Unsupervised Models for White Matter Fiber-Bundles Analysis in Multiple Sclerosis

Claudio Stamile

To cite this version:

Claudio Stamile. Unsupervised Models for White Matter Fiber-Bundles Analysis in Multiple Sclerosis. Human health and pathology. Université de Lyon; Université catholique de Louvain (1970-..), 2017. English. NNT : 2017LYSE1147 . tel-01731072

HAL Id: tel-01731072
https://theses.hal.science/tel-01731072
Submitted on 13 Mar 2018

N° d'ordre NNT : 2017LYSE1147

THÈSE DE DOCTORAT DE L'UNIVERSITÉ DE LYON
opérée au sein de
l'Université Claude Bernard Lyon 1
Department of Electrical Engineering (ESAT) - Katholieke Universiteit Leuven

École Doctorale ED205 - Interdisciplinaire Sciences Santé
Arenberg Doctoral School

Spécialité de doctorat :
Recherche clinique, innovation technologique, santé publique
Doctor of Engineering Science (PhD): Electrical Engineering

Soutenue publiquement le 11 Septembre 2017, par :
Claudio Stamile

Unsupervised Models for White Matter Fiber-Bundles Analysis in Multiple Sclerosis

---

Devant le jury composé de :

Thiran Jean-Philippe *Professeur, École polytechnique fédérale de Lausanne*	Rapporteur
Achard Sophie *Chargé de Recherche CNRS, Université Grenoble Alpes*	Rapporteure
Cotton François *Professeur des Universités - Praticien Hospitalier, Université de Lyon*	Examinateur
Guttmann Charles R.G. *Professeur Associé, Harvard Medical School*	Examinateur
Sappey-Marinier Dominique *Maître de Conférences - Praticien Hospitalier, Université de Lyon*	Directeur de thèse
Frindel Carole *Maître de Conférences, INSA Lyon*	Co-Directrice de thèse
Van Huffel Sabine *Professeure, Katholieke Universiteit Leuven*	Directrice de thèse
Maes Frederik *Professeur, Katholieke Universiteit Leuven*	Co-Directeur de thèse

UNIVERSITE CLAUDE BERNARD - LYON 1

Président de l'Université	**M. le Pr. F. FLEURY**
Président du Conseil Académique	M. le Pr. H. BEN HADID
Vice-président du Conseil d'Administration	M. le Pr. D. REVEL
Vice-président du Conseil Formation et Vie Universitaire	M. le Pr. P. CHEVALIER
Vice-président de la Commission Recherche	M. F. VALLÉE
Directrice Générale des Services	Mme D. MARCHAND

COMPOSANTES SANTE

Faculté de Médecine Lyon Est - Claude Bernard	M. le Pr. G.RODE
Faculté de Médecine et de Maïeutique Lyon Sud - Charles Mérieux	Mme la Pr. C. BURILLON
Faculté d'Odontologie	M. le Pr. D. BOURGEOIS
Institut des Sciences Pharmaceutiques et Biologiques	Mme la Pr. C. VINCIGUERRA
Institut des Sciences et Techniques de la Réadaptation	M. X. PERROT
Département de formation et Centre de Recherche en Biologie Humaine	Mme la Pr. A-M. SCHOTT

COMPOSANTES ET DEPARTEMENTS DE SCIENCES ET TECHNOLOGIE

Faculté des Sciences et Technologies	M. F. DE MARCHI
Département Biologie	M. le Pr. F. THEVENARD
Département Chimie Biochimie	Mme C. FELIX
Département GEP	M. H. HAMMOURI
Département Informatique	M. le Pr. S. AKKOUCHE
Département Mathématiques	M. le Pr. G. TOMANOV
Département Mécanique	M. le Pr. H. BEN HADID
Département Physique	M. le Pr. J-C PLENET
UFR Sciences et Techniques des Activités Physiques et Sportives	M. Y.VANPOULLE
Observatoire des Sciences de l'Univers de Lyon	M. B. GUIDERDONI
Polytech Lyon	M. le Pr. E. PERRIN
Ecole Supérieure de Chimie Physique Electronique	M. G. PIGNAULT
Institut Universitaire de Technologie de Lyon 1	M. le Pr. C. VITON
Ecole Supérieure du Professorat et de l'Education	M. le Pr. A. MOUGNIOTTE
Institut de Science Financière et d'Assurances	M. N. LEBOISNE

A Francesco e Marco.

Acknowledgements

I would like to thank the examination committee who accepted to review this thesis: Professor Jean-Philippe Thiran, Doctor Sophie Achard, Professor François Cotton, Doctor Charles Guttmann, Doctor Dominique Sappey-Marinier, Doctor Carole Frindel, Professor Sabine Van Huffel, Professor Frederik Maes.

I would like to thank my supervisors doctor Dominique Sappey-Marinier, professor Sabine Van Huffel and my co-supervisor doctor Carole Frindel and professor Frederik Maes for the patient guidance, encouragement and advice they have provided throughout my time as student. Their positive outlook and confidence in my research inspired me and gave me confidence.

I would like to thank Isabelle Magnin and Olivier Beuf directors of CREATIS. I thank also the group “Imagerie Cérébrale” and in particular Françoise Durand-Dubief, François Cotton and Salem Hannoun for all the constructive exchanges we had during these four years. Finally, I would like to thank the whole “Centre Pluridisciplinaire en Imagerie du Vivant” (CERMEP-Imagerie du Vivant) with a special thought to Jamila Lagha.

Personally, I want to say thank you to all the other ESRs in the TRANSACT project with a special thought to Victor. We really spent 4 important years together sharing our experiences and our “path” to the get the PhD.

Mais surtout merci à Gabriel. You’re really the best friend I could ever have during my PhD. You helped my like a brother and I have no words to say thank you. We shared a lot of sad and happy moments together and I still cannot understand how you have not killed me during these 4 years.

A big thank you to Adrian, a real and honest friend. I’m really happy to have friends like you in my life. I learned a lof from you, more than you can imagine and I’m really glad to remember you as part of my PhD.

Grazie a Francesco, Ferdinando ed Aldo, amici praticamente da sempre, e soprattutto cosí poco sani di mente da esserci qui, oggi, a condividere questo momento (o almeno credo dato che sto scrivendo questa dedica settimane prima della difesa. Se cosí non fosse, allora vi odio).

Infine vorrei ringraziare la mia famiglia. Nonno Amerigo e nonna Natalina con i quali non avró la fortuna di condividere tutto ció. Nonno Egidio e nonna Rosa i quali mi sono stati vicini piú di quanto loro possano immaginare. Di nonna Rosa porto sempre nella mente, quasi come un mantra, le parole che mi disse poco prima di cominciare l’universitá: “Rendici fieri". Spero di esserci riuscito in parte.

Ringrazio mia madre e mio padre che hanno sofferto le distanze manifestandolo o tacendolo secondo le occasioni. Ho imparato solo con il tempo a condividere le loro paure, quelle di un figlio che per anni va in giro per un mondo non piú tanto sereno. Vorrei ringraziare Melissa e Federico, per il loro essere stati sempre un porto sicuro nel quale nascondermi in periodi poco sereni. Ma soprattutto vorrei ringraziare Francesco a Marco, coloro a cui queste duecento pagine sono dedicate. Spero, un giorno, di riuscire a ripagare, anche in minima parte, questo mio enorme debito di lontananza. A voi, auguro di raggiungere traguardi quanto piú simili ai vostri sogni.

Per ultimo, quasi a voler rimarcare la sua meravigliosa eleganza nello starmi sempre silenziosamente vicina, vorrei ringraziare Clara, colei che mi ha permesso di arrivare qui, oggi, a scrivere queste parole. Un rapporto fatto di distanze grandi quanto i suoi magnifici occhi. Un rapporto fatto di giorni e di notti passate in aeroporto con l'agonia di salutarci o con il lacerante batticuore per incontrarci.

``I nostri incontri eran sessioni
di sguardi sorridenti
che si staccavan solo
per controllare l'ora
e tanto era fra noi lo struggimento
che spesso ci siamo presi il lusso
di non baciarci.''

Michela Mari

Abstract

Diffusion Magnetic Resonance Imaging (dMRI) is a meaningful technique for white matter (WM) fiber-tracking and microstructural characterization of axonal/neuronal integrity and connectivity. By measuring water molecules motion in the three directions of space, numerous parametric maps can be reconstructed. Among these, fractional anisotropy (FA), mean diffusivity (MD), and axial (λ_a) and radial (λ_r) diffusivities have extensively been used to investigate brain diseases. Overall, these findings demonstrated that WM and grey matter (GM) tissues are subjected to numerous microstructural alterations in multiple sclerosis (MS). However, it remains unclear whether these tissue alterations result from global processes, such as inflammatory cascades and/or neurodegenerative mechanisms, or local inflammatory and/or demyelinating lesions. Furthermore, these pathological events may occur along afferent or efferent WM fiber pathways, leading to antero- or retrograde degeneration. Thus, for a better understanding of MS pathological processes like its spatial and temporal progression, an accurate and sensitive characterization of WM fibers along their pathways is needed.

By merging the spatial information of fiber tracking with the diffusion metrics derived obtained from longitudinal acquisitions, WM fiber-bundles could be modeled and analyzed along their profile. Such signal analysis of WM fibers can be performed by several methods providing either semi- or fully unsupervised solutions.

In the first part of this work, we will give an overview of the studies already present in literature and we will focus our analysis on studies showing the interest of dMRI for WM characterization in MS.

In the second part, we will introduce two new string-based methods, one semi-supervised and one unsupervised, to extract specific WM fiber-bundles. We will show how these algorithms allow to improve extraction of specific fiber-bundles compared to the approaches already present in literature. Moreover, in the second chapter, we will show an extension of the proposed method by coupling the string-based formalism with the spatial information of the fiber-tracks.

In the third, and last part, we will describe, in order of complexity, three different fully automated algorithms to perform analysis of longitudinal changes visible along WM fiber-bundles in MS patients. These methods are based on Gaussian mixture model, nonnegative matrix and tensor factorisation respectively. Moreover, in order to validate our methods, we introduce a new model to simulate real longitudinal changes based on a generalised Gaussian probability density function. For those algorithms high levels of performances were obtained for the detection of small longitudinal changes along the WM fiber-bundles in MS patients.

In conclusion, we propose, in this work, a new set of unsupervised algorithms to perform a sensitivity analysis of WM fiber-bundle that would be useful for the characterisation of pathological alterations occurring in MS patients.

Résumé

L'imagerie de résonance magnétique de diffusion (dMRI) est une technique très sensible pour la tractographie des fibres de substance blanche et la caractérisation de l'intégrité et de la connectivité axonale. À travers la mesure des mouvements des molécules d'eau dans les trois dimensions de l'espace, il est possible de reconstruire des cartes paramétriques reflétant l'organisation tissulaire. Parmi ces cartes, la fraction d'anisotropie (FA) et les diffusivités axiale (λ_a), radiale (λ_r) et moyenne (MD) ont été largement utilisés pour caractériser les pathologies du système nerveux central. L'emploi de ces cartes paramétriques a permis de mettre en évidence la survenue d'altérations microstructurelles de la substance blanche (SB) et de la substance grise (SG) chez les patients atteints d'une sclérose en plaques (SEP). Cependant, il reste à déterminer l'origine de ces altérations qui peuvent résulter de processus globaux comme la cascade inflammatoire et les mécanismes neurodégénératifs ou de processus plus localisés comme la démyélinisation et l'inflammation. De plus, ces processus pathologiques peuvent survenir le long de faisceaux de SB afférents ou efférents, conduisant à une dégénérescence antéro- ou rétrograde. Ainsi, pour une meilleure compréhension des processus pathologiques et de leur progression dans l'espace et dans le temps, une caractérisation fine et précise des faisceaux de SB est nécessaire.

En couplant l'information spatiale de la tractographie des fibres aux cartes paramétriques de diffusion, obtenues grâce à un protocole d'acquisitions longitudinal, les profils des faisceaux de SB peuvent être modélisés et analysés. Une telle analyse des faisceaux de SB peut être effectuée grâce à différentes méthodes, partiellement ou totalement non-supervisées.

Dans la première partie de ce travail, nous dressons l'état de l'art des études déjà présentes dans la littérature. Cet état de l'art se focalisera sur les études montrant les effets de la SEP sur les faisceaux de SB grâce à l'emploi de l'imagerie de tenseur de diffusion.

Dans la seconde partie de ce travail, nous introduisons deux nouvelles méthodes, “string-based”, l'une semi-supervisée et l'autre non-supervisée, pour extraire les faisceaux de SB. Nous montrons comment ces algorithmes permettent d'améliorer l'extraction de faisceaux spécifiques comparé aux approches déjà présentes dans la littérature. De plus, dans un second chapitre, nous montrons une extension de la méthode proposée par le couplage du formalisme “string-based” aux informations spatiales des faisceaux de SB.

Dans la troisième et dernière partie de ce travail, nous décrivons trois algorithmes automatiques permettant l'analyse des changements longitudinaux le long des faisceaux de SB chez des patients atteints d'une SEP. Ces méthodes sont basées respectivement sur un modèle de mélange Gaussien, la factorisation de matrices non-négatives et la factorisation de tenseurs non-négatifs. De plus, pour valider nos méthodes, nous introduisons un nouveau modèle pour simuler des changements longitudinaux réels, basé sur une fonction de probabilité Gaussienne généralisée. Des hautes performances ont été obtenues avec ces algorithmes dans la détection de changements longitudinaux d'amplitude faible le long des faisceaux de SB chez des patients atteints de SEP.

En conclusion, nous avons proposé dans ce travail des nouveaux algorithmes non-supervisés pour une analyse précise des faisceaux de SB, permettant une meilleure caractérisation des altérations pathologiques survenant chez les patients atteints de SEP.

Samenvatting

Diffusie Magnetische Resonantie Beeldvorming (dMRI) is een zinvolle techniek voor de witte stof (WM) vezeltractografie en microstructurele karakterisering van axonale / neuronale integriteit en connectiviteit. Door de beweging van watermoleculen in de drie ruimtelijke richtingen te meten, kunnen talrijke parametrische kaarten worden gereconstrueerd. Onder anderen, fractionele anisotropie (FA), gemiddelde diffusiviteit (MD) en axiale (λ_a) en radiale (λ_r) diffusiviteiten zijn vaak gebruikt om hersenziekten te onderzoeken. Over het geheel genomen, hebben deze bevindingen aangetoond dat WM en grijze stof (GM) weefsels onderworpen zijn aan talrijke microstructurele veranderingen in multiple sclerose (MS). Het blijft echter onduidelijk of deze weefselveranderingen voortvloeien uit globale processen, zoals ontstekingscascades en / of neurodegeneratieve mechanismen, of lokale ontstekings- en / of demyeliniserende laesies. Bovendien kunnen deze pathologische gebeurtenissen optreden langs afferente of efferente WM-vezelverbindingen, wat leidt tot antero- of retrograde degeneratie. Zo is een nauwkeurige en gevoelige karakterisering van WM-vezels langs hun verbindingen nodig om beter te kunnen begrijpen op MS-pathologische processen zoals de ruimtelijke en temporale progressie.

Door het samenvoegen van de ruimtelijke informatie van vezelbundels met de diffusieparameters uit longitudinale acquisities, kunnen WM-vezelbundels worden gemodelleerd en geanalyseerd langs hun profiel. Dergelijke signaalanalyse van WM-vezels kan worden uitgevoerd door middel van verschillende werkwijzen als semi- of niet-gesuperviseerde methoden.

In het eerste deel van dit werk geven we een overzicht van de reeds aanwezige studies in de literatuur. We zullen onze analyse richten op studies die het belang van dMRI voor WM karakterisering in MS tonen.

In het tweede deel introduceren we twee nieuwe string-gebaseerde methoden (een semi-supervised en een ongesuperviseerde methode), om specifieke WM-vezelbundels te extraheren. We zullen laten zien hoe deze algoritmen toestaan om de extractie van specifieke vezelbundels te verbeteren in vergelijking met de reeds aanwezige benaderingen in de literatuur. Bovendien zullen we in het tweede hoofdstuk een uitbreiding van de voorgestelde methode laten zien door het string-gebaseerde formalisme met de ruimtelijke informatie van de vezelsporen te koppelen.

In het derde en laatste gedeelte beschrijven we drie verschillende volledig geautomatiseerde algoritmes om de analyse van longitudinale veranderingen langs WM-vezelbundels in MS-patiënten uit te voeren. Deze methoden zijn gebaseerd op Gaussiaanse mixture model, niet-negatieve matrix en tensorfactorisatie respectievelijk. Bovendien, om onze methoden te valideren, introduceren wij een nieuw model om echte longitudinale veranderingen te simuleren op basis van een gegeneraliseerd Gaussiaanse verdeling. Voor die algoritmen werden hoge niveaus van prestaties verkregen voor de detectie van kleine longitudinale veranderingen langs de WM-vezelbundels in MS-patiënten.

In conclusie, stellen we in dit werk een nieuwe reeks ongesuperviseerde algoritmen voor om een sensitiviteitsanalyse van WM-vezelbundels uit te voeren die nuttig kan zijn voor de karakterisering van pathologische veranderingen die zich voordoen in MS-patiënten.

Contents

List of Figures

I State of the art 5

1 Magnetic Resonance Imaging 6

2 Multi-Features Approach Based on Nonnegative Matrix Factorization Algorithm 99

3 Constrained Tensor Decomposition for Global Analysis of Fiber-Bundle Signal 123

List of symbols and abbreviations

Latin letters

$\vec{B}_0$	Magnetic field

Greek letters

$\lambda_{1,2,3}$	EigenValues of diffusion tensor
$\vec{\epsilon}_{1,2,3}$	EigenVectors of diffusion tensor
λ_1	Axial Diffusivity
λ_r	Radial Diffusivity

Abbreviations

AA	Angular Anisotropy
ALS	Alternating Least Squares
BBB	Blood Brain Barrier
BSS	Blind Source Separation
BTD	Block Term Decomposition
CIS	Clinically Isolated Syndrome
CC	Corpus Callosum
CNS	Central Nervous System
CSF	Cerebrospinal Fluid
CST	Cortico-Spinal Tract
CPD	Canonical Polyadic Decomposition
DI	Diffusion Imaging
dMRI	Diffusion Magnetic Resonance Imaging
DSC	Sørensen-Dice Score Coefficient
DTI	Diffusion Tensor Imaging
EDSS	Expanded Disability Status Scale
EPI	Echo-Planar Imaging
FA	Fractional Anisotropy
FOD	Fiber Orientation Density
GFA	Generalized Fractional Anisotropy
GGPDF	Generalized Gaussian Probability Density Function
GM	Grey Matter
HALS	Hierarchical Alternating Least-Squares
HAMD	Hamilton Depression Rating Scale

HC	Healthy Control
IFOF	Inferior Fronto-Occipital Fasciculi
IIT	Illinois Institute of Technology
JHU	Johns Hopkins University
KLA	Kullback-Leibler Anisotropy
LCS	Longest Common Subsequence
LOF	Density-based Local Outliers
LVS	Longitudinal Variations Simulator
MBP	Myelin Basic Protein
MD	Mean Diffusivity
MDP	Maximum Density Paths
MDF	Minimum Average Direct Flip
MLE	Maximum Likelihood Estimation
MRI	Magnetic Resonance Imaging
MS	Multiple Sclerosis
MSFC	Multiple Sclerosis Functional Composite
MUL	Multiplicative Update
NAGM	Normal Appearing Grey Matter
NAWM	Normal Appearing White Matter
NHPT	Nine-hole Peg Test
NMF	Non-negative Matrix Factorization
NTF	Non-negative Tensor Factorization
ODF	Orientation Distribution Function
PP	Primary Progressive
QB	Quick Bundle
RF	Radio Frequency
ROI	Regions of Interest
RR	Relapsing Remitting
SBED	Semi-Blind Edit Distance
SD	Standard Deviation
SP	Secondary Progressive
TBSS	Tract-Based Spatial Statistics
TCHM	Transcallosal Hand Motor Fibers
TE	Echo Time
TR	Repetition Time
WM	White Matter

Introduction

Multiple sclerosis (MS) is the most frequent disabling neurological disease in young adults with a national prevalence of 95/100 000 in France. It is a chronic demyelinating inflammatory disease of the central nervous system (CNS), mainly characterized by lesions in white matter (WM) tissue but also in grey matter (GM). Disease onset is identified by a first acute episode called clinically isolated syndrome (CIS), that evolves either into a relapsing-remitting (RR) course in about 85% of cases or into a primary progressive (PP) course in the remaining 15% of cases. RR patients will evolve into a secondary progressive (SP) course after several years. Today's neurologist challenge consists in providing new markers that can accurately characterize pathological processes and predict clinical outcomes. Achieving this goal is particularly crucial in MS since it remains without well-known etiology.

Magnetic resonance imaging (MRI) is a powerful technology to investigate the effects of MS in CNS. It is, *de facto*, an essential technique for the understanding of MS pathological mechanisms.
With the evolution of MRI, non-conventional acquisition protocols, like diffusion tensor imaging (DTI), allowed to obtain sensitive information essential for a deep characterization of WM tissue. Indeed, DTI allows to obtain: in one hand, quantitative information describing the microscopic status of the WM tissue and, in another hand, information about brain structural connectivity. Those information can be merged in order to analyze the diffusion signal changes in specific WM fiber-bundles, reconstructed from DTI data. These new approaches were used in different works showing promising results useful for the investigation of complex pathological mechanisms.

In the last years, interest in longitudinal MRI studies has grown up exponentially. Indeed, they showed that the investigation of longitudinal progression of brain damages in MS could really help to better understand the disease. However, they remain challenging especially using diffusion data. This is mainly related to the large numbers of scans requiring a high quality of acquisition reproducibility; hence a homogeneous intensity across scans and presence of methodological biases that could be itroduced during the image processing. By merging diffusion information with WM fiber-bundles, it is possible to obtain a strong and specific characterization of the WM tissue. Unfortunately, the approaches already published in literature, present certain limitations that do not allow their direct application in longitudinal settings. For instance, they allow to perform only a global analysis of the WM structure and, as consequence, they are not sensitive enough to detect small and rapid tissue alterations which typically occur in early MS patients like CIS or RR subjects.

Since longitudinal analysis of WM fiber-bundles is not an easy task and requires different data processing steps, we propose, in this work, new methods to extract and analyze longitudinal changes along WM. This work is divided in three main parts.

The first part is divided in four chapters. In the first two, we will give a general introduction about MRI, diffusion MRI (dMRI) and MS. In the third chapter, we will report a series of studies in which dMRI techniques were successfully applied to MS in order to investigate the correlations between MRI biomarkers and clinical status of the patient. In the last chapter, we will discuss in more detail the scope of this thesis.

In the second part, composed by two chapters, we will introduce the first piece of our processing pipeline for the longitudinal analysis of WM fiber-bundles. The first chapter will describe, in more detail, our proposed string-based method to automatically extract WM fiber-bundles from the whole tractogram of the brain. We will provide a complete formalization of the proposed method and an extensive validation campaign. In the second chapter, we will extend the proposed method in order to couple the information derived from the string-based formalism with the spatial coordinate of each fiber of the tractogram.

In the third, and last part, composed by three chapters, we will present, according to their order of complexity, three different algorithms for the longitudinal analysis of the signal along WM fiber-bundles. The first chapter will describe a first simple model based on the histogram analysis of the signal in each cross-section of the fiber-bundle. Moreover, we will give a better overview of the problem showing how it can overcome the limitations of the classical global method. In the second chapter, we will extend this method by proposing a non-negative matrix factorization algorithm capable to deal with information derived from multiple diffusion features and with a large number of time-points. Finally, in the third chapter, we will describe a more general model based on constrained tensor factorization. This model, thanks to the “multi-dimensional” property of the tensor, is capable to generalize the two methods previously described.

Finally, we will draw the conclusions of this work and highlight the most interesting perspectives for clinical applications.

I State of the art

Chapter 1

Magnetic Resonance Imaging

Contents

1 Magnetic Resonance Imaging

1.1 Principle of Magnetic Resonance Imaging

Magnetic Resonance Imaging (MRI) is derived from the Nuclear Magnetic Resonance (NMR). The basic idea of NMR is that certain nuclei will resonate and emit a radio signal if placed in a strong magnetic field and pulsed with a certain radiofrequency energy. In order to clarify this concept, we will proceed with our explanation starting from the principal subject of NMR: the atomic nuclei. Indeed, by studying the global effects of all the atomic nuclei composing the matter it is possible to have indirect information about the matter itself.

MRI relies upon the spin property of nuclear physics. When the spin is placed in a magnetic field (denoted with $\vec{B_0}$) the direction of the spins follows the direction of $\vec{B_0}$. The alignment of the spin with the magnetic field $\vec{B_0}$ generates a magnetization $\vec{M}$ defined as:

$$\vec{M} = \frac{\Sigma\vec{\mu}}{dV}$$

where $\vec{\mu}$ represents the magnetic moment in the magnetic field $\vec{B_0}$. Moreover, the spin precesses about that field in a motion analogous to a spinning top. The frequency of precession is governed by the Larmor equation, defined as:

$$\vec{\omega} = -\gamma\vec{B_0}.$$

where γ is the magnetogyric ratio and every nucleus has its own specific value. Under the influence of a radio frequency (RF) wave, it is possible to perturb the magnetization created by the field $\vec{B_0}$. This perturbation leads to the transition of the nuclei from their state of energy, this phenomenon is called resonance. Spontaneously, the nuclei recover their state of fundamental energy by the emission of a RF wave which will be the NMR signal. This phenomenon corresponds to the rotating magnetisation decays due to relaxation which can be subdivided into longitudinal or T1 recovery and transverse or T2 decay.

Longitudinal Recovery

The longitudinal recovery describes the regrowth of the magnetization component in the z direction. It is a relaxation time constant which is an intrinsic property of each tissue. After a 90° pulse, when all the z components are tipped into the transverse plane (M), T1 is the number of milliseconds it takes to grow to the 63% of the original orientation (M_z) (Figure 1.1). The relationship is described by the following equation:

$$M_z(t) = M_z(0)\left(1 - e^{-\frac{t}{T1}}\right) \tag{1.1}$$

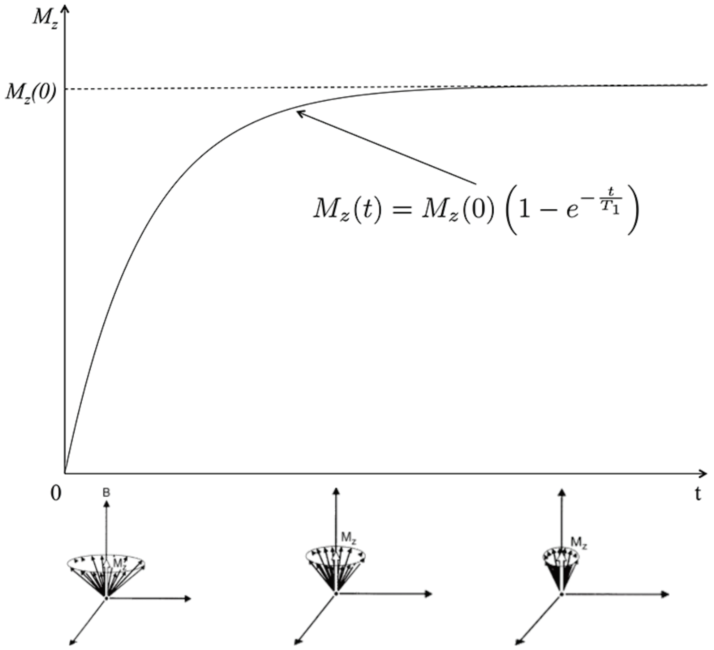


Figure 1.1: Relaxation of the longitudinal magnetization.

Transverse Decay

T2, or transverse, relaxation describes the decay of the signal in the xy plane. It occurs due to the interactions between spins as energy is released following a RF pulse. T2 decay is the number of milliseconds it takes to reduct to 37% of the magnetization in the xy (M_{xy}) plane (Figure 1.2). It is described by the equation:

$$M_{xy}(t) = M_{xy}(0)e^{-\frac{t}{T2}} \tag{1.2}$$

1.2 Conventional MRI Sequences

Since the first study [Damadian (1971)] where the author showed *in vivo* T1/T2 differences between cancerous and normal tissue, the clinical interest of MRI exponentially increased. Indeed, thanks to this noninvasive technique, *in vivo* investigation of human structures, difficult to analyze, was finally possible.

The basic MRI techniques to obtain brain images are called conventional MRI (cMRI) sequences. With this name, we usually refer to a well-defined set of standard MRI acquisition techniques that allow to obtain rather simple, yet informative anatomical *in-vivo* images of the brain, or, in general, human body. Since the beginning of MRI, two main types of sequences were used: spin-echo and gradient-echo sequences. One of the first spin-echo sequences was presented in [Hahn (1950)] (Figure 1.3). In this sequence, by tuning specific acquisition parameters, it is possible to excite a particular type of nuclei obtaining different types of information. By referring to Figure 1.3 it is possible to see how in a spin-echo sequence two main parameters can be tuned: echo-time (TE) and

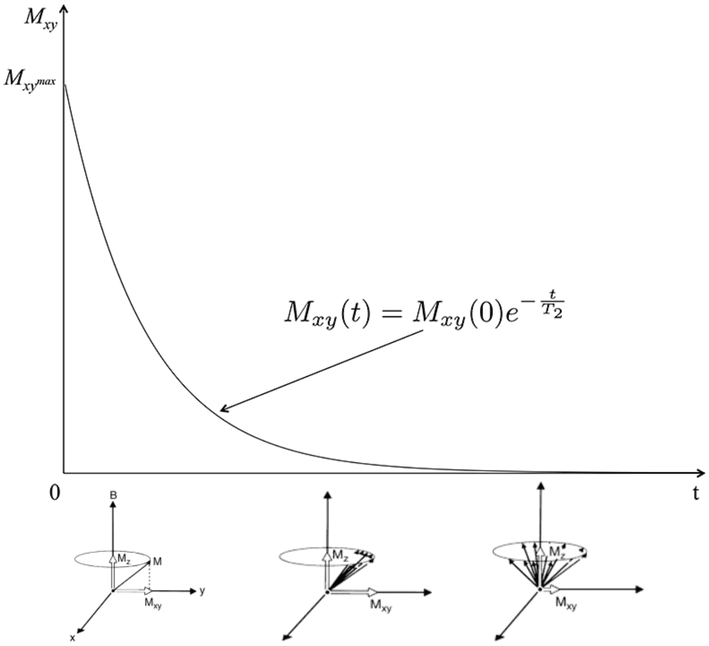


Figure 1.2: Relaxation of the transverse magnetization.

the repetition-time (TR). Specific tuning of these parameters allows to underline different characteristics of the tissue reflected by the T1 or T2 time.

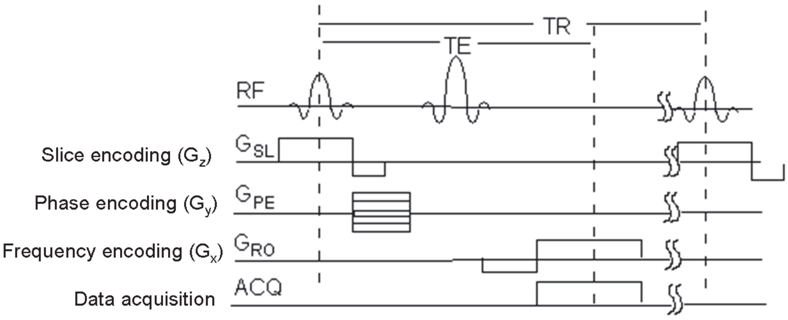


Figure 1.3: Description of the classical spin-echo sequence.

T1-Weighted Imaging

The T1-weighted sequence is obtained by tuning two parameters in the pulse sequence shown in Figure 1.3. In particular, the TR value is chosen to be less than the T1 time (usually 500 ms) and the TE value is chosen to be less than T2 (usually 30 ms). Most lesions have a prolonged T1 and they are dark in T1-weight images; hence, tumors or infarctions could be missed [Hendee and Morgan (1984)]. An interesting property of T1 weighted sequence is related to its sensibility to obtain the best contrast for paramagnetic contrast agents (e.g. a gadolinium-containing compounds). This property is extremely important especially in clinical settings where contrast agents are essential to perform a correct diagnosis, particularly in brain-related pathologies.

T2-Weighted Imaging

Like the T1-weighted imaging sequence, the T2-weighted sequence is obtained tuning the values of the TR and TE acquisition parameters. Specifically, the TR value is chosen to be greater than T1 (usually 2000 ms) and the TE values is chosen to be less than T2 (usually 100 ms). In principle, the T2-weighted images provide better contrast between pathological tissue and normal tissue, and the T1-weighted images provide better anatomical details. In T2 weighted imaging the dominant signals come from: fluid (like Cerebrospinal fluid), with high signal intensity (white), grey matter with intermediate signal intensity (grey) and white matter: hypointense compared to grey matter.

T2 FLAIR Imaging

Fluid attenuation inversion recovery (FLAIR) is an important technique which allows to remove the signal effects generated by the presence of fluids. In the T2 FLAIR images, there is a complete suppression of the cerebrospinal fluid (CSF) signal (it is dark in the obtained image) but the cerebral lesions appear intense; for this reason, T2-FLAIR images are useful to help the diagnosis of several neurodegenerative pathologies.

An example of multiple sclerosis lesions visible in T1 weighted (after gadolinium injection) and T2 FLAIR imaging is presented in Figure 1.4. In Chapter 2, we will discuss how cMRI is used to perform a complete diagnosis of multiple sclerosis.

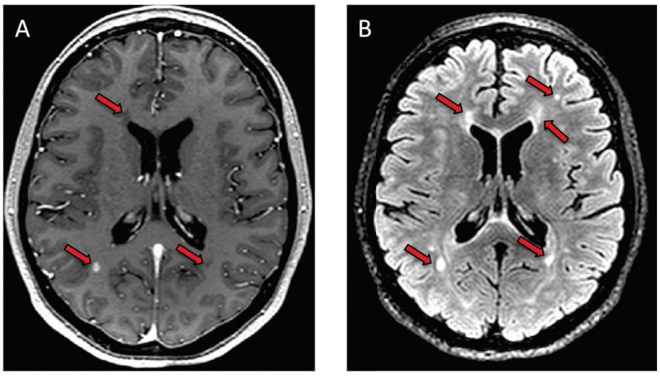


Figure 1.4: Example of multiple sclerosis lesions (red arrows) visible through: **A)** T1-weighted imaging (after gadolinium injection) **B)** T2-FLAIR imaging.

2 Diffusion MRI

2.1 Physical Meaning and Biological Interest

The idea behind the diffusion MRI (dMRI) relies on the concept of Brownian motion [Brown (1828)]. This term refers to the constant random microscopic molecular motion due to heat. So, roughly speaking, Brownian motion is the macroscopic picture emerging from a particle moving randomly in d-dimensional space due to the heat.

At a fixed temperature, the rate of diffusion was described by Einstein, in 1905, by the following equation [Einstein (1956)]:

$$r^2 = 6Dt$$

where r^2 is the mean square displacement of the molecules, t is the diffusion time and D is a constant value defined as follows:

$$D = \frac{k_B T}{6\pi\eta R}$$

k_B is the Boltzmann constant, T is the temperature of the medium, η is the dynamic viscosity of the medium and R is the radius of the spherical particle.

In our case, where the goal is to study *in vivo* the brain structure in humans, the type of diffusion being investigated is water self-diffusion, meaning the thermal motion of water molecules in a medium that itself consists mostly of water [Thomsen *et al.* (1987), Mukherjee *et al.* (2008)]. Diffusion MRI (dMRI) is a MRI technique that allows to observe the thermal motion of water molecules.

In a first analysis the link between the concept of diffusion and its use in human brain studies is not so clear. In order to simplify the transaction between the theory of the diffusion and its application in human brain investigation we provided, in Figure 1.5, a simple example. In the figure, three different types of tissue constituting the corpus callosum are shown. As it is possible to see, the movement of the water molecules in the tissue, is significantly modified by the structure of the tissue itself. Indeed, in structured tissue without myelin (described in Section 3) the diffusion is more “free” (is not subject to strong physical constraints) compared to the diffusion in tissue with partial or total myelin presence. This is a clear example that shows what is “indirectly” visible when we study diffusion in human brain. So, analysis of diffusion in human brain is an important tool to extract useful information describing the structure of the brain tissue.

2.2 Acquisition Sequence

In order to obtain *in vivo* images of the brain showing the diffusion in tissue, it is important to have an MRI sequence capable to acquire this information. In [Stejskal and Tanner (1965)] the authors developed a MRI acquisition sequence namely Pulsed Gradient Spin-Echo (PGSE), capable to acquire diffusion information from human tissue. The sequence is graphically represented in Figure 1.6

The diffusion-weighted pulse sequence is composed by the addition of a pair of diffusion gradients. Those gradients can be oriented in specific directions in order to measure the diffusivity. Gradients are created by combining the directions in the 3 dimensional space. As reported in Figure 1.6 other parameters can be used in order to measure diffusion in a

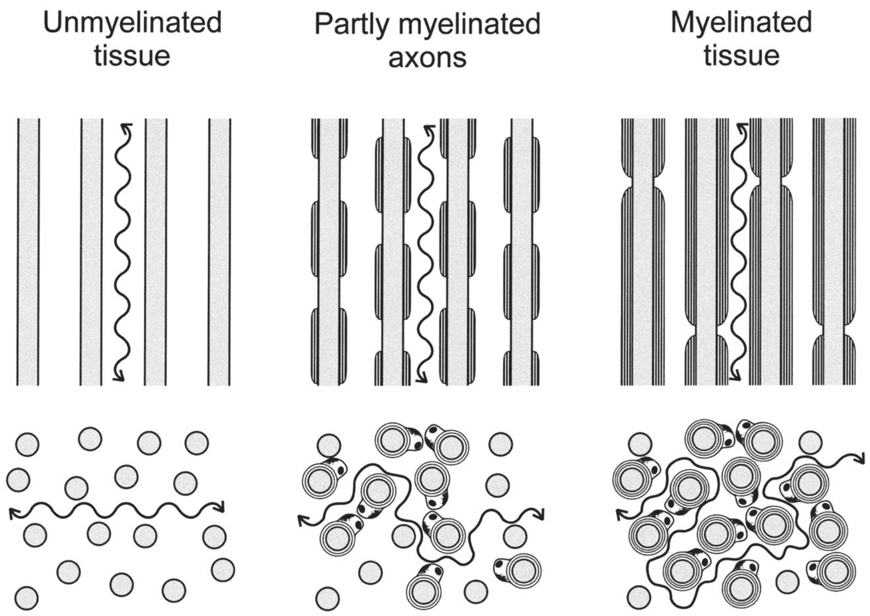


Figure 1.5: Diffusion in extracellular space of unmyelinated, partly myelinated, and myelinated corpus callosum. Top: diffusion along increasingly myelinated axons. Bottom: extracellular diffusion in direction perpendicular to orientation of axons, i.e., around axons, is compromised by a number of myelin sheaths, number of myelinated axons, and length of myelin sheaths along axons. Scheme demonstrates increased anisotropy as myelination progresses. Image and caption from [Voříšek and Syková (1997)].

specific direction. Those are the duration of each gradient (δ) and the amplitude (G) of the gradient itself.

Molecular motion thus results in loss of signal intensity due to incomplete rephasing of water proton spins, which change position between and during the applications of the 2 diffusion-sensitizing gradients [Mukherjee *et al.* (2008)]. This diffusion-weighted contrast can be fit to an exponential model:

$$S = S_0 e^{-b \cdot ADC}$$

$$\log(S) = \log(S_0) - b \cdot ADC \tag{1.3}$$

where S represents the diffusion weighted intensity in a specific voxel, S_0 is the signal intensity in the same voxel obtained without the application of diffusion gradients, and ADC is the apparent diffusion coefficient. The value of b, who represents a measure of the diffusion weighting, is defined by the following equation:

$$b = \gamma^2 G^2 \delta^2 \left(\Delta - \frac{\delta}{3}\right) \tag{1.4}$$

where γ is the gyromagnetic ratio, G is the amplitude of the diffusion gradient, δ represents the duration of each gradient and Δ is the interval between the onset of the

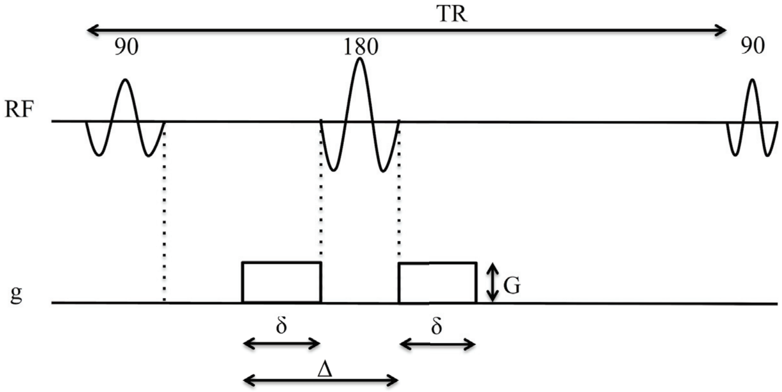


Figure 1.6: Representation of Pulsed Gradient Echo-Spin sequence. δ represents the duration of each gradient, Δ is the interval between the onset of the diffusion gradient before the refocusing pulse and that after the refocusing pulse, G is the amplitude of the diffusion gradient and RF indicates radiofrequency pulses.

diffusion gradient before the refocusing pulse and that after the refocusing pulse. Its unity is seconds per square millimetres. Typical values of b used in clinical applications range from 600 to 1500.

According to the equation of S it is then possible to obtain the value for the ADC in each voxel. The equation can be rewritten as follow:

$$ADC = \frac{\log \frac{S}{S_0}}{b}$$

ADC value is a quantitative parameter largely used to study and quantify the changes in diffusion given by the presence of different brain related pathologies [Albers (1998),Maier *et al.* (2010),Balashov and Lindzen (2012)].

2.3 Diffusion Tensor Imaging

As we showed in the last two sections, dMRI is a powerful tool to obtain a large range of interesting information by simply studying the diffusion of the water in the brain. Unfortunately, except for the ADC value, representing and exploiting dMRI information is not an easy task and a big effort in development of new mathematical models is needed. The first, and the most important formulation, is the diffusion tensor imaging (DTI) model described in [Basser *et al.* (1992),Basser *et al.* (1994)a]. The model is rather simple, yet a powerful method to obtain quantitative diffusion properties in the brain. Due to these characteristics the DTI model is still used today.

DTI starts from the assumption made in [Bloch (1946)]. The hypothesis is that the diffusion in each voxel follows a Gaussian distribution, and, as consequence, it follows just one main direction.

According to this tensor model, it is possible to rewrite equation 1.3 as follows:

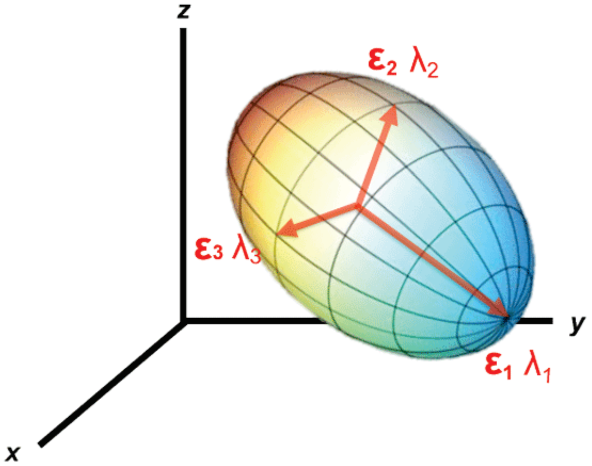


Figure 1.7: Representation of the diffusion tensor with its eigenvectors $\overrightarrow{\varepsilon_1}, \overrightarrow{\varepsilon_2}, \overrightarrow{\varepsilon_3}$ and eigenvalues $\lambda_1, \lambda_2, \lambda_3$. Image from `http://mriquestions.com/diffusion-tensor.html`.

$$\log\left(\frac{S}{S_0}\right) = -\sum_i \sum_j b_{ij} D_{ij}$$

where $b \in \mathbb{R}^{3\times3}$ is the extension of the equation 1.4 for the gradient in the 3d space:

$$b_{ij} = \gamma^2 G_i G_j \delta^2 \left(\Delta - \frac{\delta}{3}\right) \qquad i, j = x, y, z$$

The symmetric matrix $D \in \mathbb{R}^{3\times3}$ is the diffusion tensor matrix defined as follows:

$$D_{ij} = \begin{pmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{yx} & D_{yy} & D_{yz} \\ D_{zx} & D_{zy} & D_{zz} \end{pmatrix}$$

Diagonalization of this matrix allows to obtain the eigenvalues and the eigenvectors. The matrix D can be then written as:

$$D = \begin{pmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{xy} & D_{yy} & D_{yz} \\ D_{xz} & D_{yz} & D_{zz} \end{pmatrix} = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} \begin{pmatrix} \overrightarrow{\epsilon_1} \\ \overrightarrow{\epsilon_2} \\ \overrightarrow{\epsilon_3} \end{pmatrix}$$

$\lambda_1, \lambda_2, \lambda_3$ represent the eigenvalues and $\overrightarrow{\epsilon_1}, \overrightarrow{\epsilon_2}, \overrightarrow{\epsilon_3}$ represent the eigenvectors of the diffusion ellipsoid. Those values and vectors allows to obtain simple and clear information about the shape of the diffusion tensor model as shown in Figure 1.7.

2.4 DTI Derived Metrics

The obtained eigenvalues and eigenvectors are really important to understand diffusion properties of the diffusion tensor. Indeed, using the value of the eigenvalues it is possible to quantify the diffusion of each tensor. This information is useful to investigate structural properties of the tissue.

In [Kingsley (2006)b] the author reported a complete list of all the DTI derived metrics. We will start to describe those metrics according to their order of complexity. The first metric is the “axial diffusivity”. It is the value of the main eigenvalue (λ_1) and represents the part of the diffusion in a voxel which follows the principal diffusion direction. The “radial diffusivity” $\lambda_r = \frac{\lambda_2+\lambda_3}{2}$ who represents the part of the diffusion which follows the direction normal to the main eigenvector $\overrightarrow{\epsilon_1}$ identified by the eigenvector $\overrightarrow{\epsilon_2}, \overrightarrow{\epsilon_3}$. Another important metric is the “mean diffusivity” $MD = \frac{\sum_{i=1}^{3} \lambda_i}{3}$ in all three directions of a voxel. One of the most important, and used, diffusivity metric is the “fractional anisotropy” (FA):

$$FA = \sqrt{\frac{3}{2}} \frac{\sqrt{\sum_{i=1}^{3}(\lambda_i - MD)^2}}{\sqrt{\sum_{i=1}^{3} \lambda_i^2}}$$

this metrics give a quantitative measure ($0 \leq FA \leq 1$) about the anisotropy of the diffusion in a specific voxel. If in a voxel $FA = 1$, the diffusion in the specific voxel is completely anisotropic and thus it follows perfectly one direction. Otherwise, if in a voxel $FA = 0$ the diffusivity cannot be represented with a single direction since it follows all the directions in the space. In voxels with a highly structured tissue, like corpus callosum, usually we have high FA value ($FA \geq 0.8$). Contrarily, in voxels without well defined tissue, like in cerebrospinal fluid, low value sof FA are present ($FA \leq 0.2$).

Other new DTI derived metrics to measure the anisotropy are also proposed in [Prados *et al.* (2010)]. In their paper, the authors proposed the Compositional Kullback-Leibler Anisotropy (KLA) as a new anisotropy measure to study the properties of the brain tissue especially in regions in which grey and white matter components are mixed.

In order to give a practical application of the DTI derived metrics in brain investigation, in Chapter 3 we will discuss how those metrics can be used to study neurodegenerative pathologies.

2.5 Fiber Tracking

DTI is not only capable to extract quantitative maps useful to understand the effects of a disease, but it also allows to reconstruct the structure of the brain white matter (WM) thanks to the use of fiber tracking algorithms [Mori *et al.* (1999)]. Indeed, in the previous section we described how to exploit the eigenvalues in order to obtain quantitative information. In this section, we will describe how eigenvectors information can be exploited to reconstruct the inherent tissue.

We will start our dissertation by giving to the reader an intuition about the fiber reconstruction process. As we know from the previous section, the principal eigenvalue of the DT model gives, in a particular voxel, the main diffusion direction. So, following the main eigenvector of a specific voxel it is possible to follow the principal direction of the water. This direction will point to one of its contiguous voxels. The voxel reached by the diffusion direction can be used to find the direction pointing to the next contiguous one. This process can be repeated for a certain number of voxels since a termination criterion is met. According to this simple concept, it is then possible to connect, using smooth lines, those contiguous voxels in order to obtain all the fibers representing the WM structure. This process is simply described in figure 1.8. The first set of voxels, also called “seed”, used to start this iteration chain is usually selected in two different ways according to the type of tractography. For global brain tractography, usually the seed voxels are randomly selected from the whole WM. For structure analysis, like investigation of a specific WM tract, the seeds are selected by the user according to a specific anatomic knowledge i. e. atlas.

A large number of techniques have been proposed in the literature [Fillard *et al.* (2011), Jbabdi and Johansen-Berg (2011), Mangin *et al.* (2013)] and an exhaustive evaluation would be prohibitive. The algorithm we previously described belongs to a particular family of algorithms called *deterministic*. One of these algorithms was proposed in [Hagmann *et al.* (2007)]. Deterministic algorithms for tractography are quite fast and allow to obtain quite good results in terms of accuracy in WM fiber reconstruction. Limitation of this type of tractography is related to the accuracy of the path followed by the fibers. Indeed, these algorithms just follow one of the principal directions without taking into account other options that could give better results. In order to overcome this limitation, a new family of *probabilistic* algorithms was developed in [Behrens *et al.* (2003)]. Probabilistic algorithms repeat the deterministic version many times by randomly perturbing the main fiber directions each time, and produce maps of connectivity. Such maps indicate the probability that a given voxel is connected to a reference position [Fillard *et al.* (2011)].

Orientation Distribution Function

As previously said, DTI is a simple model based on the hypothesis that in each voxel the diffusion follows just one main direction [Bloch (1946)]. When we talk about brain structures, this hypothesis cannot be always satisfied due to the complex structure of the brain tissue. Indeed, it is well known that in a single voxel different types of tissue (with different diffusion directions) can be present as showed in Figure 1.9. This example provides a clear proof that the assumption made by DTI is not general enough to describe the real structure of WM tissue. A number of alternative models have been proposed to provide a more adequate description of the diffusion. One of the first models was proposed in [Tuch *et al.* (2002)]. In that paper, the authors fitted multiple diffusion tensors in order to represent the whole fiber population of fibers within each voxel. As main limitation, this method needs to know *a-priori* the number of diffusion directions in each voxel. More-

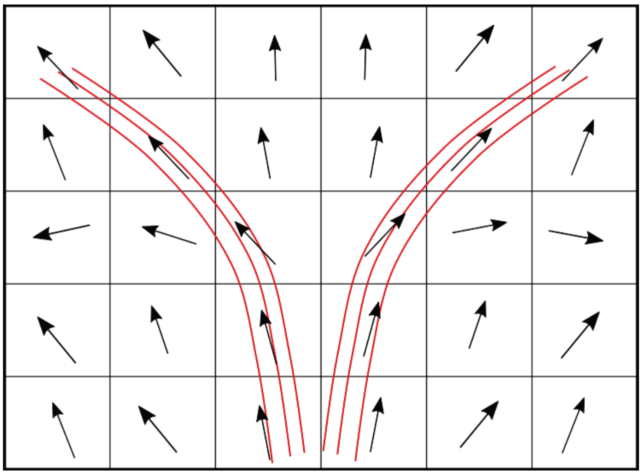


Figure 1.8: Schematic demonstrating the tractography algorithm using DTI information. Arrows represent primary eigenvectors in each voxel. Red lines are reconstructed trajectories.

over, this technique is unstable if more than two diffusion directions are fitted for each voxel [Tuch *et al.* (2002)]. Another method was proposed in [Wedeen *et al.* (2000)]. This approach allows to model the complex topology of WM fibers population by applying the Fourier-transform to diffusion data. Unfortunately, in order to compute the correct structure, this technique needs a large number of data samples. A good improvement of these models to describe the complex WM structure was proposed in [Tournier *et al.* (2004)] with the introduction of the orientation distribution function (ODF). In their paper, the authors proposed a method that is able to estimate directly the distribution of fiber orientations within a voxel from high angular resolution diffusion-weighted (HARDI) MR data without making prior assumptions regarding the number of fiber populations. Finally, with this method it was relatively simple and fast to compute the presence of multiple tissue types in a given voxel. With the introduction of this new formalism, the quality of the tractography algorithm was greatly improved as shown in Figure 1.10. Indeed, compared to the tractography algorithm based on DTI information (Figure 1.8), it is finally possible to reconstruct also fibers passing through voxels having multiple directions.

For sake of completeness, we should underline that, quantitative diffusion metrics can also be obtained from FOD. For instance, generalized fractional anisotropy (GFA) [Tuch (2004)], which is a HARDI anisotropy measure similar to the popular FA obtained from DTI [Cohen-Adad *et al.* (2011)] can be computed. Those new diffusion metrics, computed from other models different form DTI, are starting to be used in order to characterize neurodegenerative pathologies.

