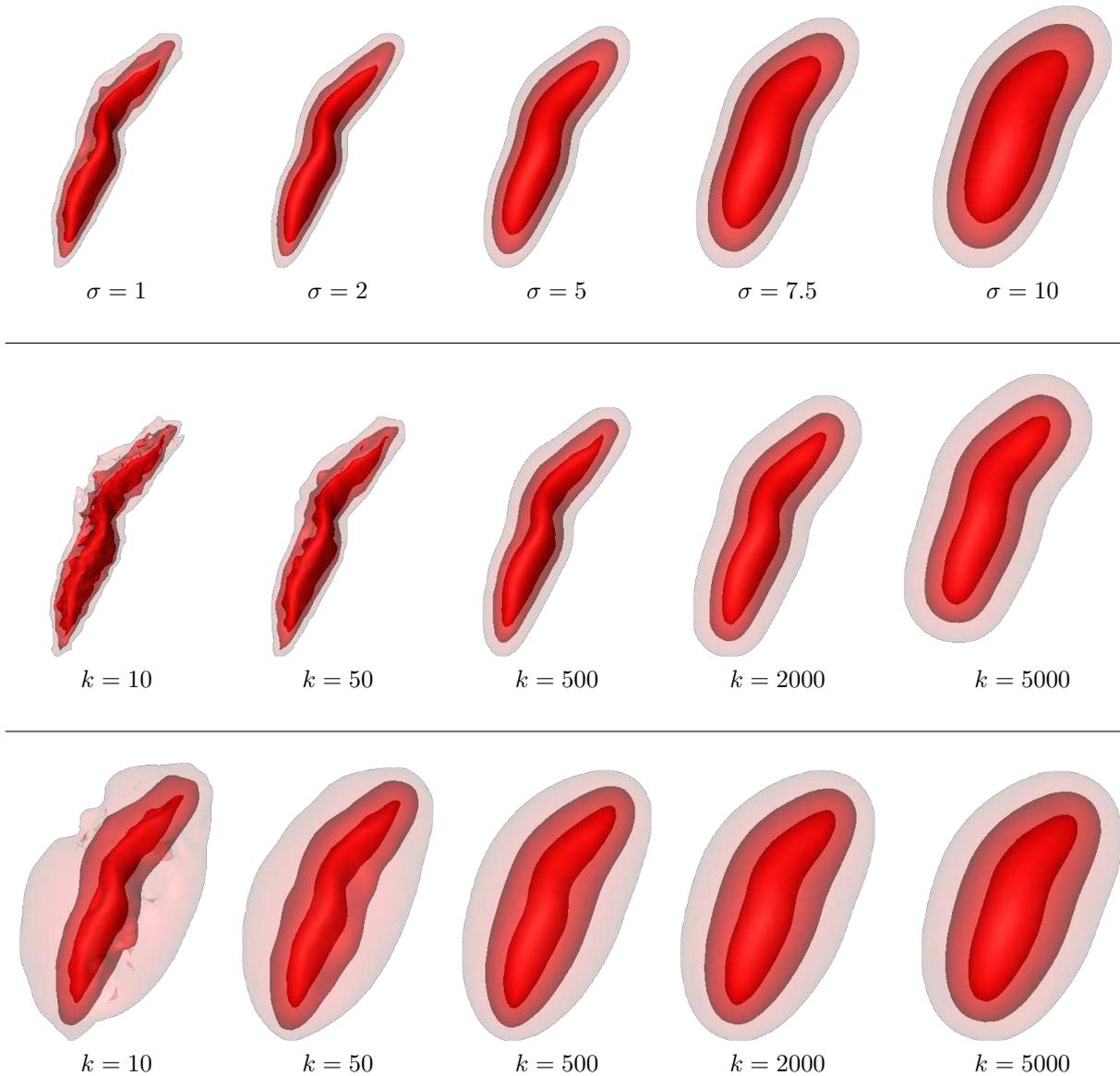


FIG. B.9 – Comparaison de trois estimateurs de densité sur des données sulcales (62 sillons centraux). Haut : estimation classique par fenêtre de Parzen (approche utilisée pour les modèles présentés dans le corps du manuscrit, avec $\sigma = 2$). Centre : estimation par k -plus proches voisins. Bas : estimation par la méthode des k -plus proches voisins généralisée (méthode combinant les deux approches précédentes).

La méthode des k -plus proches voisins généralisée estime l'incertitude spatiale en chaque point et utilise en toute logique un noyau adapté à chaque spécificité locale. Toutefois, chaque point y de l'espace est soumis à cette règle indépendamment de son voisinage. L'estimation de cette incertitude spatiale peut alors présenter des discontinuités spatiales. Une autre façon d'intégrer cette information consiste donc à considérer que chaque point y de l'espace contribue à l'estimation de la densité en tout point x à hauteur de son incertitude spatiale locale. En suivant ce principe, on obtient la méthode d'estimation par noyau variable (*Variable Kernel Estimation* [Breiman 1977]) :

$$\begin{aligned} P(x|L=l) &= \frac{1}{n} \sum_{y \in \mathcal{Y}} K_{h\sigma(y)}(x, y) \\ &= \frac{1}{nh^3(2\pi)^{\frac{3}{2}}} \sum_{y \in \mathcal{Y}} \frac{1}{d_k(y)^3} \exp\left(-\frac{\|x-y\|^2}{2h^2 d_k(y)^2}\right) \end{aligned} \quad (\text{B.99})$$

avec h un paramètre réglant le degré de lissage global de l'estimation. Contrairement à la formulation précédente, celle-ci garantit à la fois la continuité et la dérivabilité de l'estimateur. En contre-partie, le calcul effectif de cet estimateur est plus compliqué puisque pour l'estimation en x de la densité, un point éloigné de celui-ci pourrait contribuer de façon non négligeable à l'estimation. Cette contrainte complique nettement l'estimation.

Nous avons évoqué ici différentes méthodes d'estimation de densité. Il est probable que l'estimateur actuel soit proche de l'optimal pour la plupart des labels. Par contre, un certain nombre de structures, trop variables ou aux formes arborescentes pourraient profiter de telles approches. Dans notre présentation les noyaux utilisés étaient des noyaux Gaussien isotropes. La plupart des sillons ont une forme de ruban, la variabilité de la localisation n'est donc pas homogène le long de la direction principale du sillon, dans la direction de sa profondeur ou orthogonalement. Il serait envisageable d'affiner les estimateurs précédent en y incorporant un a priori tenant compte de l'anisotropie locale de cette variabilité.

B.4.3 Alternatives

Nous proposons ici 2 modélisations alternatives basées sur une représentation particulière des données sulcales. On choisit une boîte englobante telle que l'ensemble des sillons des sujets de notre base de données \mathcal{y} soient inclus. On représente alors chaque sillon ou segment sulcal par un vecteur binaire où chaque entrée de ce vecteur correspond à un voxel de la boîte englobante, de telle façon que la valeur soit égale à 1 si la structure passe par ce voxel et 0 sinon. Nous noterons \mathcal{X} un tel vecteur (par analogie avec la notation utilisée pour représenter l'ensemble des voxels représentant un segment dans le cas du modèle SPAM). Cette représentation pose des difficultés certaines à l'intégration de la prise en compte de l'incertitude spatiale, puisqu'alors les positions des voxels sont définies à une rotation et à une translation près. La transformation $R \cdot x + t$ exprimée en chaque voxel x n'a pas d'expression simple (affine par exemple) à l'échelle d'une structure \mathcal{X} . Détaillons maintenant deux approches exploitant différemment cette représentation : une basée sur une distribution de Bernoulli multivariée et une autre sur une Gaussienne multivariée.

B.4.3.1 Bernoulli multivariée

Nous allons considérer qu'en chaque voxel la probabilité d'avoir un sillon suit une probabilité de Bernoulli. C'est à dire qu'il y a présence d'une structure (valeur 1) avec une probabilité p et absence (valeur 0) avec une probabilité $1-p$ (et toute autre valeur est impossible : probabilité nulle). Notons, comme dans le cas des SPAMS $f_l(x)$ la fréquence d'apparition d'une structure portant le label l à la position spatiale x . La vraisemblance locale au point x d'une telle structure i s'écrit alors tout

simplement :

$$P_{\text{Bernoulli}}(x|L_i = l) = f_l(x) \quad (\text{B.100})$$

Pour une structure \mathcal{X} , elle s'exprime alors :

$$\begin{aligned} P_{\text{Bernoulli}}(\mathcal{X}|L_i = l) &= \prod_{x \in \mathcal{X}} P_{\text{Bernoulli}}(x|L_i = l)^x \prod_{x \in \mathcal{X}} (1 - P_{\text{Bernoulli}}(x|L_i = l))^{1-x} \\ &= \prod_{\substack{x \in \mathcal{X} \\ x=1}} f_l(x) \prod_{\substack{x \in \mathcal{X} \\ x=0}} (1 - f_l(x)) \end{aligned} \quad (\text{B.101})$$

où pour simplifier l'écriture x désigne selon les besoins la position spatiale ou la valeur du vecteur \mathcal{X} en cette position.

Dans le cas de l'estimation classique (sans prise en compte d'incertitudes spatiales) d'un modèle de SPAM, la vraisemblance locale s'écrit tout simplement : $P_{\text{spam}}(x|L_i = l) = \frac{f_l(x)}{\sum_{x \in \Omega} f_l(x)} = \frac{f_l(x)}{F_l}$, où Ω désigne l'ensemble des voxels possibles dans la boîte englobante considérée et en notant $F_l = \sum_{x \in \Omega} f_l(x)$. On notera aussi : $P_{\text{spam}}(x|L_i = l) = \frac{1-f_l(x)}{\sum_{x \in \Omega} 1-f_l(x)} = \frac{1-f_l(x)}{|\Omega| - F_l}$ avec $|\Omega|$ le nombre d'éléments de l'ensemble Ω . Cette notation désigne une sorte d'anti-SPAM, un SPAM mesurant l'absence d'un sillon plutôt que sa présence. L'introduction de ces notations permet d'écrire le modèle de Bernoulli en fonction du modèle de SPAM :

$$\begin{aligned} P_{\text{Bernoulli}}(\mathcal{X}|L_i = l) &= F_l(|\Omega| - F_l) \prod_{\substack{x \in \mathcal{X} \\ x=1}} P_{\text{spam}}(x|L_i = l) \prod_{\substack{x \in \mathcal{X} \\ x=0}} P_{\text{spam}}(x|L_i = l) \\ &= F_l(|\Omega| - F_l) P_{\text{spam}}(\mathcal{X}|L_i = l)^{|\mathcal{X}_1|} P_{\text{spam}}(\mathcal{X}|L_i = l)^{|\mathcal{X}_0|} \end{aligned} \quad (\text{B.102})$$

en notant $|\mathcal{X}_1|$ et $|\mathcal{X}_0|$ le nombre d'éléments du vecteur \mathcal{X} valant respectivement 1 et 0.

En l'état cette modélisation n'est pas utilisable car elle donne une trop grande importance à l'anti-SPAM, puisque $|\mathcal{X}_0| \gg |\mathcal{X}_1|$. On pourrait bien évidemment pondérer les contributions de chacune des deux composantes, mais faut-il réellement les équilibrer ? Dans ce modèle, on pénalise les voxels présents aux endroits où on ne s'attend pas à en trouver et également les structures absentes. Dans le cas de la labélisation d'un sillon coupé en segments, les vraisemblances seront fortement pénalisées à cause de l'absence des autres segments lors de leur évaluation individuelle. Une solution peut être de réaliser une analyse au niveau «sillon» en mesurant la vraisemblance de l'ensemble des segments portant le même label en une seule fois.