2.6 dMRI for Microstructure Imaging

In the last paragraph, we described different diffusion models which aim to estimate the diffusion direction(s) in each voxel. Their main interest is to give meaningful information to

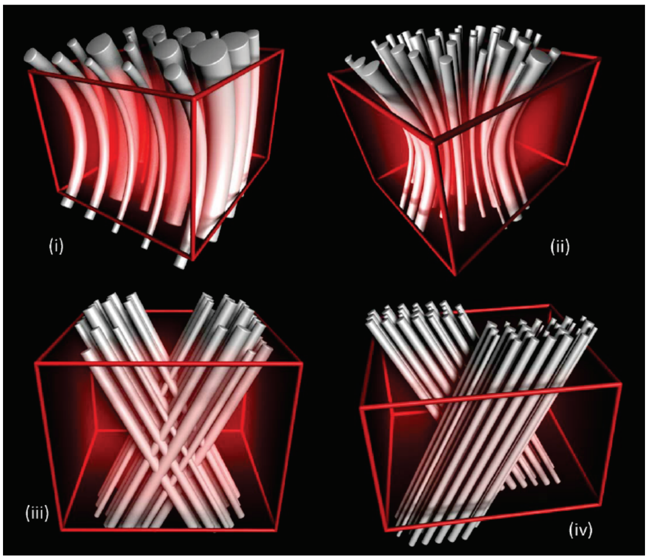


Figure 1.9: Different tissue architectures in one voxel. In *(iii)* and *(iv)* it is possible to see that the structure of the tissue, and consequently the diffusion directions does not follow just one main direction. Image from [Tournier *et al.* (2011)].

build a robust tractogram capable to reflect the WM structure of brain tissue. The problem related to studies based on tractogram analysis is related to their difficulty to generate quantitative information that can explain, for example, the "quality" or the properties of the tissue. Indeed, the number of fibers generated by the tractography algorithms can not be considered as a real quantitative measure [Jbabdi and Johansen-Berg (2011)].

In the last decade, a lot of efforts were made to extract, from diffusion data, quantitative information to explain and quantify, at the voxel level, the true underlying neuronal microstructure. The main idea of those methods, is to define "tissue specific" mathematical models capable to represent the different compartments of the neuronal microstructure. In this way, the diffusion signal measured in a voxel can be represented as the "combination" of the diffusion signals in those compartments. For example, if we analyze Figure 1.5 a simple model could be to define the diffusion signal in that voxel as the combination of the signal representing the water movement inside the axons and around the axons.

A good overview of the different mathematical models used to describe different compartments is presented in [Panagiotaki *et al.* (2012)]. In their work the authors showed, and validated, different mathematical structures useful to model intra-axonal, extra-axonal and isotropic restriction compartments. Unfortunately, some tissue parameters, such as permeability, remain elusive because mathematical models are intractable. Recently, this problem was overcome in [Nedjati-Gilani *et al.* (2017)]. In their work, the authors tried to define different compartment models by applying machine learning algorithms on simulated data with known tissue properties. Interestingly, the authors found that in-vivo results are consistent with pathology of MS lesions.

Recently, different techniques were developed in order to obtain quantitative information using compartment-based models. Two of the most successful models are Acti-

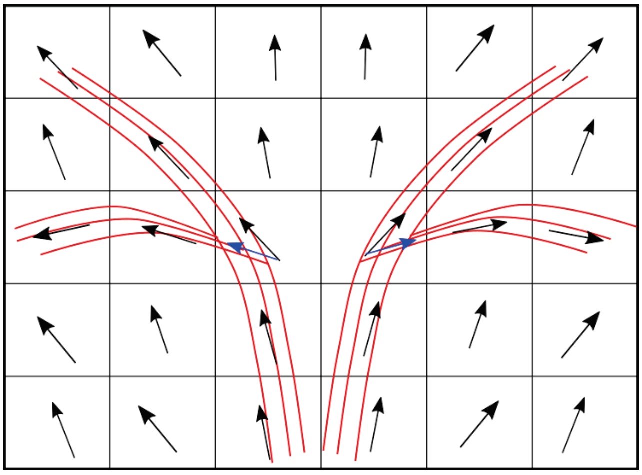


Figure 1.10: Schematic demonstrating the tractography algorithm FOD information. Arrows represent primary FOD peaks in each voxel. For voxels containing more than one population of fibers multiple principal directions have been found (blue arrows). Red lines are reconstructed trajectories.

vAx [Alexander *et al.* (2010)] and NODDI [Zhang *et al.* (2012)]. The main limitation of those models is related to the computation time needed to calculate, for a given image, the quantitative information in each voxel. In [Daducci *et al.* (2015)], the authors overcame those limitations proposing a convex optimization framework to perform quantification of micostructural information in each voxel.

Chapter 2

Multiple Sclerosis

Contents

1 Introduction

Multiple Sclerosis (MS) is a chronic disease of the central nervous system (CNS). It constitutes the leading cause of non-traumatic disability in young adults (from 20 to 40 years old) and remains without well-known etiology [Compston and Coles (2008)]. In more than two-thirds of the cases it affects mainly women. Multiple phenomena are known like demyelination, inflammation and neurodegeneration. Usually, demyelination and inflammation are considered as the initial and prominent mechanisms, while neurodegeneration is more prominent in progressive phases of MS.

2 Epidemiology

Epidemiology of MS is an interesting subject of study, indeed it allows to better understand how “geographic” factors can influence the development of MS.

A pioneering study on that field is the one described in [Kurtzke (1980)] and later updated in [Kurtzke (2000)]. These studies show the worldwide distribution of MS according to its index of prevalence in different countries.

In this section we will first discuss about the general prevalence of MS in different countries according to the results obtained in [Kurtzke (1980)] and next we will analyze some interesting results found in [Kurtzke (2000)].

As reported in [Kurtzke (1980)] and how it is shown in 2.1, the geography repartition of MS worldwide is not equally distributed. It is possible to identify “high risk” zones like: Scandinavia, Scotland, Europe, Canada and United States and “Low risk” zones like: south America or in Africa. Caucasoid people are more exposed to the risk to get MS compared to Negroid and Mongoloid people.

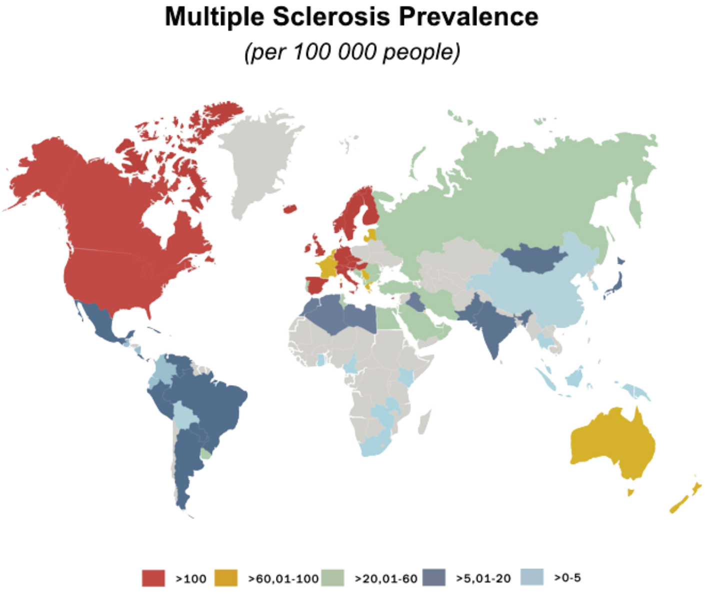


Figure 2.1: Worldwide Multiple Sclerosis prevalence. Image and data from https://www.msif.org/about-us/advocacy/atlas/.

In [Kurtzke (2000)] the authors studied the changes in the risk exposure to MS between populations migrating to different regions of risk. As a major result, the authors found that adolescents migrating before the age of 15, present the same risk factor of their original region. Contrary, adults who migrate acquire the risk factor typical of the new risk zone. These results show how the geographic partition of MS is related to environmental factors linked to a genetic susceptibility.

3 Pathophysiology

As we already said in the introduction, the main problem of MS is related to its etiology who still remains elusive [Wu and Alvarez (2011)]. Before going deeper in the discussion of the effect of MS on CNS, we need to give a brief description of the structure of the neurons.

Neurons

Neurons, are the basic units of the nervous system. They have a diameter ranging from 5 to 150 μm and they are composed by three parts: cell body, dendrites and axons (Figure 2.2). The cell body is divided in a nucleus and perinuclear cytoplasm. The cell body represents the most voluminous portion of the neuron, from the cell body dendrites and axon branch out. Dendrites are the expansion of the cytoplasmic membrane, they principally receive stimuli from axons or other neurons. Axons are responsible for the nerve impulse transmission and they are surrounded by myelin sheath cells in the CNS.

Myelin, a substance produced by oligodendrocytes, is a membrane who facilitates the nerve impulse transmission along the axons. Oligodendrocytes produce myelin just in short sections of the axon, this is why multiple cells are needed to surround the whole axon. Along the axons, certain sections, called *Nodes of Ranvier*, are not surrounded by myelin. Those small spots allow to the pulse to jump from a node to another, and thus to speed up the nerve impulse transmission along the axon. Moreover, it is important to underline that speed of impulse transmission is also correlated with the diameter of the axons, indeed, axons with larger diameter, show higher conduction speed [Preston and Shapiro (2005)].

Etiology

MS is characterized by an abnormal immune-mediated response who attacks the myelin around nerve fibers in the CNS inducing a progressive destruction of myelin. As myelin helps to speed up the nerve impulse transmission along the axon, a destruction of myelin decreases the capability of the axon to transmit the nerve impulse.

In MS, immune cells induce an inflammatory response who is the main cause of the demyelination in the axons. Inflammatory processes seem to start after a cell-mediated response. In this part, macrophages recognize the myelin basic protein (MBP) and present this to the T lymphocytes. These, after their activation, cross the blood brain barrier (BBB) and trigger the immune response and relative inflammation. Recent studies also suggest that the beginning of an antibody-mediated response, with an abnormal production of antibody for myelin destruction, plays an important role in the progression of the inflammation with relative tissue damage [Disanto *et al.* (2012)].

Demyelination in specific tissue area usually starts without axonal damage [Noseworthy *et al.* (2000), Lucchinetti *et al.* (2000), Compston and Coles (2008)], oligodendrocytes destruction and axonal damages are induced when repeated attacks appear in time.

At the beginning of the pathology, a remyelination process is opposed to the pathological demyelination one. During this process, the oligodendrocyte progenitor cells differentiate in oligodendrocytes in order to repair the damaged tissue [Goldschmidt *et al.* (2009), Brück *et al.* (2003)]. Unfortunately, the capability of the oligodendrocyte progenitor cells to differentiate oligodendrocytes is reduced in MS. This limitation influences the capability to recover the damaged tissue.

In the intermediate phase, the myelin contained in the tissue affected by the disease, is substituted with scarred tissue.

In the late part, demyelination effects are not present and the tissue area does not contain inflammatory cells. The increased permeability of the BBB and the inflammatory attacks increase the clinical effect related to the neurodegeneration and atrophy. In this part, when all the reparation mechanisms of the tissue are exhausted, the disability progression progressively increases.

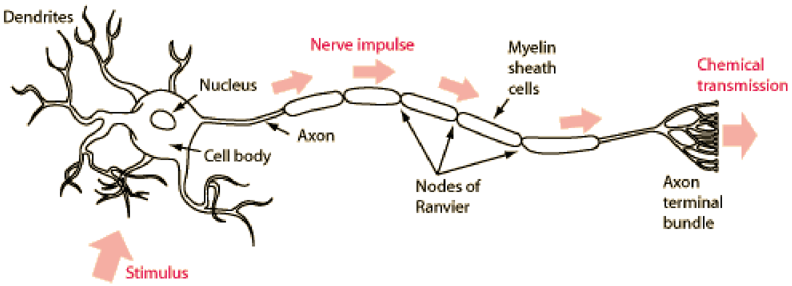


Figure 2.2: Structure of a nerve cell. Image from `http://hyperphysics.phy-astr.gsu.edu/hbase/Biology/nervecell.html`.

4 Clinical Forms

Since the progression of MS is not equal for each patient, the accurate prediction of MS evolution is still an open challenge. According to the current clinical standards four forms of MS are actually recognized (Figure 2.3) [Lublin *et al.* (1996), Lublin *et al.* (2014), McDonald *et al.* (2001)].

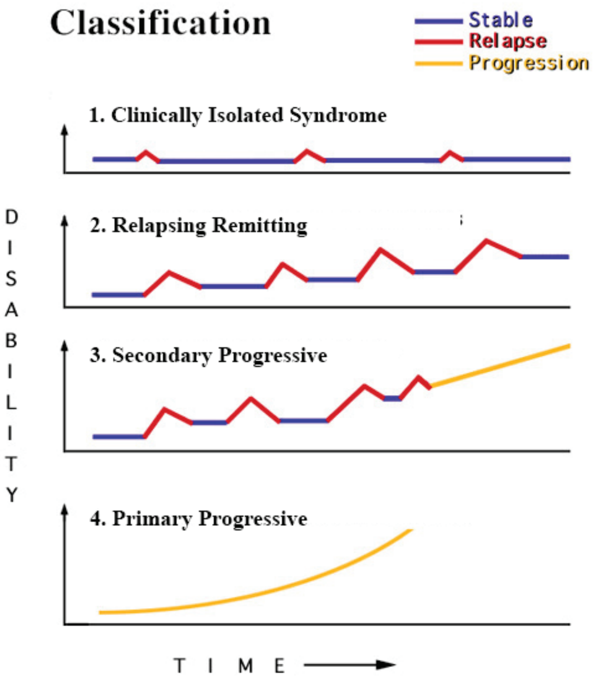


Figure 2.3: Classification of multiple sclerosis clinical forms according to the patients disability progression. Image from `http://library.med.utah.edu/kw/ms/mml/ms_class.html`

4.1 Clinically Isolated Syndrome (CIS)

The first manifestation of the MS usually happens as a Clinically Isolated Syndrome (CIS). This form of MS is the consequence of the series of one or two consecutive attacks from which the patients recovered completely, generally quite quickly, and without any progression or persistence of disability. Following this first attack, the progression to a relapsing remitting form (RR) may occur more (few months) or less rapidly (few years) depending on the patient.

4.2 Relapsing Remitting (RR)

In this form of MS, we often observe unpredictable attacks. During the presence of those new attacks new clinical symptoms appear or old clinical symptoms evolve. This phase has a variable duration and could be followed by a partial or total remission. At this stage, the pathology can be inactive for months or years.

4.3 Secondary Progressive (SP)

Patients with a RR form can then develop a secondary progressive (SP) form characterized by increased symptoms and a level of disability that increases independently of the presence of inflammatory attacks.

4.4 Primary Progressive (PP)

Primary progressive (PP) form is characterized by the absence of inflammatory attacks. The patients suffer from an accumulation of deficits and disabilities. All these effects can be stable for certain periods or can progressively degenerate in months or years.

5 Diagnosis

Since MS is characterised by multifocal effects that can change during the time, its diagnosis is not trivial. Diagnosis of MS is based on the multifocal and evolutive characteristics of the pathology. During the time, in order to help the physician to perform a correct diagnosis, different recommendation based on MRI were suggested. The most important and useful criteria for physician were identified in [McDonald *et al.* (2001)] by collecting all the clinical and radiological knowledges of the pathology. Those criteria take into account not only local effects of MS visible in a specific time-points, but also the history of the evolution of the pathology. After their publication, these criteria were largely used in clinical practice. Thanks to the advance in clinical research, these criteria were then updated in 2005 [Polman *et al.* (2005)] and 2010 [Polman *et al.* (2011)].

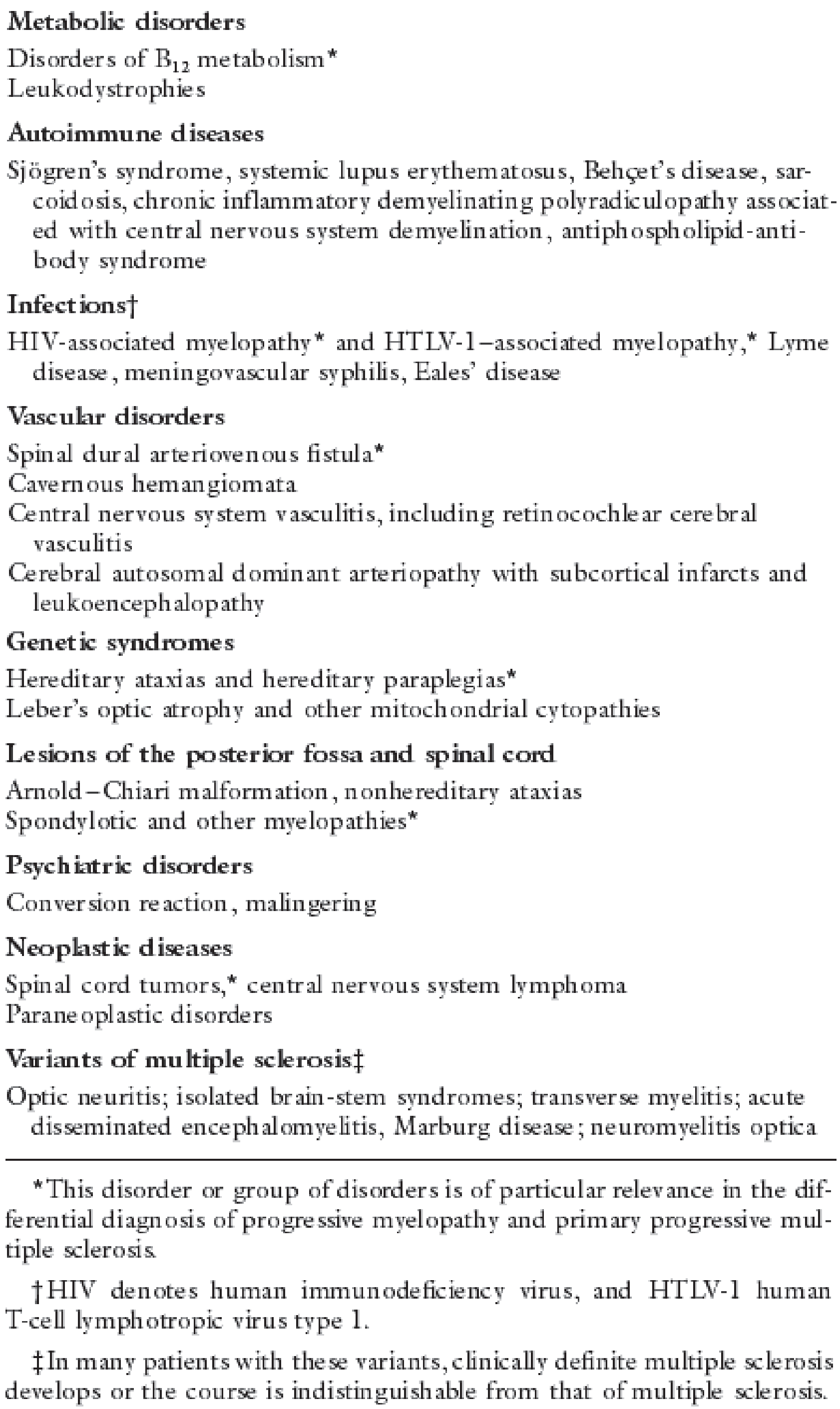
Currently, two main approaches are used to perform a correct diagnosis of MS: differential diagnosis and positive diagnosis.

5.1 Differential Diagnosis

Generally speaking, differential diagnosis is relatively simple. In order to perform a differential diagnosis, the physician need a table containing the symptoms of the pathology and check if the patient has a subset or all the symptoms. In the diagnosis of MS, this task can be really hard, indeed, due to the large variety of MS symptoms. It is not possible to clearly define a complete and robust table of criteria describing the effect of the pathology. Moreover, since the symptoms of MS evolve and change with time, the use of the differential diagnosis is not sufficient to have a general overview of the pathology. An example of a differential diagnosis table for MS, from [Noseworthy *et al.* (2000)], is shown in Figure 2.4.

5.2 Positive Diagnosis

Positive diagnosis is more complex and more effective compared to the differential diagnosis. It strongly uses MRI to perform the diagnosis since MRI is the best techniques to obtain *in vivo* images of MS lesions. Indeed, with MRI it is possible to follow in time and in space how MS lesions evolve in the brain. The positive diagnosis is based on four essential criteria: *i)* temporal lesions dissemination, *ii)* spatial dissemination of MS lesions in particular regions of the CNS, *iii)* presence of inflammatory processes in the CNS, and *iv)* absence of other progressive diseases. In order to have a full picture of the damages



Metabolic disorders

Disorders of B_{12} metabolism*
Leukodystrophies

Autoimmune diseases

Sjögren's syndrome, systemic lupus erythematosus, Behçet's disease, sarcoidosis, chronic inflammatory demyelinating polyradiculopathy associated with central nervous system demyelination, antiphospholipid-antibody syndrome

Infections†

HIV-associated myelopathy* and HTLV-1–associated myelopathy,* Lyme disease, meningovascular syphilis, Eales' disease

Vascular disorders

Spinal dural arteriovenous fistula*
Cavernous hemangiomata
Central nervous system vasculitis, including retinocochlear cerebral vasculitis
Cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy

Genetic syndromes

Hereditary ataxias and hereditary paraplegias*
Leber's optic atrophy and other mitochondrial cytopathies

Lesions of the posterior fossa and spinal cord

Arnold–Chiari malformation, nonhereditary ataxias
Spondylotic and other myelopathies*

Psychiatric disorders

Conversion reaction, malingering

Neoplastic diseases

Spinal cord tumors,* central nervous system lymphoma
Paraneoplastic disorders

Variants of multiple sclerosis‡

Optic neuritis; isolated brain-stem syndromes; transverse myelitis; acute disseminated encephalomyelitis, Marburg disease; neuromyelitis optica

---

*This disorder or group of disorders is of particular relevance in the differential diagnosis of progressive myelopathy and primary progressive multiple sclerosis.

†HIV denotes human immunodeficiency virus, and HTLV-1 human T-cell lymphotropic virus type 1.

‡In many patients with these variants, clinically definite multiple sclerosis develops or the course is indistinguishable from that of multiple sclerosis.

Figure 2.4: Differential diagnosis table of Multiple Sclerosis. Table from [Noseworthy *et al.* (2000)].

generated by MS in the brain the following three MRI sequences are needed to detect different types of lesions:

- T2 sequence which allows to detect MS lesions as hyper-signal spots

- T1 sequence where lesions are characterized by hypo-intensity signal spots
- T1 sequence acquired after injection of a contrast agent (like Gadolinium) to detect regions where disruption of the hematoencephalic barrier is present. With this sequence it is possible to differentiate between active lesions (lesions with signal hyper-intensity) and chronic lesions (lesions with signal hypo-intensity)

Information derived from MRI was used in [Barkhof *et al.* (1997)] to derive MRI based criteria for the diagnosis of MS. Those criteria were used and updated in [McDonald *et al.* (2001),Polman *et al.* (2005),Polman *et al.* (2011)]. In order to have a correct diagnosis of MS two different observations are needed: spatial and temporal dissemination of lesions. To satisfy the presence of spatial dissemination of the lesions, at least two of these four criteria must hold:

- Lesion showing hyper-intensity in T2 sequence in periventricular region
- Lesion showing hyper-intensity in T2 sequence near to the cortex
- Lesion showing hyper-intensity in T2 sequence in the supra tentorial region
- Lesion showing hyper-intensity in T2 sequence in the spinal cord

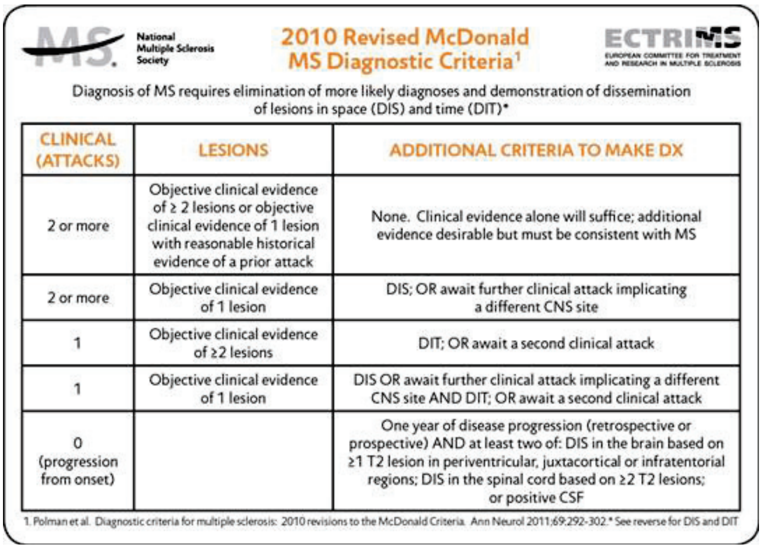
To satisfy the presence of temporal dissemination of the lesions the following criteria must hold:

- Lesion showing hyper-intensity in T1 acquired using contrast agent three months after one attack
- A new lesions showing hyper-intensity in T2 sequence three months after the first MRI exam
- If the first MRI exam was performed 3 months after the first attack, all the new lesions showing hyper-intensity in T2 sequence are considered as expression of a temporal dissemination

An overview of the spatial and temporal dissemination criteria are summarized in Figure 2.5.

6 Medical Treatment

As we already discussed, MS is a really complex disease, and, as consequence, also its treatment is complex since it strongly depends of the status of the disease when the patient undergoes the clinical exam. The classical treatment of MS is usually divided in three parts: *i)* treatment of the attacks, *ii)* treatment of the pathology, *iii)* treatment of the symptoms.



National Multiple Sclerosis Society

2010 Revised McDonald MS Diagnostic Criteria[1]

ECTRIMS European Committee for Treatment and Research in Multiple Sclerosis

Diagnosis of MS requires elimination of more likely diagnoses and demonstration of dissemination of lesions in space (DIS) and time (DIT)*

CLINICAL (ATTACKS)	LESIONS	ADDITIONAL CRITERIA TO MAKE DX
2 or more	Objective clinical evidence of ≥ 2 lesions or objective clinical evidence of 1 lesion with reasonable historical evidence of a prior attack	None. Clinical evidence alone will suffice; additional evidence desirable but must be consistent with MS
2 or more	Objective clinical evidence of 1 lesion	DIS; OR await further clinical attack implicating a different CNS site
1	Objective clinical evidence of ≥2 lesions	DIT; OR await a second clinical attack
1	Objective clinical evidence of 1 lesion	DIS OR await further clinical attack implicating a different CNS site AND DIT; OR await a second clinical attack
0 (progression from onset)		One year of disease progression (retrospective or prospective) AND at least two of: DIS in the brain based on ≥1 T2 lesion in periventricular, juxtacortical or infratentorial regions; DIS in the spinal cord based on ≥2 T2 lesions; or positive CSF

1. Polman et al. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald Criteria. Ann Neurol 2011;69:292-302.* See reverse for DIS and DIT

Figure 2.5: MS positive diagnosis criteria. Image from http://www.nationalmssociety.org/For-Professionals/Clinical-Care/Diagnosing-MS/Diagnosing-Criteria.

6.1 Treatment of the attacks

The goal of this type of treatment is to reduce the number and the frequency of the attacks and then reduce the progression of the disability by increasing the remission. This type of treatment is mainly based on the assumption of corticoids usually *methylprednisolone*. They are usually prescribed by infusion of 1 gram per day for 3 days.

6.2 Treatment of the pathology

The treatment of the pathology tries to control its evolution. The main idea is to reduce the demyelination by stimulation of the remyelination process. This treatment relies on the auto-immune nature of MS. Immunomodulators and Immunosuppressors drugs are often use to reduce the effect of MS. In the category of immunomodulators, we recall the interferon β and the Tysabri®. For the Immunosuppressors, mitoxantrone and cyclophosphamide are the most used.

6.3 Treatment of the symptoms

Treatment of the symptoms aims to improve the quality of life of the patients. They mainly help to reduce the manifestation of the symptoms without directly affecting the pathological process. For instance, for the spasticity injection of botulinum toxin or assumption of antispasmodics are used. Pain is usually treated with the assumption of analgesics or antiepileptic drugs.

7 Conclusion

In this chapter we briefly introduced and discussed MS. We gave a general overview about the pathology and its epidemiology. We discussed its physiopathology and the different criteria used for its diagnosis. Finally, we showed the four different clinical forms of the pathology and the classical treatment used to reduce the effect of the pathology.

From this general introduction, we can already see how cMRI acquisition is a powerful tool for clinicians to diagnose and study the effect of MS on the CNS. Those results encouraged researchers on MS to use more complex MRI techniques like diffusion imaging and spectroscopy to obtain more sensitive and specific information on normal appearing WM useful for a deeper investigation of the pathology.

Chapter 3

The Role of Diffusion MRI in White Matter Investigation

Contents

1 Introduction

Diffusion MRI (dMRI) is a non-invasive imaging technique that, to date, is unique in that it can be used to reveal the microstructure of the WM of human brain [Assaf *et al.* (2013)]. Due to this advantage, it is largely used to investigate the effects that neurodegenerative pathologies, like MS, have in the CNS.

The results achieved after the introduction of dMRI in the study of MS were interesting. Indeed, after the introduction of classical MRI sequences (T1, T2, T2 FLAIR) used for the MS diagnosis, poor correlations were found between lesion load measurements and patients' disability [Barkhof (2002)]. The identification of this so-called "clinico-radiological paradox" has led to several studies utilizing a multitude of MRI strategies such as magnetization transfer, spectroscopy, and DTI [Rovira *et al.* (2013)].

In case of complex diseases, like MS, dMRI seems to present a better correlation with clinical disability [Ciccarelli *et al.* (2001)]. The pathologic damage detected by dMRI derived metrics, like the one derived from DTI, in clinically eloquent NAWM regions, showed a significant contribution in estimation of MS disability [da Cruz *et al.* (2011)]. Moreover, diffusion studies indicate that the severity of damage within T2-visible lesions and in the normal appearing grey matter (NAGM), as well as in important WM tracts, has a significant effect on neurologic disability [Rovaris *et al.* (2005)]. dMRI derived metrics may thus contribute to composite MR-based scores, explaining otherwise unexplained variance in MS-related disability [Pulizzi *et al.* (2007), Mainero *et al.* (2001)]. These promising results encouraged a deeper investigation of MS using dMRI derived metrics and speed-up their use also in clinical setting. The main limitation of use of dMRI in real clinical settings is related to the acquisition time needed to perform a robust and reproducible acquisition [Tournier *et al.* (2004)], although further work is required to deliver ultrahigh-resolution imaging in clinically relevant time frames [Koenig *et al.* (2013)].

In this chapter, we will give an overview of different studies where authors focused their investigations on the effect of MS in the CNS using dMRI. We will introduce those studies according to their order of technical complexity.

In Section 2, we will start from the description of classical studies based on the analysis of diffusion metrics in pre-segmented regions of interest (ROIs) or using voxel based techniques like Tract-Based Spatial Statistics (TBSS) [Smith *et al.* (2006)]. We will focus our attention on the correlations between diffusion derived metrics and disease found by those studies.

In Section 3, an introduction on the new graph-based metrics derived from dMRI is given. More in detail, we will describe several studies where authors used structural information derived from the tractography to analyze the alterations of the CNS in MS patients.

Finally, in Section 4, we will investigate one of the last trends in medical research: the use machine learning algorithms. In more detail, we will show how the use of diffusion derived metrics as features for machine learning algorithms can be useful to help clinicians to take their decision. In particular, we will show how the sensibility of diffusion metrics is helpful to automatically classify the different MS clinical forms.

2 dMRI Studies in Pre-Segmented ROI

ROI based analysis is the principal way to analyze dMRI derived metric in specific regions of the brain. The main idea is to select brain regions and quantify the value of diffusion metrics, usually DTI derived metrics like FA, MD, λ_r, in these regions. This approach allows to perform complex studies which aim to understand how the handicap (other symptoms) caused by a pathology are related to damages detected in particular brain regions.

A second common approach, is to perform a global analysis of the WM using specific tools like TBSS [Smith *et al.* (2006)]. This analysis allows to detect, and study, diffuse effects of the pathology in the whole WM. Thanks to this approach, it is possible to have a global picture of the brain damages in relation to particular symptoms.

2.1 Region Specific Analysis

A first important study is the one described in [Filippi *et al.* (2001)]. In their paper, the authors quantified FA in lesions and NAWM regions. The regions were delineated on images of control participants and on a dataset composed by RR, SP, and PP MS patients. As major result, the author found that average FA in lesions was lower than the corresponding quantities of the NAWM. Moreover, in the overall patient sample, the authors also found that DTI-derived measures are moderately correlated with clinical disability, especially in SP MS patients. In this study, the authors suggested to use DTI as a tool for the monitoring of advanced phases of the disease.

A similar pioneering study is also presented in [Ciccarelli *et al.* (2001)]. In their paper, the authors focused their analysis in specific WM regions. More in detail, they quantified in MS patients and control subjects quantitative diffusion indices, including FA, volume ratio, and MD. The authors analyzed 30 ROIs located in normal-appearing basal ganglia, cerebellar gray matter, and supratentorial and infratentorial NAWM. In the study, interesting correlations between DTI metrics and clinical score were found. More in detail, the authors showed that both FA and MD in the cerebral peduncles were inversely correlated with the Expanded Disability Status Scale (EDSS) and pyramidal functional scores. In RR patients, a strong correlation between EDSS and FA in both supratentorial and infratentorial NAWM were also found. Finally, the authors found that in PP and SP MS patients the disease duration correlated strongly with MD in infratentorial NAWM and FA in the cerebral peduncles, respectively.

From these two important studies a large number of papers showing correlations between DTI and MS symptoms started. Those new studies started to investigate important regions in the brain like corpus callosum (CC). In [Sigal *et al.* (2012)] the authors investigated the relations between CC microstructure integrity by using DTI in RR MS patients. As major result, the authors reported that DTI parameters, namely FA, MD, $\lambda_1,\ \lambda_2,\ \lambda_3$ correlated with neurological disability. Moreover they found correlations, in CC, between λ_1 and $\lambda_2,\ \lambda_3$ and disease duration. Two studies describing relations between MS and DTI parameters quantified in CC are reported in [Rueda *et al.* (2008), Braley *et al.* (2012)]. In both papers, the authors describe how changes in FA and MD, detected in different regions of the CC, reflect axonal damage induced by MS.

Finally, we report an interesting region-based study which shows how dMRI is a powerful tool also for investigation of GM damage. In [Hannoun *et al.* (2012)b], the authors investigated DTI in deep GM. Left and right caudate and thalamus were manually segmented in RR, SP MS patients and control subjects. Volume, mean FA and MD in each structure were quantified. The authors found that FA was significantly increased in the caudate and the thalamus of patients with MS compared with controls, and was higher in SP compared with RR patients. Increased FA was associated with volume decreases of caudate and thalamus in patients with MS. Caudate FA, and, to a lesser extent, thalamic FA, were associated with functional deficits, as measured by EDSS and MSFC.

2.2 Global Brain Analysis

In the previous section, we described several published works based on regional analysis of the brain. In this section, we will focus our attention on studies based on global investigation of the brain in MS patients. All the studies described in this section, are based on TBSS [Smith *et al.* (2006)], a neuroimaging tool that allows to perform statistical voxelwise analysis of DTI metrics in different group subjects. Contrary to the previous methods, the results of this analysis will generate a map describing all the WM voxels who are significant different in the two groups of subjects. Moreover, compared to classical ROI based methods, TBSS analysis is more reproducible since it does not depend of the segmentation performed by the expert operator.

One of the first interesting papers about TBSS and MS is described in [Liu *et al.* (2012)]. In their work, the authors simply analyzed the changes visible through TBSS between RR MS patients and control subjects. From all the diffuse changes in WM, the authors identified specific regions correlating with clinical score like EDSS and disease duration. The diffuse effects of MS on WM is clearly confirmed by the results obtained. Indeed, MS patients had significant decrease of FA, and increased MD, λ_1 and λ_r in different WM regions. Significant clinical correlations between the diffusion metrics and the EDSS scores were identified, such as increasing EDSS and decreasing FA in the splenium of CC, left cingulum bundle and bilateral cortico-spinal tracts. Regions with significant negative correlations between the FA and the disease durations were identified in the whole corpus callosum and bilateral cingulum bundles.

Recently, more specific studies were conducted using TBSS to investigate particular effects of other MS related symptoms. In [Shen *et al.* (2014)] authors applied TBSS to RR MS patients and control subjects to study the presence of correlations between statistically significant regions found using TBSS and Hamilton depression rating scale (HAMD). As major finding, the authors showed a positive correlation between HAMD and FA changes in the right posterior middle cingulate gyrus, the right hippocampus, the left hypothalamus, the right precentral gyrus, and the posterior cingulate. Those results demonstrated a link between the depressive symptoms and clinically relevant brain areas in RR MS patients.

Global brain studies are also interesting for their capability to discover new predictors of disability that could help to find the best medical treatment of the patient. Two studies that show how it is possible, through global brain analysis, to identify disability predictors are reported in [Kern *et al.* (2011)] and [Asaf *et al.* (2015)]. In [Kern *et al.* (2011)] the authors focused their attention to the detection of DTI metrics capable to predict future worsening of hand function in RR MS patients. Disability of each patient was assessed by the nine-hole peg test (NHPT). Identification of important diffusion metrics was performed using a two-step processing pipeline. In the first step, the TBSS map was computed in order to detect the voxels which are significant different in the two groups. In the second step, a probabilistic tractography algorithm was performed in order to extract CST and the transcallosal hand motor fibers (TCHM). After the analysis, the authors found that λ_r of the TCHM fibers was predictive of NHPT decline over the next 12 months. The conclusion is that the TCHM fibers may play an important role in modifying the effects of MS pathology on fine motor control, and λ_r in these fibers may be a sensitive biomarker for future disability.
In the second study, described in [Asaf *et al.* (2015)], the authors aimed to elucidate the relationship of microstructural WM damage in patients with varying periods of disease duration. They acquired different control subjects and also MS patients with different disease durations. They grouped them in three groups: patients with short, moderate and long disease duration. TBSS was then applied on MS patients vs control and on the different MS groups. As expected, significant results were found between control subjects and MS patients belonging to different disease duration. The most interesting results was obtained in the comparison between short with long disease duration groups. They found significant differences in 30 different WM tracts with a particular predominance in the body of the CC. This difference show how the alterations detected at the beginning of the pathology reach a plateau within the next 5 years, and, only later in time, additional WM changes are detected. This 5 year window could be an important period to introduce specific treatments to minimize the long term effects of the pathology.

3 Analysis of MS using dMRI Connectivity Information

As we already said in Chapter 2, dMRI is not only capable to extract quantitative maps useful to understand the effects of a disease, but it allows also to reconstruct the structure of the WM tissue thanks to fiber tracking [Mori *et al.* (1999)]. This information can be used to investigate: *i)* the relationship between GM and WM degeneration in MS *ii)* alterations in brain connection.

3.1 GM and WM Relations in MS

Combination of tractography with cortical measures is an useful approach to test specific hypotheses related to clinically relevant functional systems in MS [Gorgoraptis *et al.* (2010)]. Using this methodology, an interesting study that we want to underline is the one described in [Bergsland *et al.* (2015)]. In their paper, the authors investigated the relationship between WM injury and cortical atrophy development in RR patients. In order to study the relation between WM and GM damages, the authors studied the association between focal and NAWM injury of the cortico-spinal tract (CST) with thickness and surface area measures of the primary motor cortex. The authors found that in MS patients, decreased cortical thickness was related to increased cortico-spinal tract NAWM mean, axial and radial diffusivities in addition to cortico-spinal tract lesion volume. This study showed that microstructural damages, detected by DTI, manifest their effect also in connected regions in the primary motor cortex.

A similar study was also proposed in [Tillema *et al.* (2016)]. The authors investigated the relations between the connectivity of cortical and WM lesions using novel MRI methodology including DTI. More in detail, they segmented GM and WM lesions and used DTI to reconstruct the connectivity maps between those two regions. The authors found an elevated connectivity between GM and WM lesions. These findings provide strong evidence of inter-lesional connectivity between GM and WM lesions, supporting the hypothesis of presence of intrinsic GM and WM lesions connectivity.

Other studies instead, focused their analysis to the investigation of relations between specific GM structure damages and WM. In [Anderson *et al.* (2011)] the authors assessed the cerebellar WM and GM atrophy and the degree of fibre coherence in the main cerebellar connections. The authors investigated the contribution of those atrophy measures in disability identification in RR and PP MS patients. For each group of MS patients, the authors compared tractography-derived measures from the middle and superior cerebellar peduncles, and quantitative diffusion measures like FA, MD and the other directional diffusivities. As major result, the authors found significant relationships between superior cerebellar peduncle FA and upper limb function, and between superior cerebellar peduncle FA, MD, λ_r and speed of walking. These findings indicate how a reduced fibre coherence in specific WM regions, visible through DTI, is related to motor deficit PP MS patients.

3.2 Analysis of MS using Structural Connectivity

Evolution of dMRI gave the opportunity to explore new biomarkers for investigation of WM connectivity alterations. In this section, we will give a general overview of different studies showing the application of WM connectivity to understand global WM alterations in MS.

Connectivity based measures represent a new set of interesting biomarkers to explore alteration in brain [Rubinov and Sporns (2010)]. Based on the analysis of WM fibers networks, a simple description of structural brain connectivity was introduced through the application of a geometrical graph representation [Shuman *et al.* (2013)]. This graph theory approach has become a sensitive tool to detect alterations in brain pathologies by providing both local and global characterization of WM connections [Achard *et al.* (2012)].

In [Li *et al.* (2013)], the authors performed one of the first studies using WM connectivity in MS. They compared WM connectivity graphs obtained from RR MS patients and control subjects. From the generated graphs different metrics were computed in order to quantify the "quality" of WM connectivity. The authors found structurally relevant alterations in fiber topology in early RR MS patients. Moreover, correlation between the 25-foot Walk test and the connectivity measured in the left superior frontal and the left superior temporal gyrus were found. They also showed an increase in interhemispheric and intrahemispheric white matter connectivity, probably related to compensatory mechanisms.

Similar studies were also performed in [Kocevar *et al.* (2016)] and [Nigro *et al.* (2015)]. In [Kocevar *et al.* (2016)], the authors studied how global graph-derived metrics change according to the different clinical forms of MS and how they can discriminate each clinical form. The obtained results showed that those new graph-derived biomarkers are really helpful to discriminate MS clinical forms. The authors also used this information to build an automatic classifier as we will see in the next section.

Finally, in [Nigro *et al.* (2015)] the effect of the major depression disorder, present in certain MS patients was studied using brain connectivity. The authors performed the analysis on MS patients with and without major depression disorder and control subjects as well. In this study the authors found some evidence that major depression disorder in MS may be linked with connectivity alterations at the limbic-motor interface, a group of regions that translates emotions into survival-oriented behaviors.

4 Automatic Classification of MS Clinical Forms using dMRI

Today's neurologist challenge is to predict the individual patient evolution and response to therapy based on the clinical, biological and imaging markers available from disease onset. In this section, we will focus on automatic classification of MS patients in different clinical forms based on dMRI information. We will show how different papers available in the literature tried to solve this prognostic question using a computer-based method. Due to the unknown etiology of MS and the variability of the patients' clinical history, "model based" approaches are not suitable or difficult to formalize. This limitation could be overcome using "data-driven" approaches based on machine learning algorithms. A first attempt was made in [Fiorini *et al.* (2015)]. In their paper, the authors used only clinical information to automatically classify MS patients. If this method allows to obtain good performance, it still needs clinical data to perform an accurate classification. In [Wottschel *et al.* (2015)], the authors extended this concept to MRI data. In particular, they extracted several features form cMRI data in order to differentiate between clinically-definite Multiple Sclerosis and CIS MS patients. This method reached acceptable performance but it contains a major limitation. Indeed, the extraction of the features is strongly operator dependent since it needs some manual pre-processing operation. Moreover, this method is not capable to exploit all the MRI data. In [Stamile *et al.* (2015)c], [Muthuraman *et al.* (2016)] and [Kocevar *et al.* (2016)], the authors tried to use machine learning using diffusion data without performing any kind of manual *a-priori* feature selection or definition.

In [Stamile *et al.* (2015)c], authors extracted DTI derived metrics in different WM fiber-bundles obtained from an atlas. They tried to build a classifier capable to automatically detect if DTI data, acquired at a specific time-point, belongs to control subjects, RR or SP MS patients. The authors found that specific diffusion metrics, extracted in specific WM fiber-bundles, are capable to clearly classify MS patients in the appropriate clinical form with high levels of accuracy.

In [Kocevar *et al.* (2016)] and [Muthuraman *et al.* (2016)], the authors exploited differently the dMRI data of MS patients. Indeed, in both papers, the authors, instead of using value of DTI derived metrics quantified in specific WM regions, they used graph metrics obtained from WM connectivity. In [Muthuraman *et al.* (2016)] this information was used only to discriminate RR vs CIS patients while in [Kocevar *et al.* (2016)] the information was used to classify all the clinical forms of MS patients. Interestingly, in both papers the results present high values in terms of classification accuracy. Indeed, they can automatically classify MS clinical form with a small error.

5 Conclusion

In this chapter, we gave a general overview of dMRI application in the study of WM alterations in MS patients. We showed that diffusion biomarkers give the opportunities to investigate complex mechanisms of brain diseases. We focused our analysis on previous works where dMRI was used to investigate the hidden relations between microstructural damage and disability evolution in MS.

In the first part, we showed two different ways to analyze dMRI data using region specific or global approach. The former allows to investigate correlations of specific brain structures and symptoms, whereas the latter allows to have a global picture of the diffuse effects of a neurodegenerative pathology in the whole WM. In order to show the potential of dMRI, we described several studies showing direct connection between quantitative findings and disability progression.

In the second part, a general overview of the new quantitative information that can be derived from structural architecture of WM is given. In the first section, we described how structural information derived from tractography can be coupled with quantitative information derived from DTI metrics to investigate connections between WM and GM degeneration. In the second section, we showed how this concept can be extended to create a graph describing the whole connectivity map of the brain. This map of connections can be used to extract specific metrics for the investigation of complex effects of neurodegenerative mechanisms.

Finally, in the last part, we showed how dMRI can be used to help clinicians to establish the diagnosis of MS in the specific clinical form. The results obtained with dMRI in automatic classification of MS clinical forms, underlined, what we already said in the introduction: dMRI derived metrics may contribute to composite MR-based scores, explaining otherwise unexplained variance in MS-related disability.

Chapter 4

Topic of the thesis

Longitudinal analysis of neuroimaging data is becoming an important research area. In the last few years, analysis of longitudinal data become a crucial point to better understand pathological mechanisms of complex brain diseases such as MS where WM fiber-bundles are variably altered by inflammatory and/or demyelinating events. Since pathological mechanisms remained unknown in certain brain diseases, the investigation of their temporal progression using non-invasive neuroimaging techniques is essential to better understand and predict the disease evolution and manage the therapeutic treatment [Nygaard *et al.* (2015), Mak *et al.* (2015), Vandermosten *et al.* (2015)].

As the etiology of MS as well as the pathological mechanisms including inflammatory and neurodegenerative processes, are not well understood, longitudinal studies using non-conventional MRI techniques such as DTI, providing sensitive markers, like FA and λ_r, constitute the best potential for the characterization of brain tissue alterations. For example, the analysis of grey matter (GM) structures [Mesaros *et al.* (2011)] showed the capability to evaluate the dynamics disability progression; while in white matter (WM) [Rovira *et al.* (2013)] a relationship between damaging and reparative mechanisms that occur in the lesions formation is underlined.

In [Colby *et al.* (2012)], the authors combined, for the first time, quantitative information extracted from the diffusion maps with the structural information obtained from fiber tracking. They showed that this new combination constitutes a powerful tool to investigate the diffuse effects of different pathologies. This new way to analyze WM was then used in different works [Yeatman *et al.* (2012), Mårtensson *et al.* (2013)] showing interesting results useful for the investigation of different pathological mechanisms. Unfortunately, these global approaches to analyze disease evolution are not sensitive enough to detect small and short-term (daily/weekly) longitudinal variations occurring typically in RR or CIS MS patients. A local scale approach is thus necessary to detect the presence of small "pathological" changes that could only affect a small subset of the WM fiber-bundles.

In this work, we will describe different techniques to analyze longitudinal changes occurring along WM fiber-bundles in MS patients. We will provide different methods capable to analyze at different levels of granularity longitudinal diffusion data.

Since the analysis of WM fiber-bundles is not an easy task and requires different data processing steps, we divided the processing needed in two macro areas: *i)* fiber-bundle extraction and *ii)* longitudinal fiber-bundle analysis described in Part II and IV respectively.

In Part II, we will illustrate a new string-based formalism to extract, from the whole tractography, just specific fiber-bundles. In the first chapter, we will introduce the string-based formalism used to represent WM fibers. We will also describe two algorithms, one semi-supervised and one unsupervised, to extract specific fiber-bundles. We will test our method using simulated diffusion phantoms and real data from a healthy subject. In the second chapter, we will extend the proposed algorithm in order to couple the string-based formalism previously introduced with spatial information of the fiber-tracts.

In Part IV, we will describe three different algorithms to detect longitudinal changes visible along WM fiber-bundles. More in detail, in the first chapter, we will describe a dynamic threshold-based algorithm based on a Gaussian mixture model to detect and isolate those changes. Moreover, in order to increase the sensibility of WM fiber-bundle analysis, we introduce the definition of two new populations of fibers. In the second chapter, we will extend the previous method in order to take into account more than one diffusion metric at a time and more than two time-points. The method, based on non-negative matrix factorization, is able to identify the time-points and the regions of the fiber-bundle affected by longitudinal changes. Finally, in the last chapter, we will describe a third algorithm, based on constrained tensor factorization, capable to generalize the two previous methods. Indeed, thanks to the tensor formalism it is possible to build a tensor capable to represent the complex structure of the fiber-bundle to analyze. We will also describe how we reduce the computation time needed by the algorithm thanks to parallel programming.

II Fiber-Bundle Clustering

Chapter 1

A String-Based Formalism for Fiber-Bundle Extraction

Contents

Published papers in this chapter

- F. Cauteruccio*, C. Stamile*, G. Terracina, D. Ursino, D. Sappey-Marinier, **"An automated string-based approach to extracting and characterizing White Matter fiber-bundles,"** *Computers in Biology and Medicine*, vol. 77 pp.64-75, 2016.
- F. Cauteruccio*, C. Stamile*, G. Terracina, D. Ursino, D. Sappey-Marinier, **"An Automated String-Based Approach to White Matter Fiber-Bundles Clustering,"** *in: 2015 International Joint Conference on Neural Networks (IJCNN)*, 2015, pp. 1-8.
- C. Stamile*, F. Cauteruccio*, G. Terracina, D. Ursino, G. Kocevar, D. Sappey-Marinier, **"A Model-Guided String-Based Approach to White Matter Fiber-Bundles Extraction,"** *in: Brain Informatics and Health*, LNCS vol. 9250 pp.135-144, 2015.

*These authors contributed equally to this work

1 Introduction

Reconstructing and visualizing *in vivo* White Matter (WM) fibers is a challenging issue in the investigation of brain. For instance, the knowledge of these fibers is useful to understand and predict the effects of some neurodegenerative pathologies, like multiple sclerosis [Pantano *et al.* (2002), Wilson *et al.* (2003)]. Furthermore, they can be used in neurosurgical planning to interactively guide the surgeon during an operation [Golby *et al.* (2011)]. The most accurate method to perform this task is tractography [Mori *et al.* (1999)], which is based on the analysis of the main diffusion directions of the water molecules when they move through WM tissues. This movement can be analyzed by a Magnetic Resonance Imaging (MRI) technique called Diffusion Tensor Imaging (DTI) [Basser *et al.* (1994)b].

From an anatomic point of view, particular sets of fibers (called fiber-bundles) represent different WM structures [Catani and Thiebaut de Schotten (2008)]. These connect different gray matter regions of the brain, allowing them to communicate. To analyze WM structures, it is crucial to isolate subsets of fibers belonging to the WM regions into consideration. This task is often performed manually by expert neuroanatomists that define inclusion and exclusion criteria in such a way as to delineate regions of interest and isolate specific WM fiber-bundles [Mårtensson *et al.* (2013), Hua *et al.* (2008)]. However, this way of proceeding is time consuming and operator-dependent. With regard to this fact, we observe that the amount of data to analyze when facing this problem is enormous. For instance, a simple whole brain tractography could generate up to 10^6 fibers. As a consequence, investigating the WM structures of a cohort of people becomes a big data application and cannot be performed manually.

In order to overcome this limitation, different automated approaches to extracting and, then, characterizing WM fiber-bundles have been proposed in the past literature (see, for instance, [Yeatman *et al.* (2012), O'Donnell *et al.* (2006), Zhang *et al.* (2008), Garyfallidis *et al.* (2012)]). WM fiber-bundle models, usually constructed by experts, provide a coincise representation of the bundles of interest. Then, given a set of models, the extraction and characterization of WM fiber-bundles from tractography data resorts to determining the real fiber-bundles corresponding to these models. A way to perform this last task consists of clustering WM fibers derived from tractography data and, then, applying a model-based characterization algorithm to identify fiber-bundles, starting from available models and obtained clusters. As will be clear below, this way of proceeding can guarantee the indispensable efficiency when the number of models to process is high. Nevertheless, in cases where this approach is applied as it is, fibers should be clustered on the basis of their layout in the three-dimensional space. Therefore, we are in facing a multi-dimensional clustering problem. This can be considered as a simplified version of the more complex multi-view clustering problem, which is well known for being a very difficult problem in the literature [Cai *et al.* (2013), Sun (2013)]. As a further confirmation of this fact, the approaches to clustering WM fibers, proposed in the past, present several issues, as will be explained in Section 2.

In this chapter, we aim at providing a contribution in this setting by proposing a new automated approach that, given as input a set of WM fibers, generated from streamlines produced by tractography, and a set of models, extracts fiber-bundles through clustering and a subsequent model-based characterization.

The core "ingredients" (and the main contributions) of our proposal are: *(i)* a new string-based formalism allowing an alternative representation of WM fibers, *(ii)* a new string dissimilarity metric, and *(iii)* a model-based characterization algorithm.

Our approach consists of four main steps, namely:

1. representation of WM fibers as strings;
2. computation of the dissimilarity degree for each pair of available WM fibers;
3. clustering of involved WM fibers;
4. exploitation of obtained clusters to identify the fiber-bundles corresponding to the models of interest.

In more detail, our approach represents a fiber as a sequence of m voxels in a three-dimensional space. Then, it associates a color with each voxel, specifying its space orientation. As an example, voxels parallel to the x (resp., y, z) axis are colored in red (resp., green, blue). The color of a generic voxel, not parallel to an axis, is computed by suitably mixing the colors corresponding to the axes on the basis of its orientation. As a consequence of this notation, a fiber can be represented as a sequence of colors, each expressed in the RGB color space. Finally, by discretizing this last space, our approach represents a fiber by means of a string.

Thanks to this fiber representation, our WM fiber clustering task reduces to a string clustering one, which is much simpler to face. As a matter of fact, it is sufficient: *(i)* finding a metric to measure string dissimilarity (and, in this chapter, we present a new suitable metric, which represents the second main contribution of our approach); *(ii)* constructing the dissimilarity matrix; *(iii)* applying a suitable clustering algorithm already proposed in the literature.

After the set of WM fiber clusters has been determined, our approach is able to exploit obtained clusters to extract the fiber-bundles corresponding to the models of interest. Specifically, given a set of approximate models of fiber-bundles constructed by an expert, our approach matches each model m with a corresponding cluster and, then, exploits this matching to identify the WM fiber-bundles closest to m. Observe that, once clusters are available, the matching process can be carried out efficiently, because it is not necessary to compare each model with the whole set of available WM fibers. Indeed, it is sufficient to exploit medoids to characterize clusters and, consequently, to compare each model m with the medoids of extracted clusters to determine the one closest to m. Clearly, the number of medoids is generally orders of magnitude smaller than the number of WM fibers. This provides users with the possibility to dynamically change the models of interest and to immediately obtain the corresponding bundles.

To verify the suitability of our approach, we have performed an experimental campaign. Obtained results are very satisfying, as will be illustrated in detail in Section 4.

As for a further key issue of our approach, it can be applied not only to characterize WM fiber-bundles, but also in all other cases in which it is necessary to perform multi-dimensional (and, more in general, multi-view) clustering and characterization of activities. In fact, it is sufficient to associate a color with each axis of the corresponding multi-dimensional space, to suitably color the corresponding voxel, and, after all voxels have been colored, to suitably discretize the corresponding color representation. As a consequence, the approach proposed in this chapter can be adopted in all those biomedical contexts someway requiring a multi-view clustering (think, for instance, of finding specific patterns useful to monitor the vital signs of a patient in intensive care [Ordóñez *et al.* (2008)]) and a possible model-guided characterization of obtained clusters. But, it could be adopted also in contexts very different from the biomedical ones, whenever a multi-view clustering and characterization must be carried out. Think, for instance, of the analysis of air flows or of weather perturbations in meteorology.

Interestingly, the general philosophy underlying our approach can be extended to other multi-view data applications. As an example, the search of frequent structures in a multi-dimensional space can be reduced to the search of frequent patterns in a set of strings, whereas the search of specific structures in a multi-dimensional space can be reduced to the search of specific paths in a set of strings.

Last, but not least, our dissimilarity metric allows the integrated analysis of data flows, which are (possibly) very heterogeneous but provide a multi-view representation of a certain event. For instance, in the clinical observation of the vital parameters of a patient, it is possible to simultaneously consider several measures, such as electrocardiogram, temperature, respiratory rate, etc. Thanks to our approach, all these measures can be analyzed simultaneously and compared in such a way as to evidence possible correlations. This application case shows how our approach can be exploited to integrate data coming from different domains.

This chapter is structured as follows. In Section 2, we provide an overview of the research background and the corresponding related work. In Section 3, we supply a detailed description of our approach. In Section 4, we present our experimental campaign. In section, 5 we discuss our results. Finally, in Section 6, we draw our conclusions.

2 Background and related work

In the past literature, many approaches investigating fiber-bundles from several viewpoints have been proposed. As an example, [Yeatman *et al.* (2012), Colby *et al.* (2012), Mårtensson *et al.* (2013), Reich *et al.* (2007)] evidence the importance of WM structures for better understanding and predicting the role of certain neurodegenerative pathologies (such as multiple sclerosis) in causing motor disability [Pantano *et al.* (2002), Wilson *et al.* (2003)] and other symptoms, like fatigue [Rocca *et al.* (2014), Reich *et al.* (2008)]. Other papers (e.g., [Golby *et al.* (2011)]) show how the knowledge of fiber-bundles can interactively guide a surgeon in her activity.

[Mårtensson *et al.* (2013), Hua *et al.* (2008)] present some approaches to manually extracting WM fiber-bundles. In these approaches, expert neuroanatomists define suitable inclusion and exclusion criteria allowing the delineation of the regions of interest and, ultimately, the isolation of specific WM fiber-bundles. Nevertheless, manual approaches to extracting fiber-bundles are time consuming and operator-dependent. As a consequence, they cannot be exploited to analyze large amounts of data.

To handle high volumes of data, some automatic approaches have been proposed [Yeatman *et al.* (2012), Zhang *et al.* (2008), Garyfallidis *et al.* (2012), Jin *et al.* (2014), Bénézit *et al.* (2015), Prasad *et al.* (2014)]. They can be divided in two categories, namely: *(i)* algorithms needing a priori knowledge about the location of certain WM brain regions (which are based on suitable atlases [Yeatman *et al.* (2012), Zhang *et al.* (2008), Garyfallidis *et al.* (2012), Jin *et al.* (2014), Bénézit *et al.* (2015), Prasad *et al.* (2014)]., and *(ii)* algorithms that do not need a priori knowledge [Garyfallidis *et al.* (2012)].

Some approaches of the former category (e.g., the one described in [Yeatman *et al.* (2012)]) are very simple and fast. They exploit pre-labeled WM fiber-bundle atlases on the subject's image (e.g., the atlas described in [Hua *et al.* (2008)]) and select only those fibers belonging to the regions delineated in the atlas. These approaches present several limitations. First, they can extract only fiber-bundles conforming to the ones specified in the WM atlases, which make them applicable only in few cases. Second, the quality of extracted fibers depends on the algorithm used to register the atlases on the subject's image.

The approach of [Jin *et al.* (2014)] exploits quite a sophisticated multi-stage fiber clustering scheme, based on a-priori knowledge of fibers at several levels. First, this approach carries out a length-based clustering step, aimed to filter out short fibers[1]. Then, it executes a clustering task, based on Regions of Interest (ROIs) and filters out all the fibers external to these regions. This step is based on the assumption that anatomically well known WM tracts interconnect ROIs. Finally, it performs a distance-based clustering; for this purpose, the authors define a symmetric version of the asymmetric Hausdorff distance to both compare and label fibers w.r.t. atlases.

In [Bénézit *et al.* (2015)], the authors present an approach to organizing WM fibers in the context of isolated corpus callosum dysgenesis. Their ultimate goal is comparing WM fibers of children presenting this pathology. The approach consists of a workflow pipeline. One of the first steps of this pipeline consists of an intra-subject clustering of WM tracts. This task aims at detecting groups of similar tracts that are large enough to be reliable candidates when looking for a specific bundle in case a ROI-based strategy is adopted. Clustering is intended as a pre-processing step and is applied on the geometric representation of fiber voxels. To carry out fiber comparison, a symmetrized version of the mean closest point distance is exploited. Since the proposed clustering method relies on tract geometry, the authors show that, in some cases, their approach could cluster together tracts close to each other but not belonging to the same structure.

In [Prasad *et al.* (2014)], a framework for analyzing a population of WM tracts is proposed. Its main objective is the compact representation of bundles as paths following points of maximum density (Maximum Density Paths - MDPs). In this way, the efficient comparison of WM tracts across a population can be performed. MDPs are represented by vector valued functions. The authors show that MDP representation reveals important parts of WM structure and considerably reduces the problem dimensionality. The approach of [Prasad *et al.* (2014)] is based on a set of prior knowledge from an atlas, which defines ROIs. In this context, fiber bundle clustering reduces to measuring the intersection of fibers with the ROI atlas. WM structures are represented as three-dimensional curves, which, in their turn, represent paths among the most influential regions in tractography. Shape matching among MDPs is achieved using a geodesic curve registration. The adopted clustering method is quite simple. In particular, a fiber intersection score is determined by computing the number of ROI voxels that intersect with the fiber tract. This score is used to select fibers belonging to a ROI and, thus, WM tracts.

The approach presented in [Zhang *et al.* (2008)] consists of two steps. During the first one, clustering is adopted to determine the best proximity measures, i.e., those measures best separating the involved WM structures. This process is supervised by an expert, who must tune some parameters. At the end of the first step, a template containing

---

[1] Here, the notion of "short" depends on the set of atlases to analyze.

proximity measure thresholds between sets of WM fiber-bundles is generated. The second step exploits this template to activate a clustering task aimed to select those fibers having coherent proximity values. Even this step requires the presence of an expert, who tunes some threshold values to generate the template. The approach of [Zhang *et al.* (2008)] is interesting and sophisticated. However, the role played by the expert in tuning activities is excessive in it. In fact, tuning is performed through a try-and-check activity, which interactively sets parameters until they correctly separate the involved fiber-bundles. In this approach, the method for parameter tuning should be as general as possible in such a way as to avoid the need of re-executing supervision on each new set of data.

Approaches that do not need a priori knowledge are based on specific similarity and proximity measures in $\mathbb{R}^3$ aiming at: *(i)* grouping in the same subset all fibers having the same structures; *(ii)* maximizing the possibility of distinguishing fibers having different forms. Among these approaches, we cite QuickBundles (QB) [Garyfallidis *et al.* (2012)]. QB is a simple and fast algorithm that extracts WM fiber-bundles without the need of any a priori knowledge. It is based on centroids. All the fibers comprised in the neighborhood of a centroid (whose radius is defined by an operator) are grouped in the same cluster. QB is certainly fast. However, obtained clusters could not represent anatomically plausible fiber-bundles.

As it will be clear below, one of the key components of our approach is a new string similarity metric, called Semi-Blind Edit Distance (SBED). Therefore, it appears interesting to have a quick overview to past metrics for string similarity computation.

To the best of our knowledge, no approach facing the same problem considered by SBED has been presented in the literature. Nevertheless, some variants have been investigated in the past. A very interesting approach facing a variant of the problem considered by SBED is presented in [Baker (1996)]. This approach computes similarities over parameterized strings, i.e., strings having some symbols acting as parameters that can be substituted at no cost. The approach presented in [Hazay *et al.* (2007)] aims at finding all the locations in a string s for which there exists a *global bijection* π mapping a pattern p into the substring of s that minimizes the Hamming distance. This approach allows mismatches.

[Apostolico *et al.* (2007)] considers *injective* functions, instead of bijective ones. [Baker (1999)] introduces the notion of *p-edit* distance. It handles insertions, deletions and exact p-matches, but it does not handle mismatches. The approach of [Keller *et al.* (2009)] extends the one of [Baker (1999)] by requiring the transformation function to have global validity. However, even in this case, the set of allowed edit operations is limited (for instance, substitutions are not permitted). [Greco and Terracina (2013)] presents a preliminary approach to a many-to-many mapping function. The approach of [Gawrychowski and Uznański (2016)] extends the one described in [Baker (1996)] by introducing an order-preserving matching; in this approach, the number of mismatches is limited to k.

String matching has been also investigated for the clone detection problem, i.e., for detecting if a code contains two or more cloned parts [Kamiya *et al.* (2002)]. A problem related to this last one, which has been largely investigated in the past, is the Longest Common Subsequence (LCS) one. It was studied both in the classic version and in the parameterized one [Gorbenko and Popov (2012),Blin *et al.* (2012)]. Recently, [Mendivelso and Pinzón (2015)] presented a detailed survey on parameterized matching, which contains a review of the corresponding solutions and extensions.

3 Description of the proposed approach

Our approach is aimed to extract White Matter (WM) fiber-bundles from a high number of streamlines produced by tractographies. To reach its goals, it compares models provided by experts with clusters of WM fibers generated by itself. Since the fibers to cluster are in a three-dimensional space, it is extremely difficult to carry our this task directly. For this reason, an alternative way for fiber representation becomes necessary.

Our approach consists of the following steps:

- *Step 1.* Given a set of WM fibers, it maps them into a set of strings by applying the new string-based formalism mentioned in the Introduction.
- *Step 2.* Given a set of strings representing WM fibers, it applies the new string dissimilarity metric to construct the Dissimilarity Matrix relative to them. The generic element $D[i,j]$ of this matrix indicates the dissimilarity degree of the string representations of the fibers f_i and f_j.
- *Step 3.* Given the Dissimilarity Matrix D relative to a set of WM fibers, it applies several existing clustering methods to D to group the involved WM fibers into homogeneous clusters.
- *Step 4.* Given the set of clusters thus obtained, it exploits them to extract and characterize the WM fiber-bundles through a model-based and clustering characterization technique.

In the following, we describe these steps in all details.

3.1 Representation of WM fibers as strings

The main purpose of this step is to represent a three-dimensional fiber in a different format, more compatible with clustering. In the past, several ways for representing a three dimensional line have been proposed. These different representations depend on both the context and the expected use.

In this chapter, we choose to represent a fiber as a sequence of voxels (volumetric picture elements), representing, in their turn, values on a grid in a three-dimensional space. By adopting this guideline, a generic fiber $f_i \in F$ can be represented as a sequence $f_i = (v_{i_1}, v_{i_2}, \ldots, v_{i_m})$ of voxels in the three-dimensional space. This sequence is aimed to approximate the corresponding three-dimensional line. As a consequence, the number of voxels representing a fiber is proportional to its length. Therefore, fibers having different lengths will be represented by a different number of voxels. The position of each voxel in the three-dimensional space is determined on the basis of the position of the neighboring voxels.

A color c_{i_j} is assigned to each voxel $v_{i_j} \in f_i$. This color is determined by the orientation of v_{i_j} in the three-dimensional space [Douek *et al.* (1991)]. Specifically, let $p^x_{i_j}$ (resp., $p^y_{i_j}$, $p^z_{i_j}$) be the fraction of information of v_{i_j} parallel to the x (resp., y, z) axis. A basic color $\overline{c^x}$ (resp., $\overline{c^y}$, $\overline{c^z}$) is associated with the x (resp., y, z) axis. On the basis of the standard code of colors defined in [Le Bihan *et al.* (2001)], let $\overline{c^x}$ = red, $\overline{c^y}$ = green and $\overline{c^z}$ = blue. c_{i_j} is, then, obtained by the weighted combination of $\overline{c^x}$, $\overline{c^y}$ and $\overline{c^z}$. Clearly, adopted weights are strictly related to the orientation of v_{i_j} in the three-dimensional space. For instance, if v_{i_j} is parallel to the x axis, $p^x_{i_j} = 1$, $p^y_{i_j} = 0$, $p^z_{i_j} = 0$, and, consequently, $c_{i_j} = \overline{c^x} = red$.

Thanks to this notation, a fiber $f_i \in F$ can be expressed as a sequence $\phi_i = (c_{i_1}, c_{i_2}, \ldots, c_{i_m})$ of colors. Each element c_{i_j} corresponds to the color, expressed in the RGB color space, associated with v_{i_j}. Thanks to the fiber discretization tasks described above, a fiber $f_i \in F$, through its representation ϕ_i, can be easily translated into a string using a mapping function:

$$\Upsilon : RGB \rightarrow \Sigma$$

where $|\Sigma| = s$. In our approach, we decided to adopt the algorithm described in [Arvo (2013)] for implementing Υ. Thanks to it, a fiber $f_i \in F$ can be expressed as a string in Σ^m.

The algorithm for implementing Υ belongs to the family of minimum variance quantization algorithms [Heckbert (1982)]. Quantization is a technique extensively used in image processing. It allows the reduction of the number of colors of an image. It clusters pixels into groups on the basis of the variance among the corresponding pixel values. In this way, it divides the RGB color cube into several smaller boxes and, then, maps all the colors falling within each box into the color of its center. In carrying out this task, it exploits the so-called RGB color cube, which is a three-dimensional array of all the colors that can be defined in the selected space.

There are two main quantization methods proposed in the literature, namely uniform quantization and minimum variance quantization (which is the one adopted in our approach). They differ for the technique used to divide up the RGB color cube. The former cuts up the color cube into equally-sized boxes; the latter divides the color cube into boxes of possibly different sizes, on the basis of the distribution of colors in the image.

If the number s of the boxes to obtain is an input parameter, the algorithm automatically determines the position of boxes on the basis of the variance of the color data. Once the image is partitioned into s optimally located boxes, the pixels within each box are mapped to the pixel at the center of that box. Finally, a character is associated with each box center. In this way, an alphabet Σ, representing all the s boxes generated by the algorithm, is defined.

At the end of all the activities described above, a set $F = \{f_1, f_2, \ldots, f_n\}$ of fibers can be represented as a set $T = \{t_1, t_2, \ldots, t_n\}$ of strings. Specifically, each element $t_i \in T$ is a string of Σ^m corresponding to f_i.

To formally express these transformations, we introduce a function $\tau(\cdot)$. First, it performs the transformation of f_i in a sequence of voxels; then, it associates a color with each voxel; finally, it transforms each voxel into a string. In the following, we use the notation $t_i = \tau(f_i)$ to represent the string t_i that the function $\tau(\cdot)$ returns when it takes f_i in input. The length of t_i is equal to the number of voxels exploited to represent f_i. As a consequence, if two fibers f_i and f_j have different lengths in the three-dimensional space, t_i and t_j have different lengths.

Algorithm Fibers-To-Strings (see Algorithm 1) describes the transformation of a set F of fibers into a set T of strings.

It receives a set $F = \{f_1, f_2, \ldots, f_n\}$ of WM fibers to transform, the cardinality s of the alphabet Σ of which strings belong to, the size vxS of voxels in the three-dimensional space, and the size stS of the fiber step used by the tractography algorithm. It returns a set $T = \{t_1, t_2, \ldots, t_n\}$ of strings representing the fibers in F.

The algorithm starts by computing the Step Rate stR, a parameter necessary for normalization in the next steps. For this purpose, it sets stR as the ratio of the step size stS of tractography to the norm of vxS. This computation of the norm of vxS is necessary because vxS is three-dimensional whereas stS is scalar. As a consequence, to perform the ratio, a scalar must be derived from vxS, representing it.

After the computation of stR, Fibers-To-Strings sets T to empty. Then, for each fiber $f_i \in F$, it performs some tasks devoted to obtain a string t_i over the alphabet Σ corresponding to f_i.

Specifically, t_i is initially empty. Then, on the basis of stS, a set $P_i = \{f_{i_0}, f_{i_1}, \ldots, f_{i_\nu}\}$ of three-dimensional points representing f_i are determined.

For each point f_{i_j}, $0 \leq j \leq \nu - 1$, of P_i, Fibers-To-Strings performs two tasks, namely:

- It determines the difference between f_{i_j} and $f_{i_{j+1}}$, which indicates the direction of the corresponding voxel in the three-dimensional space. Then, it normalizes this difference by dividing it by stR; let $p_i = (p_i^x, p_i^y, p_i^z)$ be the corresponding point in the three-dimensional space.

- It calls the function Υ, illustrated above, for transforming the point p_i in a character of the alphabet Σ. This character is then concatenated to t_i. The symbol $\oplus$ denotes string concatenation.

Once all the points f_{i_j}, $0 \leq j \leq \nu - 1$, of P_i have been processed, and t_i is complete, t_i is added to T.

```
Input  : a set $F = \{f_1, f_2, \dots, f_n\}$ of fibers
         the cardinality $s$ of the alphabet $\Sigma$
         the size $vxS$ of voxels in the three-dimensional space
         the size $stS$ of the tractography step
Output: a set $T = \{t_1, t_2, \dots, t_n\}$ of strings
begin
    $stR = \frac{stS}{norm(vxS)}$;
    $T = \emptyset$;
    foreach $f_i \in F$ do
        $t_i = \emptyset$;
        $P_i$ = compute_representative($f_i$);
        for $0 \leq j < |P_i|$ do
            $(p_i^x, p_i^y, p_i^z) = \left(\frac{f_{i_j}^x - f_{i_{j+1}}^x}{stR}, \frac{f_{i_j}^y - f_{i_{j+1}}^y}{stR}, \frac{f_{i_j}^z - f_{i_{j+1}}^z}{stR}\right)$;
            $t_i = t_i \oplus \Upsilon(p_i, s)$;
        end
        $T = T \cup t_i$;
    end
    return $T$;
end
```

Algorithm 1: ALGORITHM FIBERS-TO-STRINGS

FIBERS-TO-STRINGS terminates when all the fibers of F have been processed and, therefore, T is complete. It returns T as output.

3.2 Construction of the Dissimilarity Matrix

The purpose of Step 2 is the construction of the Dissimilarity Matrix D associated with F. D is a $n \times n$ matrix; its generic element $d_{ij} = D[i, j]$ is a real number in the interval $[0, 1]$ and indicates the dissimilarity degree of the string representations t_i of f_i and t_j of f_j. To obtain the value of d_{ij}, it is necessary to apply a string-based dissimilarity metric on t_i and t_j. With regard to this issue, we point out that, in our application context, classical string-based dissimilarity metrics (like the Hamming or the Levenshtein distance [Levenshtein (1965)]), which measure the minimum number of edit operations necessary to transform the first string into the second one, would not work properly. In fact, they are based on the assumption that one-to-one correspondences between the symbols of the two strings are implicitly determined simply by identity.

Actually, in our scenario, the adoption of only one-to-one correspondences would be reductive and could lead to either imprecise or wrong results. In fact, in our application context, there could be different symbols expressing the same or similar concepts. Think, for instance, of two symbols, one representing a horizontal voxel and the other denoting a slightly oblique one, derived from an approximation during discretization: an error could arise if they are considered different. Analogously, it could be extremely important to be able to match a horizontal voxel with both another horizontal voxel and/or a slightly oblique one. Clearly, to avoid over-matchings, the number of these exceptions should be limited. Finally, there may exist pairs of symbols (e.g., a horizontal and a vertical voxel) that clearly should not match. In this case, it would be necessary to constrain invalid matches.

To consider and, hopefully, generalize the previous observations, we have defined a new string-based dissimilarity metric, called SBED (Semi Blind Edit Distance). It is aimed to compute the minimum distance between two strings, provided that finding the optimal matching schema, under a set of constraints, is part of the problem. Therefore, the dissimilarity $D[i, j]$ between $t_i = \tau(f_i)$ and $t_j = \tau(f_j)$ is obtained by computing their SBED distance.

Our SBED metric takes a pair of ordered sequences of symbols (strings) as input. These play the general role of parameters, i.e., their matching with other symbols is not necessarily regulated by symbol identity.

Let Π_1 and Π_2 be two (possibly disjoint) *alphabets* of parameters. Throughout the following sections, we assume that two strings s_1 over Π_1 and s_2 over Π_2 are given.

The length of the string s_i $(i \in \{1, 2\})$, i.e., the number of symbols in it, will be denoted by $len(s_i)$. Furthermore, for each position $1 \leq j \leq len(s_i)$, the j-th symbol of s_i will be identified by $s_i[j]$.

Let $-$ be a symbol not included in $\Pi_1 \cup \Pi_2$. Then, a string $\bar{s}_i$ over $\Pi_1 \cup \Pi_2 \cup \{-\}$ is a *transposition* of s_i if it can be obtained from s_i by deleting all the occurrences of $-$. The set of all the possible transpositions of s_i is denoted by $\mathcal{TR}(s_i)$.

An *alignment* of the strings s_1 and s_2 is a pair $\langle \bar{s}_1, \bar{s}_2 \rangle$, where $\bar{s}_1 \in \mathcal{TR}(s_1)$, $\bar{s}_2 \in \mathcal{TR}(s_2)$ and $len(\bar{s}_1) = len(\bar{s}_2)$. Here, $-$ is meant to denote an insertion/deletion operation performed on s_1 or s_2.

Given an alphabet Π and an integer π such that $0 < \pi \leq |\Pi|$, *a π-partition* is a partition Φ^π of Π such that $0 < |\phi_v| \leq \pi$, for each $\phi_v \in \Phi^\pi$.

Given two alphabets Π_1 and Π_2 and two integers π_1 and π_2, *a $\langle \pi_1, \pi_2 \rangle$-matching schema* is a function $M_{\langle \pi_1, \pi_2 \rangle} : \Phi_1^{\pi_1} \times \Phi_2^{\pi_2} \rightarrow \{\mathit{true}, \mathit{false}\}$, where $\Phi_i^{\pi_i}$ $(i \in \{1, 2\})$ is a π_i-partition of Π_i and, for each $\phi_v \in \Phi_1^{\pi_1}$ (resp., $\phi_w \in \Phi_2^{\pi_2}$), there is at most one $\phi_w \in \Phi_2^{\pi_2}$ (resp., $\phi_v \in \Phi_1^{\pi_1}$) such that $M(\phi_v, \phi_w) = \mathit{true}$. This means that all the symbols in ϕ_v match with all the ones in ϕ_w. $M(\phi_v, \phi_w) = \mathit{false}$ indicates that all the symbols in ϕ_v mismatch with all the ones in ϕ_w.

A constraint χ associated with a matching schema $M_{\langle \pi_1, \pi_2 \rangle}$ is a set of unordered pairs of symbols (c_i, c_j), such that $c_i \in \Pi_1$, $c_j \in \Pi_2$ and, for each $(c_i, c_j) \in \chi$, there exists no pair (ϕ_v, ϕ_w), $\phi_v \in {\Phi_1}^{\pi_1}, \phi_w \in {\Phi_2}^{\pi_2}$, having $c_i \in \phi_1, c_j \in \phi_2$ and $M(\phi_1, \phi_2) = \mathit{true}$.

A constrained matching schema is represented by $M_{\langle \pi_1, \pi_2, \chi \rangle}$ only if $\chi \neq \emptyset$.

Let $\langle \bar{s}_1, \bar{s}_2 \rangle$ be an alignment for s_1 and s_2, let $M_{\langle \pi_1, \pi_2, \chi \rangle}$ be a $\langle \pi_1, \pi_2, \chi \rangle$-matching schema over π-partitions $\Phi_1^{\pi_1}$ and $\Phi_2^{\pi_2}$ and constraints χ, and let j be a position with $1 \leq j \leq len(\bar{s}_1) = len(\bar{s}_2)$. We say that $\langle \bar{s}_1, \bar{s}_2 \rangle$ *has a match at j* if either:

- $s_1[j] \in \Pi_1, s_2[j] \in \Pi_2$ and $s_1[j] = s_2[j]$, or
- $s_1[j] \in \phi_v, s_2[j] \in \phi_w$, $\phi_v \in \Phi_1^{\pi_1}$, $\phi_w \in \Phi_2^{\pi_2}$ and $M_{\langle \pi_1, \pi_2, \chi \rangle}(\phi_v, \phi_w) = \mathit{true}$.

The *distance* between $\bar{s}_1$ and $\bar{s}_2$ under $M_{\langle \pi_1, \pi_2, \chi \rangle}$ is the number of positions at which $\langle \bar{s}_1, \bar{s}_2 \rangle$ does not have a match.

Given two integers π_1 and π_2 such that $0 < \pi_1 \leq |\Pi_2|$ and $0 < \pi_2 \leq |\Pi_1|$, *the* $\langle \pi_1, \pi_2, \chi \rangle$*-edit distance* between s_1 and s_2 ($\mathcal{L}_{\langle \pi_1, \pi_2, \chi \rangle}(s_1, s_2)$ for short) is the minimum distance that can be obtained with any $\langle \pi_1, \pi_2, \chi \rangle$-matching schema and any alignment $\langle \bar{s}_1, \bar{s}_2 \rangle$.

Observe that, in order to properly compute $\mathcal{L}_{\langle \pi_1, \pi_2, \chi \rangle}(s_1, s_2)$, several components play a crucial role, namely:

- π_1 and π_2, which determine the (maximum) size of each partition;
- π-partitions $\Phi_1^{\pi_1}$ and $\Phi_2^{\pi_2}$; in fact there can be many π-partitions for the same set Π_1 (resp., Π_2) of π_1 (resp., π_2);
- matching schemas $M_{\langle \pi_1, \pi_2, \chi \rangle}$, which determine the way partitions of different sets can be combined via matching;
- alignments; in fact there can be many possible alignments between two strings.

The general problem of computing $\mathcal{L}_{\langle \pi_1, \pi_2, \chi \rangle}(s_1, s_2)$ is NP-Complete. To provide a feasible solution, we designed heuristics that iteratively search for the optimal matching schema. It is explained in detail in [Cauteruccio *et al.* (2015),Stamile *et al.* (2015)a].

The following example provides a practical explanation of the previous concepts.

Example 1. *Let* $s_1 = \mathtt{AAABCCDCAA}$ *and* $s_2 = \mathtt{EEFGHGGFHH}$*, which determines* $\Pi_1 = \{\mathtt{A,B,C,D}\}$ *and* $\Pi_2 = \{\mathtt{E,F,G,H}\}$*.*

For $\pi_1 = \pi_2 = 1$*, the best alignment* $\langle \bar{s}_1, \bar{s}_2 \rangle$ *that can be computed is obtained by matching* `{A}-{E}`*,* `{B}-{G}`*,* `{C}-{H}`*, and* `{D}-{F}`*. The aligment:*

$$\begin{array}{l} s_1 : \mathtt{AAABCCDDCAA} \rightarrow \mathtt{AAABCCDDCAA} \\ s_2 : \mathtt{EEFGHGGFHH} \;\;\rightarrow \mathtt{EEFGHGGFH{-}H} \\ \qquad\qquad\qquad\qquad\qquad\quad \mathtt{**\ **\ \ \ **} \end{array}$$

gives $\mathcal{L}_{\langle 1,1 \rangle}(s_1, s_2) = 5$*.*

Observe that the approach works properly even if $\Pi_1 \cap \Pi_2 = \emptyset$ *and if the input strings are of different lengths.*

If we set $\pi_1 = \pi_2 = 2$*, the best alignment is the one obtained by matching* `{B,A}-{E,H}`*, and* `{C,D}-{G,F}`*, namely:*

$$\begin{array}{l} s_1 : \mathtt{AAABCCDDCAA} \rightarrow \mathtt{AAABCCDDCAA} \\ s_2 : \mathtt{EEFGHGGFHH} \;\;\rightarrow \mathtt{{-}EEFGHGGFHH} \\ \qquad\qquad\qquad\qquad\qquad\qquad \mathtt{**\ \ *\ \ *****} \end{array}$$

which gives $\mathcal{L}_{\langle 2,2 \rangle}(s_1, s_2) = 3$*.*

Now, suppose to introduce the constraint $\chi = \{\langle \texttt{A}, \texttt{E} \rangle\}$. *For* $\pi_1 = \pi_2 = 1$, *the best alignment* $\langle \bar{s}_1, \bar{s}_2 \rangle$ *is the following:*

$$
\begin{array}{l}
s_1 : \texttt{AAABCCDDCAA} \rightarrow \texttt{AAABC-CDDCAA} \\
s_2 : \texttt{EEFGHGGFHH} \;\;\, \rightarrow \texttt{-EEFGHGGFHH} \\
\qquad\qquad\qquad\qquad\qquad\qquad\;\; \texttt{**}\ \ \ \texttt{*****}
\end{array}
$$

which still gives $\mathcal{L}_{\langle 1,1,\chi \rangle}(s_1, s_2) = 5$ *but where* `A` *and* `E` *do not match anymore. Optimal matchings are, in fact,* `{A}-{H}`, `{B}-{E}`, `{C}-{F}`, *and* `{D}-{G}`.

We are, now, able to formally express how the Dissimilarity Matrix is computed. In particular, given a set $F = \{f_1, f_2, \ldots, f_n\}$ of WM fibers and the corresponding set $T = \{t_1, t_2, \ldots, t_n\}$ of strings, such that $t_i = \tau(f_i)$, given two integers π_1 and π_2 and a constraint χ, the generic element $D[i, j]$ of the Dissimilarity Matrix D associated with F is computed as:

$$D[i, j] = \mathcal{L}_{\langle \pi_1, \pi_2, \chi \rangle}(t_i, t_j)$$

Observations

At the end of the presentation of this step, some observations about the SBED metric are in order.

First, since the possible edit operations allowed during the computation of $\mathcal{L}_{\langle \pi_1, \pi_2, \chi \rangle}(s_1, s_2)$ are insertions, deletions and substitutions, SBED allows the comparison of strings of different lengths. As for this specific case, not only our approach works properly even in presence of fibers of different lengths, but also it correctly returns a high dissimilarity value, in case of very different lengths, because a high number of insertions/deletions will be necessary. As a consequence of this fact, for instance, our approach is capable of recognizing the dissimilarity of horizontal straight fibers having very different lengths, in spite of their identical direction in the three-dimensional space.

As a further consideration, we observe that spatial metrics require a fine-grained examination of the spatial coordinates of involved fibers. In fact, they either consider the Euclidean distance between fibers or require a fine-tuned registration of these last ones before starting distance computation. Euclidean-distance-based approaches may fail to identify similar fibers within thick bundles. In fact, if the proximity threshold is high, they may put together fibers with very different shapes; on the other side, if the proximity threshold is low, they may put in different bundles fibers approximatively having the same shape. Registration-based approaches are strongly dependent on the accuracy of the registration phase, which becomes a critical step. Our metric does not require any registration phase for comparing the fibers of a certain brain, and does not depend on spatial coordinates, since it exploits directionality information only. As a consequence, when applied to compute fiber dissimilarity, it is capable of overcoming the drawbacks of

```
Input  : a set $F = \{f_1, f_2, \dots, f_n\}$ of fibers
Output: a set $Cl = \{cl_1, cl_2, \dots, cl_k\}$ of clusters
        the set $\Gamma = \{\gamma_1, \gamma_2, \dots, \gamma_k\}$ of the medoids of the clusters of $Cl$
begin
    foreach pair $(f_i, f_j)$ s.t. $f_i \in F$, $f_j \in F$ do
        $D[i, j] = \mathcal{L}_{\langle\pi_1,\pi_2,\chi\rangle}(\tau(f_i), \tau(f_j))$;
    end
    $Cl$ = cluster($F, D, \kappa$);
    foreach $cl_j \in Cl$ do
        $Cl$=($Cl \setminus cl_j$) $\cup$ split($cl_j$);
    end
    $\Gamma$ = computeMedoids($Cl$);
    return $Cl$,$\Gamma$
end
```

Algorithm 2: Algorithm WM-Fiber-Clusters

the two approach families mentioned above. As an example, fibers in a thick bundle will all be considered similar, since they approximatively share the same shape.

A further positive feature of SBED w.r.t. spatial metrics consists of its capability of avoiding the matching of two symbols representing totally different orientations (think, for instance, of a horizontal voxel and a strictly vertical one), which, again, is useful to discriminate similarities and dissimilarities of fibers despite their proximity in the three-dimensional space.

Last, but not least, our SBED-based approach is capable of supporting both expert and nonexpert people, without the need of information regarding spatial locations.

3.3 Clustering of WM fibers

This step is aimed to pre-process F by applying a clustering algorithm on it in such a way as to group together anatomically homogeneous fibers. Thanks to the Dissimilarity Matrix D, computed during Step 2, the problem of clustering three-dimensional curves can be reduced to the one of clustering a set of strings, which can be faced by means of one of the many classical clustering algorithms already proposed in the literature. For instance, some clustering algorithms that can be easily incorporated in our approach are k-means [MacQueen *et al.* (1967)], k-medoids [Kaufman and Rousseeuw (1987)], and Expectation Maximization - EM [Dempster *et al.* (1977)].

Algorithm WM-Fiber-Clusters (see Algorithm 2) describes the clustering task of our approach. It receives a set $F = \{f_1, f_2, \dots, f_n\}$ of WM fibers and returns a set $Cl = \{cl_1, cl_2, \dots, cl_k\}$ of clusters and the set $\Gamma = \{\gamma_1, \gamma_2, \dots, \gamma_k\}$ of the corresponding medoids[2].

WM-Fiber-Clusters exploits the function $\tau(\cdot)$, described in Section 3.1, to implement *Step 1* of our approach, whereas it uses the first `foreach` cycle to implement *Step 2*.

The clustering of the fibers of F is carried out by the function `cluster`, which im-

---

[2] Recall that the medoid γ of a cluster cl is the element of cl "least dissimilar" from all the other ones. Differently from other cluster representative elements (like mean), medoid is robust to noise and can always be determined, even when the context of interest does not support Euclidean distance.

plements the adopted clustering algorithm. The parameter κ is necessary if `cluster` implements a clustering algorithm requiring the number κ of desired clusters in input. The output of `cluster` is a set $Cl = \{cl_1, cl_2, \dots, cl_\kappa\}$ of clusters; specifically, the cluster $cl_j \in Cl$ contains an anatomically homogeneous subset $cl_j = \{f_{j_1}, f_{j_2}, \dots, f_{j_l}\}$ of the fibers of F.

The clusters of Cl obtained at the end of this step may still have a problem. Indeed, as pointed out in Section 3.2, SBED does not consider spatial information. As a consequence, it may happen that a cluster $cl_j \in Cl$ contains homogeneous fibers (i.e., all with similar shapes), but some of them actually distant in space (for instance, in different hemispheres). To face this issue, in our approach, each cluster cl_j undergoes a splitting phase aimed only to separate fibers very far in space. As far as this task is concerned, we point out that: *(i)* it does not require registration, because it compares fibers of the same brain; *(ii)* it is far less sensitive to space coordinates than purely spatial methods, since it is devoted to just identify very far fiber sets (for instance, fibers with homogeneous shape but belonging to different hemispheres). Function `split` carries out this task.

Finally, function `computeMedoids` returns the set $\Gamma = \{\gamma_1, \gamma_2, \dots, \gamma_k\}$ of the medoids of the clusters of Cl. The medoid $\gamma_j \in \Gamma$ is used as the representative of the cluster $cl_j \in Cl$ and is exploited to speed up the model characterization phase in *Step 4*. In fact, as it will be clear in the next section, the computation of clusters can be considered as a pre-processing phase of our fiber-bundle extraction and characterization technique. This phase must be performed once and for all for each brain. As a consequence, once it has been carried out, the derivation of the fiber-bundles associated with each model $\psi_l \in \Psi$ requires the examination of only k medoids, instead of the n available WM fibers. Interestingly, k is, generally, orders of magnitude smaller than n.

3.4 Model-based WM fiber-bundles extraction and characterization

Once clusters and their medoids are available, our approach can perform the fiber-bundle extraction and characterization activities. For this purpose, it carries out *Step 4*, which, essentially, consists of the Algorithm WM-FIBER-BUNDLES (see Algorithm 3).

This algorithm receives a set $Cl = \{cl_1, cl_2, \dots, cl_k\}$ of clusters, the set $\Gamma = \{\gamma_1, \gamma_2, \dots, \gamma_k\}$ of the medoids of the clusters of Cl, and a set $\Psi = \{\psi_1, \psi_2, \dots, \psi_q\}$ of models. Each model $\psi_l \in \Psi$ represents the approximate shape of a fiber-bundle of interest. It could be obtained in two different ways, namely: *(i)* by exploiting a spline curve to draw the profile of the fiber-bundle of interest; *(ii)* by importing the mean-line profile of the fiber-bundle of interest from an atlas of pre-labeled fiber-bundles. Interestingly, the approximate model adopted to extract a specific fiber-bundle may be adopted also to extract the same fiber-bundle from other images characterized by different resolutions, and possibly acquired from other subjects. Indeed, the representation of our model is based on its shape and is independent of its spatial location.

As previously pointed out, our approach does not require a complex registration phase.

```
Input  : a set $Cl = \{cl_1, cl_2, \ldots, cl_k\}$ of clusters
         a set $\Gamma = \{\gamma_1, \gamma_2, \ldots, \gamma_k\}$ of the medoids of the clusters of $Cl$
         a set $\Psi = \{\psi_1, \psi_2, \ldots, \psi_q\}$ of models
Output: a set $\Theta = \{\theta_1, \theta_2, \ldots, \theta_q\}$ of bundles corresponding to $\Psi$
begin
    $\Theta = \emptyset$;
    foreach $\psi_l \in \Psi$ do
        $\Delta_l = \emptyset$;
        foreach $\gamma_j \in \Gamma$ do
            $\Delta_l = \Delta_l \cup \mathcal{L}_{\langle \pi_1, \pi_2, \chi \rangle}(\tau(\psi_l), \tau(\gamma_j))$;
        end
        $\mu = argmin(\Delta_l)$;
        $\Theta = \Theta \cup cl_\mu$;
    end
    return $\Theta$
end
```

Algorithm 3: Algorithm WM-Fiber-Bundles

In fact, it only needs a smooth alignment of the reference axes of both the models and the brain. This because SBED disallows the matching of lines having completely different directions (such as a straight horizontal line and a straight vertical one), but allows the matching of lines with slightly different directions (such as a horizontal line and a slightly oblique one).

The output of WM-Fiber-Bundles is a set $\Theta = \{\theta_1, \theta_2, \ldots, \theta_q\}$ of derived fiber-bundles such that each $\theta_l \in \Theta$ contains a subset of the fibers of F.

In more detail, for each model $\psi_l \in \Psi$, WM-Fiber-Bundles generates the corresponding fiber-bundle by associating the most appropriate cluster of Cl with it. For this purpose, it constructs the set Δ_l of the distances between each medoid of Γ and ψ_l. In order to compute the distance between ψ_l and a medoid $\gamma_j \in \Gamma$, WM-Fiber-Bundles transforms both of them into two strings $s_l = \tau(\psi_l)$ and $s_j = \tau(\gamma_j)$ by applying the function $\tau(\cdot)$, described in Section 3.1. Then, it computes the SBED distance between s_l and s_j. After this, it determines the minimum value δ_μ of Δ_l. Clearly, δ_μ corresponds to a medoid $\gamma_\mu \in \Gamma$ and to a cluster $cl_\mu \in Cl$. Finally, it identifies cl_μ as the cluster of WM fibers having the closest structure and features to ψ_l and returns cl_μ as the fiber-bundle corresponding to ψ_l.

We point out, again, that, thanks to the pre-processing clustering phase, this step requires the examination of only k medoids, instead of the n available WM fibers. This feature, and the consequent efficiency of our approach, make it possible to easily apply the same models to different subjects (and, consequently, to the results of different tractographies) in a reasonable time. This way, it is possible, for instance, to compare the fiber-bundles of different people, such as healthy and ill patients.

4 Experiments

This section is devoted to present our experimental campaign conceived to test our approach. It is organized as follows: first we test the pre-processing phase (Section 4.1) and our SBED metric (Section 4.2). Then, we test the speed of our approach (Section

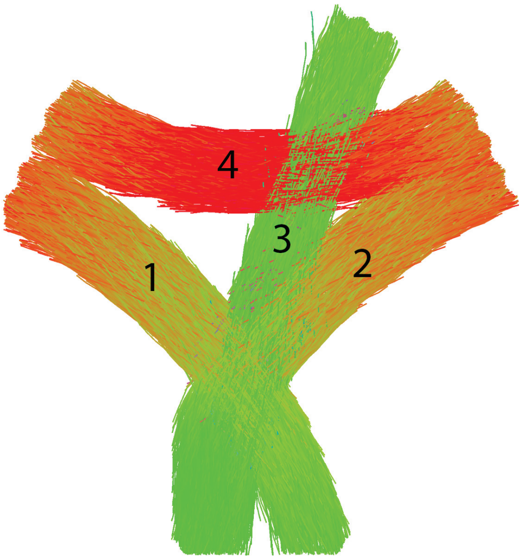


Figure 1.1: The virtual phantom used for our experimental campaign

4.3). For this purpose, we exploit suitable virtual phantoms. Finally, we test our approach on the whole for a real case study (Section 4.4).

4.1 Test of the pre-processing phase

As previously pointed out, the pre-processing phase of our approach includes the string-based representation and the clustering of available fibers. To validate this phase (measuring its effectiveness and efficiency), we performed several tests on a virtual phantom created by the phantom generator described in [Caruyer *et al.* (2014)]. Figure 1.1 graphically displays this phantom. We asked an expert to manually annotate the corresponding data. Annotations performed by him are represented in the same figure; in particular, a number is associated with each bundle. As shown in Figure 1.1, the expert defined four fiber-bundles, namely: *(1-2)* diagonal, *(3)* vertical, and *(4)* horizontal bundles, which were used as a gold standard. Indeed, our validation task required a ground truth and, in our opinion, the best possible ground truth was represented by a set of fiber-bundles manually annotated by an expert.

To perform our test activity, we compared the results of four approaches, namely: *(i)* our approach, with the usage of k-means as clustering algorithm in Step 3; *(ii)* our approach, with the exploitation of EM as clustering algorithm; *(iii) QuickBundles*, applied directly on the phantom; *(iv)* a baseline feature-based k-means, applied directly on the string-based representation of the phantom.

Whenever necessary (in particular, for cases *(i)*, *(ii)*, and *(iv)*), fibers were transformed into strings by applying the technique presented in Section 3.1; each fiber turned out to be characterized by 30 voxels. We set the cardinality of the string alphabet to 6. As for cases *(i)* and *(iv)*, we set the number of required clusters (i.e., k) to 4. As for cases *(i)* and *(ii)*, we set $\pi_1 = 1$ and $\pi_2 = 1$ for SBED and specified a constraint set χ aimed to

	Precision	**Recall**	**F-measure**	**Overall**
Our approach - k-means	0.72	0.98	0.83	0.60
Our approach - EM	0.77	0.89	0.83	0.63
QuickBundles	0.91	0.55	0.69	0.50
K-means Baseline	0.44	0.44	0.44	-0.12

Table 1.1: Results for Cluster 1

	Precision	**Recall**	**F-measure**	**Overall**
Our approach - k-means	0.75	0.97	0.85	0.65
Our approach - EM	0.76	0.92	0.83	0.63
QuickBundles	0.94	0.51	0.66	0.48
K-means Baseline	0.47	0.41	0.44	-0.05

Table 1.2: Results for Cluster 2

	Precision	**Recall**	**F-measure**	**Overall**
Our approach - k-means	0.92	0.32	0.48	0.29
Our approach - EM	0.70	0.51	0.59	0.30
QuickBundles	0.51	0.78	0.62	0.04
K-means Baseline	0.48	0.35	0.40	-0.03

Table 1.3: Results for Cluster 3

	Precision	**Recall**	**F-measure**	**Overall**
Our approach - k-means	0.93	0.99	0.96	0.92
Our approach - EM	0.98	0.94	0.96	0.92
QuickBundles	0.82	0.99	0.90	0.78
K-means Baseline	0.33	0.64	0.44	-0.63

Table 1.4: Results for Cluster 4

avoid, for instance, matches between horizontal and vertical voxels. As for case *(iii)*, we run QuickBundles with a threshold value equal to $6mm$, which empirically was proved to produce the best results. Finally, as for *(iv)*, we considered each of the 30 voxels of a fiber as a feature of the fiber itself. Therefore, we defined the input of the feature-based clustering algorithm as a set of feature vectors $w_i = (x_{i_1}, \ldots, x_{i_m})$, where each w_i corresponded to a fiber f_i and each x_{i_j} corresponded to a voxel v_{i_j} of f_i. The approach delineated by *(iv)* can be used as a baseline of comparison because the only pre-processing task it requires consists of the transformation of fibers into strings. Therefore, it is lightweight and fast, and, if it provides satisfying results, it should be preferred to the other ones.

To evaluate the accuracy of all four approaches, we compared the results obtained by them with the expert's annotation in such a way as to compute Precision, Recall, F-measure and Overall [Powers (2011)]. Obtained results are reported in Tables 1.1 – 1.4. Furthermore, qualitative and graphical representations of the clusters computed by cases *(i)* and *(iii)* are shown in Figures 1.2 and 1.3, respectively.

From the analysis of Tables 1.1 – 1.4, we can draw the following conclusions:

- The baseline approach is light and quick; however, its accuracy is so low to make it inapplicable in our context. In more detail, the values of Precision, Recall, F-measure

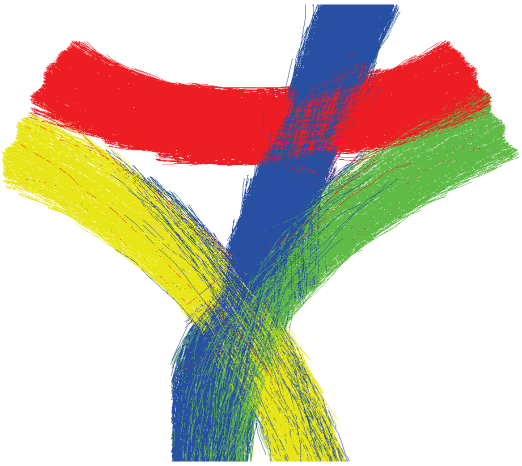


Figure 1.2: The clusters generated by our approach with the adoption of k-means as clustering algorithm

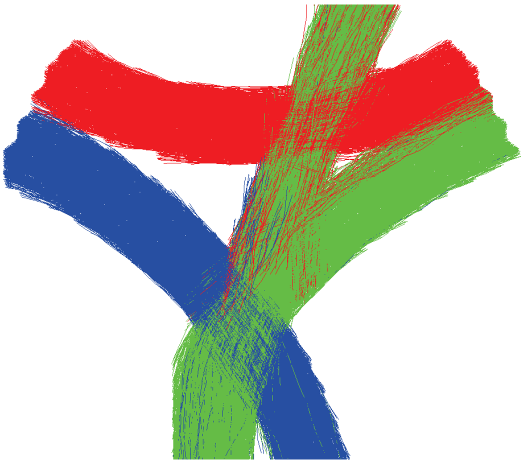


Figure 1.3: The clusters generated by QuickBundles

and Overall returned by it are worse than the ones of all the other approaches and, in any case, are unsatisfying for all clusters.

- QuickBundles guarantees a very high Precision only for Clusters 1 and 2. As for these clusters, it obtains the best Precision among the four approaches into consideration. However, this result is obtained at the price of having a very low Recall. This behavior is caused by the fact that QuickBundles ignores information about voxel directionality and considers only voxel proximity. The same reasons represent the cause of the low value of Precision obtained by QuickBundles for Cluster 3. By contrast, QuickBundles shows a high value of Recall for this last cluster, which, if related with the low value of Recall for Clusters 1 and 2, testifies the difficulty of this approach to distinguish among bundles near in space.
- If we focus on cases *(i)* and *(ii)*, which correspond to two versions of our approach, we cannot observe a substantial difference between the results obtained by applying

k-means and those returned by adopting EM. This fact represents a proof of the good robustness degree characterizing SBED.

- Results obtained for all the clusters testify that the Precision of our approach is generally higher than the one of the other approaches. At the same time, our approach can guarantee a satisfying level of completeness testified by a satisfying value of Recall. The satisfying results of our approach are further confirmed by the analysis of F-measure and Overall. In fact, QuickBundles returns a higher F-measure than the one obtained by our approach only for Cluster 3. However, for this cluster, as shown above, the Precision of QuickBundles is unacceptable, and this fact produces a very low value of Overall.

Summarizing, the previous test shows that our approach can guarantee a good balance between Precision and Recall in the extraction and characterization of fiber bundles when directionality information plays a key role.

4.2 Test of the SBED metric

In this test, we validated our approach's capability of characterizing WM fiber-bundles. The input dataset consisted of a virtual diffusion MR phantom, generated by Phantomas [Caruyer *et al.* (2014)]. This phantom accurately simulated the brain complexity with the fiber geometry used in the 2nd HARDI Reconstruction Challenge (ISBI 2013). The ground truth was obtained by requiring our expert to manually segment all the fiber-bundles in the phantom. In this way, the approximate shapes of these fiber-bundles were defined and 17 models were identified. These are illustrated in Figure 1.4. In this figure, each bundle is colored on the basis of the standard code of colors for tractography defined in [Le Bihan *et al.* (2001)] and described in Section 3.1.

To perform our validation, we measured the distance between each of the 17 models and the fibers in the phantom. To carry out this task, we applied both SBED and the classic edit distance. We compared each obtained result with the ground truth and computed Precision, Recall, F-measure and Overall [Powers (2011)] for both SBED and the edit distance. In Table 1.5, we illustrate the results obtained for each model, whereas, in Table 1.6, we present the average values of Precision, Recall, F-measure and Overall for the two distances. Before illustrating obtained results, we must preliminarily observe that, in our reference context, Precision is more important than Recall because the number of fibers generated by tractography algorithms usually does not reflect the number of real fibers of a human brain.

From the analysis of obtained results, it is possible to draw the following conclusions:

- SBED reaches a very high average Precision (i.e., 78.85%).

- It also reaches a satisfying average Recall (i.e., 36.37%) with peaks of 92%.