Pour finir ce type de modélisation et même d'autres plus poussées utilisant des mixtures de distribution de Bernoulli sont communément utilisées dans le cadre de la reconnaissance d'écriture [Juan 2004].

B.4.3.2 Gaussienne multivariée

On considère maintenant que pour chaque label l , l'ensemble des segments portant ce label, chacun représenté sous la forme vectorielle binaire proposée précédemment, suivent une distribution Gaussienne multivariée :

$$P_{\mathcal{N}}(\mathcal{X}|L_i = l, \mu_l, \Sigma_l) = \frac{1}{2\pi^{\frac{N}{2}} |\Sigma_l|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} \|\mathcal{X} - \mu_l\|_{\Sigma_l^{-1}}^2\right) \quad (\text{B.103})$$

où μ correspond à la représentation moyenne et Σ à la matrice de covariance.

Dans cette représentation, il semble y avoir une inadéquation entre la nature des objets (vecteurs binaires) et ce que le modèle est capable de générer. Cette fois encore, le cadre d'utilisation peut être restreint aux calculs de vraisemblance sur des structures. Selon ce point de vue, l'avantage de ce modèle est d'introduire la possibilité d'apprendre une structure de covariance. De cette manière, ce modèle code des informations de forme : à la position s'ajoute alors l'orientation. Dans ce cadre

l'estimation robuste du paramètre μ au point x (noté μ_x), par analogie avec celle utilisée pour les SPAMs s'exprime ainsi :

$$\mu_x = (\mathbb{E}[\mathcal{X}] * K_\sigma)(x) = \mathbb{E}[\mathcal{X} * K_\sigma](x) \quad (\text{B.104})$$

où K_σ est un noyau Gaussien d'écart-type σ . Ainsi, $\mu_x = F_l P_{\text{spam}}(x|L_i = l)$, l'expression du modèle SPAM non normalisé au point x .

Selon le même principe, un estimateur sans biais de la matrice de covariance tenant compte des incertitudes spatiales s'écrit :

$$\Sigma = \frac{1}{n-1} \sum_i (\mathcal{X}_i * K_\sigma)(\mathcal{X}_i * K_\sigma)^t \quad (\text{B.105})$$

où n correspond au nombre de structures \mathcal{X}_i sur lesquelles estimer le modèle. La matrice de covariance Σ est de dimension $|\Omega| \times |\Omega|$, de très loin supérieure au nombre de sujets de notre base, il n'est donc pas envisageable d'utiliser directement l'estimateur proposé plus tôt. Plusieurs possibilités classiques s'offrent à nous : considérer uniquement une matrice de covariance diagonale, ajouter un a priori sur Σ contraignant cette matrice à ne pas trop s'éloigner du modèle diagonal ou ne considérer dans l'estimation que les covariances impliquant des voxels voisins.

Cette approche pourrait être étendue en suivant les recommandations de [Glaunes 2004] afin d'agir sur une représentation des données sous forme de mesures discrètes (représentées chacune par une somme pondérée de Dirac) : $\sum_i w_i \delta_{x_i}$. La norme $\|\cdot\|^2$ serait alors remplacé par une norme agissant sur les mesures. Le lissage K_σ pourrait alors directement être intégré dans la définition de cette métrique. Cette vision du problème permettrait de modéliser plus finement le processus de génération des segments et de se défaire de la dépendance à la représentation des segments sous forme de vecteur binaire (qui complique fortement l'intégration de l'incertitude spatiale comme dans la partie II).

Développements informatiques

Ce chapitre thésaurise les grandes lignes du travail effectué sur le plan informatique pendant cette thèse et s'inscrit en partie dans le cadre des bouleversements logiciels qu'a connu le laboratoire LNAO pendant cette période.

C.1 Introduction

Ces dernières décennies, l'informatique est devenue un outil incontournable pour mener à bien des travaux de recherche méthodologique et les appliquer à des données réelles. La démarche scientifique de recherche s'accompagne désormais de réflexions concernant le choix des technologies (système d'exploitation, langages de programmation, environnement de développement, bibliothèques/bibliothèques de fonctions, ...) ou le choix du *design* logiciel.

Un environnement logiciel de recherche ne se conçoit évidemment pas comme un produit industrialisé, puisque ce dernier se base sur des technologies connues et maîtrisées dans un cadre d'utilisation restreint et bien défini. Dans le cadre de recherches et encore plus dans le cadre d'un travail de thèse, la compréhension du problème traité peut amener à évoluer rapidement, voire à changer complètement d'angle d'attaque (ce qui a été le cas ici au bout d'un an et demi de thèse). C'est pourquoi, sur ce type de travaux, il est important de privilégier trois facteurs : l'interaction, la souplesse et l'évolution rapide, quitte à remettre en cause la structure même du logiciel (en génie logiciel ce type d'exigence désigne les méthodes dites «agiles»). En effet, les objets (au sens programmation objet) manipulés ainsi que les interactions qui les lient (ce qui se rapproche d'une interface de programmation : *Application Programming Interface* ou API) reflètent en grande partie les schémas mentaux conceptualisant le problème traité. Ce type de développement autorise des évolutions rapides, guidées par l'objectif de recherche.

Bien entendu, cette approche ne se suffit pas à elle-même puisque le *design* logiciel ainsi défini ne peut prétendre fournir des bases assez solides pour l'élaboration de futurs travaux. De la même manière, dans le cadre de collaborations, la valorisation de nos méthodes à des applications neuroscientifiques ne peut se faire qu'après une restructuration logicielle permettant d'atteindre une qualité de «mise en production». Il est donc indispensable de retravailler régulièrement la structure logicielle lors de phases critiques du travail de recherche.

Ce chapitre sera tout d'abord consacré à une description rapide des technologies utilisées et à une justification de ces choix (section C.2). Ensuite, une partie des commandes développées au cours de cette thèse sont listées et brièvement décrites (section C.3).

C.2 Technologies utilisées

Les travaux de développement informatique réalisés pendant cette thèse font majoritairement partie du projet *Sigraph* de la plateforme logicielle *Brainvisa* [Rivière 2009]. Historiquement la majorité des logiciels et bibliothèques fournies par ce projet étaient développés en langage C++ (un langage bas niveau où le code doit d'abord être compilé avant d'être lancé, et sans possibilité native d'interactions dynamiques). Au début de ma thèse, un ensemble d'événements, catalysés par la restructuration de notre laboratoire ainsi que son déménagement (un changement favorisant

souvent d'autres changements) sont venus bouleverser ce schéma et ont précipité l'évolution rapide des comportements des chercheurs dans l'étude de leurs données.

L'évolution rapide de la puissance des machines ces 10 dernières années a poussé les langages interprétés (Java, Python, Ruby, Perl, Tcl pour ne citer que les plus connus) sur le devant de la scène informatique alors qu'ils restaient jusqu'alors relégués à des tâches d'administration ou à la gestion de services Web. Ces 5 dernières années, la communauté scientifique et en particulier le monde de la neuroimagerie a élevé le langage Python au rang de prétendant au titre de langage scientifique et s'engage désormais à concurrencer les produits commerciaux spécialisés dans le prototypage scientifique comme MatLab. Je référence ici principalement les projets NumPy, SciPy et Matplotlib qui sont au centre de cette dynamique en permettant une manipulation efficace des opérations d'algèbre linéaire, de statistique ou de visualisation, les plus communes.

Cette thèse m'a offert l'occasion d'être un acteur de cette mouvance en participant à l'évolution de la plateforme *Brainvisa* [Rivière 2009]. La mise en place d'interactions bilatérales entre le langage Python et les langages C/C++ au travers de *bindings* (moyen permettant notamment d'exporter des objets d'un langage à un autre sans copie superflue) ont permis de profiter des facilités qu'apportent les langages interprétés et faciliter le travail de recherche, au travers des possibilités d'interactivité, de modification dynamique, du gain en vitesse de développement. Ainsi, en fonction des besoins, certaines parties du code du projet *Sigraph* ont été déportées en Python pour proposer plus facilement des modélisations mathématiques alternatives, alors que d'autres algorithmes d'abord développés en Python ont été retraduits en C++ pour accélérer leur exécution. Finalement, un découpage judicieux distinguant les calculs bas niveau qui nécessitent une grande rapidité d'exécution et les manipulations d'objets haut niveau, permet de garder un logiciel efficace et très souple.

Durant la première partie de cette thèse, j'ai travaillé également à la restructuration du projet de datamining de la plateforme *Brainvisa* [Rivière 2009] : DataMind. À l'origine, écrit dans le langage de statistiques R par Édouard Duchesnay, il a été porté depuis en Python. Dès lors ce projet n'a pas cessé de s'enrichir avec l'arrivée de nouveaux étudiants (Cécilia Damon et Vincent Michel). Ceci a alors largement favorisé son utilisation au sein du laboratoire alors que le langage Python devenait le moyen le plus simple de faire communiquer les différents projets entre eux.

C.3 Commandes et fonctionnalités

Cette section décrit uniquement¹ une sélection limitée de programmes développés pendant cette thèse, choisis pour leur intérêt scientifique (méthodologique ou exploratoire) et dont l'utilité ne s'est pas démentie. Il s'agit de programmes Python, disponibles dans le projet PySigraph (répertoire `bin` ou `script`) au moment de la publication de cette thèse.

Quatre catégories de programmes se distinguent : les méthodes d'évaluation des erreurs et leur comparaison, les commandes de visualisation indispensables à toute recherche, les outils d'apprentissage et de labélisation, et enfin un outil de *grid computing* (parallélisation de tâches de calcul).

¹Les lecteurs curieux ou désireux d'interactions plus fines avec des données sulcales peuvent me contacter par email : mattieu.perrot@cea.fr

C.3.1 Mesures d'erreurs

Nom du programme	Description
<code>siErrorLightWrapper.py</code>	Calcul des erreurs de labélisation globales (E_{mass} ou E_{SI} , pondéré par la taille des structures ou non).
<code>siError.py</code>	Calcul des erreurs de labélisation locales (par sillon : $E_{local}(l)$).
<code>errors_cmp.py</code>	Compare un ensemble de résultats de labélisation obtenus avec 2 méthodes différentes afin de mesurer la significativité (au travers de p-values) des améliorations ou détériorations des résultats de façon globale (à l'échelle d'un hémisphère) ou locale (au niveau de chaque sillon).
<code>posterior_correlations.py</code>	Calcule les probabilités a posteriori locales, des mesures d'entropie et des corrélations avec la véritable labélisation, à chaque fois, pondéré ou non par la taille des structures.

C.3.2 Visualisation

Nom du programme	Description
<code>siCsvMapGraph.py</code>	Projection de données locales (un vecteur numérique par sillon ou par segment) depuis un tableau (fichier au format CSV) sur un sujet donné. Utile notamment pour représenter les résultats d'erreurs locales.
<code>display_distributions.py</code>	Traduit les distributions de probabilités des modèles locaux de localisation sous forme visuelle pour les distributions suivantes : Gaussienne, GMM, SPAM.
<code>display_local_distributions.py</code>	<i>idem</i> que la commande précédente, mais pour des distributions basées sur d'autres informations que la localisation spatiale. Pour le confort visuel, elles sont représentées, soit au niveau des centres de gravité des sillons lorsque la distribution concerne un seul label, ou sur la moyenne de 2 d'entre eux dans le cas de relations impliquant 2 labels. Cela permet de manipuler les distributions suivantes : Gaussienne, Kent, Von-Mises-Fisher, Bingham, Gaussienne.
<code>anaSulciSnapshot.py</code> <code>autocrop.py</code> <code>mosaic.py</code>	Automatisation de la création de planches au format désiré (4/3, 16/9, ...), représentant chacune un nombre pré-sélectionné de sujet par planche, permettant la visualisation d'une base de données conséquente : interfaces gris-blanc ou blanc-LCR et leurs sillons selon une orientation commune pré-déterminée (vue du dessus, de dessous, de côté, ...). Ce procédé a notamment motivé la remise en cause des labélisations manuelles de la base de données utilisée avant cette thèse.

C.3.3 Apprentissage de modèles et labélisation automatique

	Nom du programme	Description
Système historique	<code>SiLearn.py</code>	Génération des contres-exemples et apprentissage des experts de sillons et de relations (MLP, SVM) à partir d'une base de données. Cette méthode peut inclure une étape de sélection d'attributs.
	<code>SiRelax.py</code>	Reconnaissance automatique des sillons par recuit simulé sur un champ de Markov, exploitant des informations à l'échelle de sillons entiers ou de liens entre paire de sillons.
Labélisation automatique	<code>independent_tag.py</code>	Labélisation indépendante pour chaque segment en utilisant des modèles locaux décrivant chacun des sillons. Par exemple : modèles basés sur des SPAMS, modèles Gaussiens des centres de gravité des sillons, modèles directionnels (Von-Mises-Fisher, Kent) ou axiaux (Bingham).
	<code>independent_tag_with_registration.py</code>	<i>idem</i> que la commande précédente, mais avec co-estimation de paramètres de recalage (rigide, affine ou localement rigide).
	<code>annealing_tag.py</code>	Labélisation simultanée de tout un hémisphère par recuit simulé sur un champs de Markov, exploitant des informations au niveau segment et des liens entre segments sulcaux voisins.
Estimation du modèle	<code>learn_priors.py</code>	Estimation empirique de l'a priori $P(L)$ à partir d'une base de données.
	<code>learn_relations.py</code>	Estimation indépendante de divers modèles locaux de la variabilité conjointe de segments en relation, un par paire de labels. Cette covariabilité anatomique est exprimée par des informations de distances ou de directions entre segments voisins et modélisée par des modèles de type Gaussiens.
	<code>learn_spams_distributions.py</code>	Estimation indépendante pour chaque label de SPAMS modélisant la variabilité de la localisation de chaque sillon, à partir d'une base de données.
	<code>learn_registred_spams_distributions.py</code>	<i>idem</i> que la commande précédente, mais couplé au co-recalage de l'ensemble des sujets.

C.3.4 Parallélisation de tâches

Dès le début de cette thèse, nous avons été confronté à des coûts computationnels importants handicapant sérieusement la validation ou l'étude approfondie de nos modèles. S'extraire de ce carcan était vital pour dépasser notre compréhension limitée des données et fournir un processus de validation fiable sur un gros volume de données (62 sujets, 2 hémisphères chacun) pour mieux évaluer les capacités de généralisation de nos modèles.

Les temps de calcul effectifs des principaux modèles étudiés sont reportés dans le tableau C.1. L'estimation d'un modèle complet (comprenant une soixantaine de sillons) sur 62 sujets prend

		Méthodes	Temps
Labélisation de sillons	SPAM	depuis un espace commun	5s
		avec co-recalage global	6 min
		avec co-recalage local	8 min
		avec champs de Markov entre les segments voisins	5 min
	approche historique		1 h 30*

		Méthodes	Temps
Estimation des modèles	SPAM	depuis un espace commun	40 min
		avec co-recalage global	3 h
		avec co-recalage local	3 h
		avec champs de Markov entre les segments voisins	10 min**
	approche historique		10 min***

TAB. C.1 – Ces tableaux indiquent les temps de calcul concernant l’estimation des modèles (uniquement les plus utiles pour l’utilisateur final) sur une base de données de 62 sujets et la labélisation d’un nouveau sujet. Les chiffres reportés ne concernent les modèles basés sur les SPAM ainsi que l’approche «historique». * : pour une décroissance de température fixée à 0.995. ** : pour l’approche basée sur la modélisation de la distance minimale entre les segments (celle qui donne les meilleurs résultats actuellement). *** : ne comprend pas l’étape de génération des contre-exemples qui prend environ 6h.

plusieurs heures de calcul. Lors de la phase de validation, l’utilisation d’un schéma de type *leave-one-out* impacte la complexité algorithmique de plus d’un ordre de grandeur. En effet, dans ce cadre, pour chaque sujet un modèle est estimé à partir d’une base de données contenant tous les sujets sauf le sujet étudié. Ainsi, pour l’étude d’un seul type de modèle, 62 modèles (assez proches) seront effectivement estimés, un par sujet. Sans compter les tâtonnement inhérents à tout processus de recherche, plus d’une vingtaine de modèles différents ont été considérés au cours de cette thèse et validés selon le procédé chronophage évoqué plus tôt. Finalement, le temps cumulé nécessaire à l’évaluation de l’ensemble des modèles testés représente environ une année de calcul à temps plein sur une machine moderne (cpu à 2GHz, 4000 bogomips) en ne considérant qu’un seul *core*.

Cet ensemble de facteurs explique pourquoi nous nous sommes intéressés de près à la parallélisation de nos calculs sur un ensemble de machines, souvent désigné par le terme *grid computing*. Nous nous sommes tout d’abord intéressés aux serveurs de calcul mis à disposition par le CEA : le CCRT (Centre de Calcul Recherche et Technologie) mais sans succès réel puisque les contraintes imposées par le système (transfert des données et mise à jour logicielle notamment) impliquait un surcoût de maintenance à chaque analyse. De plus, cette plateforme de calcul était largement saturée par d’autres projets impliquant des temps d’attente bien supérieurs aux temps effectifs des calculs sans parallélisation. Depuis la situation a évolué puisque le nombre d’unité de calcul a été démultiplié, mais nous avons alors développé des solutions alternatives de notre côté où les problèmes de maintenance ne se posaient pas. Nous n’avons donc pas trouvé d’intérêt réel à retenter l’expérience avec le CCRT jusqu’à présent.

Au delà des difficultés pratiques liées au déménagement du centre SHFJ d’Orsay au centre Neurospin de Saclay, entre 30 à 60 machines toutes avec plusieurs processeurs et/ou cœurs (homogènes en terme de système d’exploitation et d’accessibilité réseau, mais pas en terme de puissance de calcul) étaient susceptibles d’accueillir nos calculs. Il ne s’agissait en aucun cas de machines dédiées (serveur de calcul) mais de machines utilisateurs (utilisées principalement par des chercheurs) dont la puissance est rarement mise à contribution en permanence. Notre idée a donc été de profiter de cette puissance de calcul latente afin de démultiplier d’autant la rapidité de nos analyses.

Suite au retour d’expérience de chercheurs ayant fait face à cette problématique et engagé quelques développements informatiques en ce sens, nous avons engagé une réflexion afin de poursuivre ces efforts. En particulier, cette démarche a abouti à l’élaboration de plusieurs programmes pour :

- analyser une plage d’adresse IP afin de détecter les machines fonctionnelles et les ranger par type de système d’exploitation.
- sonder une liste de machines pour obtenir leurs caractéristiques (mémoire, nombre d’unité de calcul, utilisateur de la machine).
- lancer des calculs de façon distribuée sur une liste de machine, selon plusieurs modes de fonctionnement : répétition d’une tâche, parallélisation de tâches (avec ou sans dépendance entre les tâches).

Un effort particulier a été apporté au programme de gestion de tâche afin de garantir dans la mesure du possible que l’ensemble des tâches de calcul demandées soient effectivement menées à leur terme. Cette remarque peut paraître anodine sans plus d’explication, mais il existe diverses raisons pour lesquelles une tâche (dont la qualité et le fonctionnement n’est pas à remettre en cause) peut ne pas aboutir. En effet, le logiciel qui contrôle la répartition des calculs ne peut pas garantir l’état des machines distantes utilisées pour les calculs. Un problème réseau peut perturber voire couper la connexion entraînant un arrêt de la tâche, une machine peut être éteinte ou redémarrée pendant l’exécution de la tâche, la tâche peut être explicitement (par l’utilisateur distant) ou implicitement (par le système distant) tuée. Évidemment, plus une tâche prend de mémoire, dure longtemps ou plus le nombre de tâches augmente, plus le risque est grand. Le risque est alors de devoir relancer manuellement les calculs restants jusqu’à ce qu’ils soient tous arrivés à leur terme. Ce qui nuit bien entendu à l’organisation de l’analyse et ajoute un surcoût de maintenance non négligeable. C’est pourquoi, une attention particulière a été portée à ce point : les tâches perdues ou tuées sont finalement relancées sur une autre machine (et ce uniquement si le problème n’est pas lié à la tâche elle-même). En cas de problème, le programme gérant la répartition des tâches permet de tuer les tâches distantes, même s’il s’agit de scripts *shell* complexes, réduisant au minimum les interventions manuelles fastidieuses.

Bibliographie

- [Allasonnière 2007] S. Allasonnière, Y. Amit et A. Trouvé. *Towards a coherent statistical framework for dense deformable template estimation*. Journal Of The Royal Statistical Society Series B, vol. 69, no. 1, pages 3–29, 2007. 68
- [Arsigny 2005] Vincent Arsigny, Xavier Pennec et Nicholas Ayache. *Polyrigid and Polyaffine Transformations : a Novel Geometrical Tool to Deal with Non-Rigid Deformations - Application to the registration of histological slices*. Medical Image Analysis, vol. 9, no. 6, pages 507–523, December 2005. PMID : 15948656. 71, 98
- [Art 1982] D. Art, R. Gnanadesikan et J.R. Kettenring. *Data-based metrics for hierarchical cluster analysis*. Utilitas Mathematica, vol. 21A, pages 75–99, 1982. 174
- [Ashburner 2005] J. Ashburner et K.J. Friston. *Unified segmentation*. NeuroImage, vol. 26, pages 839–851, 2005. 68
- [Ashburner 2007] John Ashburner. *A Fast Diffeomorphic Image Registration Algorithm*. NeuroImage, vol. 38, no. 1, pages 95–113, 2007. 70, 71
- [Auzias 2009] G. Auzias, J Glaunés, O. Colliot, M. Perrot, J-F. Mangin, A. Trouvé et S. Baillet. *DISCO : a Coherent Diffeomorphic Framework for Brain Registration Under Exhaustive Sulcal Constraints*. Springer, 2009. 26, 67, 71, 99, 148
- [Avants 2004] B. Avants et J. Gee. *Geodesic estimation for large deformation anatomical shape averaging and interpolation*. NeuroImage, vol. 23, pages S139–S150, 2004. 71
- [Banerjee 2005] Arindam Banerjee, Inderjit S. Dhillon, Joydeep Ghosh et Suvrit Sra. *Clustering on the Unit Hypersphere using von Mises-Fisher Distributions*. J. Mach. Learn. Res., vol. 6, pages 1345–1382, 2005. 181, 182
- [Basser 2003] P.J. Basser et S. Pajevic. *A Normal Distribution for Tensor-Valued Random Variables : Applications to Diffusion Tensor MRI*. MedImg, vol. 22, no. 7, pages 785–794, July 2003. 98
- [Basser 2007] Peter J. Basser et Sinisa Pajevic. *Spectral decomposition of a 4th-order covariance tensor : Applications to diffusion tensor MRI*. Signal Process., vol. 87, no. 2, pages 220–236, 2007. 98
- [Beg 2005] M.F. Beg, M.I. Miller, A. Trouvé et L. Younes. *Computing large deformation metric mappings via geodesic flows of diffeomorphisms*. International Journal of Computer Vision, vol. 61, no. 2, pages 139–157, 2005. 71
- [Behnke 2003] K. J. Behnke, M. E. Rettmann, D. L. Pham, S. M. Resnick D. Shen, C. Davatzikos et J. L. Prince. *Automatic Classification of Sulcal Regions of the Human Brain Cortex Using Pattern Recognition*. Proceedings of SPIE Medical Imaging 2003 : Image Processing, vol. 5032, pages 1499–1510, 2003. 25, 27
- [Bengio 1995] Yoshua Bengio. *Convergence Properties of the K-Means Algorithms*. In Advances in Neural Information Processing Systems 7, pages 585–592. MIT Press, 1995. 173
- [Besag 1975] Julian Besag. *Statistical Analysis of Non-Lattice Data*. The Statistician, vol. 24, no. 3, pages 179–195, 1975. 108
- [Besag 1986] J. Besag. *On the statistical analysis of dirty pictures*. Journal of the Royal Statistical Society. Series B, vol. 48, no. 3, pages 259–302, 1986. 47, 110
- [Besl 1992] P. J. Besl et H. D. McKay. *A method for registration of 3-D shapes*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 14, no. 2, pages 239–256, 1992. 159, 169