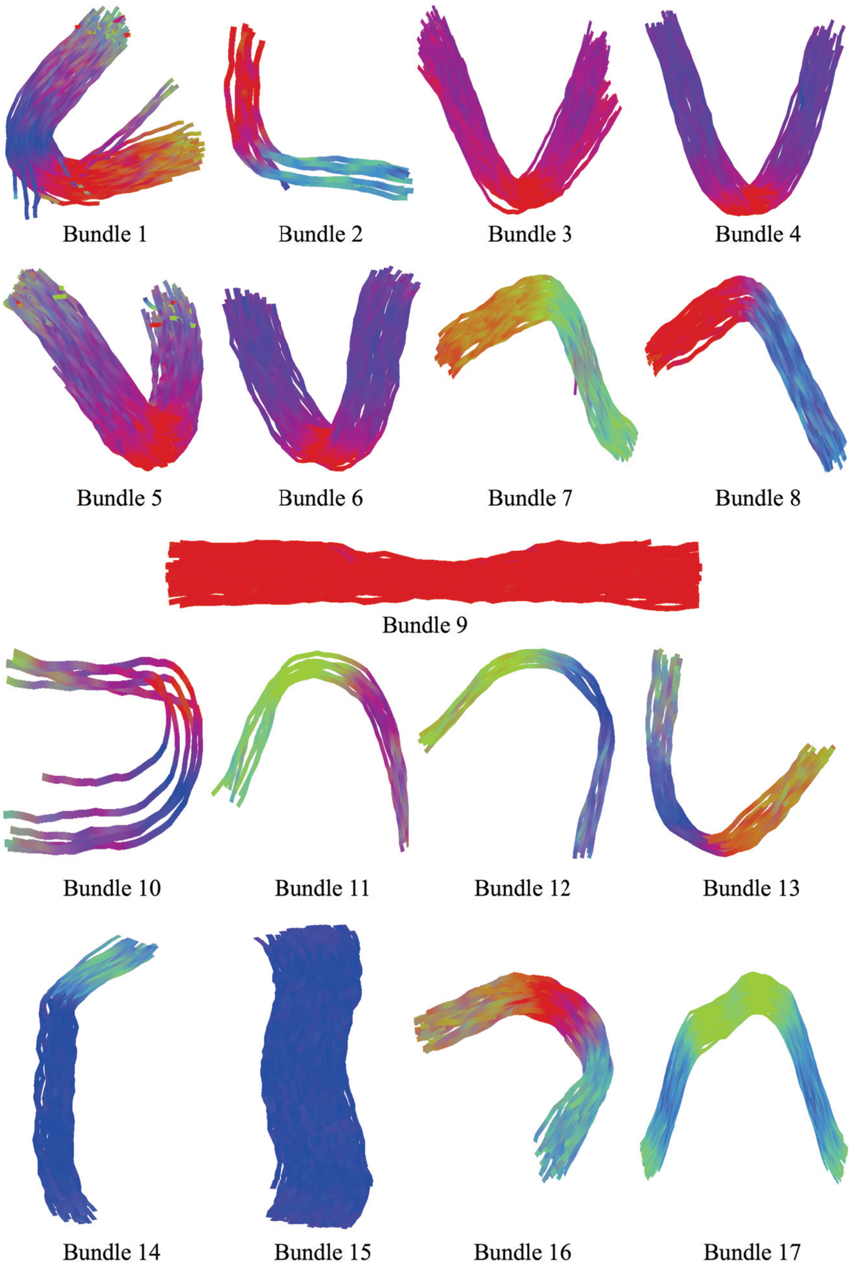


Figure 1.4: The 17 bundles identified in the diffusion MR phantom adopted in our test

- The average Precision, Recall, F-measure and Overall of SBED are higher than the ones of the edit distance (i.e., +27.60%, +5.70%, +9.69%, +17.66%, respectively).
- There are few models (i.e., 3 and 17), where both SBED and the edit distance do not work properly. These cases need deep analyses in the future, which, probably, will lead to perform some corrections on both approaches.

All the previous results allow us to conclude that the adoption of the new metric SBED (which, as specified above, represents one of the main contributions of this chapter) represents a step forward in the computation of string similarity and dissimilarity.

		Model							
		1	2	3	4	5	6	7	8
SBED	Precision	1	1	0.2081	0.4069	1	0.5194	0.9412	0.9474
	Recall	0.0220	0.1111	0.7701	0.6344	0.4886	0.6505	0.3556	0.3462
	F-measure	0.0430	0.2000	0.3277	0.4958	0.6565	0.5776	0.5162	0.5071
	Overall	0.0220	0.1111	−2.1604	−0.2903	0.4886	0.0486	0.3334	0.3270
Edit	Precision	0	0	0.3602	0.4143	0	0.5276	1	1
	Recall	0	0	0.7701	0.6236	0	0.6505	0.3556	0.3270
	F-measure	NULL	NULL	0.4908	0.4978	NULL	0.5826	0.5246	0.4928
	Overall	NULL	NULL	−0.5978	−0.2580	NULL	0.0681	0.3556	0.3270

		9	10	11	12	13	14	15	16	17
SBED	Precision	0.7333	0.6689	1	1	1	1	0.9797	1	0
	Recall	0.9245	0.6689	0.0577	0.0198	0.1154	0.0155	0.7472	0.2553	0
	F-measure	0.8179	0.6689	0.1091	0.0388	0.2069	0.0305	0.8478	0.4068	NULL
	Overall	0.5883	0.3378	0.0577	0.0198	0, 1154	0.0155	0.7317	0.2553	NULL
Edit	Precision	0.7333	0.6972	1	0	0	1	0.9797	1	0
	Recall	0.9245	0.6689	0.0192	0	0	0.0199	0.7472	0.1064	0.0007
	F-measure	0.8179	0.6828	0.0377	NULL	NULL	0.0390	0.8478	0.1923	NULL
	Overall	0.5883	0.3784	0.0192	NULL	NULL	0.0199	0.7317	0.1064	NULL

Table 1.5: Results obtained by applying SBED and the edit distance on the 17 models into consideration

	Avg. Precision	**Avg. Recall**	**Avg. F-measure**	**Avg. Overall**
SBED	0.7885	0.3637	0.4032	0.3346
Edit Distance	0.5124	0.3067	0.3063	0.1581

Table 1.6: Comparison of the average Precision, Recall, F-measure and Overall obtained by applying SBED and the edit distance

4.3 Speed test of the string-based algorithm

In carrying out this test, we started from the following considerations: *(i)* the pre-processing phase of our approach must be carried out once and for all for each brain; as a consequence, it is possible to exclude it from our test; *(ii)* one of the quickest approaches for the extraction of WM fiber-bundles is QuickBundles; as a consequence, it appears reasonable to compare our approach with it.

We applied both our approach (after the pre-processing phase) and QuickBundles to the virtual diffusion MR phantoms shown in Figures 1.1 and 1.4.

At the end of these tests, we obtained that, on average, our approach showed to be 2.54 times slower than QuickBundles. For instance, to process the phantom of Figure 1.4, QuickBundles took 0.87 seconds, whereas our approach needed 2.21 seconds.

However, in our opinion, with regard to these results, two considerations are in order. In fact: *(i)* even if slower than QuickBundles, our approach show a quickness acceptable for real cases; *(ii)* as previously pointed out, QuickBundles has an important limitation in that it is incapable of distinguishing among bundles near in space; our approach overcomes this limitation, as testified by the higher values of accuracy measures obtained in the previous tests.

As a consequence, it appears reasonable to partially sacrifice quickness in favor of accuracy (and, therefore, to choose our approach instead of QuickBundles) in all those cases in which the accuracy of results is the most important feature.

4.4 Application on a control subject

Till now, we carried out our tests on virtual phantoms. Now, it is time to test our approach as a whole and on a real case. For carrying out this task, we required the collaboration of a healthy volunteer, on whom our approach could have been applied. This volunteer underwent a MR examination on a 3 Teslas Siemens Prisma MR System (64 channels head-coil). Diffusion protocol consisted on the acquisition of 100 slices in the AC-PC plane, TR/TE = 13700/69 ms, FOV = 160×160, with a spatial resolution of 1.5 mm^3 along 45 gradient directions ($b = 3000\ s.mm^{-2}$). The Orientation Distribution Function (ODF) and the probabilistic tractography were computed using the algorithms of MRtrix [Tournier *et al.* (2012)].

We required our expert to draw approximate shapes in such a way as to extract two fiber bundles. These were Corpus Callosum (CC) forceps minor - Figure 1.5(a) - and right Cortico-Spinal Tract (CST) - Figure 1.6(a).

To carry out our validation task, first we transformed the available tractography fibers into strings. Then, we computed the Dissimilarity Matrix and performed the clustering activity (in particular, we chose EM as clustering algorithm). Finally, we carried out the extraction of fiber-bundles and, next, their characterization.

The extracted fibers are shown in Figure 1.5(b) and 1.6(b). An anatomical analysis of these figures is already sufficient to verify that our approach was capable of well extracting both forceps minor of CC (Figure 1.5(b)) and right CST (Figure 1.6(b)).

5 Discussion

After having described our approach, we want to point out that it can be easily extended from fiber-bundle extraction and characterization to several other contexts in which it is necessary to perform multi-dimensional (and, more in general, multi-view) clustering and characterization activities and/or in tasks requiring the integration of data belonging to different domains. In fact, it is sufficient to associate a color with each axis of the corresponding multi-dimensional domain to suitably color the corresponding voxel, and, after all voxels have been colored, to suitably discretize the corresponding color representation.

For instance, our approach can be adopted in all those biomedical contexts in which it is necessary to perform multi-dimensional (and, more in general, multi-view) clustering and characterization activities. Moreover, it can be adopted in those contexts somehow requiring a multi-view clustering and a possible model-guided characterization of obtained clusters. As an example, in the clinical observation of the vital parameters of a patient [Ordóñez *et al.* (2008)], it is possible to simultaneously consider several measures, such as electrocardiogram, temperature, respiratory rate, etc. Thanks to our approach, all these measures can be analyzed simultaneously and compared in such a way as to evidence possible correlations.

Furthermore, our approach could be adopted also in contexts very different from the

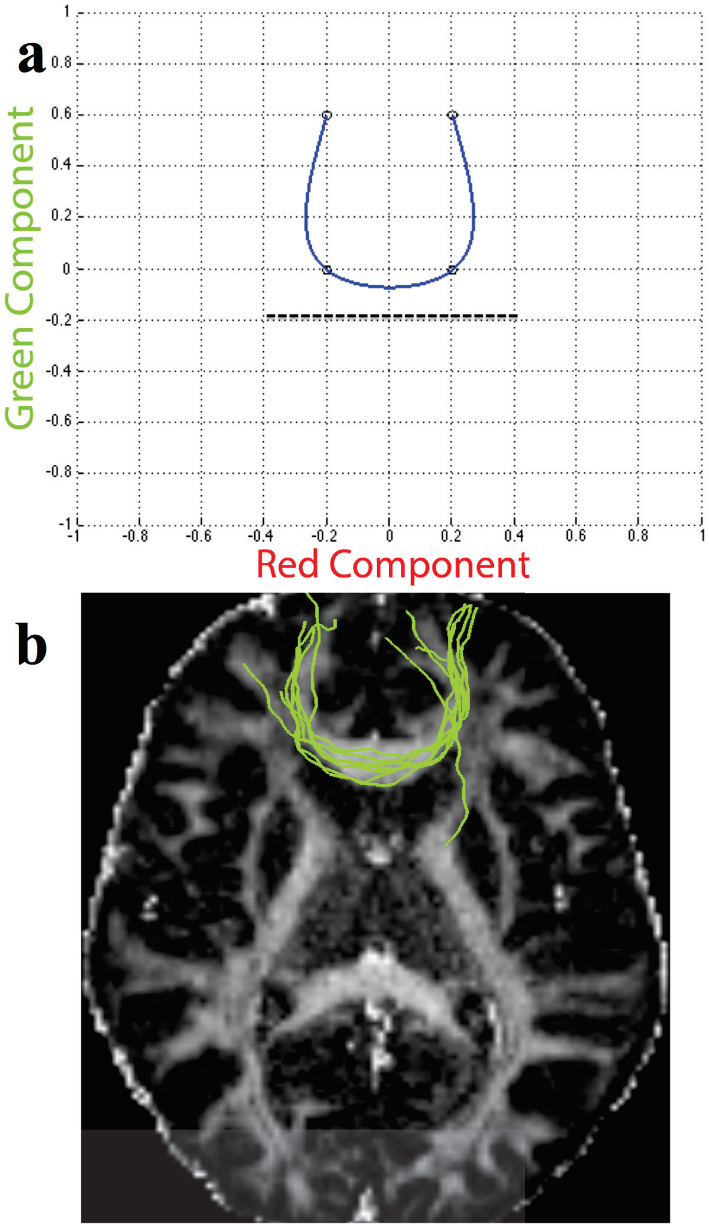


Figure 1.5: **a.** Approximate shape of Corpus Callosum (CC) and its axis of symmetry (black dotted line) drown by the operator; **b.** Extracted forcep minor of CC fibers (green)

biomedical ones, whenever multi-view clustering and characterization tasks must be carried out. Think, for instance, of the analysis of air flows or of weather perturbation in meteorology or to the discovery of hidden correlations in multi-sensor (and possibly heterogeneous) data streams related to a unique phenomenon.

Finally, the general philosophy underlying our approach can be extended to other multi-view data applications. As an example, the search of frequent structures in a multi-dimensional space can be reduced to the search of frequent patterns in a set of strings. An analogous consideration holds for the search of specific structures in a multi-dimensional space.

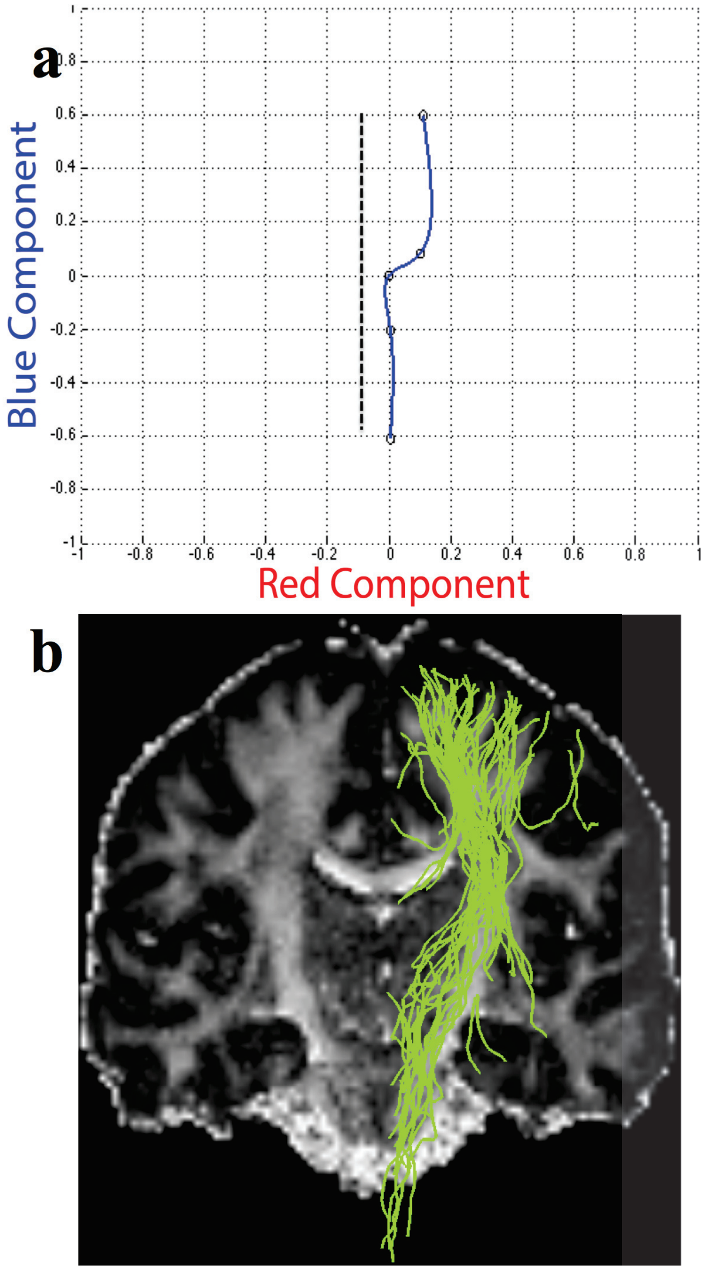


Figure 1.6: **a.** Approximate shape of Cortico-Spinal Tract (CST) and its axis of symmetry (black dotted line) drown by the operator; **b.** Extracted right CST fibers (green)

6 Conclusion

In this chapter, we have proposed a string-based approach to extracting and characterizing WM fiber-bundles. Our approach exploits a new string-based formalism, allowing an alternative representation of WM fibers, a new string similarity metric, a WM fiber clustering technique and a new model-based characterization algorithm. We have also collocated our approach in the context of related literature and we have performed an experimental campaign to test it.

Our approach overcomes different limitations of the related ones proposed in the past.

For instance, it allows a better integration of a-priori information provided by a neuroanatomist (which is not possible in QuickBundles). Indeed, the usage of a string-based model, representing the shape of a particular fiber-bundle, allows an easy extraction of just those fibers having the same structure as the provided model.

The key ideas of our approach can be applied in many other contexts, even very far from the biomedical ones. In Section 5, we have discussed some of these contexts, which we consider particularly interesting. As for future work, we plan to explore this possibility and to possibly define some approaches that apply the leading ideas presented in this chapter in the research scenarios mentioned in Section 5 and in further ones that could prove promising.

Chapter 2

Extension of String-Based Algorithm with Integration of Spatial Information

Contents

Published papers in this chapter

- F. Cauteruccio*, C. Stamile*, G. Terracina, D. Ursino, D. Sappey-Marinier, **"Integrating QuickBundle with a string-based fiber representation for extracting anatomically coherent WM fiber-bundles,"** *in: Image Analysis and Recognition (ICIAR)*, LNCS vol. 9730, pp. 633-641, 2016.
- F. Cauteruccio*, C. Stamile*, G. Terracina, D. Ursino, D. Sappey-Marinier, **"Integrating QuickBundles into a model-guided approach for extracting "anatomically-coherent" and "symmetry-aware" White Matter fiber-bundles,"** *in: 26th Italian Workshop on Neural Networks (WIRN) 2016*, to appear.

*These authors contributed equally to this work

1 Introduction

In the past, White Matter (WM) fiber-bundles have been largely investigated from several viewpoints. Different papers (e.g., [Yeatman *et al.* (2012), Colby *et al.* (2012)]) show how the analysis of WM structures is important to better understand and predict how the effects caused by certain neurodegenerative pathologies, such as multiple sclerosis, affect the brain, causing motor disability [Pantano *et al.* (2002)] and other symptoms, like fatigue [Rocca *et al.* (2014)]. Furthermore, [Golby *et al.* (2011)] shows how fiber-bundles extraction is used in neurosurgical planning to help the surgeon during an operation.

WM fiber-bundles could be extracted with the support of expert neuroanatomists, who manually delineate the regions of interest [Mårtensson *et al.* (2013)]. However, this way of proceeding is time consuming and operator dependent, avoiding the possibility to analyze data derived from the analysis of large cohorts.

To overcome such limitations, different automatic algorithms for isolating and extracting WM fiber-bundles have been proposed in the literature [Yeatman *et al.* (2012), Zhang *et al.* (2008), Garyfallidis *et al.* (2012)]. We can group them in two categories, namely: *(i)* atlas-based algorithms [Yeatman *et al.* (2012), Zhang *et al.* (2008)], which need an a priori knowledge about the location of certain WM brain regions, and *(ii)* algorithms that do not need a priori knowledge [Garyfallidis *et al.* (2012)].

Atlas-based approaches are very simple and fast. They are based on the registration of pre-labeled WM fiber-bundles atlases on the subject's image. However, these approaches suffer of different limitations, e.g., only the fiber-bundles specified in the WM atlases can be extracted, and the quality of the extracted fibers depends on the algorithm used to register the atlases. Interestingly, these approaches could integrate enhancing techniques, such as clustering [Zhang *et al.* (2008)], which are supervised and tuned by experts through some parameters.

The approaches that do not need a priori knowledge are based on the formalization of particular similarity and proximity measures in $\mathbb{R}^3$ aimed to: *(i)* group in the same subset those fibers having the same structure; *(ii)* maximize the discrimination of fibers having different forms. Among them, *QuickBundles* (QB) [Garyfallidis *et al.* (2012)] is, probably, the most famous one. Due to its simplicity, QB showed good results in terms of fiber-bundle extraction and execution time. However, as a side effect, the pure unsupervised approach used by it could lead to the extraction of anatomically incoherent regions. Indeed, the process of fiber generation adopted by QB does not take prior information from neuroanatomists into consideration. Nevertheless, this last information could play a key role for obtaining more satisfying results. As a consequence, this limitation could generate a bias in real applications, where anatomical information is important for analysis.

In this chapter, we propose an approach that integrates QB with a string-based fiber representation in such a way as to extract anatomically homogeneous WM fiber-bundles. Given a set $F = \{f_1, f_2, \ldots, f_n\}$ of WM fibers to cluster and a set $M = \{m_1, m_2, \ldots, m_k\}$ of models, our approach consists of the following steps: *(i)* application of a string-based fiber representation formalism to construct the set T (resp., V) of the strings corresponding to F (resp., M); *(ii)* construction of a matrix D such that $D[i, j]$ indicates the dissimilarity degree between the string corresponding to f_i and the one associated with m_j; *(iii)* assignment of each fiber of F to at most one model of M on the basis of D in such a way as to produce a set $B = \{b_1, b_2, \ldots, b_k\}$ of WM fiber-bundles; interestingly, at this stage, it is not possible to distinguish symmetrical structures; *(iv)* application of QB to each bundle of B for overcoming this limitation. We conducted an experimental campaign to compare the performance of our approach with that of QB. As will be clear below, obtained results are very encouraging.

This chapter is organized as follows: in Section 2, we present QB and the string-based fiber representation. In Section 3, we provide a technical description of the proposed approach. In Section 4, we illustrate the experimental campaign conducted to evaluate it. Finally, in Section 5, we draw our conclusion and delineate some possible future developments of this research.

2 Preliminaries

QuickBundles. QuickBundles (QB, for short) [Garyfallidis *et al.* (2012)] is an efficient unsupervised algorithm to cluster WM fiber-bundles. The idea behind it is simple. At each iteration, a given fiber of the tractography could be assigned to a pre-existing cluster or it could generate a new cluster. Initially, the first fiber is simply assigned to a first cluster containing only it. As for the other fibers, the assignment of a fiber to a cluster is performed according to a given threshold θ. If the distance between the current fiber and the centroid of at least one cluster is less than θ, the fiber is assigned to the cluster corresponding to the minimum distance. Otherwise, if there does not exist any cluster whose centroid has a distance from the current fiber less than θ, a new cluster is created and the fiber is assigned to it. This process is repeated until all the fibers in the tractography are assigned to a cluster. In order to measure the distance between two fibers, a new metric, called Minimum Average Direct Flip (MDF), is introduced. Differently from most classical clustering algorithms, like K-Means, in QB there is no re-assignment or updating step. So, when a fiber is assigned to one cluster, it is not possible for that fiber to change its cluster.

String-based fiber representation. In our application scenario, involved fibers are translated into strings and they represent multi-view data. As a consequence, the one-to-one matching assumption used in classical string-based distance metrics (like the Levenshtein distance [Levenshtein (1965)]) is weak and the corresponding metric could not work properly. The Semi-Blind Edit Distance (SBED, for short) [Cauteruccio *et al.* (2015), Stamile *et al.* (2015)a] was conceived to overcome this assumption and to allow the computation of the minimum edit distance between two strings, provided that finding the optimal matching schema, under a set of constraints, is part of the problem. It can be summarized as follows.

Let $\langle \bar{s}_1, \bar{s}_2 \rangle$ be an alignment (in classical terms) for s_1 and s_2, let $M_{\langle \pi_1, \pi_2, \chi \rangle}$ be a $\langle \pi_1, \pi_2, \chi \rangle$-matching schema with constraints χ. We say that $\langle \bar{s}_1, \bar{s}_2 \rangle$ has a match at j if either: *(i)* $\bar{s}_1[j] \in \Pi_1$, $\bar{s}_2[j] \in \Pi_2$ and $\bar{s}_1[j] = \bar{s}_2[j]$, or *(ii)* $\bar{s}_1[j]$ and $\bar{s}_2[j]$ match, according to $M_{\langle \pi_1, \pi_2, \chi \rangle}$. The *distance* between $\bar{s}_1$ and $\bar{s}_2$ under $M_{\langle \pi_1, \pi_2, \chi \rangle}$ is the number of positions at which $\langle \bar{s}_1, \bar{s}_2 \rangle$ does not have a match.

Given two integers π_1 and π_2, such that $0 < \pi_1 \leq |\Pi_2|$ and $0 < \pi_2 \leq |\Pi_1|$, the $\langle \pi_1, \pi_2, \chi \rangle$-edit distance between s_1 and s_2 ($\mathcal{L}_{\langle \pi_1, \pi_2, \chi \rangle}(s_1, s_2)$ for short) is the minimum edit distance that can be obtained according to any $\langle \pi_1, \pi_2, \chi \rangle$-matching schema and over any string alignment $\langle \bar{s}_1, \bar{s}_2 \rangle$ of s_1 and s_2.[1]

As an example, let $s_1 = \texttt{AAABCCDCAA}$ and $s_2 = \texttt{EEFGHGGFHH}$, which determines $\Pi_1 = \texttt{\{A,B,C,D\}}$ and $\Pi_2 = \texttt{\{E,F,G,H\}}$. For $\pi_1 = \pi_2 = 1$ and the constraint $\chi = \{\langle \texttt{A}, \texttt{E} \rangle\}$, the best alignment gives $\mathcal{L}_{\langle 1,1,\chi \rangle}(s_1, s_2) = 5$ with the optimal matching schema `{A}-{H}`, `{B}-{E}`, `{C}-{F}`, and `{D}-{G}`.

To the best of our knowledge, there is no approach in the literature facing the same problem handled by SBED, even if there are some variants. In fact, several approaches have been proposed to carry out the computation of the similarity of parameterized strings, i.e., strings where some of the symbols act as parameters that can be properly substituted at no cost (see, [Baker (1996), Hazay *et al.* (2007), Apostolico *et al.* (2007), Greco and Terracina (2013)]. Recently, a detailed survey on parameterized matching appeared in [Mendivelso and Pinzón (2015)].

3 Technical description of the proposed approach

Our approach joins together QB and SBED and aims at overcoming the main problem of the former by using the latter. Its first ingredient is a WM fiber-bundle reference

---

[1] The interested reader can find all details about SBED and algorithms for its computation in [Cauteruccio *et al.* (2015), Stamile *et al.* (2015)a].

model, which must represent an approximate shape of the fiber-bundle to extract. It could be obtained in two different ways, namely: *(i)* by exploiting a spline curve to draw the profile of the fiber-bundle of interest, or *(ii)* by importing the mean-line profile of the fiber-bundle of interest from an atlas of pre-labeled fiber-bundles. Both kinds of models can be constructed either by a generic user or with the support of an expert one. The second ingredient of our approach is a fiber representation formalism allowing fibers to be mapped on strings. Actually, a bijective mapping is needed. For this purpose, let $F = \{f_1, f_2, \ldots, f_n\}$ be a set of fibers. Here, a generic fiber $f_i \in F$ is defined as a sequence $f_i = (v_1, v_2, \ldots, v_m)$ of voxels in the three-dimensional space. We can assume, without loss of generality [Stamile *et al.* (2015)a], that all the fibers of F have the same number m of voxels. A color can be associated with each voxel $v_r \in f_i$, derived from its orientation in the space. Thus, a fiber $f_i \in F$ can be represented by using colors in the RGB color space. By discretizing the RGB space, we define a map $\Psi : RGB \rightarrow \Sigma$, where $\Sigma \subset \mathbb{N}_0$ and $|\Sigma| = s$. With the support of this map, a generic fiber f_i can now be expressed as a string in Σ^m.

Once the two main ingredients of our approach have been defined, it is possible to describe it. Specifically, let $F = \{f_1, f_2, \ldots, f_n\}$ be a set of WM fibers to cluster and let $M = \{m_1, m_2, \ldots, m_k\}$ be the set of models. Our approach consists of the following steps:

- Construction of the set $T = \{t_1, t_2, \ldots, t_n\}$ of the strings corresponding to F and of the set $V = \{v_1, v_2, \ldots, v_k\}$ of the strings corresponding to M. For this purpose, the fiber representation formalism described above is applied.

- Construction of a $n \times k$ matrix D. The element $D[i, j]$ of D indicates the dissimilarity degree computed by applying SBED on the string t_i, associated with f_i, and the string v_j, associated with m_j.

- Assignment of each fiber of F to at most one model of M as follows:

  - for each row i of D, let μ be the minimum value of this row and let j_μ be the corresponding column;
  - if μ is lesser than a certain threshold Th then f_i is assigned to m_{j_μ}; otherwise, f_i is not assigned to any model.

- At the end of this step, we have a set $B = \{b_1, b_2, \ldots, b_k\}$ of WM fiber-bundles, one for each model of M. However these bundles have a weak point. Indeed, the assignment approach above is incapable of distinguishing among symmetrical structures [Stamile *et al.* (2015)a]. To overcome this limitation, for each bundle $b_l \in B$, we apply QB to it. QB returns the same bundle b_l if it does not present a symmetrical structure. Otherwise, QB splits b_l into two symmetrical bundles b'_l and b''_l.

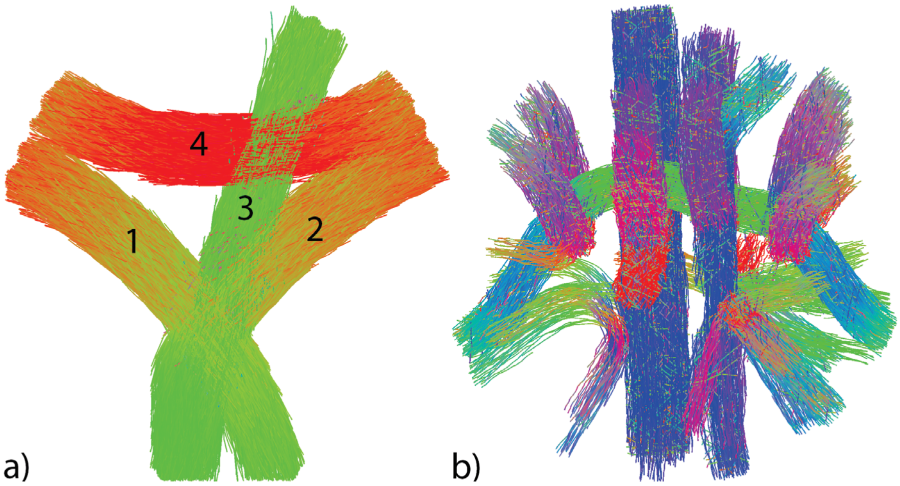


Figure 2.1: The two phantoms used in our experimental campaign

4 Experiments

Our experimental campaign consisted of two phases. The former was devoted to tune our approach. The latter aimed to compare it with the classical QB. In both cases we exploited simulated diffusion phantoms, as well as some classical performance measures, namely, Precision, Recall, F-Measure and Overall [Powers (2011)].

Phase 1: Parameter Tuning. As for this phase, the input dataset consisted of the virtual phantom shown in Figure 2.1(a) and created by Phantomas [Caruyer *et al.* (2014)]. In order to obtain the ground truth, experts segmented this phantom manually into 4 fiber-bundles, which are numbered in Figure 2.1(a). To find the best value of the input parameter Th, we considered different values of it ranging from 0.20 to 0.50. The corresponding values of Precision, Recall, F-Measure and Overall for the four models are reported in Figure 2.2. From the analysis of this figure, we can observe that, from $Th = 0.20$ to $Th = 0.36$, an increase of Th leads to an increase of at least one between Precision and Recall and to an increase of both F-Measure and Overall, which (we recall) are parameters combining Precision and Recall. Starting from $Th = 0.36$ no further increase of the values of performance measures can be observed in any model, and our approach shows a stable behavior. As a consequence, we chose to set Th to the middle of this range and we set it to 0.44.

Phase 2: Comparison with the classical QB. After having tuned Th, we applied the classical QB on the previous phantom to compare the performance of our approach (with $Th = 0.44$) with that of QB. The obtained results are reported in Table 2.1. From the analysis of this table we can observe that our approach shows a much higher average Precision, a slightly lower average Recall, a higher average F-Measure and a much higher average Overall than QB.

To obtain a (possible) confirmation of this result, we applied both our approach, with

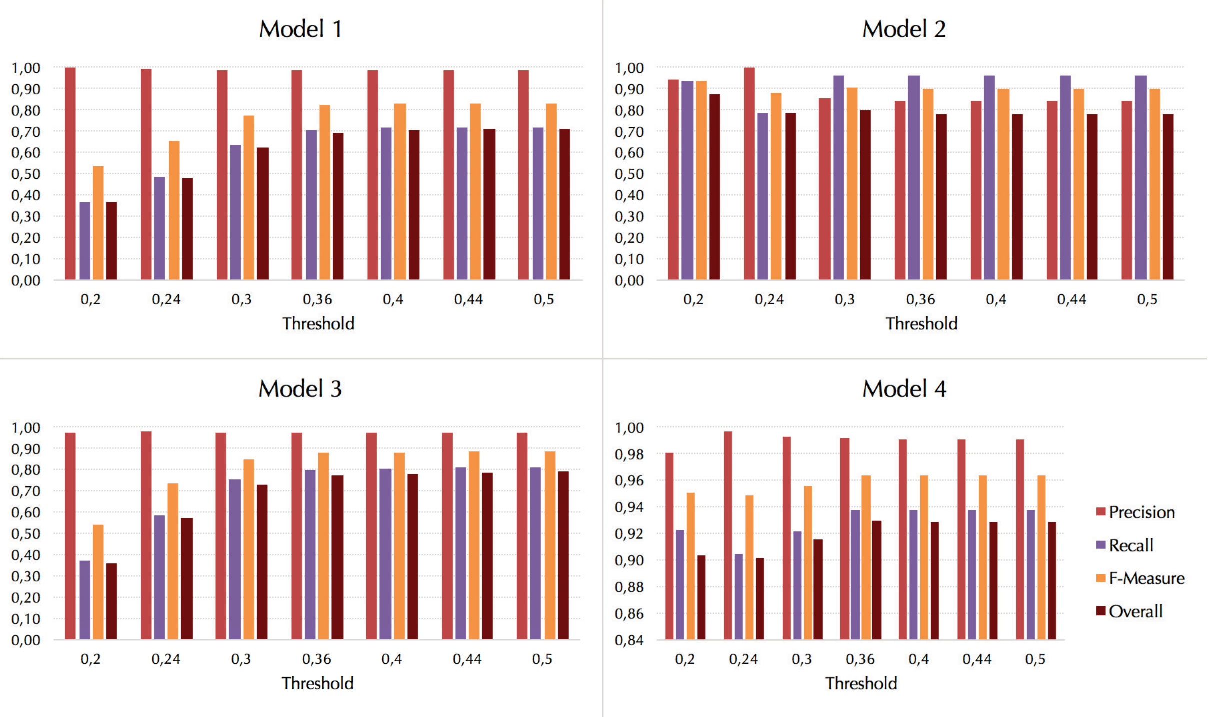


Figure 2.2: Variation of the four performance measures against the threshold Th for each model of the phantom of Figure 2.1(a)

	Our approach				**QB**			
Model	**Precision**	**Recall**	**F-Measure**	**Overall**	**Precision**	**Recall**	**F-Measure**	**Overal**
1	0.99	0.72	0.83	0.71	0.96	1.00	0.98	0.95
2	0.84	0.96	0.90	0.78	0.54	0.77	0.64	0.12
3	0.97	0.81	0.88	0.78	0.54	0.77	0.64	0.12
4	0.99	0.94	0.96	0.93	1.00	0.93	0.96	0.93
Avg values	0.95	0.85	0.89	0.80	0.76	0.87	0.80	0.54

Table 2.1: Performance values obtained by our approach and QB when applied on the phantom of Figure 2.1(a)

$Th = 0.44$, and QB on a second phantom shown in Figure 2.1(b). This consisted of a virtual diffusion MR phantom that accurately simulates the brain complexity with the fiber geometry used in the 2nd HARDI Reconstruction Challenge (ISBI 2013). Data were generated by means of Phantomas [Caruyer *et al.* (2014)]. To obtain the ground truth, experts segmented all the fiber-bundles in the phantom manually. At the end of this task, they defined the approximate shapes of these fiber-bundles; in particular, they identified 17 models. Numbering these models in Figure 2.1(b) was not possible due to the 3D nature of this image.

Because of its fully unsupervised nature, QB was capable of extracting just 8 out of the 17 fiber-bundles of the phantom. Our approach, instead, extracted all the 17 fiber bundles. As a consequence, a comparison between our approach and QB was possible only for the 8 models detected by QB. Obtained results are reported in Table 2.2.

From the analysis of this table we can observe that QB generally shows a higher value

	Our approach				**QB**			
Model	**Precision**	**Recall**	**F-Measure**	**Overall**	**Precision**	**Recall**	**F-Measure**	**Overal**
1	0.94	0.71	0.81	0.67	0.91	0.70	0.79	0.63
2	0.90	0.64	0.75	0.57	0.33	0.71	0.45	-0.76
3	0.78	0.41	0.54	0.29	0.31	0.75	0.44	-0.90
4	0.89	0.69	0.77	0.60	0.85	0.75	0.80	0.62
5	0.77	0.57	0.66	0.40	0.80	0.78	0.79	0.58
6	0.96	0.68	0.79	0.65	0.29	0.79	0.43	-1.13
7	0.94	0.29	0.45	0.27	0.14	0.82	0.24	-4.33
8	0.68	0.33	0.45	0.18	0.13	0.91	0.23	-4.96
Avg values	0.86	0.54	0.63	0.45	0.47	0.78	0.52	-1.28

Table 2.2: Performance values obtained by our approach and QB when applied on the phantom of Figure 2.1(b)

of Recall than our approach (this can be observed in 7 out of 8 models), but lower values of Precision (which were much lower in 5 cases, lower in 2 cases and slightly higher in only 1 case). As for the combined parameters F-Measure and Overall, our approach shows better results than QB in 6 out of 8 cases. Finally, if we consider the average values of these measures, we obtain that our approach shows a better Precision, a better F-Measure, a better Overall and a worse Recall than QB.

As previously pointed out, this difference of behavior is due to the nature of the structures extracted by QB. In fact, the fiber-bundles obtained by QB are not isolated but they are merged with other structures that are spatially near to them (even if they present a completely different anatomical meaning). By contrast, the fiber-bundles obtained by our approach are “purer”, since they contain only anatomically uniform fibers, corresponding to the fiber bundles of our interest. In support of this reasoning, it is well known that, in this application field, Precision is much more important than Recall.

5 Conclusion

In this chapter, we have proposed an approach that integrates QuickBundle with a string-based fiber representation for extracting anatomically coherent WM fiber-bundles. Our approach overcomes the main problem of QB, i.e., the possibility that it returns anatomically incoherent fiber clusters, in which the desired fiber-bundles are not isolated but merged with other structures spatially near to them (even if they present a completely different anatomical meaning). This work must not be considered as an ending point of our research efforts. Indeed, several developments are possible. First of all, we plan to extend our experiments from phantoms to real cases. After this, we would like to further improve QB in such a way as to correct fiber assignments to clusters when these assignments appear uncorrect in a second time. Finally, we will work on the SBED constraint optimizations in such a way as to define an approach that allows the discovery of the best constraints for improving the extraction of specific WM fiber-bundles.

III Development of Unsupervised Methods for Longitudinal Fiber-Bundle Analysis

Chapter 1

Histogram-Based Approach for Two Time-Points Analysis

Contents

Published paper in this chapter

- C. Stamile, G. Kocevar, F. Cotton, F. Durand-Dubief, S. Hannoun, C. Frindel, C. R.G. Guttmann, D. Rousseau, D. Sappey-Marinier, **“Detection of Longitudinal DTI Changes in Multiple Sclerosis Patients Based on Sensitive WM Fiber Modeling,”** PLoS ONE 11(5): e0156405. doi:10.1371/journal.pone.0156405, 2016.

1 Introduction

A major challenge of neuroimaging research consists in identifying new markers that can accurately characterize pathological processes and predict clinical outcomes. Achieving this goal is particularly crucial in Multiple Sclerosis (MS), the primary cause of neurological disability in young adults and remains without well-known etiology [Compston and Coles (2008)]. MS is a chronic demyelinating inflammatory disease of the central nervous system, characterized by white matter (WM) lesions that are well detected by conventional MRI. However, T2 lesion load is moderately correlated with the patient clinical status leading to the development of more sensitive techniques such as diffusion tensor imaging (DTI). DTI is a promising technique for white matter WM fiber-tracking and microstructural characterization of axonal/neuronal integrity and connectivity. By measuring water molecules motion in the three directions of space, numerous parametric maps can be reconstructed based on eigenvalues of the diffusion tensor. Among these, fractional anisotropy (FA), mean diffusivity (MD), and axial (λ_a) and radial (λ_r) diffusivities have extensively been used to investigate brain diseases [Basser *et al.* (1994)b, Rovaris *et al.* (2005), Kingsley (2006)a, Kingsley (2006)b] such as stroke [Lai *et al.* (2014), Zhang *et al.* (2013)], Parkinson disease [Nicoletti *et al.* (2006), Nicoletti *et al.* (2008)], brain tumors [Server *et al.* (2014), Sternberg *et al.* (2014)] and also normal aging [Cavallari *et al.* (2013), Inano *et al.* (2011)]. In MS, DTI has proved to be sensitive enough to detect microscopic changes occurring in WM lesions, normal appearing white matter (NAWM) and subcortical grey matter (GM). Indeed, several studies have demonstrated higher MD and lower FA in lesions when compared to NAWM of MS patients [Filippi *et al.* (2000), Werring *et al.* (1999), Assaf and Pasternak (2008)] and to NAWM of healthy controls [Ciccarelli *et al.* (2001), Hannoun *et al.* (2012)a]. In contrast, FA was increased in subcortical GM structures such as the caudate nuclei and thalami of MS patients that are supposed to reflect dendritic neurodegeneration mechanisms [Hannoun *et al.* (2012)b]. Overall, these findings demonstrated that WM and GM tissues are subjected to numerous microstructural alterations in MS. However, it remains unclear whether these tissue alterations result from global processes, such as inflammatory cascades and/or neurodegenerative mechanisms, or local inflammatory and/or demyelinating lesions. Furthermore, these pathological events may occur along afferent or efferent WM fiber pathways, leading to antero- or retrograde degeneration [Mukherjee *et al.* (2002)]. Thus, for a better understanding of MS pathological processes spatial progression, an accurate and sensitive characterization of WM fibers along their pathways is needed. By merging the spatial information of fiber tracking [Mori *et al.* (1999)] with the diffusion metrics derived from the tensor, WM fiber-bundles could be modeled and analyzed along their profile. Such signal analysis of WM fibers can be performed by several methods providing either semi- or automated extraction of WM fiber-bundles. Semi-automated algorithms consisted in a manual extraction of the bundle by defining a set of regions of interest (ROIs) [Colby *et al.* (2012), Mårtensson *et al.* (2013), Corouge *et al.* (2006)] based on neuroanatomical knowledge. However, this task usually performed by an expert is time consuming and operator dependent. In order to overcome such limitations, fully automated algorithms have been implemented [Yeatman *et al.* (2012), Grigis *et al.* (2013)]. These methods enable systematic, large-scale analysis of fiber bundles in large subject populations. However they remain relatively insensitive to changes affecting only a small portion of fibers within a bundle. In this work, we introduce an automated method for the analysis of WM fascicles from DTI data, and the detection of small longitudinal changes along the fiber-tracts. Based on a Gaussian mixture model, this method provides a fine cross-sectional fiber-bundle analysis allowing the differentiation of “changed” and “unchanged” fibers of the bundle.

2 Material and Methods

2.1 Subjects

Five relapsing-remitting (RR) MS patients (4 women and 1 man, mean (±SD) age: 36.8±9.5 years; media disease duration: 4.24y; max 16.5 y) (median EDSS=2.5, range=[0-4]) and one healthy control (HC) subject (age: 24 years) were included in this study. Inclusion criteria specified that studied patients were diagnosed as RR MS and present at least one new Gadolinium-enhancing lesion during the six months preceding study enrollment. All patients were not treated with disease modifying drugs for at least one year before inclusion in the study, and remained untreated during the study period. In order to limit the nephrogenic damage risks associated to Gadolinium injection, creatinine clearance was checked every 2 weeks after inclusion. A clearance higher than 60ml/min was an exclusion criterion. This study was approved by the local ethics committee (CPP Sud-Est IV) and the French national agency for medicines and health products safety (ANSM). Written informed consents were obtained from all patients and control subjects prior to study initiation.

2.2 MRI protocol

All subjects underwent a weekly examination for a period of two months (8 time-points from W_1 to W_8). MRI protocol included a DTI and a FLAIR acquisition, that were performed on a 3T Philips Achieva system (Philips Healthcare, Best, The Netherlands) with a 16-channels head-coil. The DTI image set consisted in the acquisition of 60 contiguous 2mm-thick slices parallel to the bi-commissural plane (AC-PC), and were acquired using a 2D Echo-Planar Imaging (EPI) sequence (TE/TR = 60/8210 ms, FOV = $224 \times 224 \times 120mm$) with 32 gradient directions (b = 1000 s.mm-2). The nominal voxel size at acquisition ($2 \times 2 \times 2$ mm) was interpolated to $0.875 \times 0.875 \times 2mm$ after reconstruction. The FLAIR Vista 3D sequence (TE/TR/TI = 356/8000/2400 ms, FOV=$180 \times 250 \times 250mm$) consisted in the acquisition of 576 slices of 0.43 mm thickness oriented in the AC-PC axis with a nominal voxel size of $0.6 \times 0.43 \times 0.43mm$.

2.3 Longitudinal Variations Simulation

Two time-points of the control subject (W_1 and W_2) were used to simulate longitudinal variations. 120 different lesions were simulated on the control subject's FA maps obtained at W2. All the lesions were generated in 6 different fiber-bundles, namely, left and right, Cortico-spinal tract, inferior-fronto occipital fasciculi and forceps major and minor of corpus callosum extracted from the atlas [Hua *et al.* (2008)]. Small spherical variations (radius of 2 voxels) of FA values were generated according to the following equation: $FA(x) = \alpha * FA(x)$ where α (called reduction coefficient) varies from 0 to 1, and x is a voxel belonging to the spherical region.

3 Longitudinal Fiber-bundle Analysis Methods

The processing pipeline of DTI data is composed of three steps: 1) co-registration and diffusion metrics computation, 2) tractography bundle extraction and processing, and 3) longitudinal fiber-bundle analysis (Figure 1.1). In the following, we assume that each subject underwent a longitudinal DTI examination. Each longitudinal acquisition is composed of k time-points from 1 (W_1) to k (W_k).

3.1 Co-registration and Diffusion Metrics Computation

Diffusion images were processed using the FMRIB software Library (FSL) [Jenkinson *et al.* (2012)]. Eddy current correction was first applied on the diffusion volumes using the b0 volume (b = 0 s.mm-2) as reference. The tensor model was then fitted using the FDT module of FSL Longitudinal data co-registration was performed using the method described in [Keihaninejad *et al.* (2013)] based on DTI ToolKit (DTI-TK) including the following procedures: 1) generation of a patient specific template obtained from longitudinal diffusion tensor images, 2) co-registration of the resulting template to the Illinois Institute of Technology (IIT) atlas [Varentsova *et al.* (2014)], and 3) co-registration of each time-point data into the IIT atlas space by applying the previously obtained transformations to the initial longitudinal data. The resulting images were then used to compute diffusion metrics maps (FA, MD, . . .).

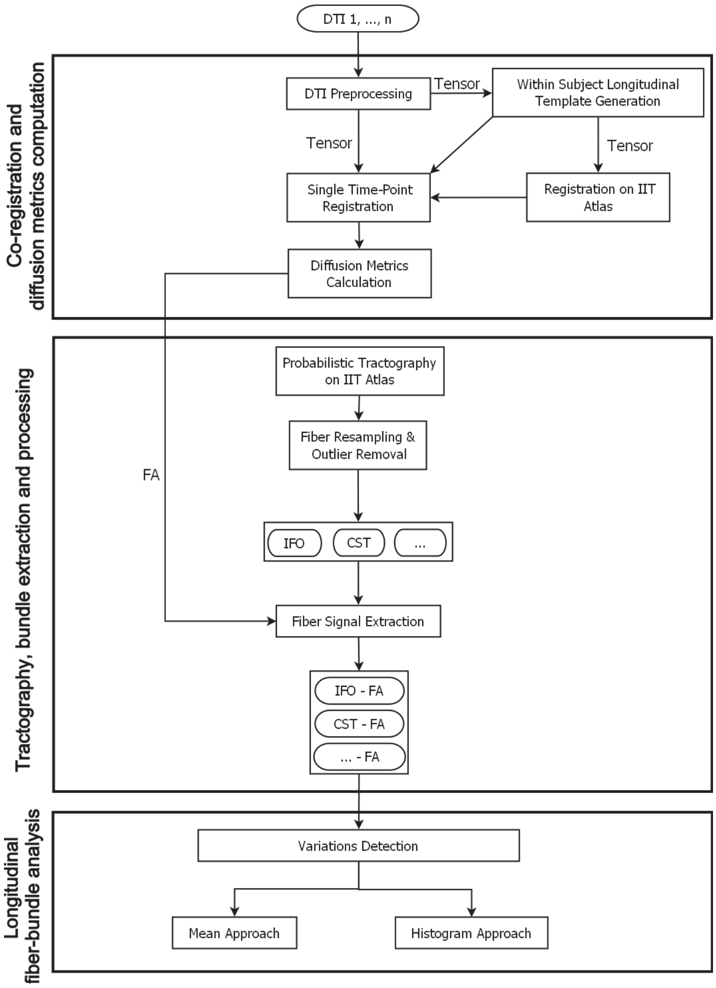


Figure 1.1: General overview of the processing pipeline for fiber-bundles longitudinal analysis: 1) Co-registration and diffusion metrics computation: DTI data were longitudinally co-registered and diffusion metrics were computed, 2) Tractography, bundle extraction and processing, 3) Longitudinal fiber-bundle analysis using both ‘mean” and “histogram” methods.

3.2 Tractography, Bundle Extraction and Processing

Probabilistic streamline tractography was performed using MRTrix [Tournier *et al.* (2012)] based on the fiber orientation density (FOD) information of the IIT Atlas. Twenty fiber bundles were extracted using 20 ROIs of the JHU atlas [Hua *et al.* (2008)] as seed and mask. Before the extraction of diffusion metrics, each fiber-bundle underwent a three-steps processing pipeline. The first step consisted in defining the start and end points of each fiber within the bundle. A classical K-Means algorithm [MacQueen *et al.* (1967)] was applied to the extracted raw fiber-bundle to generate two different clusters, R_1 for the starting points and R_2 for the ending points. Fiber points were reordered from R_1 to R_2 and fibers that did not link the two clusters were automatically removed. In a second step we resampled each fiber with $c = 100$ equidistant points (also called nodes). The third and last step consisted in the removal of fibers that were too distant from the center of the fiber-bundle. The mean fiber of the entire bundle was first computed using the method described by Klein and coworkers [Klein *et al.* (2007)]. Let F be a fiber-bundle $F = \{f_1, f_2, \dots, fm\}$ composed of m fibers of 100 points each, such as $f_i = \{p_1, \dots, p_{100}\}$ $1 \leq i \leq c$, $p_q = (x_q, y_q, z_q)$ and let $M_F = \{(p_{\mu,1}, p_{\sigma,1}), \dots, (p_{\mu,c}, p_{\sigma,c})\}$ be the mean skeleton of the fiber-bundle where $p_{\mu,i}$ and $p_{\sigma,i}$ $1 \leq i \leq c$ represent respectively the mean and the standard deviation (SD) computed from every point belonging to the $i-th$ cross-section. A fiber f_{ot} was considered an “outlier” if it contains at least one point p_r that was more than 3 SD away from the corresponding node in the mean fiber. The last step consisted in the automatic extraction of the diffusion metrics from the fiber-bundle. Based on the resampled fibers, each fiber point (x_i, y_i, z_i) was associated with the diffusion metric value of its corresponding voxel. Thus, every point of the fiber-bundle was associated with a set of diffusion metrics values allowing the characterization of the diffusion properties of the entire bundle. In this report, we focused on the FA metric, but the method is designed for any diffusion metrics that can be derived from the tensor.

3.3 Longitudinal Fiber-Bundle Analysis

Previous fiber-tract profile approaches [Mårtensson *et al.* (2013), Yeatman *et al.* (2012), Colby *et al.* (2012), O’Donnell *et al.* (2009)] were based on representing diffusion metrics along a given fiber-bundle by averaging the signal value at every cross-section of the bundle. This “mean” approach allowed first, to represent the mean and SD of any diffusion metric and second, to detect any changes along the fiber-bundle, as illustrated in Figure 1.2. In order to improve the sensitivity of the “mean” approach to detect small pathological changes along a fiber-bundle, the following “histogram” approach was developed. The histogram of FA values of a given bundle was fitted with a distribution model for every cross-section of the bundle. For a reference time-point i (W_i) and for a successive time-point $i + s$ (W_{i+s}), two histograms were independently fitted by means of a Gaussian mixture model (D_i, D_{i+q}) (Figure 1.3A1). The number of mixture components used to fit the histogram with the Gaussian model was computed by solving the following multiobjective optimization problem:

$$\underset{n\geq 1}{\text{minimize}} = \begin{cases} C(D_i(n)) = \sum_{t=1}^{n} log(D_i(x_t)) \\ C(D_{i+s}(n)) = \sum_{t=1}^{n} log(D_{i+s}(x_t)) \end{cases}$$

where D_i and D_{i+s} represent the Gaussian mixture model with n components and $\sum_{t=1}^{n} log(D(x_t))$ represents the log-likelihood of the distribution model calculated using the histogram of the cross-section. Parameters of D_i and D_{i+s} were estimated by the maximum likelihood estimation (MLE) algorithm [Myung (2001)]. The NSGA-II algorithm [Deb *et al.* (2002)] was used to solve the optimization problem. From the multiple Pareto's solutions, the one with the smallest number of mixtures was selected. Then, both distributions (D_i, D_{i+s}) were compared and local changes were detected by solving the following optimization problem:

$$\begin{aligned} \underset{\beta,\gamma}{\text{minimize}} \quad & P_D = \int_{\beta}^{\gamma} D_{i+s} \\ \text{subject to} \quad & P_F = \int_{\beta}^{\gamma} D_i \leq \zeta \\ & P_D > P_F \end{aligned} \tag{1.1}$$

where ζ is an input value called the tolerated error of the test. If no longitudinal variations were detected between W_i and W_{i+s}, the problem did not admit solutions. The obtained values β and γ were then used to differentiate fibers in two subsets. In the specific case of FA (but any other diffusivity metric can be used), fibers showing $\beta \leq FA \leq \gamma$ were labeled as "changed" while the other fibers were labeled as "unchanged" (Figure 1.3A2,B). The graphic representation of the diffusion metric values along the fiber-bundle was performed as previously described using the "mean" approach (Figure 1.3C) [Colby *et al.* (2012)]. For each subset of fibers ("changed" and "unchanged"), the mean and SD values of FA were computed for every cross-section of the bundle.

4 Results

4.1 Validation on Simulated Longitudinal Variations

The "histogram" method was applied on simulated longitudinal variations of FA (described in section 2.3) between two time-points (W_1-W_2) and was evaluated by measuring true positive (TP), true negative (TN), false positive (FP) and false negative (FN) in order to compute sensitivity, precision and F-Measure. Since our Gaussian mixture model depends on two parameters (ζ and β) the method was tested with different values of ζ, $\{0.02, 0.05, 0.08, 0.10, 0.12, 0.14\}$ and β was fixed to 0 in order to detect decreased FA changes as usually observed in pathological WM. As shown in Table 1.1, our method is strongly dependent on the parameter ζ. Low ζ values ($\zeta = 0.02$) makes the method more conservative and a large number of FN are thus detected. In contrast, high ζ value

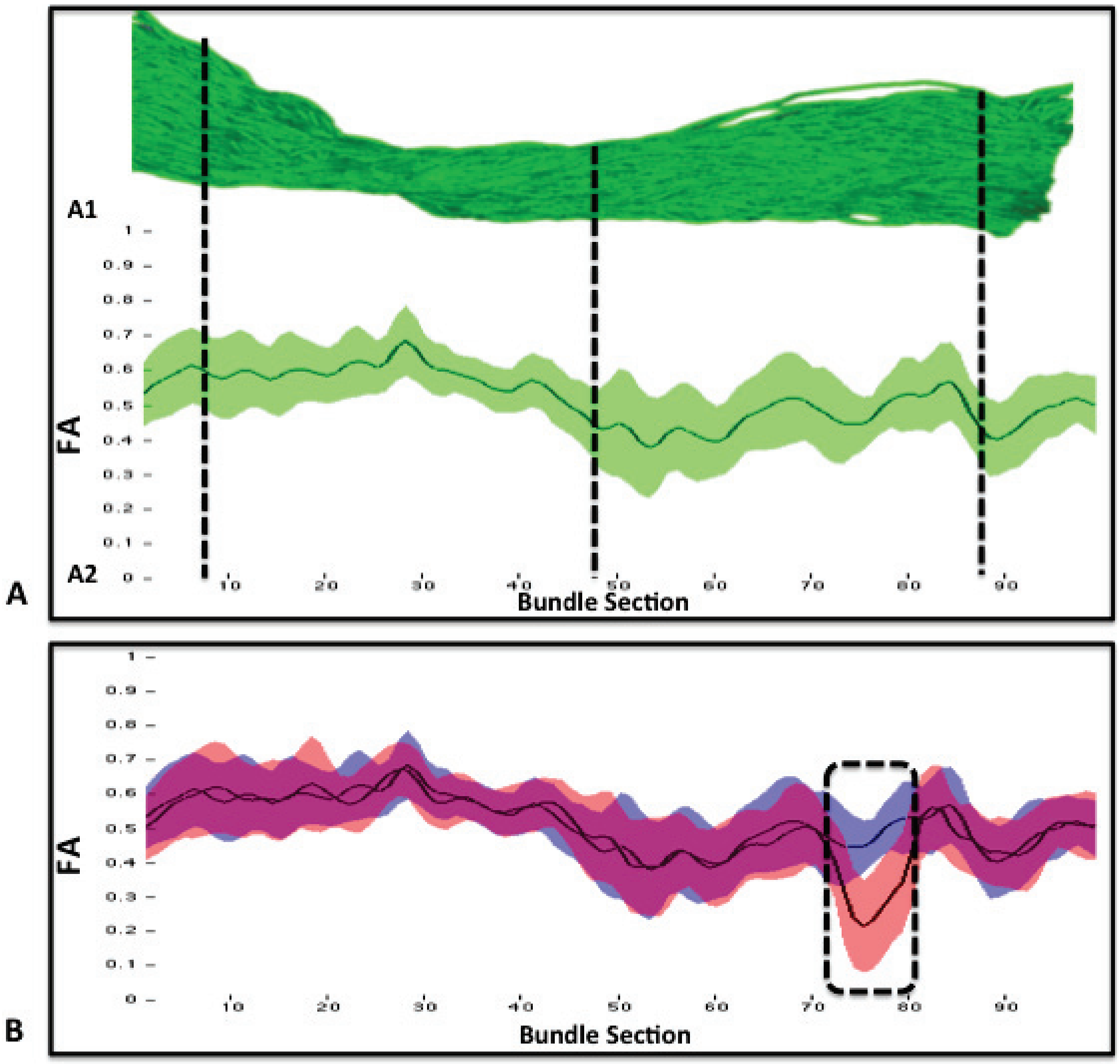


Figure 1.2: **(A1)** “Mean” cross-sectional analysis of the inferior fronto-occipital fasciculi (IFOF). **(A2)** FA values are represented by the mean (black solid line) and standard deviation (green bands) in each cross-section of the fiber-bundle. B) Longitudinal analysis of FA values between the first (blue) and fourth time-point (red) showing a significant FA decrease (no intersection in standard deviation) in several cross-sections (dashed box) of the IFOF.

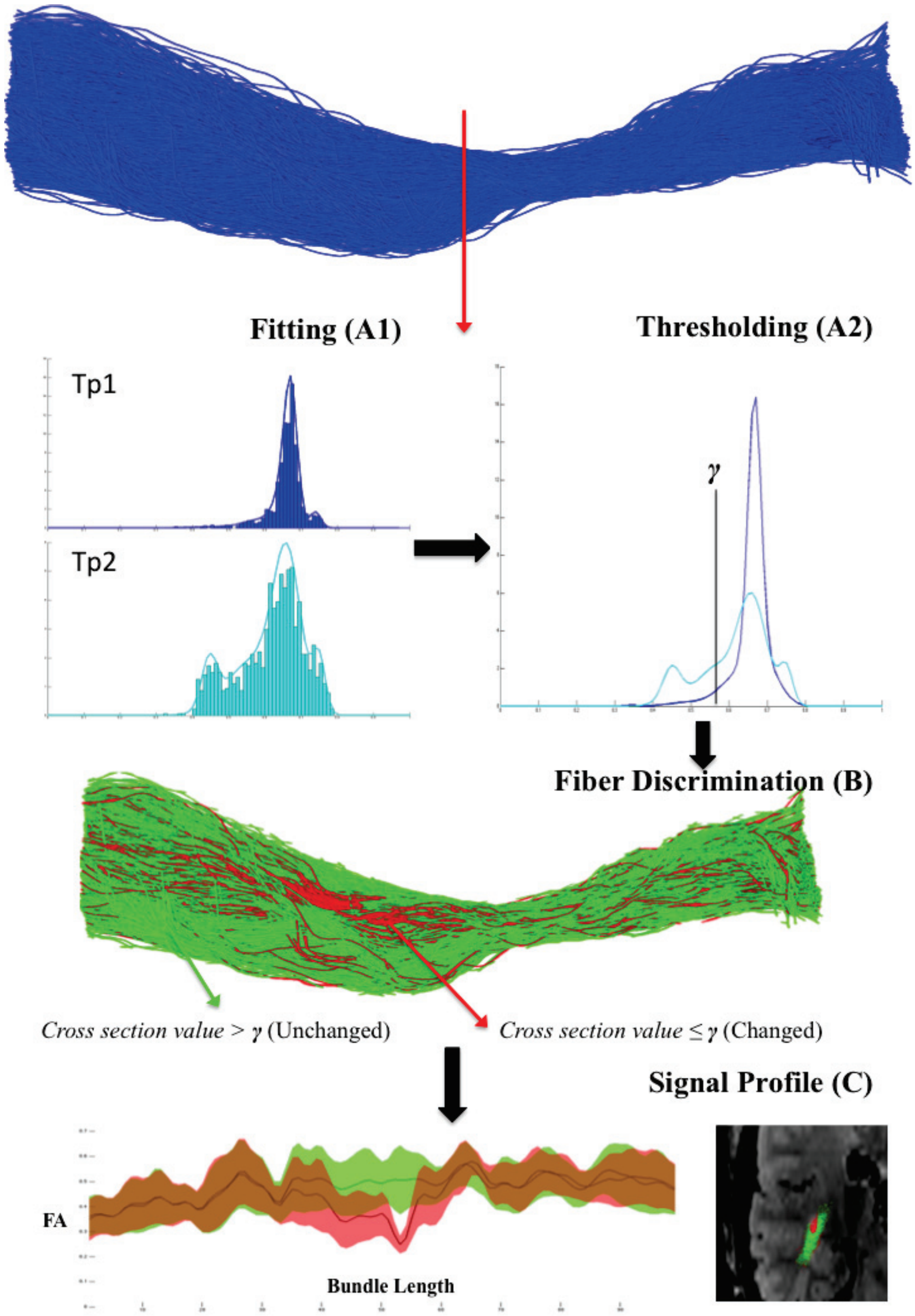


Figure 1.3: Global overview of the "histogram" approach. As first step **(A1)** the histogram of the data extracted from time point i and time-point $i+p$ in the same cross-section are fitted using Gaussian mixture model. As second step **(A2)** our method detects a pathological longitudinal variation between the two time-points in the histogram. The obtained threshold value γ is then used to differentiate between "changed" and "unchanged" fibers **(B)**. Plotted FA signal profile of the two subsets of fiber and cross-sectional view of the labeled fibers **(C)**.

($\zeta = 0.14$) results in a large number of FP. The best performance, in terms of average F-Measure (65.58%) for all the α, was reached for $\zeta = 0.12$. High performances were reached for $\alpha \geq 0.2$ showing the ability of our method to detect small pathological longitudinal variations. While low performance were obtained with $\alpha = 0.1$, probably due to FA variability. The "mean" method was also applied on these simulated variations. In order to check the presence of longitudinal changes, manual inspection of the obtained plot was performed to find cross-sections where no overlap between the confidence intervals (defined by one standard deviation) are visible in the signal profile generated from the two time-points. No differences were found between the two time-points with any value of α.

4.2 Application in MS Follow-up

DTI data of each patient were processed using our proposed pipeline including "mean" and "histogram" fiber-bundle analysis. Among the 20 fiber-bundles extracted in each patient, the Cortico-Spinal Tract (CST) and the Inferior Fronto-Occipital Fasciculi (IFOF) were selected to detect longitudinal changes. Based on our simulated results, we selected $\zeta = 0.12$. Significant changes were validated by our neurologist (FDD) and neuroradiologist (FC) experts. An example of application of our "histogram" method on the CST of a MS patient is illustrated in Figure 1.4.

4.3 Analysis of whole Fiber-Bundles

Three typical cases of longitudinal lesion evolution were selected for illustration. First, the case of two lesions preexisting at W_1 in the left CST of Patient1 is shown in Figure 1.5A. The first lesion was well detected by the "histogram" method due to its size increase, while the second lesion was not detected due to its lack of change during the follow-up period. Second, a new lesion was detected at W_6 in the right IFOF of Patient2 (Figure 1.5B). Third, the case of two lesions preexisting at W_1 and a new lesion appearing at W_7 in the same cross-section of the right IFOF of Patient2 is presented in Figure 1.5C. In these three cases, the "mean" method failed to detect any changes in contrast to our "histogram" method that allowed to differentiate "changed" fibers, characterized by significant longitudinal changes, from "unchanged" fibers.

4.4 "Changed" Fiber-Subsets Analysis

Following this previously described identification of the "changed" fiber-subsets; we iteratively applied our method to further characterize their spatial and temporal evolutions, as illustrated in the right CST of Patient1 through the W_6-W_8 period (Figure 1.6A). The analysis of the fiber-subset's signal revealed a new lesion occurring at W_6 and evolving through W_7 and W_8. The "histogram" method was able to identify at W_7 the progression and expansion of a preexisting lesion in a neighboring fiber-bundle of the CST, and to detect new changes at W_8. Also, a second preexisting lesion was detected at W_8 in another cross-section of the CST. This lesion, already present at W_1, started to evolve

	α	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
Sensitivity (%)	ζ=0.02	10.83	10.83	11.67	19.17	34.17	48.33	59.17	68.33	79.17
	ζ=0.05	15.00	15.83	25.83	36.67	52.50	63.33	72.50	78.33	85.83
	ζ=0.08	24.17	28.33	39.17	52.50	65.00	75.00	81.67	85.00	92.50
	ζ=0.10	33.33	36.67	45.83	57.50	69.17	79.17	84.17	87.50	93.33
	ζ=0.12	45.83	48.33	55.83	65,83	77.50	82.50	85.83	89.17	95.00
	ζ=0.14	55.00	56.67	63.33	71.67	81.67	49.28	88.33	92.50	96.67
Precision (%)	ζ=0.02	50.00	50.00	51.85	63.89	75.93	81.69	84.52	86.32	87.96
	ζ=0.05	50.00	51.35	63.27	70.97	77.78	80.85	82.86	83.93	85.12
	ζ=0.08	51.79	55.74	63.51	70.00	74.29	76.92	78.40	79.07	80.43
	ζ=0.10	51.28	53.66	59.14	64.49	68.60	71.43	72.66	73.43	74.67
	ζ=0.12	50.93	52.25	55.83	59.85	63.70	65.13	66.03	66.88	68.26
	ζ=0.14	50.77	51.52	54.29	57.33	60.49	61.68	62.35	63.43	64.44
F-Measure (%)	ζ=0.02	17.81	17.81	19.05	29.49	47.13	60.73	69.61	76.28	83.33
	ζ=0.05	23.08	24.20	36.69	48.35	62.69	71.03	77.33	81.03	85.48
	ζ=0.08	32.95	37.57	48.45	60.00	69.33	75.95	80.00	81.93	86.05
	ζ=0.10	40.40	43.56	51.64	60.79	68.88	75.10	77.99	79.85	82.96
	ζ=0.12	48.25	50.22	55.83	62.70	69.92	72.79	74.64	76.43	79.44
	ζ=0.14	52.80	52.80	52.80	52.80	52.80	52.80	52.80	52.80	52.80

Table 1.1: Evaluation of detection performances, as measured by the sensitivity, precision and F-Measure, on simulated longitudinal lesions in function of the reduction coefficient α using different ζ values.

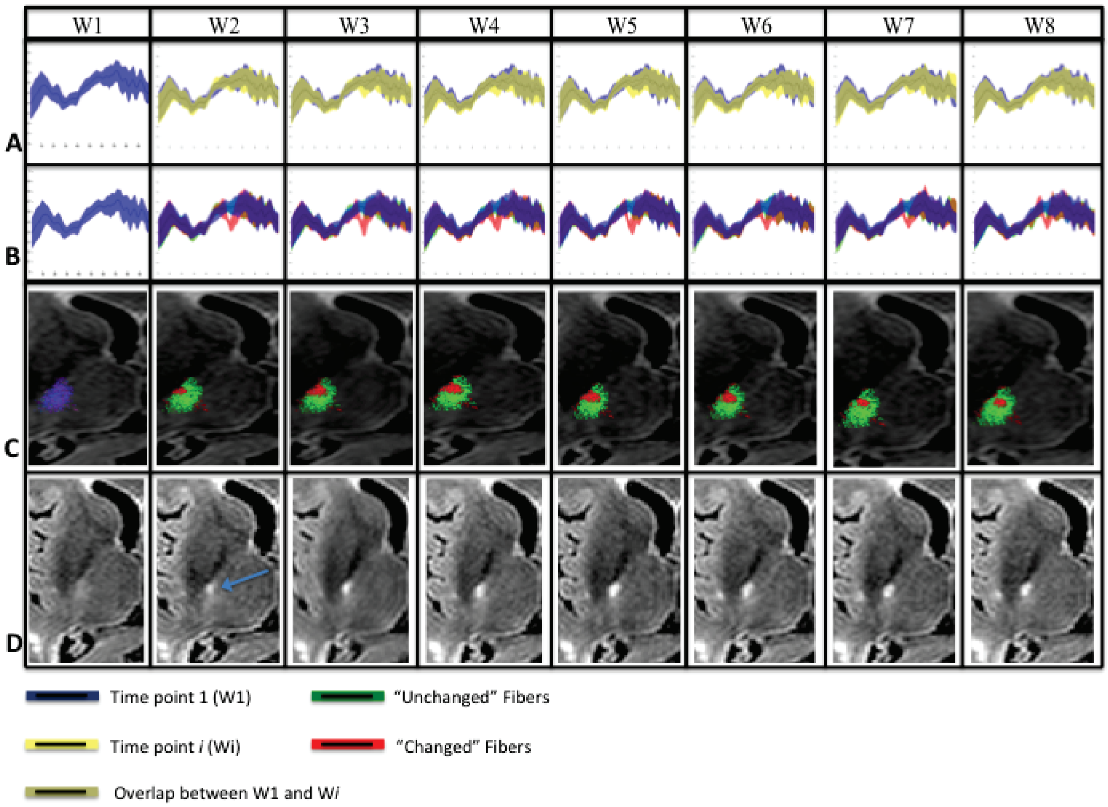


Figure 1.4: Longitudinal analyses of the FA values along the right CST of Patient1. **(A)** The "mean" method analysis showed no changes in the fiber-bundle between time-point 1 (W_1, blue) and the other 7 time-points (yellow). **(B)** The "histogram" method analysis showed significant FA changes (red) between the reference time-point W_1 (blue) and the others 7 time-points (W_2 to W_8) in different cross-sections of the fiber-bundle. **(C)** The "histogram" method allowed the distinction of "unchanged" fiber-subset (green) from "changed" fiber-subset (red) compared to the reference (W_1) fiber-bundle (blue) as shown on the cross-sectional view of the CST. **(D)** FLAIR images of Patient1 showing the corresponding lesions.

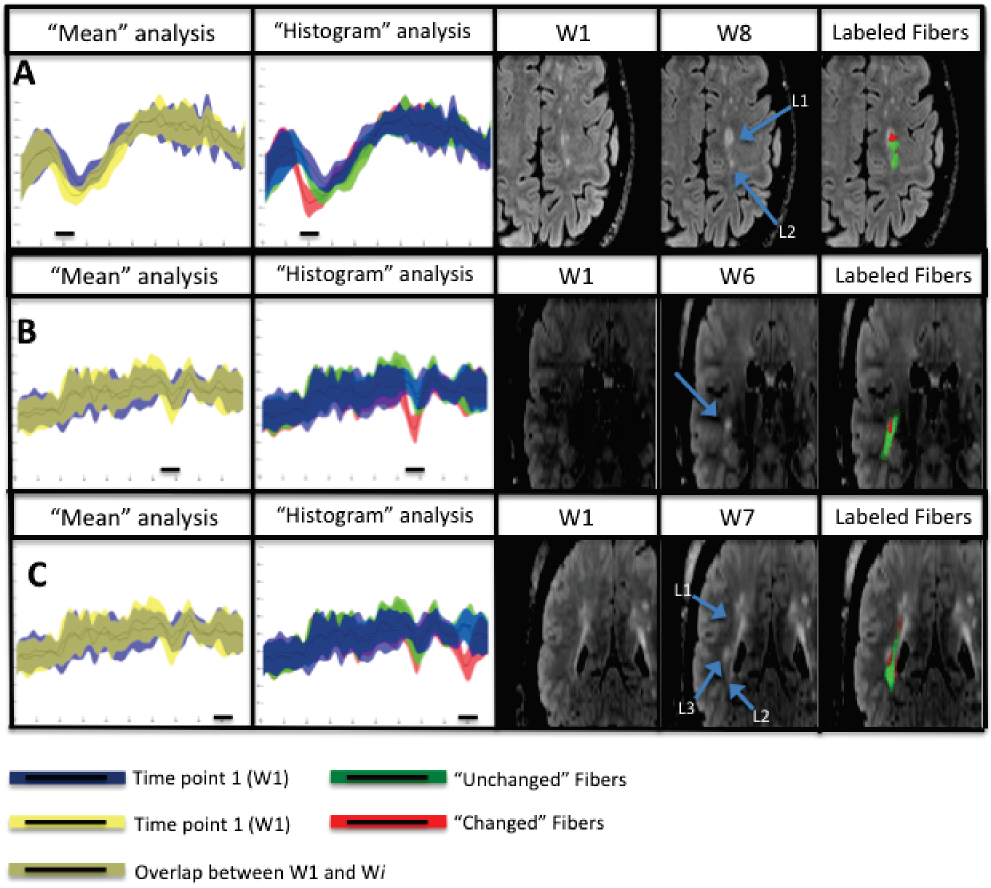


Figure 1.5: Detection of longitudinal variations by applying the "mean" and "histogram" methods: **(A)** On the left CST of Patient1 between W_1 and W_8 time-points, detecting a change in two preexisting lesions (L1, L2); **(B)** On the right IFOF of Patient2 between W_1 and W_6 detecting a new lesion; **(C)** On the right IFOF of Patient2 between W_1 and W_7 detecting a change in two preexisting lesions (L1, L2) and the apparition of a new lesion (L3). Lesions are shown on FLAIR images. Fiber-subsets labeled as "unchanged" (green) and "changed" (red) are shown on top of FLAIR images.

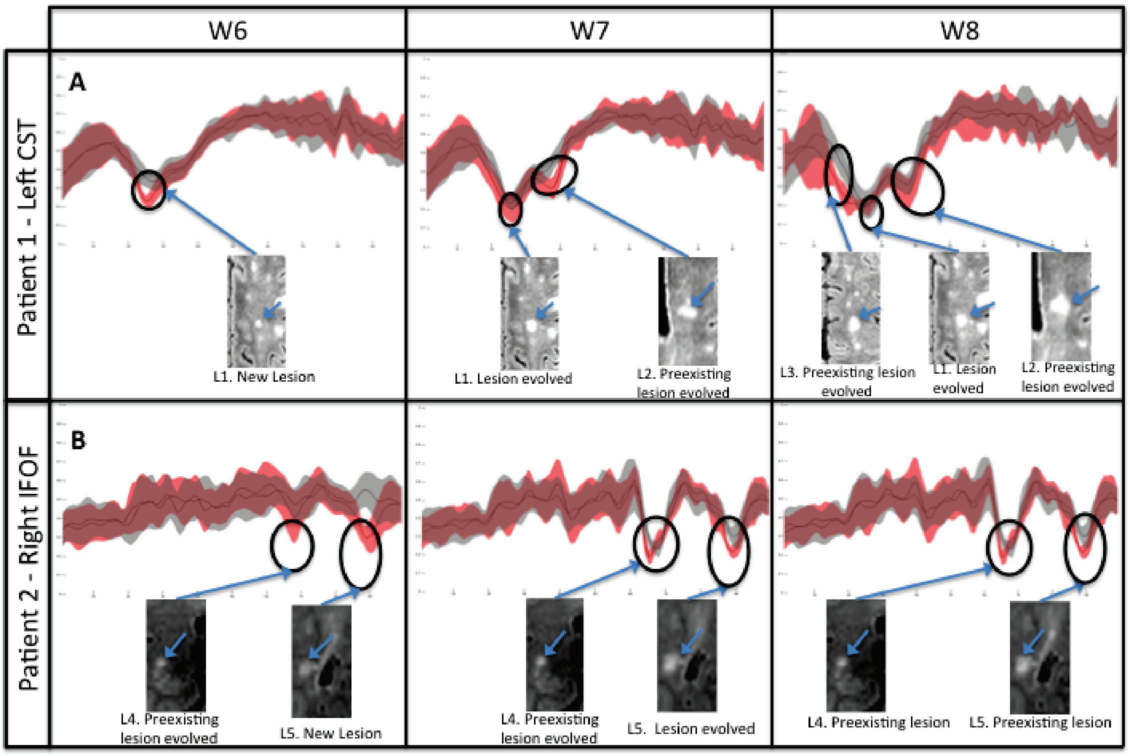


Figure 1.6: Iterative analysis of the “changed” fiber-subset of Patient1’s left CST **(A)** and of Patient2’s right IFOF **(B)** at different time-points; **(A)** Detection of a new lesion (L1) at W_6 and at W_8, and a preexisting lesion at W_7, evolving by contaminating the CST). **(B)** Detection of a preexisting lesion (L4) and a new lesion (L5) at W_6, both evolving in size and degree of FA alteration at W_7, and remaining unchanged at W_8.

at W_8. In Figure 1.6B the analysis of the “changed” fiber-subset in the right IFOF of Patient2 through the W_6-W_8 period showed the presence of a new lesion appearing at W_6 and evolved at W_7 and W_8. The “histogram” method was able to identify at W_6 a preexisting lesion and to follow its evolution through W_7 and W_8 time-points.

5 Discussion

The combination of fiber-tracking and DTI-derived measures, such as FA or other tensor metrics, offers a novel opportunity for the characterization of tissue properties along the WM fiber-tracks. In this work, we presented a new methodology providing automatic processing and detection of “changed” fibers subset of the bundle. The major interest of our “histogram” analysis method relies in its sensitivity to detect small FA changes at multiple locations along the fiber-bundle. In contrast to the “mean” analysis where local scale information contained in each cross-section of the bundle was lost, our method reduced the scale analysis, compatible with a “single” fiber scale. This scale reduction provided two main advantages. First, the detection sensitivity was greatly improved. Second, this fine scale approach allowed us to discriminate "changed" from "unchanged" fibers coexisting in a same bundle.

5.1 Clinical Interest

This new method was applied in MS patients to demonstrate its clinical interest for the characterization of MS pathological processes and its potential to detect fine longitudinal tissue damages based on diffusion metrics. For example the analysis of the fiber properties along and inside the tract showed various profiles of occurrence and evolution of lesions. Indeed, WM fibers of MS patients are subjected to several and complex pathological mechanisms such as inflammation, demyelination and Wallerian degeneration that occur in various WM regions and at different time intervals. Thus, this approach could be applied to study the propagation of tissue damage along the "changed" fiber-subsets and investigate the relationship between WM lesions and their neighboring WM tracts, as well as distant cortical and subcortical GM structures. Indeed, WM lesions might have a role in deep GM atrophy as previously shown in MS patients [Henry *et al.* (2009), Hannoun *et al.* (2012)b]. Finally, the detection of "changed" fiber-subsets based on their diffusion properties may constitute a promising tool to measure the gradient of alterations inside and along the fiber, and potentially to better understand the disease progression [Stoll *et al.* (2002)].

5.2 Methodological Limitations

Our method provided a complete, operator independent and automated processing pipeline applicable in large cohort studies. The reliability of the proposed approach stands on the accuracy and robustness of the pre- and post-processing procedures of the fiber tracking. First, each patient's time-points images were co-registered to a common diffusion atlas in order to use the atlas tractography as a reference model for the longitudinal acquisition of the subject. A major limitation stemming from the longitudinal registration pipeline used in this work could be the introduction of new biases. Specifically, regions affected by pathology might potentially be improperly registered due to the use of nonlinear registration. Second, our approach provided an automatic fiber-bundle extraction by using an atlas-based extraction method. Third, the resulting bundles were post-processed to remove improper fibers generated by the tractography algorithm. This step was performed in order to improve the reliability of the local cross-sectional fiber analysis. Despite the great interest of our method to provide a better differentiation of "changed" fiber-subsets, we should keep in mind that absolute measurement of fibers number is out of reach due to intrinsic limitations of tractography algorithms [Jones (2010)]. However, if the tracking algorithm is compatible with crossing fibers, FA and other tensor-derived measures are not. This means that, in regions containing crossing fibers, if a difference appears in the metric, it will be attributed to all fiber-bundles traversing the region. The capability of our method to detect small longitudinal alterations in FA maps was tested using simulated lesions generated on a control subject's data. This test showed that our "histogram" approach is more sensitive to detect very small spherical alterations than the "mean" approach that failed to detect any of them. These results suggest that our method enables the detection of small alterations that remained undetectable by the classical "mean" approach. The sensitivity of our proposed method appeared adequate in studying MS lesions, but will require more formal sensitivity assessment in future work, to quantify its limits and potentials in this regard. It should be noted that the proposed approach constitutes a detection mechanism; it does not perform genuine statistical analysis of the data. Indeed, the detection of change presented in this work could also been accomplished in simpler fashion: voxel-wise analysis of change could easily be combined with tractographic identification of fibers and bundles. We expect that the strength of our approach will come to bear in situations where properties (e.g. FA) need to be grouped for voxels along a specific fiber bundle or across its cross-section. One can for instance envision increasing the sensitivity of change detection within a given fiber tract by integrating change specifically along a given tract. This is straightforward to accomplish by simple adaptation of the proposed formalism. Since a number of recent reports in the literature have demonstrated that normal appearing white matter (i.e. tissue outside of overt lesions) plays an important role in MS [Steenwijk *et al.* (2014), Bergsland *et al.* (2015)], in the future we also plan to apply our method to the assessment of more subtle changes to the properties of specific tracts.

6 Conclusion

We have described a new fully automated tool for analyzing longitudinal changes in WM fiber-bundles of MS patients. Compared to previous methods developed for the characterization of fiber-tract profiles, our approach provides a fine detection of local scale longitudinal variations along the fiber-bundle, such as those expected to occur during inflammatory or neurodegenerative processes in MS patients. This new approach allows the discrimination of affected fiber-subsets within a bundle, and holds the potential for more detailed and topographically specific description of disease-induced disruption of connectivity in the brain, with implications for specific functional losses associated with disease progression.

Chapter 2

Multi-Features Approach Based on Nonnegative Matrix Factorization Algorithm

Contents

Published papers in this chapter

- C. Stamile, G. Kocevar, F. Cotton, F. Maes, D. Sappey-Marinier, S. Van Huffel **"Multiparametric Nonnegative Matrix Factorization for Longitudinal Variations Detection in White-Matter Fiber Bundles,"** *IEEE Journal of Biomedical and Health Informatics, 2016, in press.*
- C. Stamile, F. Cotton, D. Sappey-Marinier, S. Van Huffel, **"Longitudinal Neuroimaging Analysis Using Non-Negative Matrix Factorization,"** *in: Proceeding of The 12th International Conference on Signal Image Technology & Internet Systems (SITIS),* pp. 55-61, 2017.

1 Introduction

Analysis and processing of longitudinal magnetic resonance imaging (MRI) data is a crucial problem in image analysis. Since pathological mechanisms remained unknown in certain brain diseases, the investigation of their temporal progression using non-invasive neuroimaging techniques is essential to better understand and predict the disease evolution and manage the therapeutic treatment [Vandermosten *et al.* (2015),Mak *et al.* (2015)]. As the etiology of multiple sclerosis (MS) as well as the pathological mechanisms including inflammatory and neurodegenerative processes, are not well understood, longitudinal studies using advanced MRI techniques such as diffusion tensor imaging (DTI) providing sensitive markers of the underlying tissue microstructure, such as fractional anisotropy (FA) and radial diffusivity (λ_r), constitute the best potential for the characterization of brain tissue alterations. For example, the analysis of grey matter (GM) structures [Mesaros *et al.* (2011)] showed the capability to evaluate the dynamics disease progression; in white matter (WM) [Rovira *et al.* (2013)] a relationship between damaging and repairing mechanisms that occur in the lesions formation is revealed. By merging the spatial information of fiber tracking [Mori *et al.* (1999)] with the diffusion metrics derived from the diffusion tensor, it is possible to characterize the presence of "pathological" events that may occur along afferent WM fiber pathways leading to antero- or retrograde degeneration. Thus, for a better understanding of the spatial and temporal progression of MS pathological processes, an accurate and sensitive characterization of WM fibers along their pathways is needed. As shown in our previous work [Stamile *et al.* (2016)a] a global approach to analyze disease evolution was not sensitive enough to detect small and short-term (daily/weekly) longitudinal variations occurring typically in relapsing remitting (RR) MS patients. A local scale approach is thus necessary to detect the presence of small "pathological" changes that could only affect a small subset of the WM fiber-bundle. A first model was proposed in [Grigis *et al.* (2013)]. Due to its simplicity, important assumptions about classical longitudinal biases like noise and registration errors [Reuter *et al.* (2012)] are not taken into account. In order to overcome this limitation in [Stamile *et al.* (2015)b] we proposed a genetic algorithm to detect longitudinal variations on the FA histogram occurring along

WM fiber-bundles in MS patients. In this model assumptions about noise and registration errors are taken into account in order to detect pathological longitudinal variations. Unfortunately the model is limited by its computational complexity. Since it is based on a genetic algorithm the analysis of a specific region could take a large amount of time. Moreover another limitation of the models [Grigis *et al.* (2013)] [Stamile *et al.* (2015)b] is given by their incapability to manage multi-modal/parametric data and multiple time-points.

In order to overcome all these limitations, in this work we present a new fully automated method based on non-negative matrix factorization (NMF). The method allows using multi-parametric DTI derived metrics to detect small longitudinal variations occurring along WM fiber-bundles in MS patients. Since NMF is a blind source separation (BSS) technique we don't need to provide a model to describe the temporal evolution of the pathology in the brain tissues.

This paper is structured as follows. In Section 2, we provide a detailed description of our approach. In Section 3, we present our experimental campaign. In Section 4 we show our results. In Section 5 we discuss the performances, benefits and limitations of our method. Finally, in Section 6, we draw our conclusions.

2 Material and Methods

The proposed method is divided in three main parts: *i)* preprocessing of longitudinal diffusion acquisitions, *ii)* WM fiber-bundle extraction, *iii)* application of NMF and Density-based Local Outliers (LOF) algorithms to detect and delineate longitudinal variations appearing in the cross-section of the WM fiber-bundle.

2.1 Data preprocessing

As first step, each of the t time-points ($T_1 \dots T_t$) of DTI longitudinal acquisitions, are processed. Eddy current correction [Jenkinson and Smith (2001)] was first applied on the diffusion volumes using the b0 volume ($b = 0s.mm^{-2}$) as reference. The corrected volumes were then used in order to compute the tensor model using the FDT module of FSL [Jenkinson and Smith (2001)]. Longitudinal data co-registration is performed using the method described in [Keihaninejad *et al.* (2013)] based on DTI ToolKit (DTI-TK) including the following procedures: *i)* generation of a patient-specific template obtained from longitudinal diffusion tensor images, *ii)* co-registration of the resulting template to the Illinois Institute of Technology (IIT) atlas [Varentsova *et al.* (2014)], and *iii)* co-registration of each time-point data into the IIT atlas space by applying the previously obtained transformations to the initial longitudinal data. The resulting tensor image is then used to compute 6 diffusion metric maps: fractional anisotropy (FA), mean diffusivity (MD), radial diffusivity (λ_r) and the three eigenvalues of the diffusion tensor (λ_1, λ_2, λ_3). Moreover, two diffusion anisotropy measurements were computed: Compositional Kullback-Leibler Anisotropy (KLA) and Angular Anisotropy (AA) [Prados *et al.* (2010)].

2.2 Fiber-bundle extraction

Probabilistic streamline tractography was performed using MRTrix [Tournier *et al.* (2012)] based on the fiber orientation density (FOD) information of the IIT Atlas. Twenty fiber-bundles were extracted using a semi-automatic algorithm [Stamile *et al.* (2015)a] coupled with the prior knowledge extracted from the 20 regions of interest (ROI) of the JHU fiber-bundle atlas [Hua *et al.* (2008)]. In order to analyze fiber-bundles an additional step is needed. Indeed the output of the tractography could not be directly used for the analysis of the fiber-bundle since the number of points used to reconstruct the fibers varies. Moreover, start and end point of each fiber are not consistent within the same fiber-bundle. Fibers could start randomly from the two extremities of the bundle. In order to overcome those problems part of the pipeline described in [Stamile *et al.* (2016)a] was applied to process the fiber-bundle. As first step we define common start/end points of each fiber within the bundle. A classical K-Means algorithm [MacQueen *et al.* (1967)] is performed to generate two different clusters, R_1 for the starting points and R_2 for the ending points. Fiber points are then reordered from R_1 to R_2 and fibers that did not link the two clusters (broken fibers) are automatically removed. As final post-processing step each fiber is resampled with the same number $c = 100$ of points (also called nodes).

After the post-processing we can formalize the extracted fiber-bundle as set $F = \{f_1, f_2, \ldots, f_n\}$ composed of n fibers $f_i = \{\boldsymbol{p_1}, \ldots, \boldsymbol{p_c}\}$ where $\boldsymbol{p_q} = (x_q, y_q, z_q) \mid 1 \leq q \leq c$. The coordinate $\boldsymbol{p_q}$ is used to extract the voxel's value of one of the six diffusion maps ($FA(\boldsymbol{p_q})$ in case of FA) in the corresponding location of f_i. By fixing the index q in each fiber $f \in F$ it is possible to analyze the global diffusion values in a particular cross-section of F. More in detail we can collect all the FA values belonging to a given cross-section of F defining the following set: $R_q = \{FA(\boldsymbol{p_q}) \mid \boldsymbol{p_q} \in f \ \forall \ f \in F\}$ where q is the fixed index representing the cross-section to analyze.

2.3 A Non-Negative Matrix Factorization based algorithm for longitudinal change detection

The NMF based model to detect longitudinal changes is composed of two main steps: *i)* recursive application of NMF algorithm to the data, *ii)* detection of "time-point outliers" in the source matrix W.

Recursive application of NMF

NMF is a blind source separation technique [Lee and Seung (1999)], in which a data matrix V is approximately factorized into the product of a source matrix W and an abundance matrix H:

$$V \approx W \times H$$

$V \in \mathbb{R}^{n \times r}$ is a set of non-negative data, with r data points along its columns and n features per data point on its rows. The columns of $W \in \mathbb{R}^{n \times k}$ represent the k sources. $H \in \mathbb{R}^{k \times r}$

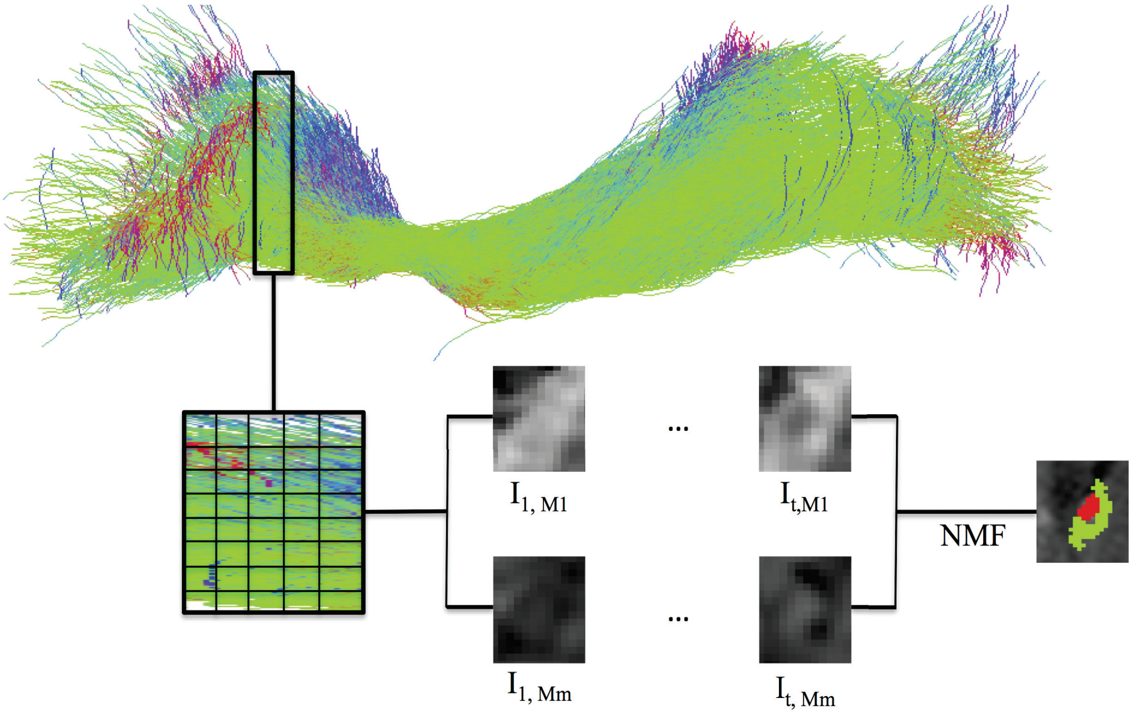


Figure 2.1: Extraction of all M_i signals in a cross-section of the fiber-bundle from time-point 1 to t and application of our NMF method to detect longitudinal variations.

contains per column the abundance of each of the k sources for one particular data point. In this way, NMF describes each point in a dataset by a linear combination of a predefined number of sources.

In particular, NMF based methods have been applied successfully in MRI-based tumor segmentation [Sajda *et al.* (2004), Ortega-Martorell *et al.* (2012), Sauwen *et al.* (2015)]. NMF aims to extract physically meaningful sources, corresponding to tissue-specific patterns. It is an unsupervised technique, i.e. it can be applied on a patient-by-patient basis without the need for any training dataset. NMF assesses the relative contribution of each tissue type within each voxel, assuming the dataset can be modeled as a linear combination of the constituent tissue types. The mathematical formulation of the basic NMF problem to perform the factorization is given below.

$$\begin{aligned} &\underset{\boldsymbol{W},\boldsymbol{H}}{\text{minimize}} && f(\boldsymbol{W},\boldsymbol{H}) = \frac{1}{2}\|V - W \times H\|_F^2 \\ &\text{subject to} && \forall\, i,j : W_{i,j}, H_{i,j} \geq 0 \end{aligned}$$

Let $\boldsymbol{v} \in \mathbb{R}^3$ be a voxel of a given image I. We define as $I(\boldsymbol{v})$ the intensity value of the image I in the voxel $\boldsymbol{v}$. Let m be the number of DTI maps used in the analysis, in our longitudinal framework, we define as I_{i,M_j} the image of DTI derived map M_j $1 \leq j \leq m$ acquired at time-point $1 \leq i \leq t$.

Since we apply the algorithm in each cross-section separately, we define with D ($|D| = s$) the set of voxels $d_j \in \mathbb{R}^3, j = 1, \ldots, s$ contained in the specific cross-section of the

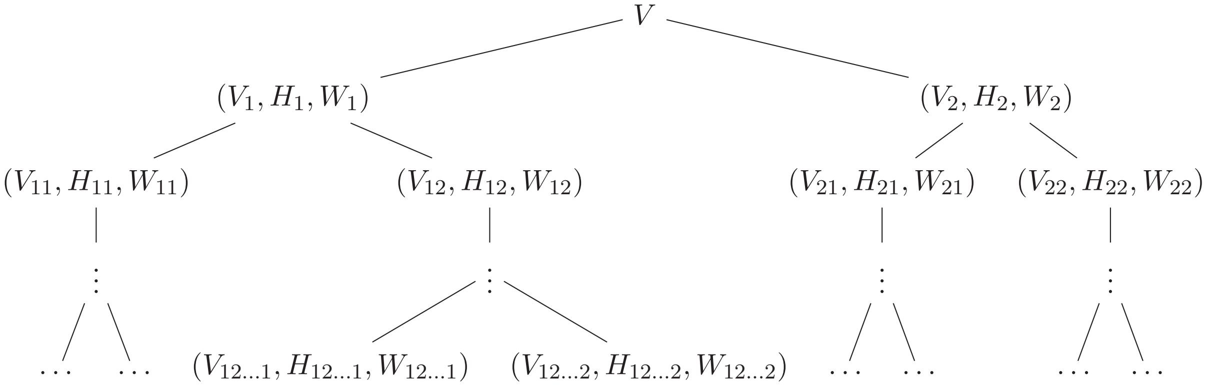


Figure 2.2: Tree generated by the recursive application of NMF.

fiber-bundle to analyze.

Our NMF method is based on a sequential application of NMF as described in algorithm 4.

```
rNMF $(N, cl, l, \omega)$;
Input  : $N$ root node containing the matrix to factorize, $cl$ current level of the tree
         (initialized to zero), $l$ total depth of the tree, $\omega$ threshold parameter
Output: $T$ tree containing the recursive factorization
if $l == cl$ then
    return $T$;
else
    $[H, W] \leftarrow NMF(N_V)$;
    $V_1 = \{v \in V[i,j] \mid \forall\ d_j \in D\ H[1, d_j] \geq \omega\}$;
    $V_2 = \{v \in V[i,j] \mid \forall\ d_j \in D\ H[2, d_j] \geq \omega\}$;
    $N_1 \leftarrow (V_1, H[1, :], W[:, 1])$;
    $N_2 \leftarrow (V_2, H[2, :], W[:, 2])$;
    $T \leftarrow$ addLeftChild$(N, N_1)$;
    $T \leftarrow$ addRightChild$(N, N_2)$;
    return rNMF$(N_1, cl + 1, l, \omega)$;
    return rNMF$(N_2, cl + 1, l, \omega)$;
end
```

Algorithm 4: Algorithm for recursive NMF

In the first level, the data matrix $V \in \mathbb{R}^{t*m \times s}$ containing the longitudinal signal information is factorized in $k = 2$ sources using NMF. The matrix is defined as:

$$V = \begin{array}{c c} & \begin{array}{cccc} d_1 & d_2 & \dots & d_s \end{array} \\ \begin{array}{c} I_{1,M_1} \\ \vdots \\ I_{1,M_m} \\ \vdots \\ I_{tM_1} \\ \vdots \\ I_{t,M_m} \end{array} & \left(\begin{array}{cccc} v^1_{1,1} & v^1_{1,2} & \dots & v^1_{1,s} \\ \vdots & \vdots & \ddots & \vdots \\ \dots & v^1_{m,2} & \dots & v^1_{m,s} \\ \vdots & \vdots & \ddots & \vdots \\ \dots & \dots & \dots & \dots \\ \vdots & \vdots & \ddots & \vdots \\ \dots & \dots & \dots & \dots \end{array} \right) \end{array}$$

V contains the voxels intensity for each of the m DTI metrics in each of the t time-points ($I_{i,M_j}\ 1 \leq i \leq t,\ 1 \leq j \leq m$) for all the voxels $d_j \in D$ extracted in a given cross section q.

The application of the first level of our hierarchical NMF model to V allows to obtain a source matrix W and an abundance matrix H. In this paper, W_1 and W_2 denote respectively the two source vectors obtained from the first and the second column of W. Similarly, H_1 and H_2 denote respectively the two abundance vectors obtained from the first and the second row of H.

Let $h_{i,j} \in H$ be the $j-th$ element of the $i-th$ abundance vector, the elements in the k (in our case $k = 2$) vectors are normalized according to the following equation:

$$h_{i,j} = \frac{h_{i,j}}{\sum_{r=1}^{k} h_{r,j}} \quad \forall\ 1 \leq j \leq s\ ,\ 1 \leq i \leq k \tag{2.1}$$

The normalized vectors are used to split the voxel set D in two sets (D_1, D_2) according to the following rule: $D_1 = \{d_j \in D \mid H[1, d_j] \geq \omega\}$ and $D_2 = \{d_j \in D \mid H[2, d_j] \geq \omega\}$ where $\omega \in \mathbb{R}$ is a given threshold. Similarly to V, we can generate two new data matrices V_1 and V_2 using respectively the voxels contained in D_1 and D_2. The NMF factorization could then be applied recursively to V_1 and V_2 in order to create a second level of matrices generated by the factorization. In general, the algorithm 4 could be applied to generate $l \in \mathbb{N}$ levels. The recursive application of NMF to each level could be formalized using a tree (Figure 2.2). Each level of the tree contains the abundance vector ($H_{i\dots}$), the source vector ($W_{i\dots}$) and the data matrix ($V_{i\dots}$) obtained from the application of NMF of the previous level data.

Detection of "time-point outliers" in the source matrix

In order to check and, eventually, isolate voxels affected by longitudinal changes, abundance and source vectors obtained at the lowest level of the hierarchy (the leaves of the tree) are used. If the analysis of the abundance vectors gives information about the single voxel, the analysis of the source vectors gives information about the contribution of each diffusion feature in each time-point. In the rest of this section we assume that V_*, H_{*1}, H_{*2}

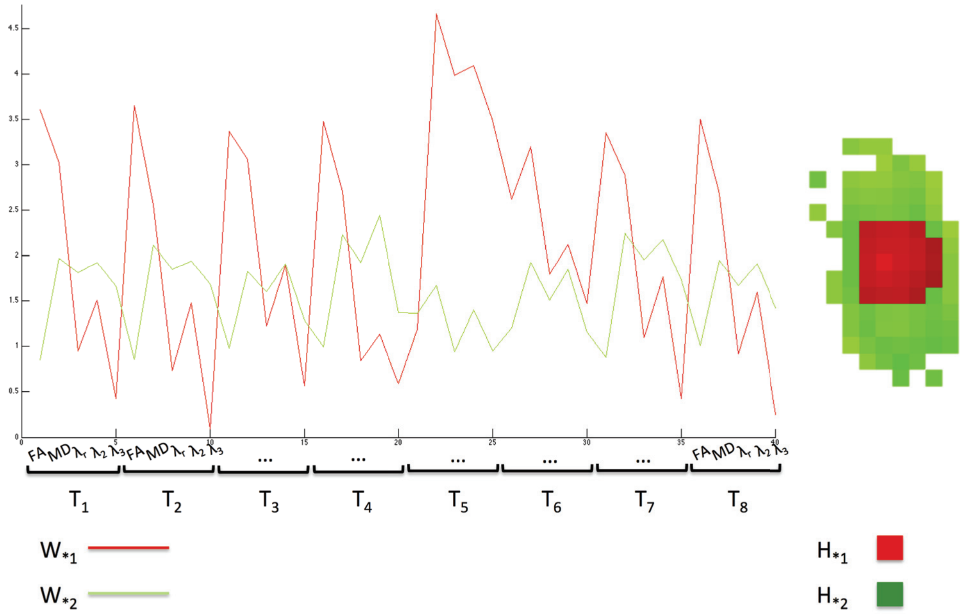


Figure 2.3: On the left, the NMF source vectors (W_{*1}, W_{*2}) of one leaf of the tree. For each of the 8 time-points with $m = 5$, diffusion metrics (FA, MD, λ_r, λ_2, λ_3) are used for a total of 40 features. The outliers' peak visible at time-point 5 (T_5) shows that the longitudinal alteration appears only at T_5. On the right, the voxels segmented using the information contained in the abundance vectors.

and W_{*1}, W_{*2} are respectively the data matrix, abundance and source vectors obtained in a generic leaf of the tree. In order to detect if, during the longitudinal follow-up, longitudinal variations are present we look for the presence of anomalies in the source vectors. More in detail, we say that a longitudinal variation appears during the follow-up if longitudinal variations are present in all the diffusion metrics belonging to certain time-points of W (Figure 2.3). We define these "changed time-points" as *outliers*. In order to detect these points, the Density-based Local Outliers (LOF) algorithm [Breunig *et al.* (2000)] is used. This clustering algorithm allows to detect outliers by computing the LOF value for each element in the cluster. The LOF value of each object represents the degree of the object to be an outlier compared to the other elements in the cluster. This value strongly depends on a single parameter *MinPts* which represents the number of nearest neighbors used in defining the local neighborhood of the object [Breunig *et al.* (2000)]. The main problem related to the LOF is the difficulty to interpret resulting LOF scores since there are no clear rules that define when a point is an outlier. In order to properly detect the outliers this value should be carefully selected for the specific dataset. For a given source vector W_g, the input of the LOF algorithm is a matrix $F \in \mathbb{R}^{t \times m}$ where $f_{i,j} \in F$ contains the contribution of the $j-th$ diffusion parameter at the $i-th$ time-point. The LOF algorithm is then applied independently to W_{*1} and W_{*2}. If outliers are detected in one of the two source vectors (W_{*1}, W_{*2}) the voxels belonging to the corresponding abundance matrix (H_{*1}, H_{*2}) having value $\geq \omega$ are marked as "evolving voxels" (Figure 2.3). These voxels delineate the region in the image affected by the longitudinal variation.

2.4 Subjects

Five relapsing-remitting (RR) MS patients (4 women and 1 man, mean (± SD) age: 36.8 ± 9.5 years; median disease duration: 4.24y; max 16.5 y) (median Expanded Disability Status Scale (EDSS)=2.5, range=[0; 4]) and one healthy control (HC) subject (age: 24 years) were included in this study. Inclusion criteria specified that studied patients were diagnosed as RR MS if they present at least one new Gadolinium-enhancing lesion during the six months preceding study enrollment. All patients had stopped their treatment for at least one year and have not started any during the study period. In order to limit the nephrogenic damage risks associated to Gadolinium injection, creatinine clearance was checked every 2 weeks after inclusion. A clearance higher than 60ml/min was an exclusion criterion. This study was approved by the local ethics committee (CPP Lyon Sud-Est IV) and the French national agency for medicine and health products safety (ANSM). Written informed consent was obtained from all patients and the control subject prior to study initiation.

2.5 MRI Protocol

All subjects underwent a weekly examination for a period of two months (8 time-points from T_1 to T_8). MRI protocol included a DTI and a FLAIR acquisition, that were performed on a 3T Philips Achieva system (Philips Healthcare, Best, The Netherlands) with a 16-channels head-coil. The DTI image set consisted of the acquisition of 60 contiguous 2mm-thick slices parallel to the bi-commissural plane (AC-PC), and were acquired using a 2D Echo-Planar Imaging (EPI) sequence (TE/TR = 60/8210 ms, FOV = 224x224x120 mm) with 32 gradient directions ($b = 1000 s.mm^{-2}$). The nominal voxel size at acquisition (2x2x2 mm) was interpolated to 0.875x0.875x2 mm after reconstruction. The FLAIR Vista 3D sequence (TE/TR/TI = 356/8000/2400 ms, FOV=180x250x250 mm) consisted of the acquisition of 576 slices of 0.43 mm thickness.

2.6 Longitudinal Variations Simulator (LVS)

As in [Grigis *et al.* (2013)], in order to generate real differences between scans such as noise induced by acquisition and changes in subject positioning, our algorithm was tested on simulated longitudinal changes generated along different fiber-bundles in the control subject described in section 5.1.

In this section we describe a new model to simulate pathological longitudinal changes on multi-parametric diffusion data. The idea behind our model is the simulation of the MS longitudinal changes using two properties: shape size and diffusion value changes. We assume that the variation takes a spherical shape of radius r that could change on time. Moreover, we assume that diffusion values of voxels belonging to this region could change during the longitudinal evolution of a factor ρ, called reduction coefficient. This coefficient ρ is used to change the voxel's diffusion values inside the spherical region in a given time-point according to the following equations:

$$\lambda_2^* = \lambda_2 + \rho * (\lambda_1 - \lambda_2) \qquad \lambda_3^* = \lambda_3 + \rho * (\lambda_1 - \lambda_3)$$

with $0 \leq \rho \leq 1$. FA, MD, and other tensor metrics are then recomputed in the given time-point using λ_1 and the new λ_2^*, λ_3^* values.