- [Best 1981] D.J. Best et N.I. Fisher. *The bias of the maximum likelihood estimators of the von Mises-Fisher concentration parameters*. Communications in Statistics - Simulation and Computation, vol. 10, no. 5, pages 493–502, 1981. 181
- [Bingham 1974] C Bingham. *An antipodally symmetric distribution on the sphere*. Annals of Statistics, vol. 2, no. 6, pages 1201–1225, 1974. 81, 184
- [Biondi 1998] A. Biondi, H. Nogueira, D. Dormont, M. Duyme, D. Hasboun, A. Zouaoui, M. Chantôme et C. Marsault. *Are the brains of monozygotic twins similar? A three-dimensional MR study*. American Journal of Neuroradiology, vol. 19, no. 7, pages 1361–1367, 08 1998. 158
- [Blanchet 2005] Juliette Blanchet, Florence Forbes et Cordelia Schmid. *Markov Random Fields for Recognizing textures modeled by Feature Vectors*. In International Conference on Applied Stochastic Models and Data Analysis, 2005. 49
- [Boisvert 2008] J. Boisvert, F. Cheriet, X. Pennec, H. Labelle et N. Ayache. *Geometric Variability of the Scoliotic Spine Using Statistics on Articulated Shape Models*. IEEE Transactions on Medical Imaging, vol. 27, no. 4, pages 557–568, April 2008. 80, 100
- [Boykov 1998] Y. Boykov, O. Veksler et R. Zabih. *Markov random fields with efficient approximations*. pages 648–655, 1998. 49
- [Boykov 2001] Y. Boykov, O. Veksler et R. Zabih. *Fast Approximate Energy Minimization via Graph Cuts*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 11, pages 1222–1239, 2001. 112
- [Bradley 1998] Paul S. Bradley et Usama M. Fayyad. *Refining initial points for K-Means clustering*. In Proc. 15th International Conf. on Machine Learning, pages 91–99. Morgan Kaufmann, San Francisco, CA, 1998. 174
- [Breiman 1977] L. Breiman, W. Meisel et E. Purcell. *Variable kernel estimates of multivariate densities*. Technometrics, vol. 19, pages 135–144, 1977. 201
- [Cachia 2003a] A. Cachia, J.-F. Mangin, D. Rivière, F. Kherif, N. Boddaert, A. Andrade, D. Papadopoulos-Orfanos, J.-B. Poline, I. Bloch, M. Zilbovicius, P. Sonigo, F. Brunelle et J. Régis. *A Primal Sketch of the Cortex Mean Curvature : a Morphogenesis Based Approach to Study the Variability of the Folding Patterns*. IEEE Trans. Med. Imag., vol. 22, no. 6, pages 754–765, 2003. 26, 28
- [Cachia 2003b] A. Cachia, J.-F. Mangin, D. Rivière, D. Papadopoulos-Orfanos, F. Kherif, I. Bloch et J. Régis. *A generic framework for parcellation of the cortical surface into gyri using geodesic Voronoi diagrams*. Medical Image Analysis, vol. 7, no. 4, pages 403–416, 2003. 148
- [Cachia 2008] A. Cachia, M.-L. Paillère-Martinot, A. Galinowski, D. Januel, R. De Beaurepaire, F. Bellivier, E. Artiges, T. Gallarda, J. Andoh, D. Bartrés-Faz, E. Duchesnay, D. Rivière, Y. Cointepas, M. Plaze, J.-F. Mangin et J.-L. Martinot. *Cortical folding abnormalities in schizophrenia patients with resistant auditory hallucinations*. Neuroimage, vol. 39, no. 3, pages 927–935, 2008. 148
- [Cachier 2001] Pascal Cachier, Jean-Francois Mangin, Xavier Pennec, Denis Rivière, Dimitri Papadopoulos-Orfanos, Jean Régis et Nicholas Ayache. *Multisubject Non-rigid Registration of Brain MRI Using Intensity and Geometric Features*. In MICCAI '01 : Proceedings of the 4th International Conference on Medical Image Computing and Computer-Assisted Intervention, pages 734–742, London, UK, 2001. Springer-Verlag. 99, 148
- [Cachier 2003] Pascal Cachier, Eric Bardenet, Didier Dormont, Xavier Pennec et Nicholas Ayache. *Iconic feature based nonrigid registration : the PASHA algorithm*. Comput. Vis. Image Underst., vol. 89, no. 2-3, pages 272–298, 2003. 67
- [Counce 1998] A. Counce et C. J. Taylor. *3D Point Distribution Models of the Cortical Sulci*. In ICCV '98 : Proceedings of the Sixth International Conference on Computer Vision, page 402, Washington, DC, USA, 1998. IEEE Computer Society. 26, 67, 123

- [Counce 2001] A. Counce et C. J. Taylor. *Building 3D sulcal models using local geometry*. Medical Image Analysis, vol. 5, pages 69–80, January 2001. 123
- [Chandler 1987] D. Chandler. Introduction to modern statistical mechanics. Oxford University Press, 1. édition, 1987. 113
- [Chang 1992] Ted Chang, Donald Richards et Christopher Bingham. *Approximating the matrix Fisher and Bingham distributions : applications to spherical regression and Procrustes analysis*. J. Multivar. Anal., vol. 41, no. 2, pages 314–337, 1992. 189
- [Chikuse 2003] Yasuko Chikuse. volume 174. 2003. 185
- [Chui 1999] Haili Chui, James Rambo, James S. Duncan, Robert Schultz et Anand Rangarajan. *Registration of Cortical Anatomical Structures via Robust 3D Point Matching*. In IPMI '99 : Proceedings of the 16th International Conference on Information Processing in Medical Imaging, pages 168–181, London, UK, 1999. Springer-Verlag. 67
- [Chui 2000] Haili Chui et Anand Rangarajan. *A Feature Registration Framework Using Mixture Models*. In MMBIA '00 : Proceedings of the IEEE Workshop on Mathematical Methods in Biomedical Image Analysis, pages 190–197, Washington, DC, USA, 2000. IEEE Computer Society. 67
- [Chui 2003] H. Chui, L. Win, R.A. Schultz, Duncan J.S. et A. Rangarajan. *A Unified Non-Rigid Feature Registration Method for Brain Mapping*. Medical Image Analysis, vol. 7, pages 112–130, 2003. 27
- [Cla 1992] *Bayesian methods for mapping disease risk*. Geographical and Environmental Epidemiology. Methods for Small Area Studies, pages 205–220, 1992. 121
- [Clouchoux 2005] C. Clouchoux, O. Coulon, D. Rivière, A. Cachia, J.-F. Mangin et J. Régis. *Anatomically constrained surface parameterization for cortical localization*. In MICCAI'05, pages 344–351, 2005. 148
- [Clouchoux 2006] C. Clouchoux, O. Coulon, J.-L. Anton, , J.-F. Mangin et J. Régis. *A new cortical surface parcellation model and its automatic implementation*. In Proc. 9th MICCAI, LNCS 4191, pages 193–200, Copenhagen,, Oct. 2006. Springer Verlag. 148
- [Collins 1994] Louis D. Collins, Peter Neelin, Terrence M. Peters et Alan C. Evans. *Automatic 3D Intersubject Registration of MR Volumetric Data in Standardized Talairach Space*. Journal of Computer Assisted Tomography, vol. 18, pages 192–205, March 1994. 70
- [Collins 1995] D.L. Collins, C.J. Holmes, T.M. Peters et A.C. Evans. *Automatic 3D Model-Based Neuroanatomical Segmentation*. Human Brain Mapping, vol. 3, no. 3, pages 190–208, 1995. 67, 70
- [Collins 1998] D. Louis Collins, Georges Le Goualher et Alan C. Evans. *Non-linear Cerebral Registration with Sulcal Constraints*. In MICCAI '98 : Proceedings of the First International Conference on Medical Image Computing and Computer-Assisted Intervention, pages 974–984, London, UK, 1998. Springer-Verlag. 27, 67
- [Comets 1992] Francis Comets. *On Consistency of a Class of Estimators for Exponential Families of Markov Random Fields on the Lattice*. The Annals of Statistics, vol. 20, no. 1, pages 455–468, 1992. 108
- [Corouge 2002] Isabelle Corouge et Christian Barillot. *Statistical Modeling of Pairs of Sulci in the Context of Neuroimaging Probabilistic Atlas*. In MICCAI '02 : Proceedings of the 5th International Conference on Medical Image Computing and Computer-Assisted Intervention-Part II, pages 655–662, London, UK, 2002. Springer-Verlag. 123
- [De Craene 2005] M. De Craene, A. du Bois d'Aische, B. Macq et S.K. Warfield. *Multi-subject variational registration for probabilistic unbiased atlas generation*. Image Processing, 2005. ICIP 2005. IEEE International Conference on, vol. 3, pages III-601–4, Sept. 2005. 68