The proposed LVS model could be summarized using the parametric function S defined as:

$$S : \begin{cases} r(t) = \mathcal{G}(t, \mu_r, \alpha_r, \beta_r) \\ \rho(t) = \mathcal{G}(t, \mu_\rho, \alpha_\rho, \beta_\rho) \end{cases}$$

the function $\mathcal{G}$ is the generalized Gaussian probability density function (GGPDF) defined as follows:

$$\mathcal{G}(x, \mu, \alpha, \beta) = \frac{\beta}{2\alpha\Gamma(\frac{1}{\beta})} e^{-(\frac{|x-\mu|}{\alpha})^\beta}$$

where $x, \mu, \alpha, \beta \in \mathbb{R}$ with $\alpha, \beta \geq 0$ and Γ denotes the gamma function. This particular distribution includes the normal Gaussian distribution $\mathcal{N}$ when $\beta = 2$ (with mean μ and variance $\frac{\alpha^2}{2}$) and it includes also the Laplace distribution when $\beta = 1$.

Since the values of $r(t)$ and $\rho(t)$ given by the function S are such that $0 \leq r \leq 1$ the range of values is rescaled in order to match the interval $[0, \psi]$ for r and $[0, \xi]$ for ρ.

Using the parametric function S we can generate several longitudinal lesions in different cross-sections of a given fiber-bundle. Since the longitudinal behavior of a lesion is not a-priori predictable, the μ_r, α_r, β_r and $\mu_\rho, \alpha_\rho, \beta_\rho$ parameters are randomly selected in order to simulate different kinds of longitudinal variations. Several examples of longitudinal simulated lesions are visible in Figure 2.4.

The function S could also be generalized in order to simulate appearing and disappearing lesions. The model could be easily extended by using a mixture of GGPDFs. The function $\mathcal{G}$ could then be replaced with $\mathcal{M}$ defined as:

$$\mathcal{M}(x, \bar{\mu}, \bar{\alpha}, \bar{\beta}) = \sum_{i=1}^{v} \iota_i \frac{\beta_i}{2\alpha_i\Gamma(\frac{1}{\beta_i})} e^{-(\frac{|x-\mu_i|}{\alpha_i})^{\beta_i}}$$

where v is the number of reappearing peaks of the lesions and $\iota_i \in \mathbb{R} \mid \sum_{i=1}^{v} \iota_i = 1$ is a weighting factor given to each mixture component.

The method could also be extended with other lesion shapes. For instance, an ellipsoid could be used to model the shape of the variation. In this case the function S could be extended to $S_{ellipsoid}$. This new function uses three different radii r_1, r_2, r_3 respectively representing the size of the three axes of the ellipsoid.

In the rest of this paper when we refer to the function S, we assume as GGPDF the function $\mathcal{G}$ and as lesion shape a sphere of radius r.

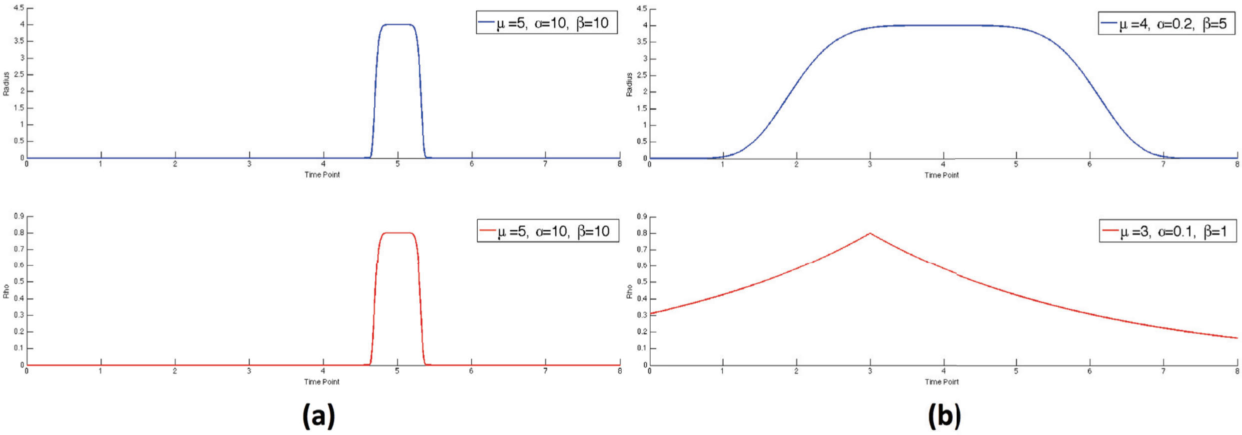


Figure 2.4: Example of function S corresponding to longitudinal simulated variations evolving in shape (blue function) and in reduction coefficient (red function). **a)** Single time-point appearing variation. **b)** Variation with longitudinal stable shape (radius r) and evolving diffusion changes ρ.

3 Experiments

3.1 Simulation of longitudinal variations

100 different longitudinal variations were simulated on the control subject's diffusion maps obtained using the model described in section 2.6. All the variations were generated along 10 different fiber-bundles, namely, left and right, cortico-spinal tract, inferior-fronto occipital fasciculi, cingulum, and forceps major and minor of corpus callosum. In order to simulate small variations, we fixed $\psi = 2$ and $\xi = 0.5$. Since different NMF implementations performed differently according to the domain of application, as reported in [Li *et al.* (2012),Sauwen *et al.* (2015)], in this work three different algorithms to compute NMF are used and compared: alternating least-squares algorithm [Gillis and Glineur (2012)] (ALS), hierarchical alternating least-squares algorithm (HALS) [Cichocki and Phan (2009)] and multiplicative update (MUL) [Lee and Seung (1999)]. For each NMF analysis, the source and the aboundance matrices were initialized randomly, the NMF procedure was repeated 20 times and the output with the best objective function value was withheld.

The capability to detect the presence of longitudinal variations was tested. For each algorithm three different tests were performed: *i)* binary classification for longitudinal detection, *ii)* find the time-points affected by the longitudinal alterations and *iii)* delineate the regions affected by the longitudinal changes. Each test was performed with $l = 1$ and $l = 2$ only. Due to the number of voxels in each cross-section, with $l > 2$ we do not obtain enough voxels to apply NMF.

In order to find the best values of the threshold ω, *MinPts* and the LOF different tests were performed using a range of values. The ω-interval ($0.5 \leq \omega \leq 1$) was divided into 11 steps, as a finer step size did not significantly improve the results any further. *MinPts* values range from 1 to 8 (the total number of time-points) and the LOF interval was given by $\{2, 2.5, 2.8, 3, 3.5, 3.8, 4, 4.5, 4.8\}$.

Binary classification for longitudinal detection

In order to quantify how many simulated longitudinal variations were correctly identified by our method, we tested the capability of our method to detect if in a given cross-section longitudinal changes are present during the follow-up. True positives (TP) and false negatives (FN) were computed to assess the Recall ($\frac{TP}{TP+FN}$).

Longitudinal change detection

We tested the capacity of our method to detect the time-points affected by the longitudinal variations. More in detail, we check for each simulated variation, whether each time-point is correctly classified as "normal" or "outlier". TP, false positives (FP) and FN were then used to compute the *F-Measure* ($\frac{2*TP}{2*TP+FP+FN}$) for the variation. Since multiple longitudinal variations were simulated along the fiber-bundles, mean ($\overline{F - Measure}$) and standard deviation ($\sigma(F - Measure)$) of *F-Measure* were computed.

Longitudinal change delineation

We tested the capability of our method to well delineate the region affected by the longitudinal changes. In other words, we tested the capability of our algorithm to detect the simulated spherical regions in which we simulated the signal changes. In order to quantify the quality of the delineation, the Sørensen-Dice Score Coefficient (DSC) [Zou *et al.* (2004)] was computed according to the following equation:

$$DSC = \frac{2 * |A \cap B|}{|A| + |B|}$$

where A is the voxel set containing the regions with the simulated longitudinal variation and B is the voxel set with the region detected by our method. According to this index, we can have three different cases:

1. $DSC = 0$: No overlap
2. $0 < DSC < 1$: Partial overlap
3. $DSC = 1$: Complete overlap

Since multiple longitudinal variations were simulated along the fiber-bundles, mean ($\overline{DSC}$) and standard deviation ($\sigma(DSC)$) of DSC were computed.

In order to generalize the capability of the proposed method to detect longitudinal changes, two steps are performed. In the first step, a nested search of the parameters of the algorithm is performed, on simulated variations, for each combination of features and for each NMF algorithm. In the second step, the method was applied on a new set of simulated variations using the best parameters discovered in the first step.

3.2 Application on MS patients follow-up

Two RR MS patients (see section 5.1) were selected due to the presence of visible longitudinal alterations assessed by our neuroradiologist (FC). DTI data of each patient were processed using our proposed pipeline described in section 2. Among the 20 fiber-bundles (contained in the JHU atlas), the Cortico-Spinal Tract (CST) was selected by our neuroradiologist (FC) because of the presence of longitudinal MS alterations.

4 Results

4.1 Simulated longitudinal data

Best results for all three experiments, described in the previous section 3, are reported in Table 2.1 for $l = 1$ and in Table 2.2 for $l = 2$. In these tables we show how, for each combination of diffusion parameters and for each NMF algorithm, different performances were obtained. In the tables we denote with “All” the case obtained when all 7 diffusion parameters (FA, MD, λ_2, λ_3, λ_r, KLA, AA) are used as features.

For $l = 1$, best results in terms of Recall are obtained using HALS algorithm with KLA as feature (0.97). Best $\overline{DSC}$ (0.67) is obtained using HALS algorithm and λ_2, λ_3 as features while best $\overline{F - Measure}$ (0.78) is obtained using ALS algorithm and FA as feature. Compared to the other algorithms and features, ALS globally shows a better capability to delineate the longitudinal evolutions with high values of Recall (0.82), $\overline{F - Measure}$ (0.70) and $\overline{DSC}$ (0.60) reached using λ_2, λ_3 as features and with the following parameters: $\omega = 0.55$, $MinPts = 3$ and $LOF = 2.5$. With respect to the other two algorithms MUL shows comparable performances.

For $l = 2$, similar results are obtained for HALS, ALS and MUL having respectively 0.73, 0.80, 0.75 as Recall, 0.70, 0.68, 0.74 as $\overline{F - Measure}$ and 0.57, 0.59, 0.61 as $\overline{DSC}$ with λ_2, λ_3 as features. Moreover, as for $l = 1$, HALS algorithm shows the highest performances in terms of Recall (0.99) with the following three features' sets *i)* KLA *ii)* KLA, FA and *iii)* KLA, AA. Also ALS algorithm reaches a high value of Recall (0.99) but just having as feature KLA, FA.

In Tables 2.3, 2.4 and 2.5 we show how the time-point detection and delineation performances change according to ω, LOF and $MinPts$ parameters.

We tested our method with different parameters by using MUL algorithm and two levels ($l = 2$) with a feature vector composed of FA, MD, λ_r, λ_2 and λ_3. Values with $7 \leq MinPts \leq 8$ are not shown in the tables since for these parameters the values are all zero. As expected, the ω parameter mainly influences the $\overline{DSC}$ performances. Low values of ω ($0.5 \leq \omega \leq 0.6$) show high $\overline{DSC}$ values, while, large values of ω ($\omega \geq 0.65$) yield a degradation of the $\overline{DSC}$ performances. This behavior is justified by the fact that the ω parameter is used as threshold to decide which voxels should be selected from the abundance vectors.

Table 2.1: Best values of Recall, mean ($\overline{F-Measure}$, $\overline{DSC}$) and standard deviation ($\sigma(F-Measure)$, $\sigma(DSC)$) in parenthesis of *F-Measure* and DSC for longitudinal time-points identification using one level ($l = 1$) and different combinations of diffusion parameters

		Recall	$\overline{F-Measure}$	$\overline{DSC}$
HALS	**All**	0.68	0.72 (0.34)	0.50 (0.27)
	KLA	0.97	0.65 (0.25)	0.25 (0.10)
	KLA,FA	0.71	0.72 (0.32)	0.29 (0.10)
	KLA,AA	0.70	0.56 (0.20)	0.24 (0.12)
	FA,MD,λ_r,λ_2,λ_3	0.67	0.56 (0.35)	0.39 (0.25)
	FA	0.88	0.74 (0.32)	0.17 (0.26)
	MD	0.71	0.52 (0.32)	0.63 (0.34)
	FA,MD	0.64	0.52 (0.37)	0.31 (0.25)
	FA,MD,λ_r	0.70	0.54 (0.36)	0.37 (0.24)
	λ_2,λ_3	0.72	0.66 (0.37)	0.67 (0.30)
ALS	**All**	0.71	0.52 (0.20)	0.34 (0.15)
	KLA	0.93	0.71 (0.20)	0.26 (0.11)
	KLA,FA	0.85	0.60 (0.18)	0.30 (0.16)
	KLA,AA	0.80	0.59 (0.20)	0.35 (0.15)
	FA,MD,λ_r,λ_2,λ_3	0.70	0.57 (0.36)	0.38 (0.25)
	FA	0.80	0.78 (0.30)	0.25 (0.32)
	MD	0.72	0.64 (0.32)	0.52 (0.35)
	FA,MD	0.65	0.52 (0.35)	0.35 (0.20)
	FA,MD,λ_r	0.68	0.53 (0.35)	0.38 (0.24)
	λ_2,λ_3	0.82	0.70 (0.36)	0.60 (0.33)
MUL	**All**	0.62	0.54 (0.25)	0.34 (0.15)
	KLA	0.87	0.61 (0.22)	0.32 (0.18)
	KLA,FA	0.77	0.53 (0.20)	0.30 (0.15)
	KLA,AA	0.79	0.57 (0.24)	0.23 (0.12)
	FA,MD,λ_r,λ_2,λ_3	0.69	0.54 (0.35)	0.41 (0.28)
	FA	0.86	0.68 (0.34)	0.23 (0.32)
	MD	0.74	0.62 (0.33)	0.55 (0.31)
	FA,MD	0.63	0.48 (0.38)	0.33 (0.22)
	FA,MD,λ_r	0.67	0.55 (0.35)	0.37 (0.22)
	λ_2,λ_3	0.71	0.74 (0.33)	0.61 (0.33)

Table 2.2: Best values of Recall, mean ($\overline{F - Measure}$, $\overline{DSC}$) and standard deviation ($\sigma(F - Measure)$, $\sigma(DSC)$) in parenthesis of *F-Measure* and DSC for longitudinal time-points identification using two levels ($l = 2$) and different combination of diffusion parameters

		Recall	$\overline{F - Measure}$	$\overline{DSC}$
HALS	**All**	0.72	0.69 (0.25)	0.52 (0.24)
	KLA	0.99	0.48 (0.20)	0.25 (0.12)
	KLA,FA	0.99	0.50 (0.22)	0.25 (0.11)
	KLA,AA	0.99	0.52 (0.21)	0.22 (0.09)
	FA,MD,λ_r,λ_2,λ_3	0.79	0.66 (0.34)	0.54 (0.29)
	FA	0.89	0.60 (0.33)	0.21 (0.26)
	MD	0.82	0.54 (0.31)	0.40 (0.29)
	FA,MD	0.72	0.68 (0.33)	0.23 (0.23)
	FA,MD,λ_r	0.72	0.67 (0.35)	0.46 (0.27)
	λ_2,λ_3	0.73	0.70 (0.26)	0.57 (0.26)
ALS	**All**	0.75	0.62 (0.22)	0.36 (0.10)
	KLA	0.94	0.55 (0.23)	0.30 (0.12)
	KLA,FA	0.99	0.49 (0.20)	0.30 (0.19)
	KLA,AA	0.97	0.52 (0.21)	0.29 (0.15)
	FA,MD,λ_r,λ_2,λ_3	0.75	0.68 (0.32)	0.48 (0.30)
	FA	0.88	0.63 (0.32)	0.18 (0.24)
	MD	0.86	0.56 (0.35)	0.47 (0.31)
	FA,MD	0.74	0.66 (0.33)	0.34 (0.28)
	FA,MD,λ_r	0.72	0.67 (0.34)	0.39 (0.34)
	λ_2,λ_3	0.80	0.68 (0.33)	0.59 (0.29)
MUL	**All**	0.86	0.64 (0.21)	0.40 (0.20)
	KLA	0.98	0.48 (0.22)	0.26 (0.12)
	KLA,FA	0.98	0.51 (0.25)	0.27 (0.15)
	KLA,AA	0.94	0.42 (0.21)	0.26 (0.11)
	FA,MD,λ_r,λ_2,λ_3	0.75	0.66 (0.33)	0.51 (0.27)
	FA	0.97	0.52 (0.25)	0.29 (0.25)
	MD	0.82	0.53 (0.32)	0.43 (0.28)
	FA,MD	0.73	0.60 (0.35)	0.34 (0.27)
	FA,MD,λ_r	0.71	0.69 (0.32)	0.42 (0.30)
	λ_2,λ_3	0.75	0.74 (0.30)	0.61 (0.25)

Table 2.3: Recall for affected time-points identification using MUL algorithm with two levels ($l = 2$) and using FA,MD,λ_r,λ_2,λ_3 ($m = 5$) as features

	Recall																													
	$\omega \geq 0.50$						**$\omega \geq 0.55$**						**$\omega \geq 0.60$**						**$\omega \geq 0.65$**						**$\omega \geq 0.70$**					
MinPts / LOF	**1**	**2**	**3**	**4**	**5**	**6**	**1**	**2**	**3**	**4**	**5**	**6**	**1**	**2**	**3**	**4**	**5**	**6**	**1**	**2**	**3**	**4**	**5**	**6**	**1**	**2**	**3**	**4**	**5**	**6**
2	0.97	0.97	0.94	0.86	0.63	0.42	0.95	0.95	0.94	0.91	0.68	0.38	0.94	0.94	0.88	0.81	0.52	0.29	0.90	0.90	0.85	0.79	0.54	0.19	0.86	0.86	0.81	0.70	0.44	0.23
2.5	0.86	0.86	0.83	0.78	0.51	0.32	0.88	0.88	0.86	0.81	0.51	0.26	0.86	0.86	0.79	0.73	0.39	0.22	0.83	0.83	0.77	0.65	0.36	0.15	0.77	0.77	0.73	0.64	0.35	0.15
2.8	0.82	0.82	0.78	0.75	0.48	0.29	0.85	0.85	0.81	0.76	0.45	0.22	0.82	0.82	0.75	0.69	0.35	0.18	0.80	0.80	0.74	0.63	0.30	0.14	0.73	0.73	0.69	0.62	0.29	0.13
3	0.78	0.78	0.76	0.72	0.45	0.26	0.84	0.84	0.79	0.75	0.41	0.19	0.78	0.78	0.72	0.66	0.35	0.18	0.75	0.75	0.69	0.60	0.25	0.13	0.71	0.71	0.67	0.60	0.27	0.12
3.5	0.75	0.75	0.73	0.68	0.41	0.20	0.78	0.78	0.73	0.68	0.35	0.16	0.70	0.70	0.67	0.61	0.29	0.13	0.63	0.63	0.60	0.55	0.21	0.11	0.65	0.65	0.61	0.53	0.22	0.10
3.8	0.70	0.70	0.68	0.64	0.39	0.20	0.75	0.75	0.71	0.64	0.30	0.16	0.68	0.68	0.62	0.56	0.27	0.09	0.59	0.59	0.57	0.51	0.18	0.11	0.58	0.58	0.57	0.47	0.18	0.09
4	0.70	0.70	0.68	0.64	0.38	0.17	0.71	0.71	0.67	0.60	0.30	0.16	0.66	0.66	0.62	0.56	0.24	0.09	0.55	0.55	0.54	0.46	0.17	0.10	0.54	0.54	0.53	0.43	0.17	0.08
4.5	0.64	0.64	0.60	0.57	0.32	0.15	0.69	0.69	0.65	0.54	0.26	0.14	0.58	0.58	0.56	0.48	0.22	0.08	0.47	0.47	0.46	0.41	0.15	0.10	0.51	0.51	0.48	0.42	0.14	0.08
4.8	0.61	0.61	0.56	0.51	0.27	0.13	0.66	0.66	0.64	0.53	0.24	0.13	0.56	0.56	0.55	0.47	0.17	0.08	0.47	0.47	0.46	0.40	0.14	0.08	0.51	0.51	0.48	0.39	0.12	0.08

Table 2.4: Changes in mean F-Measure ($\overline{F - Measure}$) for affected time-points identification using MUL algorithm with two levels ($l = 2$) and using FA,MD,λ_r,λ_2,λ_3 ($m = 5$) as features

	$\overline{F - Measure}$																													
	$\omega \geq$ **0.50**						$\omega \geq$ **0.55**						$\omega \geq$ **0.60**						$\omega \geq$ **0.65**						$\omega \geq$ **0.70**					
MinPts / LOF	**1**	**2**	**3**	**4**	**5**	**6**	**1**	**2**	**3**	**4**	**5**	**6**	**1**	**2**	**3**	**4**	**5**	**6**	**1**	**2**	**3**	**4**	**5**	**6**	**1**	**2**	**3**	**4**	**5**	**6**
2	0.58	0.58	0.60	0.65	0.59	0.60	0.55	0.55	0.59	0.62	0.58	0.60	0.55	0.55	0.60	0.66	0.62	0.63	0.54	0.54	0.60	0.64	0.59	0.78	0.57	0.57	0.62	0.72	0.64	0.57
2.5	0.64	0.64	0.67	0.73	0.66	0.59	0.58	0.58	0.61	0.67	0.60	0.60	0.63	0.63	0.69	0.72	0.66	0.67	0.58	0.58	0.66	0.75	0.65	0.78	0.61	0.61	0.66	0.72	0.62	0.63
2.8	0.65	0.65	0.70	0.74	0.65	0.62	0.60	0.60	0.64	0.70	0.64	0.63	0.65	0.65	0.71	0.73	0.68	0.71	0.61	0.61	0.69	0.74	0.71	0.76	0.62	0.62	0.69	0.71	0.66	0.69
3	0.68	0.68	0.72	0.74	0.66	0.62	0.61	0.61	0.65	0.70	0.67	0.65	0.66	0.66	0.71	0.76	0.68	0.71	0.61	0.61	0.70	0.76	0.68	0.78	0.63	0.63	0.69	0.73	0.68	0.71
3.5	0.68	0.68	0.72	0.73	0.68	0.67	0.63	0.63	0.67	0.70	0.67	0.69	0.69	0.68	0.71	0.75	0.71	0.71	0.66	0.66	0.73	0.75	0.70	0.74	0.66	0.66	0.72	0.76	0.72	0.70
3.8	0.68	0.68	0.71	0.73	0.66	0.67	0.64	0.64	0.68	0.73	0.70	0.69	0.69	0.69	0.74	0.75	0.73	0.72	0.66	0.66	0.70	0.73	0.74	0.74	0.69	0.69	0.72	0.78	0.72	0.72
4	0.68	0.68	0.71	0.72	0.65	0.68	0.66	0.66	0.70	0.74	0.74	0.73	0.69	0.69	0.72	0.74	0.78	0.72	0.70	0.70	0.73	0.78	0.75	0.72	0.69	0.69	0.72	0.77	0.74	0.75
4.5	0.66	0.66	0.71	0.73	0.70	0.69	0.66	0.66	0.68	0.74	0.71	0.73	0.74	0.74	0.75	0.80	0.79	0.75	0.73	0.73	0.76	0.77	0.77	0.72	0.69	0.69	0.74	0.76	0.79	0.75
4.8	0.66	0.66	0.71	0.73	0.70	0.68	0.66	0.66	0.67	0.75	0.69	0.73	0.73	0.73	0.74	0.81	0.84	0.75	0.73	0.73	0.75	0.76	0.77	0.75	0.68	0.68	0.74	0.76	0.79	0.75

Table 2.5: Changes in mean DSC ($\overline{DSC}$) for affected time-points delineation using MUL algorithm with two levels ($l = 2$) and using FA,MD,λ_r,λ_2,λ_3 ($m = 5$) as features

	$\overline{DSC}$																													
	$\omega \geq$ **0.50**						$\omega \geq$ **0.55**						$\omega \geq$ **0.60**						$\omega \geq$ **0.65**						$\omega \geq$ **0.70**					
LOF \ *MinPts*	**1**	**2**	**3**	**4**	**5**	**6**	**1**	**2**	**3**	**4**	**5**	**6**	**1**	**2**	**3**	**4**	**5**	**6**	**1**	**2**	**3**	**4**	**5**	**6**	**1**	**2**	**3**	**4**	**5**	**6**
2	0.49	0.49	0.51	0.56	0.51	0.54	0.40	0.40	0.43	0.46	0.46	0.49	0.34	0.34	0.37	0.40	0.43	0.45	0.25	0.25	0.27	0.28	0.26	0.38	0.14	0.14	0.14	0.17	0.19	0.21
2.5	0.54	0.54	0.56	0.61	0.56	0.59	0.43	0.43	0.46	0.51	0.48	0.51	0.38	0.39	0.41	0.43	0.44	0.50	0.26	0.26	0.28	0.31	0.30	0.33	0.15	0.15	0.15	0.17	0.19	0.20
2.8	0.55	0.55	0.58	0.61	0.55	0.59	0.45	0.45	0.48	0.54	0.51	0.51	0.39	0.39	0.43	0.45	0.45	0.51	0.27	0.27	0.29	0.31	0.27	0.34	0.14	0.14	0.16	0.17	0.20	0.21
3	0.57	0.57	0.58	0.61	0.55	0.58	0.44	0.44	0.48	0.53	0.53	0.55	0.41	0.41	0.44	0.46	0.45	0.51	0.27	0.27	0.29	0.31	0.28	0.34	0.14	0.14	0.16	0.18	0.22	0.23
3.5	0.57	0.57	0.60	0.62	0.60	0.66	0.46	0.46	0.50	0.53	0.55	0.58	0.41	0.41	0.43	0.45	0.50	0.49	0.30	0.30	0.31	0.31	0.32	0.37	0.16	0.16	0.17	0.18	0.23	0.22
3.8	0.60	0.60	0.62	0.61	0.60	0.66	0.47	0.47	0.51	0.56	0.53	0.58	0.40	0.40	0.44	0.45	0.51	0.53	0.31	0.31	0.32	0.32	0.35	0.37	0.18	0.18	0.18	0.20	0.24	0.24
4	0.60	0.60	0.62	0.61	0.60	0.70	0.49	0.49	0.54	0.57	0.55	0.58	0.42	0.42	0.44	0.45	0.53	0.53	0.32	0.32	0.33	0.34	0.36	0.36	0.18	0.18	0.19	0.20	0.25	0.25
4.5	0.58	0.58	0.61	0.60	0.60	0.72	0.48	0.48	0.52	0.56	0.60	0.60	0.45	0.45	0.46	0.48	0.54	0.56	0.32	0.32	0.33	0.34	0.37	0.36	0.18	0.18	0.20	0.20	0.27	0.25
4.8	0.58	0.57	0.61	0.61	0.62	0.73	0.50	0.50	0.52	0.57	0.60	0.61	0.45	0.45	0.46	0.48	0.60	0.56	0.33	0.33	0.33	0.34	0.39	0.43	0.18	0.18	0.20	0.20	0.32	0.25

In addition, LOF and $MinPts$ parameters are used for the LOF outlier detection algorithm, and they mainly influence the Recall and $\overline{F-Measure}$ performances. With $1 \leq MinPts \leq 4$ and $2 \leq LOF \leq 3.8$ high values of Recall and $\overline{F-Measure}$ are obtained.

4.2 MS patients follow-up

In order to show how the combination of the five diffusion parameters could be helpful in application on real data, a feature vector composed by FA, MD, λ_r, λ_2 and λ_3 ($m = 5$) was used. Based on the results obtained in the simulated validation, we used HALS as algorithm since it globally reached, for these five features, the best results in the three tests. The following parameters $l = 2, LOF = 2.8$, $MinPts = 4$ were selected.

As illustrated in Figure 2.5 and Figure 2.6, both small and large “pathological” longitudinal variations occurring along the two WM fiber-bundles were correctly identified using the proposed method. In both figures, it is possible to see how MS modifies, longitudinally, the diffusion values in a specific WM region. With our method it is possible to identify where these regions are located along specific WM fiber-bundles and which part of these bundles are affected by the longitudinal changes. We can thus extract two types of information: *i):* segmentation of the regions affected by the longitudinal variations and *ii):* the delineation of the time-points affected by the longitudinal variations. Detection of significant changes were validated by our neuroradiologist (FC).

5 Discussion

We described a new fully automated tool to analyze longitudinal changes in the WM fiber-bundles of MS patients. Particularly, we developed a new NMF based method to detect local scale longitudinal variations caused by rapid inflammatory processes in MS patients. The fully automated method described in this paper is divided in three major steps: *i)* preprocessing of longitudinal diffusion acquisitions, *ii)* WM fiber-bundles extraction, *iii)* application of NMF and LOF cluster algorithm to detect and delineate longitudinal variations appearing in the cross-section of the given WM fiber-bundle. Moreover, in order to test the capability of our method to detect and delineate longitudinal changes, a new simulation paradigm is introduced. The new simulation method LVS allows to replicate real longitudinal variations occurring in diffusion data in MS patients. The method uses the longitudinal data of a control subject to simulate longitudinal variations in shape and signal by following a GGPDF distribution.

Results on simulated data show the capability of our method to detect and delineate the presence of longitudinal changes in cross-sections of WM fiber-bundles. The experiment section was also enriched by the analysis of the features set and the input parameters of our method. This process allowed us to see how the performance of our method changes according to the parameters, and which range of parameters gives the best results in terms

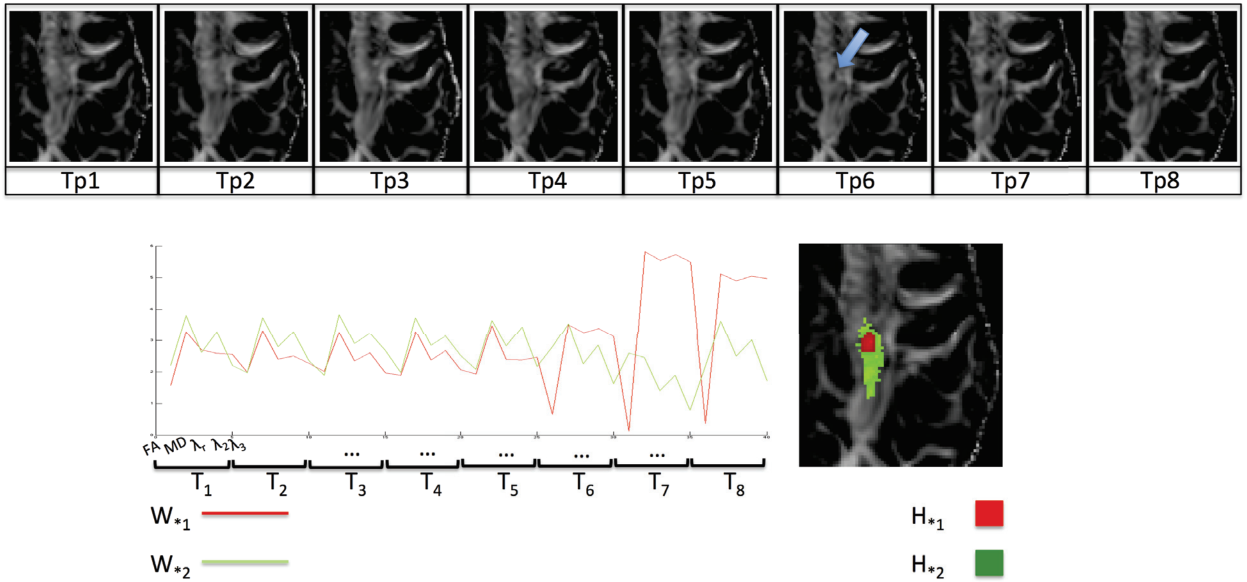


Figure 2.5: Longitudinal variations detected (in red) by the application of our method on left CST of an MS patient. On top, the longitudinal evolution of the FA signal is visible. Starting from time-point 6 a variation, given by the presence of a new MS lesion, is visible. On bottom, the application of our NMF based method shows how in the source matrix irregularities are visible starting from time-point 6 to time-point 8. In parallel the abundance matrix shows the delineation of the region affected by the longitudinal variation (red) and the regions not affected by the longitudinal variation (green).

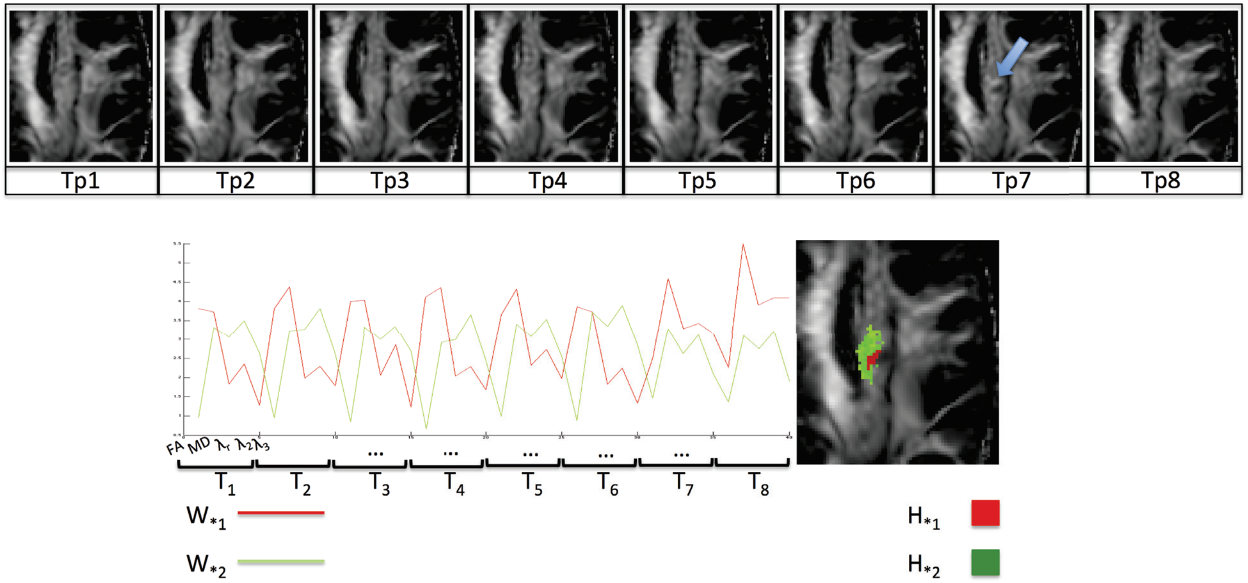


Figure 2.6: Longitudinal variations detected (in red) by the application of our method on left CST of an MS patient. On top, the longitudinal evolution of the FA signal is visible. Starting from time-point 7 a variation, given by the presence of a new MS lesion, is visible. On bottom, the application of our NMF based method shows how in the source matrix irregularities are visible starting from time-point 7 to time-point 8. In parallel the abundance matrix shows the delineation of the region affected by the longitudinal variation (red) and the regions not affected by the longitudinal variation (green).

of detection and delineation. As a result of this validation step, we obtained that good improvements were reached using two levels ($l = 2$) compared to the one level analysis $l = 1$. Moreover this validation process gave us the possibility to explore how certain parameters like LOF and $MinPts$ influence performances like Recall and $\overline{F - Measure}$.

Those parameters can also be adjusted in order to adapt the proposed method to different situations. For instance, for the analysis of thin fiber-bundles, where the number of the voxels in the cross-section is small, the number of levels in the tree can be set to one ($l = 1$).

The features used to perform the analysis play also an important role on the final results. It is possible to see how the introduction of KLA and AA allows to obtain excellent results in terms of $\overline{F - Measure}$ compared to the “standard” DTI derived metrics for $l = 1$ and $l = 2$. Globally the optimal results are reached using λ_2, λ_3 as feature set. The introduction of other diffusion parameters does not drastically improve the performance. Moreover, from the results obtained on the simulated data, we can observe two principal characteristics. First, ALS is the best algorithm to detect if longitudinal changes are present in the follow-up. Second, MUL is the best to detect the time-points affected by the changes and segment the affected voxels.

One limitation present in this work is related to the low value of $\overline{DSC}$ reached by our method. In future work, we plan to overcome this limitation by using structural information derived from classical MRI sequences like T1, T2 FLAIR, etc..

As final experimental step, the proposed method was applied in real MS patients showing its capability to detect and delineate the temporal and spatial changes observed by the neuroradiologist.

Another interesting added value of our method is its capability to easily include new time-points acquired later in time. Indeed, due to the robustness of the registration pipeline used in this work [Keihaninejad *et al.* (2013)] the effects of certain longitudinal biases [Reuter *et al.* (2012)], like atrophy, are minimized.

Compared to the other methods already presented in the literature [Stamile *et al.* (2016)a, Stamile *et al.* (2015)b, Grigis *et al.* (2013)], the proposed method allows to overcome multiple limitations. First, it allows to analyze multiple DTI maps taking into account more than 2 time-points. Second, it not only allows to delineate the regions affected by the longitudinal changes, but also to show which time-points are affected by those changes. Third, the method is unsupervised. Hence there is no need to formulate particular hypotheses about the distribution and/or evolution of the diffusion parameters. Finally it could be easily extended to include other modalities without modifying the algorithm and without formulating new hypotheses about the modalities’ signal distribution and/or evolution.

6 Conclusion

In this work a new pipeline for longitudinal analysis of changes along WM fiber-bundles was presented. We showed how NMF, in combination with LOF, is a powerful tool to discover the presence of longitudinal variations. The method was tested on simulated data, generated with a new simulation paradigm, and on real data.

Results suggested that the proposed method is a promising tool to longitudinal analysis of fiber-bundles in neurodegenerative diseases like MS.

Encouraged by those preliminary results, as for future work, we plan to extend the application of our method to perform multimodal analysis. The idea is to apply the algorithm on different MRI modalities (MRS, T2, T1, etc..) and other diffusion metrics to quantify anisotropy [Prados *et al.* (2010)] like Aitchison Anisotropy, shape anisotropy and others in order to discover the optimal set of features to perform longitudinal analysis.

Chapter 3

Constrained Tensor Decomposition for Global Analysis of Fiber-Bundle Signal

Contents

Published papers in this chapter

- C. Stamile, F. Cotton, F. Maes, D. Sappey-Marinier, S. Van Huffel, **"White Matter Fiber-Bundle Analysis Using Non-Negative Tensor Factorization,"** *in: Image Analysis and Recognition (ICIAR)*, LNCS vol. 9730, pp. 650-657, 2016.
- C. Stamile, F. Cotton, D. Sappey-Marinier, S. Van Huffel, **"Constrained Tensor Decomposition for Longitudinal Analysis of Diffusion Imaging Data,"** *IEEE Journal of Biomedical and Health Informatics*, *submitted*, 2017.

1 Introduction

Magnetic resonance imaging (MRI) is becoming the reference technique to assess brain related pathology. The evolution of fast magnetic acquisition techniques, allowed to use MRI to perform short term longitudinal follow-up of single patients in order to assess the pathological course.

Among all the available MRI biomarkers, diffusion tensor imaging (DTI) or, in general, diffusion imaging (DI), is one of the most interesting techniques to analyze structural changes in the brain. More in detail, in the last decade, different papers showed how image biomarkers derived from DI are sensible for the characterization of brain pathologies like brain tumors and multiple sclerosis (MS) [Gerstner and Sorensen (2011), Kocevar *et al.* (2016), Hannoun *et al.* (2012)b]. Moreover, DI was successfully applied in longitudinal follow-up of brain related pathology showing a good capability to enhance the presence of small longitudinal variations occurring in white matter (WM). These results are encouraging in exploiting DI in follow-up of relapsing remitting (RR) MS patients where small longitudinal changes occur in WM during the evolution of the pathology.

Main challenge in DI analysis is derived from the large number of information that is possible to derive. Indeed DI provides both structural and quantitative information. The former can be derived from the application of a tractography algorithm that allows to reconstruct the architecture of WM fibers [Mori *et al.* (1999)]. The latter is obtained by using a specific model to analyze diffusion like DTI. This information allows to obtain quantitative measures helpful to asses the properties of the tissue. If in this setting longitudinal information is added, it is quite easy to see how the corresponding data analysis problem starts to become quite complex.

Recently, different efforts were made in order to perform such complex analysis. In [Grigis *et al.* (2013)], the authors showed a method to analyze longitudinal changes in WM. Main limitation of this method is related to its capability to perform a longitudinal analysis restricted to two time-points. Moreover, it can not take into account multiparametric data at the same time. In our previous works, we overcome the limitation of this method [Stamile *et al.* (2016)a, Stamile *et al.* (2017), Stamile *et al.* (2016)b]. In particular, in [Stamile *et al.* (2016)b] we developed a method based on non-negative matrix factorization to perform longitudinal analysis in MS patients in cross-sections of the WM fiber-bundles.

Since the major constraint in using a matrix factorization is given by the fixed number of the dimensions, this method was applied only to one specific cross-section of the bundle at time. As result, only local information contained in a specific cross-section was taken into account during the analysis, ignoring the information available along different fibers within the bundle.

In the last years, tensor factorization techniques [Sorber *et al.* (2015)] emerged as a convincing tool to perform data fusion. More in detail, their capability to analyze data, that can be represented using more than two dimensions, made them a perfect tool easy to use in scenarios where more than two dimensions are needed to represent the data.

In this chapter, we propose a new tensor-based framework to analyze longitudinal changes in MS patients occurring along WM fiber-bundles. Our main goal is to detect which subset of fibers within a WM fiber-bundle and which sections of those fibers contain "pathological" longitudinal changes and which are the time-points affected by those changes. We developed a complete fully automated pipeline to register and extract WM fiber-bundles from MRI data and analyze them using tensor factorization. Moreover, in order to optimize the computational time needed to perform the tensor factorization, we developed a parallel pipeline to perform such analysis.

This chapter is structured as follows. In Section 2, we provide a detailed description of the processing pipeline used in this chapter to extract WM fiber-bundles and register longitudinal data. In Section 3, we describe how our problem to analyze longitudinal changes in WM fiber-bundle can be formalized using tensors. In Section 4 we show how to speed-up the proposed framework using parallel programming. In Section 5, we present our experimental campaign. In Section 6 we show our results. In Section 7 we discuss the performances, benefits and limitations of our method. Finally, in Section 8, we draw our conclusions.

2 Data processing pipeline

Before describing or data processing pipeline, we need to state some assumptions made in this chapter. More in detail, we assume that each patient has longitudinal MRI acquisition. In general, we denote with s the number of acquisitions for a single patient. Moreover, for each patient we assume that diffusion data are available at each time-point.

2.1 Data registration

As first step, each of the s time-points $(T_1 \ldots T_s)$ of DTI longitudinal acquisitions, are processed. Eddy current correction [Jenkinson *et al.* (2012)] was first applied on the diffusion volumes using the b0 volume ($b = 0s.mm^{-2}$) as reference. The corrected volumes were then used in order to compute the tensor model using the FDT module of FSL [Jenkinson *et al.* (2012)]. Longitudinal data co-registration is performed using the method described in [Keihaninejad *et al.* (2013)] based on DTI ToolKit (DTI-TK) including the following procedures: *i)* generation of a patient-specific template obtained from longitudinal diffusion tensor images, *ii)* co-registration of the resulting template to the Illinois Institute of Technology (IIT) atlas [Varentsova *et al.* (2014)], and *iii)* co-registration of each time-point data into the IIT atlas space by applying the previously obtained transformations to the initial longitudinal data. The resulting tensor image is then used to compute 6 diffusion metric maps: fractional anisotropy (FA), mean diffusivity (MD), radial diffusivity (λ_r) and the three eigenvalues of the diffusion tensor (λ_1, λ_2, λ_3). Moreover, two diffusion anisotropy measurements were computed: Compositional Kullback-Leibler Anisotropy (KLA) and Angular Anisotropy (AA) [Prados *et al.* (2010)].

2.2 Fiber-bundle extraction

Probabilistic streamline tractography was performed using MRTrix [Tournier *et al.* (2012)] based on the fiber orientation density (FOD) information of the IIT Atlas. Twenty fiber-bundles were extracted using a semi-automatic algorithm [Stamile *et al.* (2015)a] coupled with the prior knowledge extracted from the 20 regions of interest (ROI) of the JHU fiber-bundle atlas [Hua *et al.* (2008)]. In order to analyze fiber-bundles an additional step is needed. Indeed, the output of the tractography could not be directly used for the analysis of the fiber-bundle since the number of points used to reconstruct the fibers varies. Moreover, start and end point of each fiber are not consistent within the same fiber-bundle. Fibers could start randomly from the two extremities of the bundle. In order to overcome those problems part of the pipeline described in [Stamile *et al.* (2016)a] was applied to process the fiber-bundle. As first step we define common start/end points of each fiber within the bundle. A classical K-Means algorithm [MacQueen *et al.* (1967)] is performed to generate two different clusters, R_1 for the starting points and R_2 for the ending points. Fiber points are then reordered from R_1 to R_2 and fibers that did not link the two clusters (broken fibers) are automatically removed. As final post-processing step each fiber is resampled with the same number $q = 100$ of points (also called nodes or cross-sections).

2.3 Fiber-bundle formalization

After the post-processing we can formalize the extracted fiber-bundle as set $F = \{f_1, f_2, \ldots, f_v\}$ composed of v fibers $f_j = \{\boldsymbol{p_1}, \ldots, \boldsymbol{p_q}\}$ where $\boldsymbol{p_i} = (x_i, y_i, z_i) \mid 1 \leq i \leq q$. The coordinate $\boldsymbol{p_i}$ is used to extract the voxel's value of one of the six diffusion maps ($FA(\boldsymbol{p_i})$ in case of FA) in the corresponding location of f_j. With this formalization it is possible to analyze the signal along each fiber within the bundle. For instance, the FA signal profile along a fiber $f_j \in F$ can be regrouped in the set $E_j = \{FA(\boldsymbol{p_1}), \ldots, FA(\boldsymbol{p_q})\}$. This subdivision is graphically described in Figure 3.1. A common way to represent the signal profile along the whole fiber-bundle is to average all the signal profiles along each fiber $f_j \in F$. The global profile is then expressed using mean and standard deviation of the signal along the fiber bundle as presented in Figure 3.2.

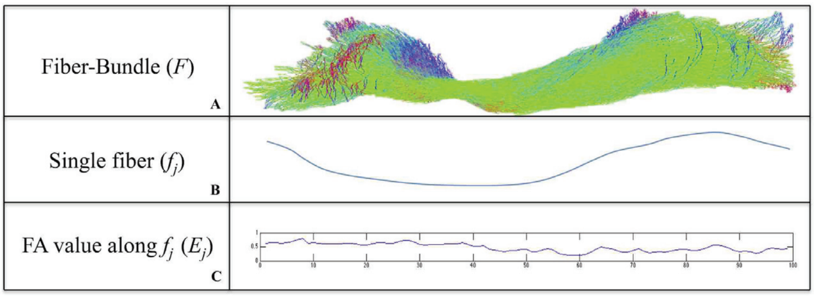


Figure 3.1: **A)** Original fiber-bundle F. **B)** A single fiber $f_j \in F$. **C)** FA signal extract along the fiber f_j.

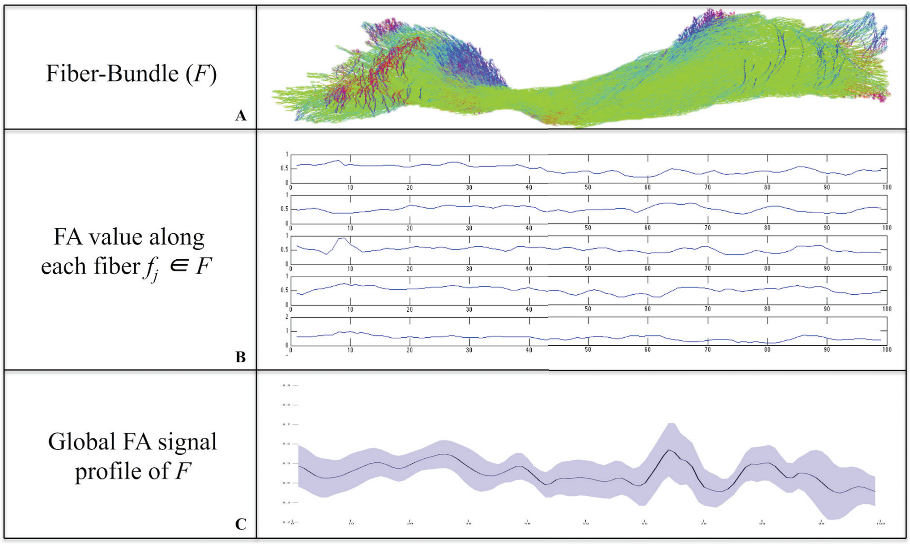


Figure 3.2: **A)** Original fiber-bundle F. **B)** FA signal along each fiber $f_j \in F$. **C)** Mean FA signal (black line) with standard deviation (coloured band) representing the global signal profile along F.

Similarly, taking all the cross-sections $\boldsymbol{p}$ with the index i in each fiber $f \in F$ it is possible to analyze the global diffusion values in a particular cross-section of F. More in detail we can collect all the FA values belonging to a given cross-section of F defining the following set: $S_i = \{FA(\boldsymbol{p_i}) \mid \boldsymbol{p_i} \in f \; \forall \; f \in F\}$ where i is the fixed index representing the cross-section to analyze.

3 Fiber-Bundle as tensor

Before the introduction of our tensor-based formalism we need to define the notation used in this chapter. More in detail, if not clearly stated, we will use the notation described in [Kolda and Bader (2009)]. We denote with lower case letters e.g. a scalar values, with bold lower case letters e.g. $\boldsymbol{a}$ 1-dimensional vectors, with boldface uppercase letter e.g. $\boldsymbol{A}$ matrices and with boldface Euler script letters $\boldsymbol{\mathcal{A}}$ tensors.

In this chapter we describe a new method to analyze longitudinal changes visible along

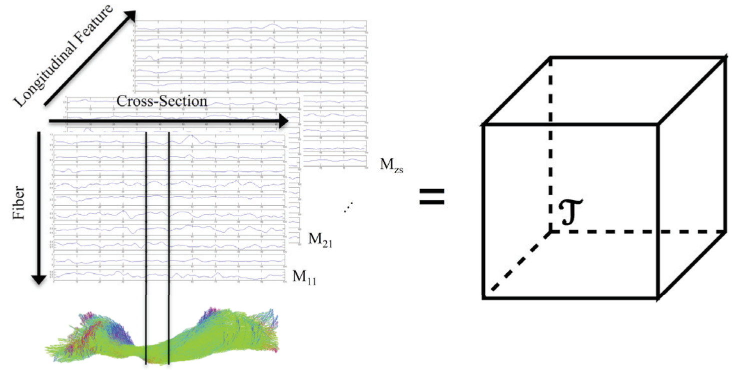


Figure 3.3: Tensorization of longitudinal diffusion features along a fiber-bundle (cortico-spinal tract in this case).

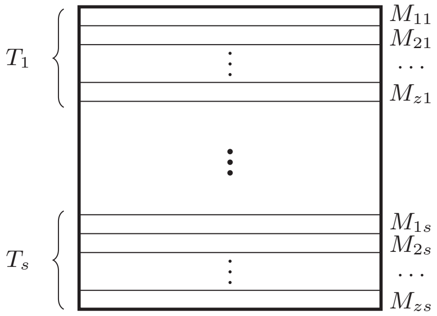


Figure 3.4: Representation of the 3^{rd} mode of the tensor $\boldsymbol{\mathcal{T}}$

WM fiber-bundles using a tensor-based model. In Figure 3.3 we show the model of our tensor to describe the signal along a given fiber-bundle. The third-order tensor $\boldsymbol{\mathcal{T}} \in \mathbb{R}^{v \times q \times m}$ is generated by concatenating all the longitudinal features extrapolated along all the fibers within a bundle.

More in detail, let z be the number of features we want to extrapolate along the fiber-bundle and let s be the number of acquired time-points. Since we have multiple features and multiple time-points, we define with M_{ij} $1 \leq i \leq z$, $1 \leq j \leq s$ the $i-th$ feature extrapolated at the $j-th$ time-point. Using this information the tensor $\boldsymbol{\mathcal{T}} \in \mathbb{R}^{v \times q \times m}$, with $m = s * z$, is built. In Figure 3.4 we show a graphical representation describing how the z features of each time-point are stacked in the 3^{rd} mode of the tensor $\boldsymbol{\mathcal{T}}$.

This tensor-based representation gives us the possibility to analyze, in this one structure, local and global aspects of a fiber-bundle.

3.1 Tensor factorization using canonical polyadic decomposition

The canonical polyadic decomposition (CPD) decomposition factorizes a tensor into a sum of rank-one tensors. For example, given a third-order tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{v \times q \times m}$ we wish to

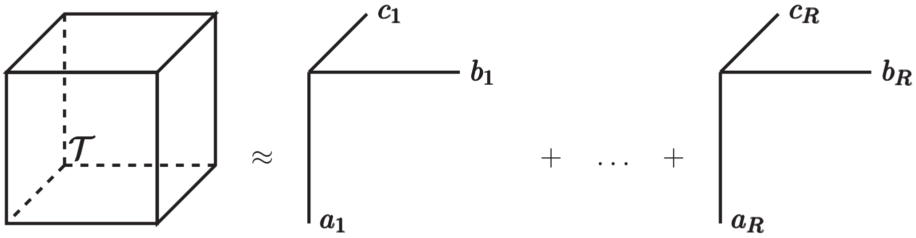


Figure 3.5: Canonical polyadic decomposition

decompose it as:

$$\boldsymbol{\mathcal{X}} = \sum_{r=1}^{R} \boldsymbol{a_r} \circ \boldsymbol{b_r} \circ \boldsymbol{c_r} + \boldsymbol{\mathcal{E}} \tag{3.1}$$

where R is a positive integer, $\boldsymbol{a_r} \in \mathbb{R}^v, \boldsymbol{b_r} \in \mathbb{R}^q, \boldsymbol{c_r} \in \mathbb{R}^m \ \forall\ 1 \leq r \leq R$ are the mode-1, mode-2 and mode-3 vectors respectively and $\boldsymbol{\mathcal{E}} \in \mathbb{R}^{v \times q \times m}$ is the residual tensor. The symbol "$\circ$" represents the outer product.

The rank of a tensor $\boldsymbol{\mathcal{X}}$, denoted $rank(\boldsymbol{\mathcal{X}})$, is defined as the smallest number of rank-one tensors that generate $\boldsymbol{\mathcal{X}}$ as their sum. In other words, this is the smallest number of components in an exact CP decomposition, where "exact" means that there is equality in equation 3.1 with residual tensor $\boldsymbol{\mathcal{E}}$ set to zero. An exact CP decomposition with $R = rank(\boldsymbol{\mathcal{X}})$ components is called the canonical polyadic decomposition (CPD).

The CPD problem can be formalized as follows:

$$\min_{\hat{\boldsymbol{\mathcal{X}}}} \|\boldsymbol{\mathcal{X}} - \hat{\boldsymbol{\mathcal{X}}}\|_F \quad \text{with} \quad \hat{\boldsymbol{\mathcal{X}}} = \sum_{r=1}^{R} \boldsymbol{a_r} \circ \boldsymbol{b_r} \circ \boldsymbol{c_r} = [\![\boldsymbol{A}; \boldsymbol{B}; \boldsymbol{C}]\!] \tag{3.2}$$

In our specific case, since non-negative values are present in the analysis, we imposed a non-negativity constraint to the factorization. More in detail, in order to factorize our tensor, the following optimization problem is solved.

$$\min_{\boldsymbol{A \geq 0, B \geq 0, C \geq 0}} \|\boldsymbol{\mathcal{T}} - \boldsymbol{A} \circ \boldsymbol{B} \circ \boldsymbol{C}\| \tag{3.3}$$

The formulation of the tensor factorization using CPD is graphically described in Figure 3.5.

In this case, the factor matrix $\boldsymbol{A} \in \mathbb{R}^{v \times R}$ contains information describing the contribution of each fiber during the factorization. Similarly, the factor matrix $\boldsymbol{B} \in \mathbb{R}^{q \times R}$ represents the contribution of each fiber-bundle, and finally the factor matrix $\boldsymbol{C} \in \mathbb{R}^{m \times R}$ contains the contribution of each feature extracted in a specific time-point.

In this chapter we used the information contained in the $\boldsymbol{A}$ and $\boldsymbol{B}$ factor matrices in order to discriminate fibers and cross-section affected by longitudinal changes. The factor matrix $\boldsymbol{C}$ is used to identify which are, among the R sources, the ones containing "pathological" longitudinal changes. The idea behind the automatic selection of the sources containing information about the longitudinal “pathological” changes is described in Section 3.3.

3.2 Rank estimation for factorization

An important issue in tensor factorization is the estimation of the rank for the CPD factorization of a tensor $\boldsymbol{\mathcal{T}}$. The literature on estimation of decomposition rank from the tensor is limited. Tensorlab package [Vervliet *et al.* (2016)] has a method *rankest*, which estimates the rank based on the L-curve of the number of rank-one terms in a CPD [Castellanos *et al.* (2002)]. In a real case scenario, this method can overestimate the real rank of the data. Indeed, the estimated rank tends to be larger than the real rank of components present in the image, due to noise or biological variations. In this chapter, since the proposed method is capable to automatically identify which are the sources of interest, (as we will describe in Section 3.3), we use *rankest* function to estimate the rank of the factorization. We prefer to overestimate the real rank by using the *rankest* function, in order to avoid loss of information after the factorization.

3.3 Detection of longitudinal changes from tensor factorization

In order to detect the sources containing the changes generated by longitudinal variations, we generalize the method we previously proposed in Chapter 2. In our previous work, we introduced a model to detect longitudinal changes in sources obtained by factorizing data using a non-negative matrix factorization (NMF). Here we generalize the previous model to the tensor case where each component to analyze is represented by a one dimensional vector.

After the factorization of the tensor $\boldsymbol{\mathcal{T}}$ in R components, we obtained the vectors $\boldsymbol{a_r} \in \mathbb{R}^v, \boldsymbol{b_r} \in \mathbb{R}^q, \boldsymbol{c_r} \in \mathbb{R}^m \ \forall\ 1 \leq r \leq R$. In order to detect if a component i $1 \leq i \leq R$ captures abnormal changes contained in certain features in specific time-points its vector $\boldsymbol{c_i}$ is used. We recall that the vector $\boldsymbol{c_i}$ contains the information of all the m features extracted in the s time-points as follows:

$$\boldsymbol{c_i} = [\underbrace{c^i_1, \cdots, c^i_z}_{T_1}, \cdots, \underbrace{c^i_{s*z-z}, \cdots, c^i_{s*z}}_{T_s}]$$

In this vector, each block of z consecutive elements represents the contribution of each diffusion feature, extracted in a specific time-point i. Starting from the vector $\boldsymbol{c_i}$ the goal is to detect if these $i-th$ components contains longitudinal variations. More in detail, we say that a longitudinal variation appears in the $i-th$ component if longitudinal variations are present in all the diffusion metrics belonging to certain time points of T_i. We define these “changed time points” as *outliers*. In order to clarify this concept, we reported in Figure 3.6 a graphical example of outliers contained in component vector $\boldsymbol{c_i}$. In the rest of this section we will show how we can: *i)* detect the outliers in the component vector $\boldsymbol{c_i}$,

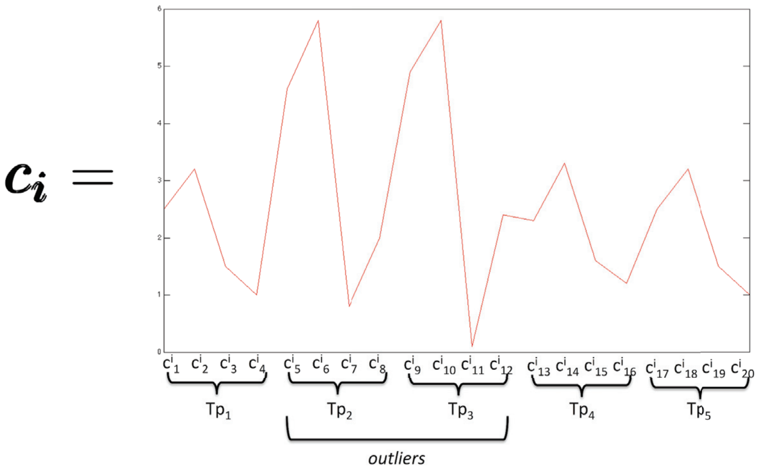


Figure 3.6: Graphical example showing a plot of a component vector $\boldsymbol{c_i}$ with $s = 5$ and $z = 4$. The vector contains outliers values from c^i_5 to c^i_{12} corresponding to time-point 2 and time-point 3. Detection of those outliers time-points allows to understand if the $i-th$ component "captures" longitudinal alterations. Moreover outlier detection in $\boldsymbol{c_i}$ allows to detect time-points containing longitudinal pathological changes.

ii) exploit the information derived from the outlier analysis in order to extract fibers and cross-sections affected by longitudinal pathological changes.

Since the tensor factorization generates R components, this check is performed for all $\boldsymbol{c_i}$ $1 \leq i \leq R$. Detection of outliers was performed using a density-based local outliers (LOF) algorithm [Breunig *et al.* (2000)].
The LOF algorithm allows to detect outliers by computing the LOF value for each element. The LOF value of each object represents the degree of the object to be an outlier compared to the other elements in the cluster. This value strongly depends on a single parameter $MinPts$, which represents the number of nearest neighbours used in defining the local neighborhood of the object [Breunig *et al.* (2000)]. The main problem related to the LOF is the difficulty to interpret resulting LOF scores since there are no clear rules that define when a point is an outlier. In order to properly detect the outliers, this value should be carefully selected for the specific dataset. A value is defined as "outlier" if and only if the LOF value is greater than a fixed threshold value ω.

In order to use the LOF algorithm to detect if $\boldsymbol{c_i}$ contains *outliers time-points*, we *reshape* (using the name of the MATLAB function) the vector in order to create the following matrix:

$$\boldsymbol{\hat{C}_i} = \begin{array}{c} \begin{pmatrix} c^i_1 & \dots & c^i_z \\ \vdots & \ddots & \vdots \\ c^i_{s*z-z} & \dots & c^i_{s*z} \end{pmatrix} \end{array} \begin{array}{c} T_1 \\ \vdots \\ T_s \end{array}$$

the matrix $\boldsymbol{\hat{C}_i} \in \mathbb{R}^{s \times z}$ is a simple reshaped version of the original vector $\boldsymbol{c_i}$ where each of the s rows represents a time-point defined by its diffusion features.

The matrix $\boldsymbol{\hat{C}_i}$ can now be used by the LOF algorithm in order to detect if *outlying* *time-points* are present in the corresponding $i-th$ component.

As a result of the application of the LOF algorithm on the matrix $\hat{\boldsymbol{C_i}}$ a vector $\boldsymbol{l_i} \in \mathbb{R}^s$ is generated. This vector contains real values l^i_j $1 \leq j \leq s$ representing the LOF value corresponding to the $j-th$ row of the matrix $\hat{\boldsymbol{C_i}}$. The LOF value of each row in $\hat{\boldsymbol{C_i}}$ allows us to understand if a specific time-point in the $i-th$ component is an outlier. Or, in other words, this value allows to detect if the $i-th$ component of the tensor factorization captured "longitudinal pathological variations" appearing in the follow-up. We say that the $i-th$ component of the factorization contains pathological longitudinal changes if the following condition holds:

$$\exists\ l^i_j \in \boldsymbol{l_i} \mid l^i_j > \omega,\ 1 \leq j \leq s \tag{3.4}$$

where ω is the LOF threshold that allows to define whether an element is an outlier or not. By applying this procedure to each component it is possible to obtain two types of information: *i)* the components who "capture" the longitudinal pathological changes and *ii)* the time-points where the pathological longitudinal changes appear. This procedure is summarized in Algorithm 5. In detail, the algorithm take the component vectors $\boldsymbol{c_1}, \dots, \boldsymbol{c_s}$ generated from the tensor factorization and the LOF threshold ω. The algorithm applies the LOF algorithm to each reshaped component vector in order to detect which components and which time-points contain longitudinal changes. The algorithm stores all the time-points and the components containing the longitudinal changes in the sets U and Y respectively.

The information derived from the proposed algorithm can be used in order to detect which fibers (represented by the component matrix $\boldsymbol{A}$) and which cross-sections (represented by the component $\boldsymbol{B}$) are affected by longitudinal pathological changes.

More in detail, in order to verify whether a fiber f_i presents longitudinal changes the following condition is checked. Let $a_{y[i]}$ be the $i-th$ element of the component vector $\boldsymbol{a_y}$ obtained from the tensor factorization and let Y be the set containing the components detected as "outliers" by the LOF algorithm. We say that f_i is affected by longitudinal pathological changes if the following condition holds:

$$\exists y \in Y \mid a_{y[i]} > a_{k[i]} \forall 1 \leq k \leq R, k \neq y \tag{3.5}$$

Roughly speaking, this condition checks whether the fiber f_i has its maximal contribution in one of the components detected as "outliers" by the LOF algorithm.

Similarly, in order to check if a specific cross-section j $(1 \leq j \leq q)$ of the fiber-bundle is affected by longitudinal changes, the following condition is also checked. Let $b_{y[j]}$ be the $j-th$ element of the component vector $\boldsymbol{b_y}$ obtained from the tensor factorization and let Y be the set containing the components detected as "outliers" by the LOF algorithm. We say that the cross-section j is affected by longitudinal changes if the following condition

```
CmpTpDet ($c_1, \dots, c_R, \omega$);
Input  : the component vectors $c_1, \dots, c_R$ obtained from the tensor factorization
         a positive real value $\omega > 0$ representing the LOF threshold
Output: a set $Y$ with the components detected as containing longitudinal changes
         a set $U$ containing the time-points affected by the longitudinal changes
begin
    $Y \leftarrow \emptyset$;
    $U \leftarrow \emptyset$;
    foreach $i \in \{1, \dots, R\}$ do
        $\hat{C}_i$ = RESHAPE($c_i$);
        $l_i$ = LOF($\hat{C}_i$);
        foreach $j \in 1, \dots, s$ do
            if $l^i_j > \omega$ then
                $U \leftarrow U \cup j$;
            end
        end
        $Y \leftarrow Y \cup i$;
    end
    return $Y$, $U$
end
```

Algorithm 5: Algorithm for detection of components and time-points affected by longitudinal changes

```
FibDet ($a_1, \dots, a_R, Y, v$);
Input  : the component vectors $a_1, \dots, a_R$ obtained from the tensor factorization
         the set $Y$ with the components detected as containing longitudinal changes by Algorithm 5
         number of fibers $v$
Output: A set $P$ with the fibers containing longitudinal pathological changes
begin
    $H \leftarrow \emptyset$;
    foreach $i \in \{1, \dots, q\}$ do
        if $\exists y \in Y \mid a_{y[i]} > a_{k[i]} \forall 1 \leq k \leq R, k \neq y$ then
            $P \leftarrow P \cup i$
        end
    end
    return $P$
end
```

Algorithm 6: Algorithm for detection of fibers affected by longitudinal changes

holds:

$$\exists y \in Y \mid b_{y[j]} > b_{k[j]} \forall 1 \leq k \leq R, k \neq y \tag{3.6}$$

Like for the fibers, this condition checks whether the cross-section j has its maximal contribution in one of the components detected as “outliers” by the LOF algorithm.

The procedures used for the detection of all the fibers and cross-sections affected by the longitudinal changes are described in Algorithm 6 and Algorithm 7 respectively. Both the algorithms check, for each fiber and cross-section, whether the conditions defined in equation 3.5 and 3.6 are satisfied. The algorithms store all the fibers and the cross-sections detected as containing the pathological longitudinal changes in the sets P and H respectively.

```
CsDet ($\boldsymbol{b_1}, \dots, \boldsymbol{b_R}, Y, q$);
Input  : the component vectors $\boldsymbol{b_1}, \dots, \boldsymbol{b_R}$ obtained from the tensor factorization
         the set $Y$ with the components detected as containing longitudinal changes by Algorithm 5
         number of cross-sections $q$
Output: A set $H$ with the cross-sections containing longitudinal pathological changes
begin
    $H \leftarrow \emptyset$;
    foreach $i \in \{1, \dots, q\}$ do
        if $\exists y \in Y \mid b_{y[i]} > b_{k[i]} \forall 1 \leq k \leq R, k \neq y$ then
            $H \leftarrow H \cup i$
        end
    end
    return $H$
end
```

Algorithm 7: Algorithm for detection cross-section affected by longitudinal changes

4 Parallel implementation of the proposed method

In order to reduce the computation time needed to perform the factorization of the fiber-bundle, we parallelized the algorithm according to the “divide et impera” programming paradigm. Instead of computing the CPD and LOF on the whole fiber-bundle, we split the fiber-bundle in small subsets of fibers. In order to split the whole fiber-bundle in K different sub fiber-bundles, for each pair of fibers $f_a, f_b \in F$ we compute the Minimum Average Direct Flip (MDF) metric [Garyfallidis *et al.* (2012)]. The MDF metric computes the distance between two fibers composed by q points according to the following equations:

$$d_{direct}(f_a, f_b) = d(f_a, f_b) = \frac{1}{q}\sum_{i=1}^{q} \|\boldsymbol{p_i^a} - \boldsymbol{p_i^b}\| \tag{3.7}$$

$$d_{flipped}(f_a, f_b) = d(f_a, f_b^{\#}) = d(f_a^{\#}, f_b) \tag{3.8}$$

$$MDF(f_a, f_b) = min(d_{direct}(f_a, f_b), d_{flipped}(f_a, f_b)) \tag{3.9}$$

where $\|\boldsymbol{p_i^a} - \boldsymbol{p_i^b}\|$ represents the Euclidean distance between the $i-th$ point of f_a and f_b. In equation 3.8, $f_a^{\#}$ denotes the flipped version of f_a, e.g. $f_a = \{\boldsymbol{p_1}, \boldsymbol{p_2}, \dots, \boldsymbol{p_q}\}$, $f_a^{\#} = \{\boldsymbol{p_q}, \boldsymbol{p_{q-1}}, \dots, \boldsymbol{p_1}\}$. The MDF distance is a metric on the space of fibers, it respects the triangle inequality and it is fast to compute [Garyfallidis *et al.* (2012)].

Let v the total number of fibers, we can compute the MDF distance for each pair of fibers building the positive symmetric matrix $\boldsymbol{M} \in \mathbb{R}^{v \times v}$ where $m_{ij} \in \boldsymbol{M}$ represents the MDF distance computed between the fibers f_i and f_j. The matrix $\boldsymbol{M}$ is then used as distance matrix for the K-medoids clustering algorithm in order to generate a set $Cl = \{F_1, F_2, \dots, F_K\}$ of K clusters (Algorithm 8). In our prospective, a cluster $F_i \in Cl$ represents a sub fiber-bundle generated from the original fiber-bundle F such that $\bigcup_{i=1}^{K} F_i = F$.

After their computation, each sub fiber-bundle is assigned to a different processor in order to perform the sub fiber-bundle analysis. In this analysis, each processor, independently, computes the CPD for the assigned sub-bundle. The results of this factorization

```
WmCl ($F, K$);
Input  : a set $F = \{f_1, f_2, \dots, f_v\}$ of fibers
         a positive integer $K > 1$ representing the number of clusters to generate
Output: a set $Cl = \{F_1, F_2, \dots, F_K\}$ of clusters
begin
    $M \leftarrow$ zeros(v, v);
    foreach pair $(f_i, f_j)$ s.t. $f_i \in F, f_j \in F$ do
        $M[i, j] = MDF(f_i, f_j)$;
    end
    $Cl =$ KMedoids($F, M, K$);
    return $Cl$
end
```

Algorithm 8: Algorithm for Cluster Generation

are then used to detect the time points, the fibers and the cross-sections affected by the longitudinal pathological changes using Algorithm 5, Algorithm 6 and Algorithm 7 respectively. The results obtained by each processor are finally merged into different *synchronized* shared variables (Algorithm 9). At the end of this process the algorithm will generate three sets U, P and H representing the time points, the fibers and the cross-sections affected by pathological longitudinal changes respectively.

The parallel execution of the algorithm could be graphically represented as a diagram (Figure 3.7). Each branch represents the computation flow independently executed by each of the K processors. At the end of the parallel execution, each processor adds its local solution to the complete solution set.

```
WmTf ($F, \omega, K, v, q$);
Input  : $F$ set containing all the fibers of a bundle, $\omega$ LOF threshold parameter, $K$ number of parallel
         processors, $v$ total number of fibers, $q$ total number of cross-sections
Output: $U$ set containing the time-points affected by the longitudinal changes
        $P$ set of fibers containing longitudinal changes
        $H$ set of cross-section containing longitudinal changes.
$U \leftarrow \emptyset$;
$P \leftarrow \emptyset$;
$H \leftarrow \emptyset$;
$F_1, \dots, F_K \leftarrow$ WmCl($F, K$);
foreach $F' \in \{F_1, \dots, F_K\}$ do
    StartNewProcess;
    $\boldsymbol{\mathcal{T}} \leftarrow GenTensor(F')$;
    $R \leftarrow rankest(\boldsymbol{\mathcal{T}})$;
    $\boldsymbol{a_1}, \dots, \boldsymbol{a_R}, \boldsymbol{b_1}, \dots, \boldsymbol{b_R}, \boldsymbol{c_1}, \dots, \boldsymbol{c_R} \leftarrow \min_{\substack{\boldsymbol{a_r} \geq \boldsymbol{0}, \boldsymbol{b_r} \geq \boldsymbol{0}, \boldsymbol{c_r} \geq \boldsymbol{0}, \\ 1 \leq r \leq R}} \|\boldsymbol{\mathcal{T}} - \sum_{r=1}^{R} \boldsymbol{a_r} \circ \boldsymbol{b_r} \circ \boldsymbol{c_r}\|$;
    $Y, U' \leftarrow$ CmpTpDet($\boldsymbol{c_1}, \dots, \boldsymbol{c_R}, \omega$);
    synchronized;
    $U \leftarrow U \cup U'$;
    $P \leftarrow P \cup$ FibDet($\boldsymbol{a_1}, \dots, \boldsymbol{a_R}, Y, v$);
    $H \leftarrow H \cup$ CsDet($\boldsymbol{b_1}, \dots, \boldsymbol{b_R}, Y, q$);
    EndOfProcess;
end
```

Algorithm 9: Parallel Algorithm for WM Analysis using Tensor Factorization

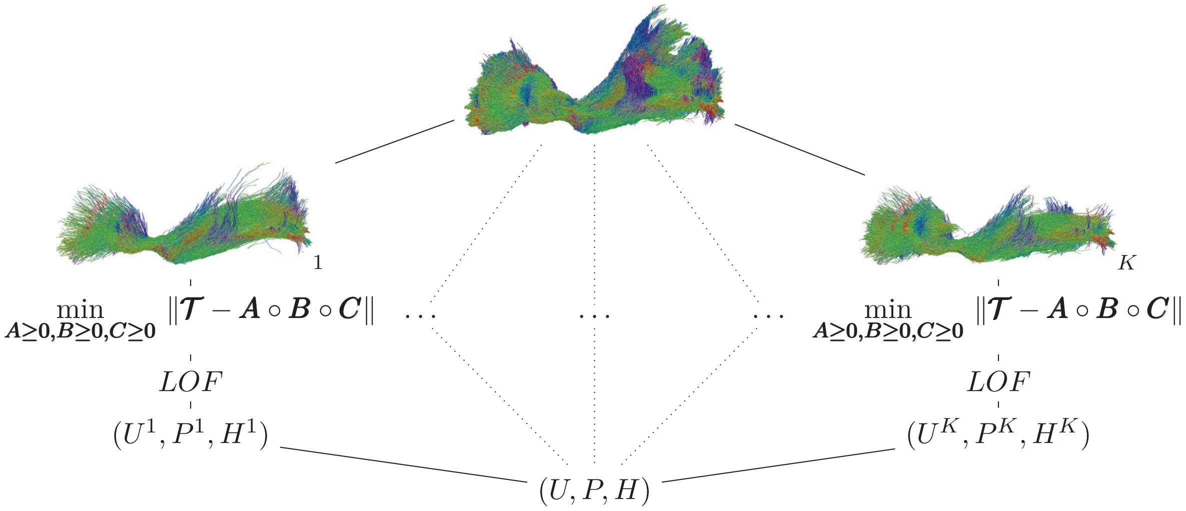


Figure 3.7: Graphical representation of parallel execution of the proposed algorithm

5 Experiments

5.1 Subjects

Four relapsing-remitting (RR) MS patients (3 women and 1 man, mean (± SD) age: 36.8 ± 9.5 years; median disease duration: 4.24y; max 16.5 y) (median Expanded Disability Status Scale (EDSS)=2.5, range=[0; 4]) and one healthy control (HC) subject (age: 24 years) were included in this study. Inclusion criteria specified that studied patients were diagnosed as RR MS if they present at least one new Gadolinium-enhancing lesion during the six months preceding study enrolment. All patients had stopped their treatment for at least one year and have not started any during the study period. In order to limit the nephrogenic damage risks associated to Gadolinium injection, creatinine clearance was checked every 2 weeks after inclusion. A clearance higher than 60ml/min was an exclusion criterion. This study was approved by the local ethics committee (CPP Sud-Est IV) and the French national agency for medicine and health products safety (ANSM). Written informed consent was obtained from all patients and the control subject prior to study initiation.

5.2 MRI protocol

All subjects underwent a weekly examination for a period of two months (8 time-points from T_1 to T_8). MRI protocol included a DTI and a FLAIR acquisition, that were performed on a 3T Philips Achieva system (Philips Healthcare, Best, The Netherlands) with a 16-channels head-coil. The DTI image set consisted of the acquisition of 60 contiguous 2mm-thick slices parallel to the bi-commissural plane (AC-PC), and were acquired using a 2D Echo-Planar Imaging (EPI) sequence (TE/TR = 60/8210 ms, FOV = 224x224x120 mm) with 32 gradient directions ($b = 1000 s.mm^{-2}$). The nominal voxel size at acquisition (2x2x2 mm) was interpolated to 0.875x0.875x2 mm after reconstruction. The FLAIR Vista 3D sequence (TE/TR/TI = 356/8000/2400 ms, FOV=180x250x250 mm) consisted of the acquisition of 576 slices of 0.43 mm thickness.

5.3 Experiments on simulated longitudinal variations

100 different longitudinal variations were simulated on the control subject's diffusion maps. All the variations were generated along 10 different fiber-bundles, namely, left and right cortico-spinal tract, inferior-fronto occipital fasciculi, cingulum, and forceps major and minor of corpus callosum. In order to simulate the longitudinal variations on diffuse data, the method we described in 2.6 of Chapter 2 was used. More in detail, we randomly selected several cross-sections and fibers within a specific bundle. For each voxel belonging to these regions, a generalized Gaussian probability density function (GGPDF) with parameters $\mu_r, \alpha_r, \beta_r, \mu_\rho, \alpha_\rho, \beta_\rho$ (see 2.6 of Chapter 2 for a complete explanation of those parameters) randomly chosen were used to simulate the longitudinal changes. Three different tests were performed: *i)* detection of fibers containing longitudinal changes, *ii)* detection of cross-sections affected by longitudinal alterations and *iii)* identification of abnormal time-points.

Performance measurements used in this chapter are based on the analysis of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) instances classified during the testing phase. Performances of longitudinal variations were evaluated using accuracy (Ac), precision (Pr) and sensitivity (Se) respectively defined as: $Ac = \frac{TP+TN}{P+N}$, $Pr = \frac{TP}{TP+FP}$ and $Se = \frac{TP}{TP+FN}$. Since multiple tests were performed, mean and standard deviation of accuracy, precision and sensitivity ($\overline{Ac} \pm \sigma_{Ac}$, $\overline{Pr} \pm \sigma_{Pr}$, $\overline{Se} \pm \sigma_{Se}$) were computed.

In order to find the best values of the *MinPts* and the LOF different tests were performed using a range of values. *MinPts* values range from 1 to 8 (the total number of time-points) and the ω interval was given by $\{2, 4, 6, 8, 10, 12\}$.

To generalize the capability of the proposed method to detect longitudinal changes, two steps are performed. In the first step, a nested search of the parameters of the algorithm is performed, on simulated variations, for each combination of features. In the second step, the method was applied on a new set of simulated variations using the best parameters discovered in the first step.

Moreover, in order to show the improvement obtained using tensor factorization, we compared the proposed method with our previous NMF-based method (Section 2.6 of Chapter 2). Since the NMF algorithm allows just a local cross-sectional analysis of the fiber-bundle, we compared the capability of the two methods to identify only cross-sections and time-points affected by longitudinal pathological changes.

5.4 Experiments on real MS follow-up data

Four RR MS patients (see section 5.1) were selected due to the presence of visible longitudinal alterations assessed by our neuroradiologist (FC). DTI data of each patient were processed using our proposed pipeline described in section 2 and 3. Among the 20 fiber-bundles (contained in the JHU atlas), the Cortico-Spinal Tract (CST), fronto-occipital fasciculus (IFOF) and superior longitudinal fasciculus (SLF) were selected because of the presence of longitudinal MS alterations.

6 Results

6.1 Detection of affected fibers, cross-sections and time-points on simulated data

Performances obtained for the detection, in simulated data, of affected fibers within a bundle are reported in table 3.1. The table reports the value of accuracy, precision and sensitivity obtained in the task of the extraction of fibers affected by longitudinal changes. The tests were performed using different diffusion features. Best results were obtained with using as diffusion features $\langle \lambda_2, \lambda_3 \rangle$ with an accuracy, precision and sensitivity respectively of 0.97, 0.79 and 0.40. Satisfying results were also obtained using $\langle \lambda_2, \lambda_3, FA, MD \rangle$ and $\langle \lambda_2, \lambda_3, FA \rangle$ having respectively an accuracy of 0.77 and 0.79, a precision of 0.73 and 0.76 and a sensitivity of 0.20 and 0.23. Worst results were obtained using $\langle KLA, AA \rangle$ and $\langle KLA, FA, AA \rangle$ having respectively an accuracy of 0.71 and 0.77, a precision of 0.69 and 0.69 and a sensitivity of 0.16 and 0.15.

Performances obtained in the detection of affected cross-sections within a bundle are reported in table 3.2. Like the previous case, globally, the best results were achieved using $\langle \lambda_2, \lambda_3 \rangle$ as feature set with an accuracy, precision and sensitivity respectively of 0.63, 0.98 and 0.95. The features set composed by $\langle \lambda_2, \lambda_3, FA \rangle$ show low values in terms of accuracy (0.40) and sensitivity (0.56) but the highest level of precision (1.00). Like for the tests performed for the detection of fibers affected by longitudinal changes, worst results were achieved by the features set $\langle KLA, AA \rangle$ and $\langle KLA, FA, AA \rangle$. Moreover, in table 3.2, we show the performances obtained with the tensor-based and NMF methods to detect cross-sections, of a fiber-bundle, affected by pathological longitudinal changes. From the table it is possible to see how the NMF method method always outperforms tensor factorization in accuracy. More in detail, NMF reaches the best accuracy (0.74) with $\langle \lambda_2, \lambda_3, FA \rangle$ and $\langle \lambda_2, \lambda_3, FA, MD \rangle$. Similar results are visible in sensitivity, indeed NMF always outperforms tensor factorization except for the feature set $\langle \lambda_2, \lambda_3 \rangle$ where tensor factorization reaches the best performance. Contrarily, for the accuracy, the tensor factorization outperforms, in all the different diffusion features used, the NMF method.

Performances obtained in the detection of time-points affected by longitudinal changes are reported in table 3.3. Best results were achieved using $\langle \lambda_2, \lambda_3 \rangle$ as features with an accuracy, precision and sensitivity respectively of 0.84, 0.93 and 0.96. Contrary to the previous case, the second best features set is $\langle \lambda_2, \lambda_3, FA, MD \rangle$ having accuracy, precision and sensitivity respectively of 0.73, 0.87 and 0.81. In this case $\langle \lambda_2, \lambda_3, FA \rangle$ globally shows results comparable to $\langle KLA, AA \rangle$ and $\langle KLA, FA, AA \rangle$. The largest difference is in the accuracy who reaches 0.60 for $\langle \lambda_2, \lambda_3, FA \rangle$ and 0.50 for both $\langle KLA, AA \rangle$ and $\langle KLA, FA, AA \rangle$. Moreover, as additional experiment, in table 3.3 we compared the performances in detection of time-points affected by longitudinal changes obtained using the proposed tensor factorization algorithm with the NMF method. From the table it is possible to see how the proposed tensor factorization algorithm always outperforms the NMF method in all the performance measurements and in all the combination of features used in the experiment.

Regarding the parameters used to perform the analysis used by the tensor factorization algorithm, the best results were achieved with $MinPts = 3$ and $\omega = 8$.

Table 3.1: Results of mean and standard deviation (in parenthesis) of accuracy, precision and sensitivity for detection of longitudinal affected fibers using different combination of diffusion parameters.

	Accuracy	**Precision**	**Sensitivity**
λ_2,λ_3	**0.97 (0.06)**	**0.79 (0.08)**	**0.40 (0.11)**
λ_2,λ_3, FA	0.77 (0.29)	0.73 (0.11)	0.20 (0.02)
λ_2,λ_3, FA, MD	0.79 (0.32)	0.76 (0.15)	0.23 (0.05)
KLA,AA	0.71 (0.23)	0.69 (0.13)	0.16 (0.01)
KLA,FA,AA	0.77 (0.20)	0.66 (0.15)	0.15 (0.04)

Table 3.2: Results obtained using tensor and non-negative matrix factorization. Results are reported with mean and standard deviation (in parenthesis) of accuracy, precision and sensitivity for detection of longitudinal affected cross-sections using different combination of diffusion parameters.

	Tensor Factorization			**Non-Negative Matrix Factorization**		
	Accuracy	**Precision**	**Sensitivity**	**Accuracy**	**Precision**	**Sensitivity**
λ_2,λ_3	0.63 (0.10)	0.98 (0.02)	**0.95 (0.15)**	0.65 (0.12)	0.60 (0.08)	0.92 (0.10)
λ_2,λ_3, FA	0.40 (0.12)	**1.00 (0.01)**	0.56 (0.12)	**0.74 (0.11)**	0.68 (0.04)	0.90 (0.09)
λ_2,λ_3, FA, MD	0.47 (0.13)	0.97 (0.03)	0.74 (0.23)	**0.74 (0.08)**	0.68 (0.12)	0.89 (0.11)
KLA,AA	0.30 (0.10)	0.96 (0.03)	0.64 (0.30)	0.59 (0.07)	0.55 (0.12)	0.91 (0.10)
KLA,FA,AA	0.32 (0.12)	0.96 (0.03)	0.63 (0.28)	0.56 (0.11)	0.53 (0.09)	0.85 (0.05)

6.2 Evaluation of parallel implementation

In order to underline the difference between serial and parallel implementation, we show the computational time obtained by using the two implementations. More in detail, the implementations were run 10 times on the same task and the computational time (in terms of mean and standard deviation) were measured. For the parallel implementation

Table 3.3: Results obtained using tensor and non-negative matrix factorization. Results are reported with mean and standard deviation (in parenthesis) of accuracy, precision and sensitivity for detection of longitudinal affected time-points using different combination of diffusion parameters.

	Tensor Factorization			**Non-Negative Matrix Factorization**		
	Accuracy	**Precision**	**Sensitivity**	**Accuracy**	**Precision**	**Sensitivity**
λ_2,λ_3	**0.84 (0.12)**	**0.93 (0.09)**	**0.96 (0.13)**	0.72 (0.30)	0.86 (0.18)	0.78 (0.31)
λ_2,λ_3, FA	0.60 (0.20)	0.76 (0.10)	0.54 (0.16)	0.57 (0.36)	0.75 (0.17)	0.53 (0.35)
λ_2,λ_3, FA, MD	0.73 (0.30)	0.87 (0.18)	0.81 (0.32)	0.67 (0.36)	0.84 (0.18)	0.78 (0.33)
KLA,AA	0.50 (0.30)	0.71 (0.21)	0.82 (0.30)	0.50 (0.36)	0.67 (0.19)	0.56 (0.37)
KLA,FA,AA	0.50 (0.32)	0.74 (0.19)	0.75 (0.20)	0.50 (0.36)	0.66 (0.19)	0.57 (0.34)

the same tasks were executed with different parallel processors ($2 \leq K \leq 7$). Results in terms of execution time are reported in Figure 3.8. Since the serial execution time does not depend on the number of parallel process it is always stable. An important decrease in computational time is already clearly visible starting with $K = 2$ where the serial implementation takes 899.8 ± 14.0 seconds while the parallel one takes 550.3 ± 10.3 seconds. We define with speed-up factor the ratio between the mean computational time of the serial implementation and the parallel implementation. Already with $K = 2$, the speed-up factor, is 1.63. Moreover, it is possible to see how by increasing the number of processors the computational time decreases. The minimum and maximum speed-up factors are 1.63 and 2.6 reached for $K = 2$ and $K = 4$ respectively. Results show how the improvement in computational performance is clear using the parallel implementation.

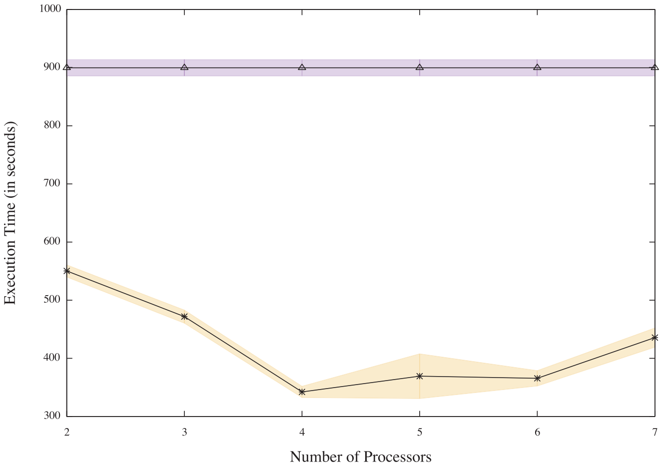


Figure 3.8: Mean and standard deviation, computed from the 10 different runs, of the computation time for Serial ($\triangle \pm$ 🟪) and Parallel ($* \pm$ 🟨) implementations.

6.3 Detection of affected fibers, cross-sections and time-points on real data

The proposed method was also tested on real data. A feature vector composed by $\langle \lambda_2, \lambda_3 \rangle$ ($M = 2$) was used. Based on the results obtained in the previous sections, $MinPts = 3$, $\omega = 8$ and $K = 4$ were selected as parameters.

Results of application on MS real data are illustrated in Figures 3.9-3.12. In all the figures, the mean FA signal profile along the fibers, within the bundle, detected by our method as affected by longitudinal variations are reported. Time-points detected as containing longitudinal changes are reported with the symbol "*" in the name. Cross-sections detected as affected by those changes are underlined with black lines.

In Figure 3.9 and Figure 3.10 we report the FA signal profile of two different MS subjects. The signal profile refers to the CST in the subset of fibers detected as containing longitudinal pathological changes. It is possible to see how in those subsets of fibers, in the time-points and in the cross-sections detected as longitudinal changed, the FA signal goes down due to the presence of small longitudinal MS alterations.

Same behaviour is also visible in the other two MS subjects reported in Figure 3.11 and Figure 3.12 respectively. The former represents the inferior fronto-occipital fasciculus. In the cross-sections and in the time-points identified by our method, longitudinal changes in FA are visible. More in detail, the FA signal goes down in the detected cross-sections due to the presence of MS lesions. In Figure 3.12, we reported the superior longitudinal fasciculus. Like for the other fiber-bundles, in the subset of fibers and cross-sections identified as affected by longitudinal changes, alterations in FA signal profile are visible.

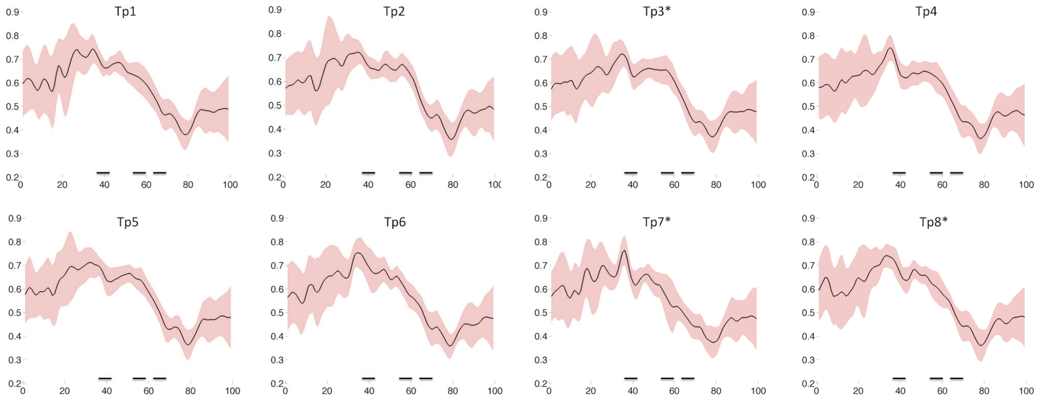


Figure 3.9: Mean FA signal profile (of cortico-spinal tract) along the subset of fibers identified by our method as affected by longitudinal changes. Time-points detected as containing longitudinal changes, detected by our method, contain the symbol “*” in the name. Cross-sections detected as affected by longitudinal pathological changes are underlined with black lines.

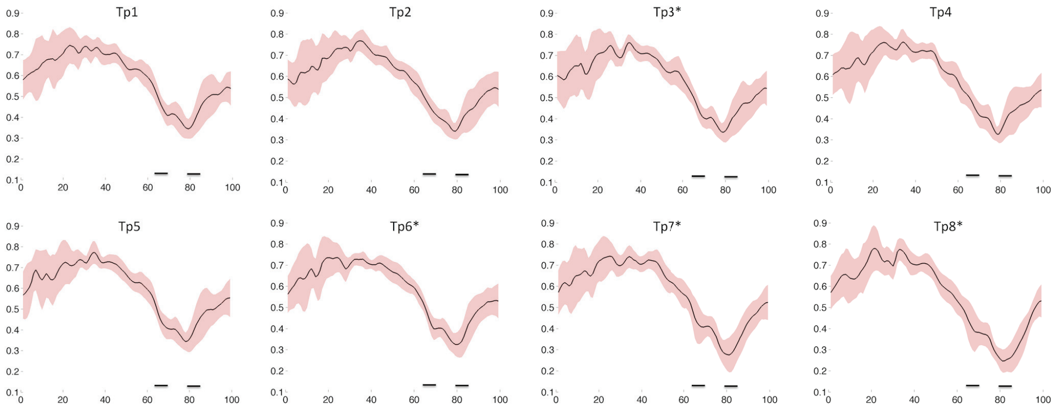


Figure 3.10: Mean FA signal profile (of cortico-spinal tract) along the subset of fibers identified by our method as affected by longitudinal changes. Time-points detected as containing longitudinal changes, detected by our method, contain the symbol “*” in the name. Cross-sections detected as affected by longitudinal pathological changes are underlined with black lines.

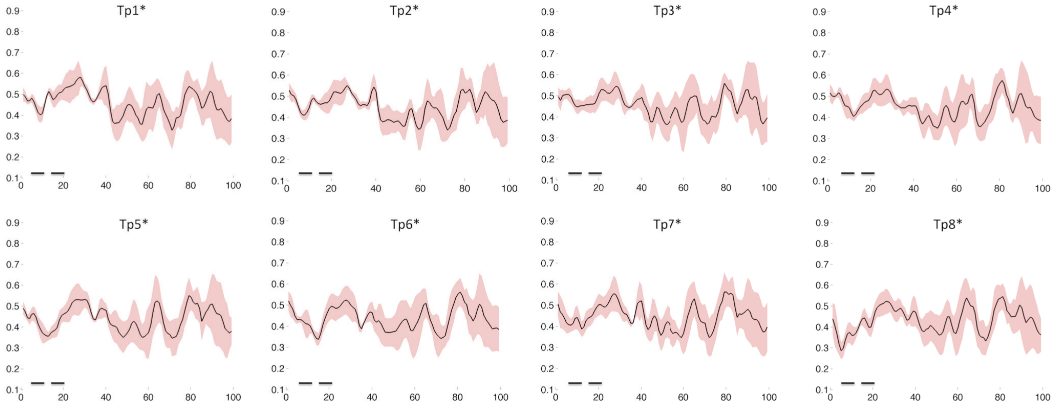


Figure 3.11: Mean FA signal profile (of inferior fronto-occipital fasciculus) along the subset of fibers identified by our method as affected by longitudinal changes. Time-points detected as containing longitudinal changes, detected by our method, contain the symbol “*” in the name. Cross-sections detected as affected by longitudinal pathological changes are underlined with black lines.

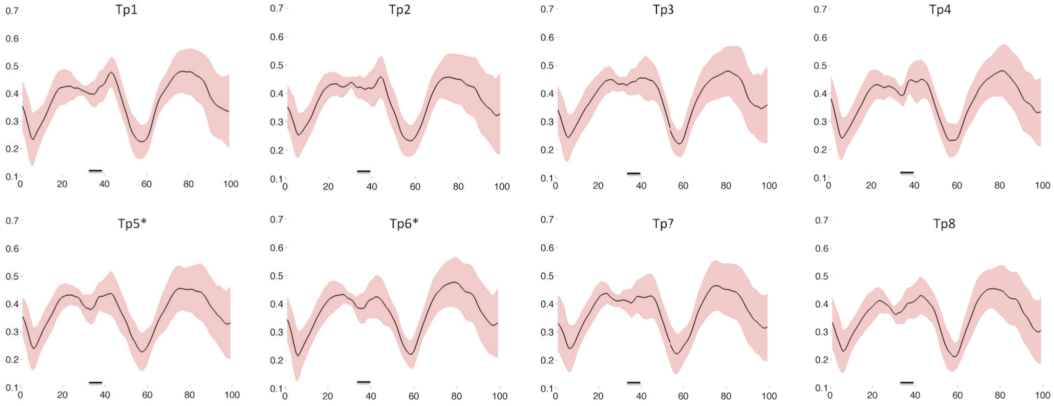


Figure 3.12: Mean FA signal profile (of superior longitudinal fasciculus) along the subset of fibers identified by our method as affected by longitudinal changes. Time-points detected as containing longitudinal changes, detected by our method, contain the symbol “*” in the name. Cross-sections detected as affected by longitudinal pathological changes are underlined with black lines.

7 Discussion

In this chapter we described a new tensor-based method to automatically analyze longitudinal changes in WM fiber-bundles of MS patients. More in detail, we provided a complete pipeline capable to automatically extract fiber-bundles and register longitudinal data to a common template. As major message, we formalized the problem using a tensor-based formalism to detect local scale longitudinal variations caused by rapid inflammatory processes in MS patients.

Moreover, we improved the computational speed of our method by providing a parallel implementation of the algorithm. We used the “divide-et-impera” paradigm to subdivide the main problem in sub-problems. We then merged the sub-solutions of each sub-problem into an unique final solution. In order to perform a better parallelization, we split the fiber-bundle in small sub-bundles using a modified version of K-medoids. As first step, a global distance matrix, based on the MDF metric, was computed between each pair of fibers within a bundle. The obtained matrix was then used as distance matrix for the K-medoids algorithm. The tensor decomposition pipeline was then applied independently to the data extracted in each sub-bundle. The results obtained by the factorization in each sub-problem were then merged in an unique solution. The decrease in computation time, between serial and parallel implementation, is well described in the test we made (Section 6.2). In particular using 4 parallel processors ($K = 4$) we got a speed-up factor of 2.6 (from 899.8 seconds to 342.2 seconds). It should be noted that the number K of parallel processors mainly depends on two factors: *i)* the number of parallel processors available on the machine used to run the algorithm, *ii)* the number of fibers within the bundle. Indeed, if K is large (thus the value $\lvert\lceil \frac{v}{K} \rceil\rvert$ is small) the number of fibers in each process could not be enough to compute an accurate tensor decomposition.

Results on simulated data, generated using GGPDF, show the capability of our method to detect and delineate the presence of longitudinal changes in both fibers and cross-sections of WM fiber-bundles. The experiment section (Section 5) was also enriched by the analysis of the features used to compute the factorization of the longitudinal data. The diffusion features used to perform the analysis play also an important role on the final results. Best results were obtained using $\langle \lambda_2, \lambda_3 \rangle$ as features. The introduction of other diffusion parameters does not drastically improve the performances.

As final experimental step, the proposed method was applied to real MS patients showing its capability to detect and delineate time-points, fibers and cross-sections identified, by the neuroradiologist, as affected by longitudinal changes. With these experiments, we showed that also on real data it is possible to identify regions containing longitudinal pathological changes. This gave us the possibility to extract two types of information: *i)* segmentation of the regions (fibers and cross-sections) affected by the longitudinal variations and *ii)* the delineation of the time-points affected by the longitudinal variations. Detection of longitudinal changes were validated by our neuroradiologist (FC).

Another crucial point of investigation is related to the rank selection. Indeed, in this chapter we used a “brute force” way to compute the rank using the *rankest* function provided in Tensorlab [Vervliet *et al.* (2016)]. This process usually tends to overestimate the real number of components and, as results, noise or data variability can be modelled by the factorization. The main consequence related to this overestimation is a decrease in performances obtained in detection of “pathological” components. This problem suggest how it is important to find a specific “problem dependent” algorithm for the rank estimation.

An important limitation is derived by the low performances, in terms of sensitivity, obtained in the extraction of the fibers affected by longitudinal changes. The low performances could be explained by the CPD model we applied to perform the factorization.

Indeed, CPD uses only rank-1 terms to factorize the tensor. With this restriction, representation of biological variations typical of MRI data could not be perfectly estimated. It is important to observe that, in our reference context, Precision should be privileged over Sensitivity because the number of fibers generated by tractography algorithms usually does not reflect the number of real fibers of a human brain.

An interesting added value of our method is its capability to easily include new time-points acquired later in time. Indeed, due to the robustness of the registration pipeline used in this chapter [Keihaninejad *et al.* (2013)] the effects of certain longitudinal biases [Reuter *et al.* (2012)], like atrophy, are minimized. Moreover, since the values acquired in a new time-point are simply stacked in the 3^{rd} mode of the tensor, no modification is needed in the algorithm. Furthermore, this framework is easy to extend, features derived from other modalities (like T1, T2, etc...) can be concatenated in the 3^{rd} mode of the tensor (as described in Figure 3.4) without requiring additional work.

The proposed method can be seen as a general framework capable to extend our previous method based on the NMF [Stamile *et al.* (2016)b]. Indeed, due to the capability of the tensor to represent high dimensional data, fiber-bundles can be analyzed globally without generating local view (like in cross-sectional analysis). In the comparison we performed with the NMF method, we saw that the better performances, in cross-section analysis, were obtained by the NMF method compared to the tensor factorization. These results are due to the local analysis performed by NMF. Indeed, this method analyze each cross-section independently performing a specific local analysis focused on cross-section. In the comparison of the performances obtained in time-point detection, the proposed tensor factorization algorithm outperformed the NMF algorithm. This show how analysis of multiple information at the same time, really helps to improve the detection of small longitudinal changes in WM fiber-bundles.

Moreover, compared to [Stamile *et al.* (2016)a, Stamile *et al.* (2017), Grigis *et al.* (2013)], the proposed method allows to *i)* analyze multiple DTI features taking into account more than 2 time-points *ii)* delineate regions, fibers and detect time-points affected by longitudinal changes. Furthermore, it is important to underline that the method is unsupervised. Hence there is no need to formulate particular hypotheses about the distribution and/or evolution of the diffusion parameters.

8 Conclusion

In this chapter, we proposed a new tensor-based framework for the analysis of longitudinal changes occurring along WM fiber-bundles. We described how constrained tensor factorization is a potential tool to analyze multi-dimensional data that can not be fully described using a simple matrix representation.

In order to reach our goal, two main challenges related to tensor factorization were analyzed and solved, namely rank selection and computational time for the factorization. The former was solved by estimating L-curve error of the CPD with increasing number of rank-one terms. The latter was improved by splitting the problem in sub-tasks using the "divide et impera" paradigm.

The method was tested and validated on simulated data and real data. We generated simulated data using our previously proposed simulation paradigm . For the real data, we used a dataset containing MS subjects affected by small longitudinal changes visible in a short (weekly) interval. Results suggested that the proposed method is a promising tool to longitudinal analysis of fiber-bundles in neurodegenerative diseases like MS.

As for future work, we plan to extend the application of our method to perform multimodal analysis. The idea is to apply the algorithm on different MRI modalities (MRS, T2, T1, etc.) combined. We also plan to improve the detection of cross-sections affected by longitudinal changes by introducing regularization for the component matrix representing cross-sections. Moreover, as we reported in Section 7, CPD could be too restrictive for some applications as it does not model all variability in the data [Hunyadi *et al.* (2014)]. We plan to increase the performance, especially for the cross-section component, using block term decomposition (BTD) instead of CPD. Indeed, using a BTD model, it is possible to model the variability on the data by performing a so-called rank $(L_r, L_r, 1)$ BTD [Sorber *et al.* (2013)].

IV Conclusions and Perspectives

Chapter 1

Conclusions

1 Main Contributions

Processing of longitudinal MRI data is a crucial challenge to better understand pathological mechanisms of complex brain diseases such as MS where WM fiber-bundles are variably altered by inflammatory events. In this work, we developed a set of supervised and unsupervised methods to extract and analyze longitudinal changes appearing along WM fiber-bundles.

2 Developed Methods

In this work, the following scientific contributions were made:

- **Fiber-Bundle Clustering:** An automated approach for WM fiber-bundles clustering through shape-based model characterization. The key novelties of our approach are: a new string-based formalism, allowing an alternative representation of WM fibers, a new string dissimilarity metric, a WM fiber clustering technique, and a new model-based characterization algorithm. Thanks to these novelties, the complex problem of WM fiber-bundle extraction and characterization reduces to a much simpler and well-known string extraction and analysis problem.
- **Histogram-Based Approach:** The algorithm allows to analyze longitudinal DTI data along WM fiber-bundle. It detects small longitudinal changes along the fiber-tracts. Based on a Gaussian mixture model, this method provides a fine cross-

sectional fiber-bundle analysis allowing the differentiation of "changed" and "unchanged" fibers of the bundle.

- **Multi-Features NMF Algorithm:** The method improves the algorithm previous described using a combination of non-negative matrix factorization and LOF algorithms. The method allows to detect and delineate longitudinal variations appearing in multiple time-points and in multiple features along cross-sections of the WM fiber-bundle.

- **Constrained Tensor Decomposition Framework:** It is a new tensor-based framework capable to detect longitudinal changes appearing along fiber-bundles in dMRI data in MS patients. The algorithm allows to identify at the same time which subset of fibers (within a bundle) and which cross-sections of these fibers contain "pathological" longitudinal changes. Moreover, it allows to delineate the time-points in which the longitudinal pathological changes appear.

3 Discussion

In this thesis we developed two principal axes, namely fiber-bundle clustering and unsupervised methods for longitudinal fiber-bundle analysis. For each of those points we developed specific algorithms and performed different experiments to validate the methods.

Fiber-Bundle Clustering

We presented a new string-based algorithm to automatically extract WM fiber-bundles from the whole tractogram. The proposed approach exploits a new string-based formalism useful to obtain an alternative representation of WM fibers. We also enriched this new representation proposing a new string similarity metric, namely SBED, used as distance measure to perform unsupervised WM fiber clustering. In order to validate the proposed algorithm, we collocated our approach in the context of related literature and we performed an experimental campaign to test it. As major result, the proposed string-based algorithm overcomes different limitations of the related ones proposed in the past. For instance, it allows a better integration of a-priori information provided by a neuroanatomist (which is not possible in QuickBundles). Indeed, the usage of a string-based model, representing the shape of a particular fiber-bundle, allows to easily extract just the subset of fiber having the same shape as the provided model. Moreover, we improved the proposed method by coupling the proposed string-based formalism with the spatial information of the tractogram.