- [Dempster 1977] A. P. Dempster, N. M. Laird et D. B. Rubin. *Maximum Likelihood from Incomplete Data via the EM Algorithm*. Journal of the Royal Statistical Society. Series B (Methodological), vol. 39, no. 1, pages 1–38, 1977. 36, 73, 127, 196
- [Descombes 1997] Xavier Descombes, Robin D. Morris, Josiane Zerubia et Marc Berthod. *Maximum Likelihood Estimation of Markov Random Field Parameters Using Markov Chain Monte Carlo Algorithms*. In EMMCVPR '97 : Proceedings of the First International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition, pages 133–148, London, UK, 1997. Springer-Verlag. 108
- [Do Carmo 1992] M. Do Carmo. *Riemannian geometry*. Birkhauser, 1992. 175
- [Downs 1972] D.T. Downs. *Orientation statistics*. Biometrika, vol. 59, no. 3, pages 665–676, 1972. 178, 185, 186
- [Dubois 2008] J. Dubois, M. Benders, A. Cachia, F. Lazeyras, R. Ha-Vinh Leuchter, S.V. Sizonenko, C. Borradori-Tolsa, J.-F. Mangin et P.S. Hüppi. *Mapping the early cortical folding process in the preterm newborn brain*. Cereb. Cortex, vol. 18, no. 6, pages 1444–54, 2008. 142, 157
- [Duchesnay 2007] E. Duchesnay, A. Cachia, A. Roche, D. Rivière, Y. Cointepas, D. Papadopoulos-Orfanos, M. Zilbovicius, J. Martinot, J. Régis et J. Mangin. *Classification based on cortical folding patterns*. IEEE Trans Med Imaging, vol. 26, no. 4, pages 553–65, 2007. 148
- [Durrleman 2007] Stanley Durrleman, Xavier Pennec, Alain Trounev et Nicholas Ayache. *Measuring Brain Variability via Sulcal Lines Registration : a Diffeomorphic Approach*. In Nicholas Ayache, Sébastien Ourselin et Anthony Maeder, editors, Proc. Medical Image Computing and Computer Assisted Intervention (MICCAI), volume 4791 of LNCS, pages 675–682, Brisbane, Australia, October 2007. Springer. PMID : 18051117. 26
- [Edelman 1998] Alan Edelman, Tom Ás, A. Arias, Steven et T. Smith. *The geometry of algorithms with orthogonality constraints*. SIAM J. Matrix Anal. Appl, vol. 20, pages 303–353, 1998. 175
- [Essen 1997] D.C.V. Essen. *A tension-based theory of morphogenesis and compact wiring in the central nervous system*. nat, vol. 385, pages 313–318, jan 1997. 4, 25, 151
- [Evans 1993] A.C. Evans, D.L. Collins, S.R. Mills, E.D. Brown, R.L. Kelly et T.M. Peters. *3D statistical neuroanatomical models from 305 MRI volumes*. IEEE Nuclear Science Symposium and Medical Imaging Conference, vol. 3, pages 1813–1817, 1993. 70
- [Evans 1994] A.C. Evans et et al. *Three-dimensional correlative imaging : Applications in human brain mapping*, chapter 14, pages 145–161. Functional Neuroimaging, 1994. 42
- [Fielding 2000] Mark Fielding. *Simulated Annealing With An Optimal Fixed Temperature*. SIAM J. on Optimization, vol. 11, no. 2, pages 289–307, 2000. 109
- [Fillard 2007a] Pierre Fillard, Vincent Arsigny, Xavier Pennec, Kiralee M. Hayashi, Paul M. Thompson et Nicholas Ayache. *Measuring Brain Variability by Extrapolating Sparse Tensor Fields Measured on Sulcal Lines*. Neuroimage, vol. 34, no. 2, pages 639–650, January 2007. Also as INRIA Research Report 5887, April 2006. PMID : 17113311. 26
- [Fillard 2007b] Pierre Fillard, Xavier Pennec, Paul Thompson et Nicholas Ayache. *Evaluating Brain Anatomical Correlations via Canonical Correlation Analysis of Sulcal Lines*. Research Report RR-6241, INRIA, 2007. 123, 142
- [Fischl 2004] B. Fischl, A. van der Kouwe, C. Destrieux, E. Halgren, F. Ségonne, D. H. Salat, E. Busa, L. J. Seidman, J. Goldstein, D. Kennedy, V. Caviness, N. Makris, B. Rosen et A. M. Dale. *Automatically parcellating the human cerebral cortex*. Cereb Cortex, vol. 14, no. 1, pages 11–22, January 2004. 25, 67, 115, 116

- [Fletcher 2003] P.T. Fletcher, C. Lu et S. Joshi. *Statistics of shape via principal component analysis on Lie groups*. In Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR'03), Los Alamitos, CA, pages 95–101, 2003. 84
- [Fletcher 2007] P. Thomas Fletcher et Sarang Joshi. *Riemannian geometry for the statistical analysis of diffusion tensor data*. Signal Process., vol. 87, no. 2, pages 250–262, 2007. 175
- [Fodor 2002] Imola Fodor. *A Survey of Dimension Reduction Techniques*. Technical report, 2002. 15
- [Forgy 1965] E. Forgy. *Cluster analysis of multivariate data : Efficiency vs. interpretability of classifications*. Biometrics, vol. 21, page 768, 1965. 173
- [Fouquet 2009] Marine Fouquet, Matthieu Perrot, Edouard Duchesnay, Brigitte Landeau, Florence Mezng, Vincent de la Sayette, Fausto Viader, Jean-Claude Baron, Béatrice Desgranges, Francis Eustache, Jean-François Mangin et Gaël Chetelat. *How the morphology of brain sulci may help in detecting early Alzheimer's disease*. Alzheimer's and Dementia, vol. 5, no. 4, Supplement 1, pages P32 – P33, 2009. Alzheimer's Association International Conference on Alzheimer's Disease, Alzheimer's Association International Conference on Alzheimer's Disease. 148
- [Fréchet 1948] M. Fréchet. *Les éléments aléatoires de nature quelconque dans un espace distancié*. Ann. Inst. H. Poincaré, vol. 10, page 215–310, 1948. 175
- [Geman 1984] Stuart Geman et Donald Geman. *Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 6, no. 6, pages 721–741, November 1984. 105
- [Gholipour 2007] Ali Gholipour, Nasser D. Kehtarnavaz, Richard Briggs, Michael Devous et Kaundinya S. Gopinath. *Brain Functional Localization : A Survey of Image Registration Techniques*. IEEE Trans. Med. Imaging, vol. 26, no. 4, pages 427–451, 2007. 67
- [Gidas 1987] B. Gidas. *Consistency of maximum likelihood and pseudo-likelihood estimators for Gibbs distributions*. In Workshop Stochastic Differential Systems with Applications, IMA, Univ. Minnesota. New York : Springer-Verlag, 1987. 108
- [Girosi 1995] Federico Girosi, Michael Jones et Tomaso Poggio. *Regularization theory and neural networks architectures*. Neural Comput., vol. 7, no. 2, pages 219–269, March 1995. 41
- [Glaunes 2004] Joan Glaunes, Alain Trounev et Laurent Younes. *Diffeomorphic Matching of Distributions : A New Approach for Unlabelled Point-Sets and Sub-Manifolds Matching*. cvpr, vol. 02, pages 712–718, 2004. 62, 203
- [Glaunès 2006] Joan Glaunès et Sarang Joshi. *Template estimation from unlabeled point set data and surfaces for Computational Anatomy*. In X. Pennec et S. Joshi, editors, Proc. of the International Workshop on the Mathematical Foundations of Computational Anatomy (MFCA-2006), pages 29–39, 1st of October 2006. 62
- [Goualher 1997] Georges Le Goualher, Christian Barillot et Yves Bizais. *Modeling Cortical Sulci with Active Ribbons*. IJPRAI, vol. 11, no. 8, pages 1295–1315, 1997. 27, 55
- [Goualher 1998] G. Le Goualher, D. L. Collins, C. Barillot et A. C. Evans. *Automatic Identification of Cortical Sulci Using a 3D Probabilistic Atlas*. In MICCAI '98 : Proceedings of the First International Conference on Medical Image Computing and Computer-Assisted Intervention, volume 1496, pages 509–518, 1998. 27, 42, 55, 67, 120
- [Gower 1975] J. Gower. *Generalized procrustes analysis*. Psychometrika, vol. 40, no. 1, pages 33–51, March 1975. 26, 169
- [Hammersley 1971] J. M. Hammersley et P. Clifford. *Markov field on finite graphs and lattices*. 1971. 105

- [Hastie 2001] T. Hastie, R. Tibshirani et J. Friedmann. *The element of statistical learning*. Springer, 2001. 18
- [Hill 1977] G.W. Hill. *New approximations to the von Mises distribution*. *Biometrika*, vol. 64, page 655, 1977. 181
- [Horn 1987] Berthold K. P. Horn. *Closed-form solution of absolute orientation using unit quaternions*. *Journal of the Optical Society of America. A*, vol. 4, no. 4, pages 629–642, Apr 1987. 172
- [Huang 2002] Fuchun Huang et Yosihiko Ogata. *Generalized Pseudo-Likelihood Estimates for Markov Random Fields on Lattice*. *Annals of the Institute of Statistical Mathematics*, vol. 54, no. 1, pages 1–18, March 2002. 108
- [Hurdal 2008] Monica K. Hurdal, Juan B. Gutierrez, Christian Laing et Deborah A. Smith. *Shape analysis for automated sulcal classification and parcellation of MRI data*. *J. Comb. Optim.*, vol. 15, no. 3, pages 257–275, 2008. 26, 39
- [Hussin 2008] A. G. Hussin et I. B. Mohamed. *Efficient approximation for the von Mises concentration parameter*. *Asian Journal of Mathematics and Statistics*, vol. 1, no. 3, pages 165–169, 2008. 181
- [Im 2009] K. Im, H.J. Jo, J.-F. Mangin, A.C. Evans, S.I. Kim et J.-M. Lee. *Spatial Distribution of Deep Sulcal Landmarks and Hemispherical Asymmetry on the Cortical Surface*. *Cerebral Cortex*, 2009. 26
- [Jammalamadaka 2001] S.Rao Jammalamadaka et A Sengupta. *Topics in circular statistics (series on multivariate analysis)*. 2001. 178
- [Johnson 1966] R.M. Johnson. *The minimal transformation to orthonormality*. *Psychometrika*, vol. 31, pages 61–66, 1966. 85
- [Joshi 2004] S. Joshi, Brad Davis, B Matthieu Jomier et Guido Gerig B. *Unbiased diffeomorphic atlas construction for computational anatomy*. *Neuroimage*, vol. 23, pages 151–160, 2004. 68, 71
- [Joshi 2005] Anand A. Joshi, David W. Shattuck, Paul M. Thompson et Richard M. Leahy. *A Framework for Registration, Statistical Characterization and Classification of Cortically Constrained Functional Imaging Data*. In *IPMI*, pages 186–196, 2005. 67
- [Juan 2004] Alfons Juan, José García-Hernández et Enrique Vidal. *EM Initialisation for Bernoulli Mixture Learning*. In *SSPR/SPR*, pages 635–643, 2004. 202
- [Jupp 1979] P.E. Jupp et K.V. Mardia. *Maximum Likelihood Estimators for the Matrix Von Mises-Fisher and Bingham Distributions*. *Ann. Statist.*, vol. 7, no. 3, pages 599–606, 1979. 185, 188, 189
- [Jupp 1987] P.E. Jupp, B.D. Spurr, G.J. Nichols et J.P.P. Hirst. *Statistical estimation of the apex of a sediment distribution system from paleocurrent data*. *Mathematical Geology*, vol. 19, no. 4, pages 319–333, 1987. 178
- [Jupp 1989] P.E. Jupp et K.V. Mardia. *A unified view of the theory of directional statistics, 1975-1988*. *Int. Statistical Review*, vol. 57, no. 3, pages 261–294, 1989. 178
- [Jupp 1995] P.E. Jupp. *Some applications of directional statistics to astronomy*. *Multivariate statistics and matrices in statistics*, vol. 3, pages 123–133, 1995. 178
- [Kabani 1998] N. J. Kabani, D. MacDonald, C. J. Holmes et A. C. Evans. *3D anatomical atlas of the human brain*. *NeuroImage*, vol. 7, no. 4, 1998. 25
- [Kao 2007] Chiu-Yen Kao, Michael Hofer, Guillermo Sapiro, Josh Stern, Kelly Rehm et David A. Rottenberg. *A Geometric Method for Automatic Extraction of Sulcal Fundi*. *IEEE Trans. Medical Imaging*, vol. 26, no. 4, pages 530–540, 2007. 26

- [Keller 2009] M. Keller, M. Lavielle, M. Perrot et A. Roche. *Anatomically Informed Bayesian Model Selection for fMRI Group Data Analysis (accepted)*. In 12th MICCAI, London, U.K., 2009. 148
- [Kent 1982] John T. Kent. *The Fisher-Bingham Distribution on the Sphere*. Journal of Royal Statistical Society, vol. 44, pages 71–80, 1982. 182, 183
- [Khatri 1977] C.G Khatri et K.V. Mardia. *The von Mises-Fisher matrix distribution in orientation statistics*. J. Roy. Statist. Soc., vol. Ser. B 39, pages 95–106, 1977. 178, 185
- [Kirkpatrick 1983] S. Kirkpatrick, C. D. Gelatt et M. P. Vecchi. *Optimization by Simulated Annealing*. Science, Number 4598, 13 May 1983, vol. 220, 4598, pages 671–680, 1983. 110
- [Klein 2005] A. Klein et J. Hirsch. *Mindboggle : a scatterbrained approach to automate brain labeling*. Neuroimage, vol. 24, no. 2, pages 261–280, January 2005. 27
- [Klein 2009] Arno Klein, Jesper Andersson, Babak Ardekani, John Ashburner, Brian Avants, Ming-Chang Chiang, Gary Christensen, D. Louis Collins, Pierre Hellier, Joo Hyun Song, Mark Jenkinson, Claude Lepage, Daniel Rueckert, Paul Thompson, Tom Vercauteren, Roger Woods, John Mann et Ramin Parsey. *Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration*. NeuroImage, vol. 46, no. 3, pages 786–802, 01 2009. 97, 98
- [Klingenberg 1982] W. Klingenberg. *Riemannian geometry*. Walter de Gruyter, 1982. 175
- [Kobashi 2007] S. Kobashi, S. Sueyoshi, K. Kondo et Y. Hata. *Automated Gyruis Labeling Using Knowledge-based Fuzzy Inference Systems*. System of Systems Engineering, 2007. SoSE '07. IEEE International Conference on, pages 1–6, 16-18 April 2007. 26
- [Koev 2006] Plamen Koev et Alan Edelman. *The efficient evaluation of the hypergeometric function of a matrix argument*. Math. Comp., vol. 75, pages 833–846, 2006. 185, 189
- [Kolmogorov 2002] Vladimir Kolmogorov et Ramin Zabih. *What Energy Functions Can Be Minimized via Graph Cuts ?* In ECCV '02 : Proceedings of the 7th European Conference on Computer Vision-Part III, pages 65–81, London, UK, 2002. Springer-Verlag. 112
- [Krieger Lassen 1994] N.C. Krieger Lassen, D. Juul Jensen et K. Conradsen. *On the statistical analysis of orientation data*. Acta Crystallogr, vol. A50, pages 741–748, 1994. 178
- [Kumar 2006] S. Kumar et M. Hebert. *Discriminative Random Fields*. International Journal of Computer Vision, vol. 68, no. 2, pages 179–201, 2006. 107
- [Lafferty 2001] John Lafferty, Andrew McCallum et Fernando Pereira. *Conditional Random Fields : Probabilistic Models for Segmenting and Labeling Sequence Data*, 2001. 107
- [Le Cun 1988] Y. Le Cun. *A theoretical framework for back-propagation*. In Proceedings of the 1988 Connectionist Models Summer School, pages 21–28. Citeseer, 1988. 13
- [Lefèvre 2009] J. Lefèvre, F. Leroy, S. Khan, J. Dubois, P.S. Huppi, S. Baillet et J.-F. Mangin. *Identification of Growth Seeds in the Neonate Brain through Surfacic Helmholtz Decomposition*. In Proc. 21th IPMI, LNCS-5636, pages 176–187, Williamsburg, VA, July 2009. Springer. 150
- [León 2006] Carlos A. León, Jean-Claude Massé et Louis-Paul Rivest. *A statistical model for random rotations*. J. Multivar. Anal., vol. 97, no. 2, pages 412–430, 2006. 178, 185
- [Li 2001] Stan Z. Li. *Markov random field modeling in image analysis*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2001. 105
- [Lin 2003] H. Lin, C. Lin et R. Weng. *A note on Platt's probabilistic outputs for support vector machines*, 2003. 16
- [Lohmann 1998] G. Lohmann. *Extracting line representations of sulcal and gyral patterns in MR images of the human brain*. IEEE Trans Med Imaging, vol. 17, no. 6, 1998. 26

- [Lohmann 2000] G. Lohmann et D. Y. von Cramon. *Automatic labelling of the human cortical surface using sulcal basins*. Medical Image Analysis, vol. 4, no. 3, pages 179–188, 2000. 26, 67, 98, 123, 157, 158
- [Lohmann 2008] G. Lohmann, D.Y. von Cramon et A.C.F. Colchester. *Deep Sulcal Landmarks Provide an Organizing Framework for Human Cortical Folding*. Cerebral Cortex, vol. 18, no. 6, pages 1415–1420, 2008. 26, 55, 99
- [Lui 2006] Lok Ming Lui, Yalin Wang, Tony F. Chan et Paul M. Thompson. *Automatic Landmark Tracking and its Application to the Optimization of Brain Conformal Mapping*. In CVPR '06 : Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pages 1784–1792, Washington, DC, USA, 2006. IEEE Computer Society. 26
- [Lui 2007] Lok Ming Lui, Yalin Wang, Tony F. Chan et Paul Thompson. *Landmark constrained genus zero surface conformal mapping and its application to brain mapping research*. Appl. Numer. Math., vol. 57, no. 5-7, pages 847–858, 2007. 67
- [Lyttelton 2007] O. Lyttelton, M. Boucher, S. Robbins et A. Evans. *An unbiased iterative group registration template for cortical surface analysis*. Neuroimage, vol. 34, no. 4, pages 1535–1544, 2007. 68
- [M. Radoš 2006] I. Kostović M. Radoš M. Judaš. *In vitro MRI of brain development*. European Journal of Radiology, vol. 57, no. 2, pages 187–198, 2006. 28
- [MacQueen 1967] J. B. MacQueen. *Some Methods for Classification and Analysis of MultiVariate Observations*. In L. M. Le Cam et J. Neyman, editors, Proc. of the fifth Berkeley Symposium on Mathematical Statistics and Probability, volume 1, pages 281–297. University of California Press, 1967. 173
- [Madsen 2005] R. E. Madsen, D. Kauchak et C. Elkan. *Approximating The Dirichlet Compound Multinomial Distribution*. In Neural Information Processing Systems, dec 2005. 49
- [Maintz 1998] J. B. Maintz et M. A. Viergever. *A survey of medical image registration*. Med Image Anal, vol. 2, no. 1, pages 1–36, March 1998. 67
- [Mangin 1995a] J.-F. Mangin. *Mise en correspondance d'images médicales 3D multi-modalités multi-individus pour la corrélation anatomo-fonctionnelle cérébrale*. PhD thesis, ENST Paris, Paris , France, Mar. 1995. 4, 27
- [Mangin 1995b] J.-F. Mangin, V. Frouin, I. Bloch, J. Régis et J. Lopez-Krahe. *From 3D Magnetic Resonance Images to Structural Representations of the Cortex Topography Using Topology Preserving Deformations*. Journal of Mathematical Imaging and Vision, vol. 5, pages 297–318, 1995. 28
- [Mangin 1995c] J.-F. Mangin, F. Tupin, V. Frouin, I. Bloch, R. Rougetet, J. Régis et J. Lopez-Krahe. *Deformable topological models for segmentation of 3D medical Images*. In Y. Bizais, C. Barillot et R. Di Paola, editors, Proc. 14th IPMI, Computational Imaging and Vision, pages 153–164, Ile de Berder, Brest, France, June 1995. Kluwer Academic Publishers. 29
- [Mangin 1995d] Jean-François Mangin, Vincent Frouin, Isabelle Bloch, Jean Régis et Jaime López-Krahe. *From 3D magnetic resonance images to structural representations of the cortex topography using topology preserving deformations*. J. Math. Imaging Vis., vol. 5, no. 4, pages 297–318, 1995. 29, 121
- [Mangin 1996] J-F Mangin, J. Regis et V. Frouin. *Shape Bottlenecks and Conservative Flow Systems*. Mathematical Methods in Biomedical Image Analysis, IEEE Workshop on, vol. 0, page 0319, 1996. 29
- [Mangin 1998] J.-F. Mangin, O. Coulon et V. Frouin. *Robust Brain Segmentation Using Histogram Scale-Space Analysis and Mathematical Morphology*. Lecture Notes in Computer Science, vol. 1496, pages 1230–??, 1998. 29

- [Mangin 2003a] J.-F. Mangin, D. Rivière, A. Cachia, D. Papadopoulos-Orfanos, D. L. Collins, A. C. Evans et J. Régis. *Object-based strategy for morphometry of the cerebral cortex*. In IPMI, Ambleside, UK, LNCS 2732, Springer Verlag, pages 160–171, 2003. 28, 148
- [Mangin 2003b] Jean-Francois Mangin, Fabrice Poupon, Denis Rivière, Arnaud Cachia, D. Louis Collins, Alan C. Evans et Jean Régis. *3D Moment Invariant Based Morphometry*. In Randy E. Ellis et Terry M. Peters, editors, MICCAI (2), volume 2879 of *Lecture Notes in Computer Science*, pages 505–512. Springer, 2003. 11, 39, 117
- [Mangin 2005] J.-F. Mangin. *Une vision structurelle de l'analyse des images cérébrales*. Habilitation thesis, Université Paris-Sud, Orsay, France, 2005. 4
- [Mani 2008] Meena Mani, Anuj Srivastava et Christian Barillot. *The Labeling of Cortical Sulci using Multidimensional Scaling*. The Midas Journal, 2008. <http://hdl.handle.net/10380/1502>. 26
- [Mardia 1975] K.V. Mardia. *Statistics of directional data (with discussion)*. J. Roy. Statist. Soc., vol. Ser. B 37, pages 349–393, 1975. 178
- [Mardia 1999] K. V. Mardia. *Directional statistics and shape analysis*. J. Appl. Statist, vol. 26, pages 949–957, 1999. 178
- [Mardia 2000] Kanti V. Mardia et Peter E. Jupp. *Directional statistics*. Wiley, 2 sub édition, 2000. 81, 178, 180, 181
- [Marroquin 1987] J. Marroquin, S. Mitter et Poggio T. *Probabilistic Solution of Ill-Posed Problems in Computational Vision*. Technical report, Cambridge, MA, USA, 1987. 109
- [Mazziotta 1995] John C. Mazziotta, Arthur W. Toga, Alan Evans, Peter Fox et Jack Lancaster. *A Probabilistic Atlas of the Human Brain : Theory and Rationale for Its Development : The International Consortium for Brain Mapping (ICBM)*. NeuroImage, vol. 2, no. 2, Part 1, pages 89 – 101, 1995. 68
- [Mazziotta 2001] J. Mazziotta, A. Toga, A. Evans, P. Fox, J. Lancaster, K. Zilles, R. Woods, T. Paus, G. Simpson, B. Pike, C. Holmes, L. Collins, P. Thompson, D. MacDonald, M. Iacoboni, T. Schormann, K. Amunts, N. Palomero-Gallagher, S. Geyer, L. Parsons, K. Narr, N. Kabani, G. Le Goualher, D. Boomsma, T. Cannon, R. Kawashima et B. Mazoyer. *A probabilistic atlas and reference system for the human brain : International Consortium for Brain Mapping (ICBM)*. Philos Trans R Soc Lond B Biol Sci, vol. 356, no. 1412, pages 1293–1322, August 2001. 70
- [Mccallum 2003] Andrew Mccallum. *Efficiently inducing features of conditional random fields*. In Nineteenth Conference on Uncertainty in Artificial Intelligence (UAI03), 2003. 107
- [Minka 2003] Thomas P. Minka. *Estimating a Dirichlet distribution*. unpublished paper, 2003. 48, 49, 191
- [Murphy 1999] K. Murphy, Y. Weiss et M. I. Jordan. *Loopy belief propagation for approximate inference : An empirical study*. In Proceedings of Uncertainty in AI, volume 9, 1999. 111
- [Murray 1994] Richard M. Murray, Zexiang Li et Shankar S. Sastry. *A mathematical introduction to robotic manipulation*. CRC, March 1994. 84
- [Murray 2004] Iain Murray et Zoubin Ghahramani. *Bayesian learning in undirected graphical models : approximate MCMC algorithms*. In AUAI '04 : Proceedings of the 20th conference on Uncertainty in artificial intelligence, pages 392–399, Arlington, Virginia, United States, 2004. AUAI Press. 108
- [Møller 2004] Jesper Møller, A. N. Pettitt, K. K. Berthelsen, R. W. Reeves, An Markov, Monte Carlo et J. Møller. *An efficient Markov chain Monte Carlo method for distributions with intractable normalising constants*. Biometrika, vol. 93, 2004. 108

- [Neal 1993] Radford M. Neal. *Probabilistic Inference Using Markov Chain Monte Carlo Methods*. Technical report, Department of Computer Science, University of Toronto, 1993. 143
- [Nwogu 2008] I. Nwogu et J.J. Corso. *Labeling Irregular Graphs with Belief Propagation*. In IW-CIA08, pages xx–yy, 2008. 121
- [Ono 1990] Michio Ono, Stefan Kubik et Chad D. Abarnathey. Atlas of the cerebral sulci. Thieme Medical Publishers, 1990. 157
- [Onstott 1980] T.C. Onstott. *Application of the Bingham distribution function in paleomagnetic studies*. Journal of Geophysical Research, vol. 85, pages 1500–1510, 1980. 184
- [Ourselin 2000] S. Ourselin, A. Roche, S. Prima et N. Ayache. *Block matching : a general framework to improve robustness of rigid registration of medical images*. In S.L. Delp, A.M. DiGioia et B. Jaramaz, editors, 3rd International Conference on Medical Image Computing and Computer-Assisted Intervention, MICCAI'2000, volume 1935 of *Lecture Notes in Computer Science*, pages 557–566, Pittsburgh, USA, October 2000. Springer. 70
- [Parisi 1988] G. Parisi. Statistical field theory. Addison Wesley, 1988. 113
- [Parzen 1962] Emanuel Parzen. *On Estimation of a Probability Density Function and Mode*. The Annals of Mathematical Statistics, vol. 33, no. 3, pages 1065–1076, 1962. 43
- [Pearl 1982] J. Pearl. *Reverend Bayes on inference engines : A distributed hierarchical approach*. In Proceedings of the American Association of Artificial Intelligence National Conference on AI, pages 133–136, Pittsburgh, PA, 1982. 111
- [Penneç 1997] Xavier Penneç et Jean-Philippe Thirion. *A Framework for Uncertainty and Validation of 3D Registration Methods based on Points and Frames*. Int. Journal of Computer Vision, vol. 25, no. 3, pages 203–229, December 1997. 80, 82
- [Penneç 1999] X. Penneç. *Probabilities and statistics on Riemannian manifolds : Basic tools for geometric measurements*. In IEEE Workshop on Nonlinear Signal and Image Processing, volume 4. Citeseer, 1999. 175
- [Penneç 2006] X. Penneç, P. Fillard et N. Ayache. *A Riemannian framework for tensor computing*. International Journal of Computer Vision, vol. 66, no. 1, pages 41–66, 2006. 175
- [Perrot 2005] Matthieu Perrot. *Utilisation des Support Vector Machines pour l'identification des sillons du cortex*, septembre 2005. 7, 13, 16
- [Perrot 2008] Matthieu Perrot, Denis Rivière et Jean-François Mangin. *Identifying cortical sulci from localization, shape and local organization*. In ISBI, pages 420–423, 2008. 27, 51, 52, 57, 58, 67, 91, 100, 110, 132, 157
- [Perrot 2009] M. Perrot, D. Rivière, A. Tucholka et J.-F. Mangin. *Joint Bayesian Cortical Sulci Recognition and Spatial Normalization*. In 21st. IPMI, pages 176–187, Williamsburg, U.S.A, july 2009. 27, 148
- [Pieczynski 2000] Wojciech Pieczynski et Abdel nasser Tebbache. *Pairwise Markov random fields and segmentation of textured images*. In Machine Graphics and Vision, pages 705–718, 2000. 106
- [Pieczynski 2002] Wojciech Pieczynski, Dalila Benboudjema et Pierre Lanchantin. *Statistical image segmentation using Triplet Markov fields*. In SPIE's International Symposium on Remote Sensing, pages 22–27. SPIE, 2002. 107
- [Platt 2000] J. Platt. *Probabilistic outputs for support vector machines and comparison to regularized likelihood methods*. pages 61–74, 2000. 16
- [Pohl 2006] Kilian M. Pohl, John Fisher, Grimson, Ron Kikinis et William M. Wells. *A Bayesian model for joint segmentation and registration*. NeuroImage, vol. 31, no. 1, pages 228–239, May 2006. 68

- [Poupon 2006] C. Poupon, F. Poupon, L. Allirol et J.-F. Mangin. *A database dedicated to anatomic-functional study of human brain connectivity*. In 12th HBM Neuroimage, number 646, Florence, Italie, 2006. 158
- [Powell 1964] M. J. D. Powell. *An efficient method for finding the minimum of a function of several variables without calculating derivatives*. The Computer Journal, vol. 7, no. 2, pages 155–162, February 1964. 82, 185, 189
- [Powell 1992] M. J. D. Powell. *The Theory of Radial Basis Function Approximation in 1990*. In Will A. Light, editor, *Advances in Numerical Analysis Vol. 2 : Wavelets, Subdivision Algorithms, and Radial Basis Functions*, chapter Kap. 3, pages 105–210. Clarendon Press, 1992. NAM-Bibliothek N/609. 194
- [Prentice 1986] M.J. Prentice. *Orientation statistics without parametric assumptions*. J. Roy. Stat. Soc., vol. Ser B. 48, no. 2, pages 214–222, 1986. 178, 189
- [Rathi 2009] Y. Rathi, O. Michailovich, M.E. Shenton et Bouix S. *Directional functions for orientation distribution estimation*. Med Image Anal, vol. 13, no. 3, pages 432–444, 06 2009. 178
- [Régis 2005] J. Régis, J.-F. Mangin, T. Ochiai, V. Frouin, D. Rivière, A. Cachia, M. Tamura et Y. Samson. *"Sulcal root" generic model : a hypothesis to overcome the variability of the human cortex folding patterns*. Neurol Med Chir (Tokyo), vol. 45, pages 1–17, 2005. 26, 29, 99, 148, 157, 158
- [Rettmann 2005] M.E. Rettmann, D. Tosun, X. Tao, S.M. Resnick et J.L. Prince. *Program for Assisted Labeling of Sulcal Regions (PALS) : description and reliability*. Neuroimage, vol. 24, no. 2, pages 398–416, 2005. 4, 27
- [Rivest 2008] Louis-Paul Rivest, Sophie Baillargeon et Michael Pierrynowski. *A Directional Model for the Estimation of the Rotation Axes of the Ankle Joint*. Journal of the American Statistical Association, vol. 103, no. 483, pages 1060–1069, September 2008. 178
- [Rivière 2002] D. Rivière, J.-F. Mangin, D. Papadopoulos-Orfanos, Martinez J.-M., V. Frouin et J. Régis. *Automatic recognition of cortical sulci of the Human Brain using a congregation of neural networks*. Medical Image Analysis, vol. 6, no. 2, pages 77–92, 2002. 5, 13, 16, 27, 28, 51, 53, 57, 59, 61, 67, 91, 100, 104, 115, 117, 120, 132, 147, 148, 150
- [Rivière 2000] D. Rivière. *Apprentissage de la variabilité inter-individuelle de l'anatomie corticale cérébrale pour la reconnaissance automatique des sillons*. PhD thesis, Université d'évry Val d'Essonne, évry , France, Sep. 2000. 4, 7, 10, 11, 13, 28, 29, 104
- [Rivière 2009] D. Rivière, D. Geffroy, I. Denghien, N. Souedet et Y. Cointepas. *BrainVISA : an extensible software environment for sharing multimodal neuroimaging data and processing tools*. In Hum. Brain Mapp., 2009. 4, 28, 32, 38, 69, 148, 158, 159, 205, 206
- [Royackkers 1999] Nicolas Royackkers, Michel Desvignes, Houssam Fawal et Marinette Revenu. *Detection and Statistical Analysis of Human Cortical Sulci*. NeuroImage, vol. 10, pages 10–625, 1999. 26, 123
- [Schönemann 1966] P.H. Schönemann. *A generalized solution of the orthogonal Procrustes problem*. Psychometrika, vol. 31, pages 1–10, 1966. 85
- [Schwartzman 2005] Armin Schwartzman, Robert F Dougherty et Jonathan E Taylor. *Cross-subject comparison of principal diffusion direction maps*. Magnetic Resonance in Medicine, vol. 53, no. 6, pages 1423–1431, Jun 2005. 184
- [Seunarine 2007] K.K. Seunarine, P.A. Cook, M.G. Hall, K.V. Embleton, G.J.M. Parker et D.C. Alexander. *Exploiting peak anisotropy for tracking through complex structures*. pages 1–8, 2007. 178

- [Shattuck 2008] D. W. Shattuck, M. Mirza, V. Adisetiyo, C. Hojatkashani, G. Salamon, K. L. Narr, R. A. Poldrack, R. M. Bilder et A. W. Toga. *Construction of a 3D probabilistic atlas of human cortical structures*. *Neuroimage*, vol. 39, no. 3, pages 1064–1080, February 2008. 68
- [Shattuck 2009] DW Shattuck, AA Joshi, D Pantazis, E Kan, RA Dutton, ER Sowell, PM Thompson, AW Toga et RM Leahy. *Semi-automated method for delineation of landmarks on models of the cerebral cortex*. *Journal of Neuroscience Methods*, vol. 178, no. 2, pages 385–392, 2009. 4, 26
- [Shen 2001] Dinggang Shen et Christos Davatzikos. *HAMMER : Hierarchical Attribute Matching Mechanism for Elastic Registration*. *Mathematical Methods in Biomedical Image Analysis, IEEE Workshop on*, vol. 0, page 29, 2001. 25
- [Shi 2007] Yonggang Shi, Zhuowen Tu, Allan Reiss, Rebecca Dutton, Agatha Lee, Albert Galaburda, Ivo Dinov, Paul Thompson et Arthur Toga. *Joint Sulci Detection Using Graphical Models and Boosted Priors*. pages 98–109. 2007. 31, 67, 115
- [Shi 2008] Y.G. Shi, P.M. Thompson, I. Dinov et A.W. Toga. *Hamilton-Jacobi Skeleton on Cortical Surfaces*. *MedImg*, vol. 27, no. 5, pages 664–673, May 2008. 26, 31, 115
- [Shi 2009] Yonggang Shi, Zhuowen Tu, A.L. Reiss, R.A. Dutton, A.D. Lee, A.M. Galaburda, I. Dinov, P.M. Thompson et A.W. Toga. *Joint Sulcal Detection on Cortical Surfaces With Graphical Models and Boosted Priors*. *Medical Imaging, IEEE Transactions on*, vol. 28, no. 3, pages 361–373, March 2009. 31, 67, 115, 116, 117, 133
- [Stacy 1962] E.W. Stacy. *A generalization of the Gamma Distribution*. *Ann. Math. Statist.*, vol. 33, pages 1187–1192, 1962. 126
- [Sun 2007] Zhong Yi Sun, Denis Rivière, Fabrice Poupon, Jean Régis et Jean-Francois Mangin. *Automatic Inference of Sulcus Patterns Using 3D Moment Invariants*. In Nicholas Ayache, Sébastien Ourselin et Anthony J. Maeder, editors, MICCAI (1), volume 4791 of *Lecture Notes in Computer Science*, pages 515–522. Springer, 2007. 11, 39, 117
- [Sun 2009] Zhong Yi Sun, Matthieu Perrot, Alan Tucholka, Denis Rivière et Jean-François Mangin. *Constructing a Dictionary of Human Brain Folding Patterns*. In MICCAI, volume LNCS 5762, pages 117–124, 2009. 103, 148, 159
- [Sutton 2005] Charles Sutton et Andrew Mccallum. *Piecewise Training for Undirected Models*. In *Neural Information Processing Systems*, 2005. 108
- [Sutton 2007] Charles Sutton et Andrew McCallum. *Piecewise pseudolikelihood for efficient training of conditional random fields*. In *ICML '07 : Proceedings of the 24th international conference on Machine learning*, pages 863–870, New York, NY, USA, 2007. ACM. 108
- [Szeliski 2008] Richard Szeliski, Ramin Zabih, Daniel Scharstein, Olga Veksler, Vladimir Kolmogorov, Aseem Agarwala, Marshall Tappen et Carsten Rother. *A Comparative Study of Energy Minimization Methods for Markov Random Fields with Smoothness-Based Priors*. *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 6, pages 1068–1080, 2008. 109
- [Talairach 1988] Jean Talairach et Pierre Tournoux. *Co-planar stereotaxic atlas of the human brain*. 1988. 31, 69
- [Tanabe 2007] Akihiro Tanabe, Kenji Fukumizu, Shigeyuki Oba, Takashi Takenouchi et Shin Ishii. *Parameter estimation for von Mises-Fisher distributions*. *Comput. Stat.*, vol. 22, no. 1, pages 145–157, 2007. 182
- [Tanaka 1999] H Tanaka. *Circular asymmetry of the paleomagnetic directions observed at low latitude volcanic sites*. *Earth, Planets Space*, vol. 51, no. 12, pages 1279–1286, 1999. 184
- [Tao 2001] Xiaodong Tao, Xiao Han, Maryam E. Rettmann, Jerry L. Prince et Christos Davatzikos. *Statistical Study on Cortical Sulci of Human Brains*. In *IPMI '01 : Proceedings of the 17th*

- International Conference on Information Processing in Medical Imaging, pages 475–487, London, UK, 2001. Springer-Verlag. 123
- [Tao 2002] X. Tao, J. L. Prince et C. Davatzikos. *Using a Statistical Shape Model to Extract Sulcal Curves on the Outer Cortex of the Human Brain*. IEEE Trans. Med. Imaging, vol. 21, no. 5, pages 513–524, 2002. 26
- [Tarsitano 2003] A. Tarsitano. *Mahalanobis metrics for k-means algorithms*. Convegno intermedio SIS, pages 9–11, 2003. 174
- [Taylor 1992] C. J. Taylor, D. H. Cooper et J. Graham. *Training models of shape from sets of examples*. In In Proc. British Machine Vision Conference, pages 9–18, 1992. 26
- [Taylor 1994] Camillo Taylor, , Camillo J. Taylor, Camillo J. Taylor, David J. Kriegman et David J. Kriegman. *Minimization on the Lie Group $SO(3)$ and Related Manifolds*. Technical report, Yale Univ, 1994. 84
- [Thompson 1997] P. M. Thompson et A. W. Toga. *Detection, visualization and animation of abnormal anatomic structure with a deformable probabilistic brain atlas based on random vector field transformations*. Med Image Anal, vol. 1, pages 271–294, Sep 1997. 26, 68
- [Thompson 2000] P. M. Thompson, R. P. Woods, M. S. Mega et A. W. Toga. *Mathematical/computational challenges in creating deformable and probabilistic atlases of the human brain*. Hum Brain Mapp, vol. 9, no. 2, pages 81–92, February 2000. 67
- [Tosun 2003] Duygu Tosun, Maryam E. Rettmann et Jerry L. Prince. *Mapping Techniques for Aligning Sulci across Multiple Brains*. In Randy E. Ellis et Terry M. Peters, editors, MICCAI (2), volume 2879 of *Lecture Notes in Computer Science*, pages 862–869. Springer, 2003. 67
- [Tosun 2008] D. Tosun et J.L. Prince. *A Geometry-Driven Optical Flow Warping for Spatial Normalization of Cortical Surfaces*. IEEE Transactions on Medical Imaging, vol. 27, no. 12, pages 1739–1753, December 2008. 26, 67
- [Tu 2007] Zhuowen Tu, Songfeng Zheng, Alan L. Yuille, Allan L. Reiss, Rebecca A. Dutton, Agatha D. Lee, Albert M. Galaburda, Ivo D. Dinov, Paul M. Thompson et Arthur W. Toga. *Automated Extraction of the Cortical Sulci Based on a Supervised Learning Approach*. IEEE Trans. Med. Imaging, vol. 26, no. 4, pages 541–552, 2007. 26, 39
- [Turtinen 2006] M. Turtinen et M. Pietikainen. *Contextual Analysis of Textured Scene Images*. In BMVC06, page II :849, 2006. 121
- [Uwano 2008] I. Uwano, M. Kameda, T. Inoue, H. Nishimoto, S. Fujiwara, R. Hirooka et A. Ogawa. *Computer-assisted identification of the central sulcus in patients with brain tumors using MRI*. J Magn Reson Imaging, vol. 27, pages 1242–1249, Jun 2008. 149
- [Vaillant 1997] Marc Vaillant et Christos Davatzikos. *Finding Parametric Representations of the Cortical Sulci Using an Active Contour Model*. Medical Image Analysis, vol. 1, pages 295–315, 1997. 27
- [Vaillant 1999] Marc Vaillant et Christos Davatzikos. *Hierarchical matching of cortical features for deformable brain image registration*. In IPMI, pages 182–195. Springer-Verlag, 1999. 68
- [Van Essen 2005] D. C. Van Essen. *A Population-Average, Landmark- and Surface-based (PALS) atlas of human cerebral cortex*. Neuroimage, vol. 28, no. 3, pages 635–662, November 2005. 67
- [Vapnik 1998] Vladimir N. Vapnik. *Statistical learning theory*. Wiley-Interscience, September 1998. 13, 35
- [Vivodtzev 2005] Fabien Vivodtzev, Lars Linsen, Bernd Hamann, Kenneth I. Joy et Bruno A. Olshausen. *Brain Mapping Using Topology Graphs Obtained by Surface Segmentation*, 2005. 27, 120

- [Wall 2004] Melanie M. Wall. *A close look at the spatial structure implied by the CAR and SAR models*. Journal of Statistical Planning and Inference, vol. 121, no. 2, pages 311–324, April 2004. 121
- [Wallach 2004] Hanna M. Wallach. *Conditional Random Fields : An Introduction*. 2004. 107
- [Watson 1965] Watson. *Equatorial Distributions on a Sphere*. Biometrika, vol. 52, pages 193–201, 1965. 183
- [Watson 1968] G.S. Watson. *Orientation Statistics in the Earth Sciences*. Technical report, Bull. Geol. Inst. Univ, 1968. 178
- [Weiskopf 2004] D. Weiskopf. *Dye Advection Without the Blur : A Level-Set Approach for Texture-Based Visualization of Unsteady Flow*. Computer Graphics Forum (Proc. Eurographics), vol. 2004, page 2004, 2004. 141
- [Welker 1990] W. Welker. *Why does cerebral cortex fissure and fold ? A review of determinants of gyri and sulci*. Cerebral cortex, vol. 8, pages 3–136, 1990. 142
- [Wilcoxon 1945] F. Wilcoxon. *Individual Comparisons by Ranking Methods*. Biometrics Bull., vol. 1, no. 6, pages 80–83, 1945. 53
- [Winkler 2006] Gerhard Winkler. Image analysis, random fields and markov chain monte carlo methods : A mathematical introduction (stochastic modelling and applied probability). Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006. 105
- [Wong 1998] Tzu-Tsung Wong. *Generalized Dirichlet distribution in Bayesian analysis*. Applied Mathematics and Computation, vol. 97, no. 2-3, pages 165–181, December 1998. 191
- [Wong 2005] Tzu-Tsung Wong. *A Bayesian Approach Employing Generalized Dirichlet Priors in Predicting Microchip Yields*. Journal of the Chinese Institute of Industrial Engineers, vol. 22, pages 210–217, 2005. 191, 192
- [Wood 1993] A.T.A Wood. *Estimation of the concentration parameters of the Fisher matrix distribution on $SO(3)$ and the Bingham distribution on S^q , $q \geq 2$* . Austral. J. Statist., vol. 35, pages 69–79, 1993. 178, 189
- [Yang 2007] Faguo Yang et Frithjof Kruggel. *Optimization Algorithms for Labeling Brain Sulci Based on Graph Matching*. In ICCV, pages 1–7, 2007. 52, 67, 117
- [Yang 2009] F. Yang et F. Kruggel. *A Graph Matching Approach for Labeling Brain Sulci Using Location, Orientation, and Shape*. Neurocomputing, in print, 2009. 27, 67, 116, 117
- [Yedidia 2001] J. Yedidia, W.T. Freeman et Y. Weiss. *Bethe free energy, Kikuchi approximations, and belief propagation algorithms*. In SCTV01, pages xx–yy, 2001. 111
- [Yedidia 2005] J. S. Yedidia, W. T. Freeman et Y. Weiss. *Constructing free-energy approximations and generalized belief propagation algorithms*. Information Theory, IEEE Transactions on, vol. 51, no. 7, pages 2282–2312, 2005. 111
- [Yeo 2008a] B. Yeo, M. Sabuncu, R. Desikan, B. Fischl et P. Golland. *Effects of registration regularization and atlas sharpness on segmentation accuracy*. 10 2008. 68
- [Yeo 2008b] B. T. Yeo, Mert Sabuncu, Tom Vercauteren, Nicholas Ayache, Bruce Fischl et Polina Golland. *Spherical Demons : Fast Surface Registration*. In MICCAI '08 : Proceedings of the 11th international conference on Medical Image Computing and Computer-Assisted Intervention - Part I, pages 745–753, Berlin, Heidelberg, 2008. Springer-Verlag. 67
- [Younes 1988] L. Younes. *Estimation and Annealing for Gibbsian Fields*. Annales de l'Institut Henri Poincaré, Section B, Calcul des Probabilités et Statistique, vol. 24, pages 269–294, 1988. 108

- [Zeng 1999] Xiaolan Zeng, Lawrence H. Staib, Robert T. Schultz, Hemant Tagare, L. Win et James S. Duncan. *A New Approach to 3D Sulcal Ribbon Finding from MR Images*. In MIC-CAI '99 : Proceedings of the Second International Conference on Medical Image Computing and Computer-Assisted Intervention, pages 148–157, London, UK, 1999. Springer-Verlag. 27
- [Zhang 2001] Y Zhang, M Brady et S Smith. *Segmentation of Brain MR Images through a Hidden Markov Random Field Model and the Expectation Maximization Algorithm*. IEEE Trans. on Medical Imaging, vol. 20, no. 1, pages 45–57, 2001. 106
- [Zhang 2009] F. Zhang, E. Hancock, C. Goodlett et G. Gerig. *Probabilistic white matter fiber tracking using particle filtering and von Mises-Fisher sampling*. Med Image Anal, 02 2009. 178
- [Zhou 1999] Yong Zhou, Paul M. Thompson et Arthur W. Toga. *Extracting and Representing the Cortical Sulci*. IEEE Comput. Graph. Appl., vol. 19, no. 3, pages 49–55, 1999. 27
- [Zhu 2002] Song Chun Zhu et Xiuwen Liu. *Learning in Gibbsian Fields : How Accurate and How Fast Can It Be*. IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 24, pages 1001–1006, 2002. 108