Unsupervised Methods for Longitudinal Fiber-Bundle Analysis

Histogram-Based Approach

We introduced a new simple fully automated tool for analyzing longitudinal changes in WM fiber-bundles of MS patients. Compared to previous methods, already described in literature, for the characterization of fiber-tract profiles, our approach provides a fine detection of local scale longitudinal variations along the fiber-bundle. Those small local scale changes are indeed expected to occur during inflammatory or neurodegenerative processes in MS patients. This new approach allows, for the first time, the discrimination between healthy and affected fibers within a bundle. The method was tested on simulated and real clinical data. High levels of F-Measure were obtained on simulated data. Experiments on cortico-spinal tract and inferior fronto-occipital fasciculi on patients with MS included in a weekly follow-up protocol highlighted the greater sensitivity of this fiber scale approach to detect small longitudinal alterations.

From the results we obtained we saw how this method holds the potential for more detailed and topographically specific description of disease-induced disruption of connectivity in the brain, with implications for specific functional losses associated with disease progression.

Multi-Features NMF Algorithm

In order to overcome several limitations present in the previous methods, like two time-points analysis, we developed a new unsupervised algorithm based on NMF in combination with LOF. Like the previous method, we tested our algorithm on simulated and on real data. In order to validate our algorithm, we introduced a new model to simulate real longitudinal changes based on a generalized Gaussian probability density function. High level of performances were obtained for the detection of small longitudinal changes along the WM fiber-bundles in MS patients.

Constrained Tensor Decomposition Framework

Finally, we built a general framework based on constrained tensor factorization. This new formalism is capable to analyze all the information contained in the WM fiber-bundle without create local view of the data. Indeed, using tensor-based formalism, it is possible to exploit the “higher dimensionality” of the model in order to describe complex high dimensional data. Moreover, since the tensor factorization could be computationally intensive, we optimized the computational time by using parallel programming. The original problem of the longitudinal fiber-bundle analysis was subdivided in small sub-problems solved in parallel. The sub-solutions of each sub-problem were then merged in a unique solution. With this new implementation we obtained a speed-up factor of 2.6 (from 899.8 seconds to 342.2 seconds). As the for the previous methods we tested the performances of this algorithm on simulated and real data. Moreover, in order to show the power of tensor-based methods, we compared its performances with the ones obtained with the NMF-based algorithm. The comparison shows that tensor factorization, thanks to its capability to maintain the high dimensional structure of the data, outperforms the NMF-based algorithm.

Chapter 2

Perspectives

All the algorithms proposed in this work can be seen as a starting point for a more general and robust pipeline to analyze longitudinal signal changes along WM fiber-bundles. Indeed, several efforts are still needed in order to improve the algorithms developed in this work.

Fiber-Bundle Clustering

For the fiber-bundles extraction algorithm, we plan to investigate a possible improvement of QB in such a way as to make it capable of correcting the assignment of a fiber to a cluster when this action, in a second time, appears incorrect. Moreover, instead using a fully unsupervised algorithm, we plan to use the shape information obtained from string models to train supervised machine learning algorithms on both real and simulated data.,Finally, in order to find the best configuration for SBED, we plan to optimize the usage its constraints in such a way as to favor the extraction of specific WM fiber-bundles.

Histogram-Based Approach

For the histogram method used for longitudinal analysis, encouraged by the obtained results, we plan to extend the application of our method to perform multimodal analysis and not only on the FA metrics. Indeed, in order to improve the analysis, the idea is to apply the algorithm on different MRI modalities (MRS, T2, T1, etc..) and other diffusion metrics like FOD derived metrics.

Multi-Features NMF Algorithm

For this algorithm an interesting area of investigation is the study of the effects of different regularization operators like l_1- or l_2- norms on the final solution. Indeed, since MS lesions show nice properties like regular closed shape or intensity gradient between the center and the border, the application of specific regularization operators can be useful to incorporate such information in the desired solution.

Constrained Tensor Decomposition Framework

The last proposed algorithm based on constrained tensor decomposition is really a novelty in brain studies and its perspectives, at this time, are quite large. For instance, one interesting point of investigation is the application of different tensorization techniques to build the tensor from the data. For instance, well known tensorization techniques like Hankelization, Löwnerization, decimation or segmentation are currently used to generate tensors starting from raw data. Those techniques allow to generate low rank structures that can preserve interesting properties that can be used during and after the factorization. Moreover, another interesting point, is the analysis of other factorization techniques like BTD, or Tucker decomposition. Indeed, they can be exploited to better describe the typical variations occurring in complex data like MRI.

To conclude, all these new techniques, to analyze small longitudinal changes in WM fiber-bundles, can be used to perform better studies about the pathological mechanisms of several neurodegenerative pathologies. For instance, these new algorithms will provide a better characterization of brain alterations and particularly in the entire afferent visual pathway from the retina to the visual cortex of CIS patients. This information will constitute a great benefit in the clinical setting for evidence-based decision making for personalized treatment of CIS patients to reduce disease severity at an early time-point.

Bibliography

[Achard *et al.* (2012)] Achard, S., Delon-Martin, C., Vértes, P. E., Renard, F., Schenck, M., Schneider, F., Heinrich, C., Kremer, S., and Bullmore, E. T. (2012). Hubs of brain functional networks are radically reorganized in comatose patients. *Proceedings of the National Academy of Sciences*, 109(50):20608–20613.

[Albers (1998)] Albers, G. W. (1998). Diffusion-weighted mri for evaluation of acute stroke. *Neurology*, 51(3 Suppl 3):S47–S49.

[Alexander *et al.* (2010)] Alexander, D. C., Hubbard, P. L., Hall, M. G., Moore, E. A., Ptito, M., Parker, G. J., and Dyrby, T. B. (2010). Orientationally invariant indices of axon diameter and density from diffusion mri. *Neuroimage*, 52(4):1374–1389.

[Anderson *et al.* (2011)] Anderson, V., Wheeler-Kingshott, C., Abdel-Aziz, K., Miller, D., Toosy, A., Thompson, A., and Ciccarelli, O. (2011). A comprehensive assessment of cerebellar damage in multiple sclerosis using diffusion tractography and volumetric analysis. *Multiple Sclerosis Journal*, 17(9):1079–1087.

[Apostolico *et al.* (2007)] Apostolico, A., Erdős, P. L., and Lewenstein, M. (2007). Parameterized matching with mismatches. *Journal of Discrete Algorithms*, 5(1):135–140.

[Arvo (2013)] Arvo, J. (2013). *Graphics gems II*. Elsevier.

[Asaf *et al.* (2015)] Asaf, A., Evan, S., and Anat, A. (2015). Injury to white matter tracts in relapsing–remitting multiple sclerosis: A possible therapeutic window within the first 5 years from onset using diffusion-tensor imaging tract-based spatial statistics. *NeuroImage: Clinical*, 8:261–266.

[Assaf *et al.* (2013)] Assaf, Y., Alexander, D. C., Jones, D. K., Bizzi, A., Behrens, T. E., Clark, C. A., Cohen, Y., Dyrby, T. B., Huppi, P. S., Knoesche, T. R., *et al.* (2013). The connect project: combining macro-and micro-structure. *Neuroimage*, 80:273–282.

[Assaf and Pasternak (2008)] Assaf, Y. and Pasternak, O. (2008). Diffusion tensor imaging (dti)-based white matter mapping in brain research: a review. *Journal of molecular neuroscience*, 34(1):51–61.

[Baker (1996)] Baker, B. S. (1996). Parameterized pattern matching: Algorithms and applications. *J. Comput. Syst. Sci.*, 52(1):28–42.

[Baker (1999)] Baker, B. S. (1999). Parameterized diff. In *Proceedings of the Tenth Annual ACM-SIAM Symposium on Discrete Algorithms, 17-19 January 1999, Baltimore, Maryland.*, pages 854–855. ACM/SIAM.

[Balashov and Lindzen (2012)] Balashov, K. E. and Lindzen, E. (2012). Acute demyelinating lesions with restricted diffusion in multiple sclerosis. *Multiple Sclerosis Journal*, 18(12):1745–1753.

[Barkhof (2002)] Barkhof, F. (2002). The clinico-radiological paradox in multiple sclerosis revisited. *Current opinion in neurology*, 15(3):239–245.

[Barkhof *et al.* (1997)] Barkhof, F., Filippi, M., Miller, D. H., Scheltens, P., Campi, A., Polman, C. H., Comi, G., Ader, H. J., Losseff, N., and Valk, J. (1997). Comparison of mri criteria at first presentation to predict conversion to clinically definite multiple sclerosis. *Brain*, 120(11):2059–2069.

[Basser *et al.* (1992)] Basser, P., Mattiello, J., and LeBihan, D. (1992). Diagonal and off-diagonal components of the self-diffusion tensor: their relation to and estimation from the nmr spin-echo signal. *Proc. 11th Annu. Meet. SMRM, Berlin*, 1:1222.

[Basser *et al.* (1994)a] Basser, P. J., Mattiello, J., and LeBihan, D. (1994a). Estimation of the effective self-diffusion tensor from the nmr spin echo. *Journal of Magnetic Resonance, Series B*, 103(3):247–254.

[Basser *et al.* (1994)b] Basser, P. J., Mattiello, J., and LeBihan, D. (1994b). Mr diffusion tensor spectroscopy and imaging. *Biophysical journal*, 66(1):259–267.

[Behrens *et al.* (2003)] Behrens, T., Johansen-Berg, H., Woolrich, M., Smith, S., Wheeler-Kingshott, C., Boulby, P., Barker, G., Sillery, E., Sheehan, K., Ciccarelli, O., *et al.* (2003). Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. *Nature neuroscience*, 6(7):750–757.

[Bergsland *et al.* (2015)] Bergsland, N., Laganà, M. M., Tavazzi, E., Caffini, M., Tortorella, P., Baglio, F., Baselli, G., and Rovaris, M. (2015). Corticospinal tract integrity is related to primary motor cortex thinning in relapsing-remitting multiple sclerosis. *Multiple sclerosis (Houndmills, Basingstoke, England)*, 21:1771–1780.

[Blin *et al.* (2012)] Blin, G., Jiang, M., and Vialette, S. (2012). The longest common subsequence problem with crossing-free arc-annotated sequences. In *String Processing and Information Retrieval - 19th International Symposium, SPIRE 2012, Cartagena de Indias, Colombia, October 21-25, 2012. Proceedings*, volume 7608 of *Lecture Notes in Computer Science*, pages 130–142. Springer.

[Bloch (1946)] Bloch, F. (1946). Nuclear induction. *Physical review*, 70(7-8):460.

[Braley *et al.* (2012)] Braley, T. J., Lee, Y. H., Mohan, S., Segal, B. M., Berini, S., and Srinivasan, A. (2012). Differences in diffusion tensor imaging–derived metrics in the corpus callosum of patients with multiple sclerosis without and with gadolinium-enhancing cerebral lesions. *Journal of computer assisted tomography*, 36(4):410–415.

[Breunig *et al.* (2000)] Breunig, M. M., Kriegel, H.-P., Ng, R. T., and Sander, J. (2000). Lof: Identifying density-based local outliers. *SIGMOD Rec.*, 29(2):93–104.

[Brown (1828)] Brown, R. (1828). A brief account of microscopical observations made in the months of june, july and august 1827 on the particles contained in the pollen of plants.

[Brück *et al.* (2003)] Brück, W., Kuhlmann, T., and Stadelmann, C. (2003). Remyelination in multiple sclerosis. *Journal of the neurological sciences*, 206(2):181–185.

[Bénézit *et al.* (2015)] Bénézit, A., Hertz-Pannier, L., Dehaene-Lambertz, G., Monzalvo, K., Germanaud, D., Duclap, D., Guevara, P., Mangin, J.-F., Poupon, C., Moutard, M.-L., *et al.* (2015). Organising white matter in a brain without corpus callosum fibres. *Cortex*, 63:155–171.

[Cai *et al.* (2013)] Cai, X., Nie, F., and Huang, H. (2013). Multi-view k-means clustering on big data. In *IJCAI 2013, Proceedings of the 23rd International Joint Conference on Artificial Intelligence, Beijing, China, August 3-9, 2013*, pages 2598–2604. IJCAI/AAAI.

[Caruyer *et al.* (2014)] Caruyer, E., Daducci, A., Descoteaux, M., Houde, J.-C., Thiran, J.-P., and Verma, R. (2014). Phantomas: a flexible software library to simulate diffusion mr phantoms. In *ISMRM*.

[Castellanos *et al.* (2002)] Castellanos, J. L., Gómez, S., and Guerra, V. (2002). The triangle method for finding the corner of the L-curve. *Applied Numerical Mathematics. An IMACS Journal*, 43(4):359–373.

[Catani and Thiebaut de Schotten (2008)] Catani, M. and Thiebaut de Schotten, M. (2008). A diffusion tensor imaging tractography atlas for virtual in vivo dissections. *Cortex; a journal devoted to the study of the nervous system and behavior*, 44:1105–1132.

[Cauteruccio *et al.* (2015)] Cauteruccio, F., Stamile, C., Terracina, G., Ursino, D., and Sappey-Mariniery, D. (2015). An automated string-based approach to white matter fiber-bundles clustering. In *Neural Networks (IJCNN), 2015 International Joint Conference on*, pages 1–8. IEEE.

[Cavallari *et al.* (2013)] Cavallari, M., Moscufo, N., Skudlarski, P., Meier, D., Panzer, V. P., Pearlson, G. D., White, W. B., Wolfson, L., and Guttmann, C. R. (2013). Mobility impairment is associated with reduced microstructural integrity of the inferior and superior cerebellar peduncles in elderly with no clinical signs of cerebellar dysfunction. *NeuroImage: Clinical*, 2:332–340.

[Ciccarelli *et al.* (2001)] Ciccarelli, O., Werring, D., Wheeler-Kingshott, C., Barker, G., Parker, G., Thompson, A., and Miller, D. (2001). Investigation of ms normal-appearing brain using diffusion tensor mri with clinical correlations. *Neurology*, 56(7):926–933.

[Cichocki and Phan (2009)] Cichocki, A. and Phan, A. H. (2009). Fast local algorithms for large scale nonnegative matrix and tensor factorizations. *IEICE Transactions*, 92-A(3):708–721.

[Cohen-Adad *et al.* (2011)] Cohen-Adad, J., Descoteaux, M., and Wald, L. L. (2011). Quality assessment of high angular resolution diffusion imaging data using bootstrap on q-ball reconstruction. *Journal of Magnetic Resonance Imaging*, 33(5):1194–1208.

[Colby *et al.* (2012)] Colby, J. B., Soderberg, L., Lebel, C., Dinov, I. D., Thompson, P. M., and Sowell, E. R. (2012). Along-tract statistics allow for enhanced tractography analysis. *Neuroimage*, 59(4):3227–3242.

[Compston and Coles (2008)] Compston, A. and Coles, A. (2008). Multiple sclerosis. *Lancet (London, England)*, 372:1502–1517.

[Corouge *et al.* (2006)] Corouge, I., Fletcher, P. T., Joshi, S., Gouttard, S., and Gerig, G. (2006). Fiber tract-oriented statistics for quantitative diffusion tensor mri analysis. *Medical image analysis*, 10(5):786–798.

[Daducci *et al.* (2015)] Daducci, A., Canales-Rodríguez, E. J., Zhang, H., Dyrby, T. B., Alexander, D. C., and Thiran, J.-P. (2015). Accelerated microstructure imaging via convex optimization (amico) from diffusion mri data. *NeuroImage*, 105:32–44.

[Damadian (1971)] Damadian, R. (1971). Tumor detection by nuclear magnetic resonance. *Science*, 171:1151–1153.

[Deb *et al.* (2002)] Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197.

[Dempster *et al.* (1977)] Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society. Series B (methodological)*, pages 1–38.

[Disanto *et al.* (2012)] Disanto, G., Morahan, J., Barnett, M., Giovannoni, G., and Ramagopalan, S. (2012). The evidence for a role of b cells in multiple sclerosis. *Neurology*, 78(11):823–832.

[Douek *et al.* (1991)] Douek, P., Turner, R., Pekar, J., Patronas, N., and Le Bihan, D. (1991). Mr color mapping of myelin fiber orientation. *Journal of computer assisted tomography*, 15:923–929.

[Einstein (1956)] Einstein, A. (1956). *Investigations on the Theory of the Brownian Movement*. Courier Corporation.

[Filippi *et al.* (2000)] Filippi, M., Iannucci, G., Cercignani, M., Rocca, M. A., Pratesi, A., and Comi, G. (2000). A quantitative study of water diffusion in multiple sclerosis lesions and normal-appearing white matter using echo-planar imaging. *Archives of neurology*, 57(7):1017–1021.

[Filippi *et al.* (2001)] Filippi, M., Cercignani, M., Inglese, M., Horsfield, M., and Comi, G. (2001). Diffusion tensor magnetic resonance imaging in multiple sclerosis. *Neurology*, 56(3):304–311.

[Fillard *et al.* (2011)] Fillard, P., Descoteaux, M., Goh, A., Gouttard, S., Jeurissen, B., Malcolm, J., Ramirez-Manzanares, A., Reisert, M., Sakaie, K., Tensaouti, F., *et al.* (2011). Quantitative evaluation of 10 tractography algorithms on a realistic diffusion mr phantom. *Neuroimage*, 56(1):220–234.

[Fiorini *et al.* (2015)] Fiorini, S., Verri, A., Tacchino, A., Ponzio, M., Brichetto, G., and Barla, A. (2015). A machine learning pipeline for multiple sclerosis course detection from clinical scales and patient reported outcomes. In *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, pages 4443–4446. IEEE.

[Garyfallidis *et al.* (2012)] Garyfallidis, E., Brett, M., Correia, M. M., Williams, G. B., and Nimmo-Smith, I. (2012). Quickbundles, a method for tractography simplification. *Frontiers in neuroscience*, 6:175.

[Gawrychowski and Uznański (2016)] Gawrychowski, P. and Uznański, P. (2016). Order-preserving pattern matching with k mismatches. *Theoretical Computer Science*, 638:136–144.

[Gerstner and Sorensen (2011)] Gerstner, E. R. and Sorensen, A. G. (2011). Diffusion and diffusion tensor imaging in brain cancer. *Seminars in radiation oncology*, 21:141–146.

[Gillis and Glineur (2012)] Gillis, N. and Glineur, F. (2012). Accelerated multiplicative updates and hierarchical ALS algorithms for nonnegative matrix factorization. *Neural Computation*, 24(4):1085–1105.

[Golby *et al.* (2011)] Golby, A. J., Kindlmann, G., Norton, I., Yarmarkovich, A., Pieper, S., and Kikinis, R. (2011). Interactive diffusion tensor tractography visualization for neurosurgical planning. *Neurosurgery*, 68(2):496.

[Goldschmidt *et al.* (2009)] Goldschmidt, T., Antel, J., König, F., Brück, W., and Kuhlmann, T. (2009). Remyelination capacity of the ms brain decreases with disease chronicity. *Neurology*, 72(22):1914–1921.

[Gorbenko and Popov (2012)] Gorbenko, A. and Popov, V. (2012). The longest common parameterized subsequence problem. *Applied Mathematical Sciences*, 6(58):2851–2855.

[Gorgoraptis *et al.* (2010)] Gorgoraptis, N., Wheeler-Kingshott, C. A., Jenkins, T. M., Altmann, D. R., Miller, D. H., Thompson, A. J., and Ciccarelli, O. (2010). Combining tractography and cortical measures to test system-specific hypotheses in multiple sclerosis. *Multiple Sclerosis Journal*, 16(5):555–565.

[Greco and Terracina (2013)] Greco, G. and Terracina, G. (2013). Frequency-based similarity for parameterized sequences: Formal framework, algorithms, and applications. *Information Sciences*, 237:176–195.

[Grigis *et al.* (2013)] Grigis, A., Noblet, V., Blanc, F., Heitz, F., de Seze, J., Kremer, S., and Armspach, J.-P. (2013). Longitudinal change detection: inference on the diffusion tensor along white matter pathways. *Medical image analysis*, 17(3):375–386.

[Hagmann *et al.* (2007)] Hagmann, P., Kurant, M., Gigandet, X., Thiran, P., Wedeen, V. J., Meuli, R., and Thiran, J.-P. (2007). Mapping human whole-brain structural networks with diffusion mri. *PloS one*, 2(7):e597.

[Hahn (1950)] Hahn, E. L. (1950). Spin echoes. *Physical review*, 80(4):580.

[Hannoun *et al.* (2012)a] Hannoun, S., Bagory, M., Durand-Dubief, F., Ibarrola, D., Comte, J.-C., Confavreux, C., Cotton, F., and Sappey-Marinier, D. (2012a). Correlation of diffusion and metabolic alterations in different clinical forms of multiple sclerosis. *PLoS One*, 7(3):e32525.

[Hannoun *et al.* (2012)b] Hannoun, S., Durand-Dubief, F., Confavreux, C., Ibarrola, D., Streichenberger, N., Cotton, F., Guttmann, C., and Sappey-Marinier, D. (2012b). Diffusion tensor–mri evidence for extra-axonal neuronal degeneration in caudate and thalamic nuclei of patients with multiple sclerosis. *American Journal of Neuroradiology*, 33(7):1363–1368.

[Hazay *et al.* (2007)] Hazay, C., Lewenstein, M., and Sokol, D. (2007). Approximate parameterized matching. *ACM Transactions on Algorithms (TALG)*, 3(3):29.

[Heckbert (1982)] Heckbert, P. (1982). *Color image quantization for frame buffer display*, volume 16. ACM.

[Hendee and Morgan (1984)] Hendee, W. R. and Morgan, C. J. (1984). Magnetic resonance imaging part ii - clinical applications. *Western Journal of Medicine*, 141(5):638.

[Henry *et al.* (2009)] Henry, R. G., Shieh, M., Amirbekian, B., Chung, S., Okuda, D. T., and Pelletier, D. (2009). Connecting white matter injury and thalamic atrophy in clinically isolated syndromes. *Journal of the neurological sciences*, 282:61–66.

[Hua *et al.* (2008)] Hua, K., Zhang, J., Wakana, S., Jiang, H., Li, X., Reich, D. S., Calabresi, P. A., Pekar, J. J., van Zijl, P. C., and Mori, S. (2008). Tract probability maps in stereotaxic spaces: analyses of white matter anatomy and tract-specific quantification. *Neuroimage*, 39(1):336–347.

[Hunyadi *et al.* (2014)] Hunyadi, B., Camps, D., Sorber, L., Paesschen, W. V., Vos, M. D., Huffel, S. V., and Lathauwer, L. D. (2014). Block term decomposition for modelling epileptic seizures. *EURASIP J. Adv. Sig. Proc.*, 2014:139.

[Inano *et al.* (2011)] Inano, S., Takao, H., Hayashi, N., Abe, O., and Ohtomo, K. (2011). Effects of age and gender on white matter integrity. *American Journal of Neuroradiology*, 32(11):2103–2109.

[Jbabdi and Johansen-Berg (2011)] Jbabdi, S. and Johansen-Berg, H. (2011). Tractography: where do we go from here? brain connectivity 1 (3), 169–183.

[Jenkinson *et al.* (2012)] Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., and Smith, S. M. (2012). Fsl. *Neuroimage*, 62(2):782–790.

[Jenkinson and Smith (2001)] Jenkinson, M. and Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Medical image analysis*, 5:143–156.

[Jin *et al.* (2014)] Jin, Y., Shi, Y., Zhan, L., Gutman, B. A., de Zubicaray, G. I., McMahon, K. L., Wright, M. J., Toga, A. W., and Thompson, P. M. (2014). Automatic clustering of white matter fibers in brain diffusion mri with an application to genetics. *Neuroimage*, 100:75–90.

[Jones (2010)] Jones, D. K. (2010). Challenges and limitations of quantifying brain connectivity in vivo with diffusion mri. *Imaging in Medicine*, 2(3):341–355.

[Kamiya *et al.* (2002)] Kamiya, T., Kusumoto, S., and Inoue, K. (2002). Ccfinder: a multilinguistic token-based code clone detection system for large scale source code. *IEEE Transactions on Software Engineering*, 28(7):654–670.

[Kaufman and Rousseeuw (1987)] Kaufman, L. and Rousseeuw, P. (1987). *Clustering by means of medoids.* North-Holland.

[Keihaninejad *et al.* (2013)] Keihaninejad, S., Zhang, H., Ryan, N. S., Malone, I. B., Modat, M., Cardoso, M. J., Cash, D. M., Fox, N. C., and Ourselin, S. (2013). An unbiased longitudinal analysis framework for tracking white matter changes using diffusion tensor imaging with application to alzheimer's disease. *NeuroImage*, 72:153–163.

[Kern *et al.* (2011)] Kern, K. C., Sarcona, J., Montag, M., Giesser, B. S., and Sicotte, N. L. (2011). Corpus callosal diffusivity predicts motor impairment in relapsing–remitting multiple sclerosis: a tbss and tractography study. *Neuroimage*, 55(3):1169–1177.

[Kingsley (2006)a] Kingsley, P. B. (2006a). Introduction to diffusion tensor imaging mathematics: Part i. tensors, rotations, and eigenvectors. *Concepts in Magnetic Resonance Part A*, 28(2):101–122.

[Kingsley (2006)b] Kingsley, P. B. (2006b). Introduction to diffusion tensor imaging mathematics: Part ii. anisotropy, diffusion-weighting factors, and gradient encoding schemes. *Concepts in Magnetic Resonance Part A*, 28(2):123–154.

[Klein *et al.* (2007)] Klein, J., Hermann, S., Konrad, O., Hahn, H. K., and Peitgen, H.-O. (2007). Automatic quantification of dti parameters along fiber bundles. *Bildverarbeitung fÃijr die Medizin 2007*.

[Kocevar *et al.* (2016)] Kocevar, G., Stamile, C., Hannoun, S., Cotton, F., Vukusic, S., Durand-Dubief, F., and Sappey-Marinier, D. (2016). Graph theory-based brain connectivity for automatic classification of multiple sclerosis clinical courses. *Frontiers in neuroscience*, 10:478.

[Koenig *et al.* (2013)] Koenig, K. A., Sakaie, K. E., Lowe, M. J., Lin, J., Stone, L., Bermel, R. A., Beall, E. B., Rao, S. M., Trapp, B. D., and Phillips, M. D. (2013). High spatial and angular resolution diffusion-weighted imaging reveals forniceal damage related to memory impairment. *Magnetic resonance imaging*, 31(5):695–699.

[Kurtzke (1980)] Kurtzke, J. F. (1980). Geographic distribution of multiple sclerosis: an update with special reference to europe and the mediterranean region. *Acta Neurologica Scandinavica*, 62(2):65–80.

[Kurtzke (2000)] Kurtzke, J. F. (2000). Multiple sclerosis in time and space-geographic clues to cause. *Journal of neurovirology*, 6(2):S134.

[Lai *et al.* (2014)] Lai, C., Zhou, H.-C., Ma, X.-H., and Zhang, H.-X. (2014). Quantitative evaluation of the axonal degeneration of central motor neurons in chronic cerebral stroke with diffusion tensor imaging. *Acta Radiologica*, 55(1):114–120.

[Lee and Seung (1999)] Lee, D. D. and Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, 401:788–791.

[Levenshtein (1965)] Levenshtein, V. I. (1965). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics. Doklady*, 10:707–710.

[Le Bihan *et al.* (2001)] Le Bihan, D., Mangin, J. F., Poupon, C., Clark, C. A., Pappata, S., Molko, N., and Chabriat, H. (2001). Diffusion tensor imaging: concepts and applications. *Journal of magnetic resonance imaging : JMRI*, 13:534–546.

[Li *et al.* (2012)] Li, Y., Sima, D. M., Cauter, S. V., Himmelreich, U., Pi, Y., and Huffel, S. V. (2012). Simulation study of tissue type differentiation using non-negative matrix factorization. In *BIOSIGNALS 2012 - Proceedings of the International Conference on Bio-inspired Systems and Signal Processing, Vilamoura, Algarve, Portugal, 1-4 February, 2012.*, pages 212–217. SciTePress.

[Li *et al.* (2013)] Li, Y., Jewells, V., Kim, M., Chen, Y., Moon, A., Armao, D., Troiani, L., Markovic-Plese, S., Lin, W., and Shen, D. (2013). Diffusion tensor imaging based network analysis detects alterations of neuroconnectivity in patients with clinically early relapsing-remitting multiple sclerosis. *Human brain mapping*, 34(12):3376–3391.

[Liu *et al.* (2012)] Liu, Y., Duan, Y., He, Y., Yu, C., Wang, J., Huang, J., Ye, J., Parizel, P. M., Li, K., and Shu, N. (2012). Whole brain white matter changes revealed by multiple diffusion metrics in multiple sclerosis: a tbss study. *European journal of radiology*, 81(10):2826–2832.

[Lublin *et al.* (1996)] Lublin, F. D., Reingold, S. C., *et al.* (1996). Defining the clinical course of multiple sclerosis results of an international survey. *Neurology*, 46(4):907–911.

[Lublin *et al.* (2014)] Lublin, F. D., Reingold, S. C., Cohen, J. A., Cutter, G. R., Sørensen, P. S., Thompson, A. J., Wolinsky, J. S., Balcer, L. J., Banwell, B., Barkhof, F., *et al.* (2014). Defining the clinical course of multiple sclerosis the 2013 revisions. *Neurology*, 83(3):278–286.

[Lucchinetti *et al.* (2000)] Lucchinetti, C., Bruck, W., Parisi, J., Scheithauer, B., Rodriguez, M., Lassman, H., *et al.* (2000). Heterogeneity of multiple sclerosis lesions: implications for the pathogenesis of demyelination. *Annals of neurology*, 47(6):707–717.

[MacQueen *et al.* (1967)] MacQueen, J. *et al.* (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 281–297. Oakland, CA, USA.

[Maier *et al.* (2010)] Maier, S. E., Sun, Y., and Mulkern, R. V. (2010). Diffusion imaging of brain tumors. *NMR in biomedicine*, 23(7):849–864.

[Mainero *et al.* (2001)] Mainero, C., De Stefano, N., Iannucci, G., Sormani, M., Guidi, L., Federico, A., Bartolozzi, M., Comi, G., and Filippi, M. (2001). Correlates of ms disability assessed in vivo using aggregates of mr quantities. *Neurology*, 56(10):1331–1334.

[Mak *et al.* (2015)] Mak, E., Su, L., Williams, G. B., and O'Brien, J. T. (2015). Neuroimaging correlates of cognitive impairment and dementia in parkinson's disease. *Parkinsonism & related disorders*, 21:862–870.

[Mangin *et al.* (2013)] Mangin, J.-F., Fillard, P., Cointepas, Y., Le Bihan, D., Frouin, V., and Poupon, C. (2013). Toward global tractography. *NeuroImage*, 80:290–296.

[McDonald *et al.* (2001)] McDonald, W. I., Compston, A., Edan, G., Goodkin, D., Hartung, H.-P., Lublin, F. D., McFarland, H. F., Paty, D. W., Polman, C. H., Reingold, S. C., *et al.* (2001). Recommended diagnostic criteria for multiple sclerosis: guidelines from the international panel on the diagnosis of multiple sclerosis. *Annals of neurology*, 50(1):121–127.

[Mendivelso and Pinzón (2015)] Mendivelso, J. and Pinzón, Y. J. (2015). Parameterized matching: Solutions and extensions. In Holub, J. and Zdárek, J., editors, *Proceedings of the Prague Stringology Conference 2015, Prague, Czech Republic, August 24-26, 2015*, pages 118–131. Department of Theoretical Computer Science, Faculty of Information Technology, Czech Technical University in Prague.

[Mesaros *et al.* (2011)] Mesaros, S., Rocca, M. A., Pagani, E., Sormani, M. P., Petrolini, M., Comi, G., and Filippi, M. (2011). Thalamic damage predicts the evolution of primary-progressive multiple sclerosis at 5 years. *AJNR. American journal of neuroradiology*, 32:1016–1020.

[Mori *et al.* (1999)] Mori, S., Crain, B. J., Chacko, V. P., and Van Zijl, P. (1999). Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging. *Annals of neurology*, 45(2):265–269.

[Mukherjee *et al.* (2002)] Mukherjee, P., Miller, J. H., Shimony, J. S., Philip, J. V., Nehra, D., Snyder, A. Z., Conturo, T. E., Neil, J. J., and McKinstry, R. C. (2002). Diffusion-tensor mr imaging of gray and white matter development during normal human brain maturation. *American Journal of Neuroradiology*, 23(9):1445–1456.

[Mukherjee *et al.* (2008)] Mukherjee, P., Berman, J., Chung, S., Hess, C., and Henry, R. (2008). Diffusion tensor mr imaging and fiber tractography: theoretic underpinnings. *American journal of neuroradiology*, 29(4):632–641.

[Muthuraman *et al.* (2016)] Muthuraman, M., Fleischer, V., Kolber, P., Luessi, F., Zipp, F., and Groppa, S. (2016). Structural brain network characteristics can differentiate cis from early rrms. *Frontiers in neuroscience*, 10.

[Myung (2001)] Myung, I. J. (2001). Tutorial tutorial on maximum likelihood estimation.

[Mårtensson *et al.* (2013)] Mårtensson, J., Nilsson, M., Ståhlberg, F., Sundgren, P. C., Nilsson, C., van Westen, D., Larsson, E.-M., and Lätt, J. (2013). Spatial analysis of diffusion tensor tractography statistics along the inferior fronto-occipital fasciculus with application in progressive supranuclear palsy. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 26(6):527–537.

[Nedjati-Gilani *et al.* (2017)] Nedjati-Gilani, G. L., Schneider, T., Hall, M. G., Cawley, N., Hill, I., Ciccarelli, O., Drobnjak, I., Wheeler-Kingshott, C. A. G., and Alexander, D. C. (2017). Machine learning based compartment models with permeability for white matter microstructure imaging. *NeuroImage*, 150:119–135.

[Nicoletti *et al.* (2006)] Nicoletti, G., Lodi, R., Condino, F., Tonon, C., Fera, F., Malucelli, E., Manners, D., Zappia, M., Morgante, L., Barone, P., *et al.* (2006). Apparent diffusion coefficient measurements of the middle cerebellar peduncle differentiate the parkinson variant of msa from parkinson's disease and progressive supranuclear palsy. *Brain*, 129(10):2679–2687.

[Nicoletti *et al.* (2008)] Nicoletti, G., Tonon, C., Lodi, R., Condino, F., Manners, D., Malucelli, E., Morelli, M., Novellino, F., Paglionico, S., Lanza, P., *et al.* (2008). Apparent diffusion coefficient of the superior cerebellar peduncle differentiates progressive supranuclear palsy from parkinson's disease. *Movement disorders*, 23(16):2370–2376.

[Nigro *et al.* (2015)] Nigro, S., Passamonti, L., Riccelli, R., Toschi, N., Rocca, F., Valentino, P., Nisticò, R., Fera, F., and Quattrone, A. (2015). Structural "connectomic" alterations in the limbic system of multiple sclerosis patients with major depression. *Multiple Sclerosis Journal*, 21(8):1003–1012.

[Noseworthy *et al.* (2000)] Noseworthy, J. H., Lucchinetti, C., Rodriguez, M., and Weinshenker, B. G. (2000). Multiple sclerosis. *New England Journal of Medicine*, 343(13):938–952. PMID: 11006371.

[Nygaard *et al.* (2015)] Nygaard, G. O., Celius, E. G., Sigrid, A., Sowa, P., Gustavsen, M. W., Fjell, A. M., Landrø, N. I., Walhovd, K. B., and Harbo, H. F. (2015). A longitudinal study of disability, cognition and gray matter atrophy in early multiple sclerosis patients according to evidence of disease activity. *PloS one*, 10(8):e0135974.

[O'Donnell *et al.* (2006)] O'Donnell, L. J., Kubicki, M., Shenton, M. E., Dreusicke, M. H., Grimson, W. E. L., and Westin, C. F. (2006). A method for clustering white matter fiber tracts. *AJNR. American journal of neuroradiology*, 27:1032–1036.

[O'Donnell *et al.* (2009)] O'Donnell, L. J., Westin, C.-F., and Golby, A. J. (2009). Tract-based morphometry for white matter group analysis. *NeuroImage*, 45:832–844.

[Ordóñez *et al.* (2008)] Ordóñez, P., desJardins, M., Feltes, C., Lehmann, C. U., and Fackler, J. C. (2008). Visualizing multivariate time series data to detect specific medical conditions. In *AMIA 2008, American Medical Informatics Association Annual Symposium, Washington, DC, USA, November 8-12, 2008*. AMIA.

[Ortega-Martorell *et al.* (2012)] Ortega-Martorell, S., Lisboa, P. J. G., Vellido, A., Simões, R. V., Pumarola, M., Julià-Sapé, M., and Arús, C. (2012). Convex non-negative matrix factorization for brain tumor delimitation from mrsi data. *PloS one*, 7:e47824.

[Panagiotaki *et al.* (2012)] Panagiotaki, E., Schneider, T., Siow, B., Hall, M. G., Lythgoe, M. F., and Alexander, D. C. (2012). Compartment models of the diffusion mr signal in brain white matter: a taxonomy and comparison. *Neuroimage*, 59(3):2241–2254.

[Pantano *et al.* (2002)] Pantano, P., Mainero, C., Iannetti, G. D., Caramia, F., Di Legge, S., Piattella, M. C., Pozzilli, C., Bozzao, L., and Lenzi, G. L. (2002). Contribution of corticospinal tract damage to cortical motor reorganization after a single clinical attack of multiple sclerosis. *NeuroImage*, 17:1837–1843.

[Polman *et al.* (2005)] Polman, C. H., Reingold, S. C., Edan, G., Filippi, M., Hartung, H.-P., Kappos, L., Lublin, F. D., Metz, L. M., McFarland, H. F., O'Connor, P. W., *et al.* (2005). Diagnostic criteria for multiple sclerosis: 2005 revisions to the "mcdonald criteria". *Annals of neurology*, 58(6):840–846.

[Polman *et al.* (2011)] Polman, C. H., Reingold, S. C., Banwell, B., Clanet, M., Cohen, J. A., Filippi, M., Fujihara, K., Havrdova, E., Hutchinson, M., Kappos, L., *et al.* (2011). Diagnostic criteria for multiple sclerosis: 2010 revisions to the mcdonald criteria. *Annals of neurology*, 69(2):292–302.

[Powers (2011)] Powers, D. M. (2011). Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation.

[Prados *et al.* (2010)] Prados, F., Boada, I., Prats-Galino, A., Martín-Fernández, J. A., Feixas, M., Blasco, G., Puig, J., and Pedraza, S. (2010). Analysis of new diffusion tensor imaging anisotropy measures in the three-phase plot. *Journal of magnetic resonance imaging : JMRI*, 31:1435–1444.

[Prasad *et al.* (2014)] Prasad, G., Joshi, S. H., Jahanshad, N., Villalon-Reina, J., Aganj, I., Lenglet, C., Sapiro, G., McMahon, K. L., de Zubicaray, G. I., Martin, N. G., Wright, M. J., Toga, A. W., and Thompson, P. M. (2014). Automatic clustering and population analysis of white matter tracts using maximum density paths. *NeuroImage*, 97:284–295.

[Preston and Shapiro (2005)] Preston, D. and Shapiro, B. (2005). Electromyography and neuromuscular disorders clinical-electrophysiologic correlations. boston oxford johannesburg melbourne new delhi singapore; basic nerve conduction study. *Butterworth-Heinemann*, 1:4–10.

[Pulizzi *et al.* (2007)] Pulizzi, A., Rovaris, M., Judica, E., Sormani, M. P., Martinelli, V., Comi, G., and Filippi, M. (2007). Determinants of disability in multiple sclerosis at various disease stages: a multiparametric magnetic resonance study. *Archives of neurology*, 64(8):1163–1168.

[Reich *et al.* (2007)] Reich, D. S., Smith, S. A., Zackowski, K. M., Gordon-Lipkin, E. M., Jones, C. K., Farrell, J. A. D., Mori, S., van Zijl, P. C. M., and Calabresi, P. A. (2007). Multiparametric magnetic resonance imaging analysis of the corticospinal tract in multiple sclerosis. *NeuroImage*, 38:271–279.

[Reich *et al.* (2008)] Reich, D. S., Zackowski, K. M., Gordon-Lipkin, E. M., Smith, S. A., Chodkowski, B. A., Cutter, G. R., and Calabresi, P. A. (2008). Corticospinal tract abnormalities are associated with weakness in multiple sclerosis. *AJNR. American journal of neuroradiology*, 29:333–339.

[Reuter *et al.* (2012)] Reuter, M., Schmansky, N. J., Rosas, H. D., and Fischl, B. (2012). Within-subject template estimation for unbiased longitudinal image analysis. *NeuroImage*, 61:1402–1418.

[Rocca *et al.* (2014)] Rocca, M. A., Parisi, L., Pagani, E., Copetti, M., Rodegher, M., Colombo, B., Comi, G., Falini, A., and Filippi, M. (2014). Regional but not global brain damage contributes to fatigue in multiple sclerosis. *Radiology*, 273(2):511–520.

[Rovaris *et al.* (2005)] Rovaris, M., Gass, A., Bammer, R., Hickman, S., Ciccarelli, O., Miller, D., and Filippi, M. (2005). Diffusion mri in multiple sclerosis. *Neurology*, 65(10):1526–1532.

[Rovira *et al.* (2013)] Rovira, A., Auger, C., and Alonso, J. (2013). Magnetic resonance monitoring of lesion evolution in multiple sclerosis. *Therapeutic advances in neurological disorders*, 6:298–310.

[Rubinov and Sporns (2010)] Rubinov, M. and Sporns, O. (2010). Complex network measures of brain connectivity: uses and interpretations. *Neuroimage*, 52(3):1059–1069.

[Rueda *et al.* (2008)] Rueda, F., Hygino Jr, L. C., Domingues, R. C., Vasconcelos, C. C., Papais-Alvarenga, R. M., and Gasparetto, E. L. (2008). Diffusion tensor mr imaging evaluation of the corpus callosum of patients with multiple sclerosis. *Arquivos de neuro-psiquiatria*, 66(3A):449–453.

[Sajda *et al.* (2004)] Sajda, P., Du, S., Brown, T. R., Stoyanova, R., Shungu, D. C., Mao, X., and Parra, L. C. (2004). Nonnegative matrix factorization for rapid recovery of constituent spectra in magnetic resonance chemical shift imaging of the brain. *IEEE transactions on medical imaging*, 23:1453–1465.

[Sauwen *et al.* (2015)] Sauwen, N., Sima, D. M., Van Cauter, S., Veraart, J., Leemans, A., Maes, F., Himmelreich, U., and Van Huffel, S. (2015). Hierarchical non-negative matrix factorization to characterize brain tumor heterogeneity using multi-parametric mri. *NMR in biomedicine*, 28:1599–1624.

[Server *et al.* (2014)] Server, A., Graff, B. A., Josefsen, R., Orheim, T. E., Schellhorn, T., Nordhøy, W., and Nakstad, P. H. (2014). Analysis of diffusion tensor imaging metrics for gliomas grading at 3t. *European journal of radiology*, 83(3):e156–e165.

[Shen *et al.* (2014)] Shen, Y., Bai, L., Gao, Y., Cui, F., Tan, Z., Tao, Y., Sun, C., and Zhou, L. (2014). Depressive symptoms in multiple sclerosis from an in vivo study with tbss. *BioMed research international*, 2014.

[Shuman *et al.* (2013)] Shuman, D. I., Narang, S. K., Frossard, P., Ortega, A., and Vandergheynst, P. (2013). The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Processing Magazine*, 30(3):83–98.

[Sigal *et al.* (2012)] Sigal, T., Shmuel, M., Mark, D., Gil, H., and Anat, A. (2012). Diffusion tensor imaging of corpus callosum integrity in multiple sclerosis: correlation with disease variables. *Journal of Neuroimaging*, 22(1):33–37.

[Smith *et al.* (2006)] Smith, S. M., Jenkinson, M., Johansen-Berg, H., Rueckert, D., Nichols, T. E., Mackay, C. E., Watkins, K. E., Ciccarelli, O., Cader, M. Z., Matthews, P. M., *et al.* (2006). Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. *Neuroimage*, 31(4):1487–1505.

[Sorber *et al.* (2013)] Sorber, L., Barel, M. V., and Lathauwer, L. D. (2013). Optimization-based algorithms for tensor decompositions: Canonical polyadic decomposition, decomposition in rank-$(l_r, l_r, 1)$ terms, and a new generalization. *SIAM Journal on Optimization*, 23(2):695–720.

[Sorber *et al.* (2015)] Sorber, L., Barel, M. V., and Lathauwer, L. D. (2015). Structured data fusion. *IEEE Journal of Selected Topics in Signal Processing*, 9(4):586–600.

[Stamile *et al.* (2015)a] Stamile, C., Cauteruccio, F., Terracina, G., Ursino, D., Kocevar, G., and Sappey-Marinier, D. (2015a). A model-guided string-based approach to white matter fiber-bundles extraction. In *International Conference on Brain Informatics and Health*, pages 135–144. Springer.

[Stamile *et al.* (2015)b] Stamile, C., Kocevar, G., Cotton, F., Hannoun, S., Durand-Dubief, F., Frindel, C., Rousseau, D., and Sappey-Marinier, D. (2015b). A longitudinal model for variations detection in white matter fiber-bundles. In *Proc. Signals and Image Processing (IWSSIP) 2015 Int. Conf. Systems*, pages 57–60.

[Stamile *et al.* (2015)c] Stamile, C., Kocevar, G., Durand-Dubief, F., Frindel, C., Cotton, F., Hannoun, S., Rousseau, D., and Sappey-Marinier, D. (2015c). A random forest based method for classification of white matter fiber-bundles in multiple sclerosis. In *GRETSI, 8-11 September 2015, Lyon, France.*

[Stamile *et al.* (2016)a] Stamile, C., Kocevar, G., Cotton, F., Durand-Dubief, F., Hannoun, S., Frindel, C., Guttmann, C. R. G., Rousseau, D., and Sappey-Marinier, D. (2016a). A sensitive and automatic white matter fiber tracts model for longitudinal analysis of diffusion tensor images in multiple sclerosis. *PloS one*, 11:e0156405.

[Stamile *et al.* (2016)b] Stamile, C., Kocevar, G., Cotton, F., Maes, F., Sappey-Marinier, D., and Huffel, S. V. (2016b). Multi-parametric non-negative matrix factorization for longitudinal variations detection in white matter fiber-bundles. *IEEE Journal of Biomedical and Health Informatics*, PP(99):1.

[Stamile *et al.* (2017)] Stamile, C., Kocevar, G., Cotton, F., and Sappey-Marinier, D. (2017). A genetic algorithm-based model for longitudinal changes detection in white matter fiber-bundles of patient with multiple sclerosis. *Computers in Biology and Medicine*.

[Steenwijk *et al.* (2014)] Steenwijk, M. D., Daams, M., Pouwels, P. J. W., Balk, L. J., Tewarie, P. K., Killestein, J., Uitdehaag, B. M. J., Geurts, J. J. G., Barkhof, F., and Vrenken, H. (2014). What explains gray matter atrophy in long-standing multiple sclerosis? *Radiology*, 272:832–842.

[Stejskal and Tanner (1965)] Stejskal, E. O. and Tanner, J. E. (1965). Spin diffusion measurements: spin echoes in the presence of a time-dependent field gradient. *The journal of chemical physics*, 42(1):288–292.

[Sternberg *et al.* (2014)] Sternberg, E., Lipton, M., and Burns, J. (2014). Utility of diffusion tensor imaging in evaluation of the peritumoral region in patients with primary and metastatic brain tumors. *American Journal of Neuroradiology*, 35(3):439–444.

[Stoll *et al.* (2002)] Stoll, G., Jander, S., and Myers, R. R. (2002). Degeneration and regeneration of the peripheral nervous system: from augustus waller's observations to neuroinflammation. *Journal of the peripheral nervous system : JPNS*, 7:13–27.

[Sun (2013)] Sun, S. (2013). A survey of multi-view machine learning. *Neural Computing and Applications*, 23(7-8):2031–2038.

[Thomsen *et al.* (1987)] Thomsen, C., Henriksen, O., and Ring, P. (1987). In vivo measurement of water self diffusion in the human brain by magnetic resonance imaging. *Acta Radiologica*, 28(3):353–361.

[Tillema *et al.* (2016)] Tillema, J.-M., Weigand, S. D., Mandrekar, J., Shu, Y., Lucchinetti, C. F., Pirko, I., and Port, J. D. (2016). In vivo detection of connectivity between cortical and white matter lesions in early ms. *Multiple Sclerosis Journal*, page 1352458516671027.

[Tournier *et al.* (2004)] Tournier, J.-D., Calamante, F., Gadian, D. G., and Connelly, A. (2004). Direct estimation of the fiber orientation density function from diffusion-weighted mri data using spherical deconvolution. *NeuroImage*, 23(3):1176–1185.

[Tournier *et al.* (2011)] Tournier, J.-D., Mori, S., and Leemans, A. (2011). Diffusion tensor imaging and beyond. *Magnetic resonance in medicine*, 65(6):1532–1556.

[Tournier *et al.* (2012)] Tournier, J., Calamante, F., Connelly, A., *et al.* (2012). Mrtrix: diffusion tractography in crossing fiber regions. *International Journal of Imaging Systems and Technology*, 22(1):53–66.

[Tuch (2004)] Tuch, D. S. (2004). Q-ball imaging. *Magnetic resonance in medicine*, 52(6):1358–1372.

[Tuch *et al.* (2002)] Tuch, D. S., Reese, T. G., Wiegell, M. R., Makris, N., Belliveau, J. W., and Wedeen, V. J. (2002). High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity. *Magnetic resonance in medicine*, 48(4):577–582.

[Vandermosten *et al.* (2015)] Vandermosten, M., Vanderauwera, J., Theys, C., De Vos, A., Vanvooren, S., Sunaert, S., Wouters, J., and Ghesquière, P. (2015). A dti tractography study in pre-readers at risk for dyslexia. *Developmental cognitive neuroscience*, 14:8–15.

[Varentsova *et al.* (2014)] Varentsova, A., Zhang, S., and Arfanakis, K. (2014). Development of a high angular resolution diffusion imaging human brain template. *NeuroImage*, 91:177–186.

[Vervliet *et al.* (2016)] Vervliet, N., Debals, O., Sorber, L., Van Barel, M., and De Lathauwer, L. (2016). Tensorlab 3.0.

[Voříšek and Syková (1997)] Voříšek, I. and Syková, E. (1997). Evolution of anisotropic diffusion in the developing rat corpus callosum. *Journal of neurophysiology*, 78(2):912–919.

[Wedeen *et al.* (2000)] Wedeen, V., Reese, T., Tuch, D., Weigel, M., Dou, J., Weiskoff, R., and Chessler, D. (2000). Mapping fiber orientation spectra in cerebral white matter with fourier-transform diffusion mri. In *Proceedings of the 8th Annual Meeting of ISMRM, Denver*, page 82.

[Werring *et al.* (1999)] Werring, D., Clark, C., Barker, G., Thompson, A., and Miller, D. (1999). Diffusion tensor imaging of lesions and normal-appearing white matter in multiple sclerosis. *Neurology*, 52(8):1626–1626.

[Wilson *et al.* (2003)] Wilson, M., Tench, C. R., Morgan, P. S., and Blumhardt, L. D. (2003). Pyramidal tract mapping by diffusion tensor magnetic resonance imaging in multiple sclerosis: improving correlations with disability. *Journal of neurology, neurosurgery, and psychiatry*, 74:203–207.

[Wottschel *et al.* (2015)] Wottschel, V., Alexander, D., Kwok, P., Chard, D., Stromillo, M., De Stefano, N., Thompson, A., Miller, D., and Ciccarelli, O. (2015). Predicting outcome in clinically isolated syndrome using machine learning. *NeuroImage: Clinical*, 7:281–287.

[Wu and Alvarez (2011)] Wu, G. F. and Alvarez, E. (2011). The immunopathophysiology of multiple sclerosis. *Neurologic clinics*, 29(2):257–278.

[Yeatman *et al.* (2012)] Yeatman, J. D., Dougherty, R. F., Myall, N. J., Wandell, B. A., and Feldman, H. M. (2012). Tract profiles of white matter properties: automating fiber-tract quantification. *PloS one*, 7(11):e49790.

[Zhang *et al.* (2008)] Zhang, S., Correia, S., and Laidlaw, D. H. (2008). Identifying white-matter fiber bundles in dti data using an automated proximity-based fiber-clustering method. *IEEE transactions on visualization and computer graphics*, 14:1044–1053.

[Zhang *et al.* (2012)] Zhang, H., Schneider, T., Wheeler-Kingshott, C. A., and Alexander, D. C. (2012). Noddi: practical in vivo neurite orientation dispersion and density imaging of the human brain. *Neuroimage*, 61(4):1000–1016.

[Zhang *et al.* (2013)] Zhang, Y., Wan, S., and Zhang, X. (2013). Geniculocalcarine tract disintegration after ischemic stroke: a diffusion tensor imaging study. *American Journal of Neuroradiology*, 34(10):1890–1894.

[Zou *et al.* (2004)] Zou, K. H., Warfield, S. K., Bharatha, A., Tempany, C. M. C., Kaus, M. R., Haker, S. J., Wells, W. M., Jolesz, F. A., and Kikinis, R. (2004). Statistical validation of image segmentation quality based on a spatial overlap index. *Academic radiology*, 11:178–189.

[da Cruz *et al.* (2011)] da Cruz, L. C. H., Batista, R. R., Domingues, R. C., and Barkhof, F. (2011). Diffusion magnetic resonance imaging in multiple sclerosis. *Neuroimaging Clinics of North America*, 21(1):71–88.

[Keller *et al.* (2009)] Keller, O., Kopelowitz, T., and Lewenstein, M. (2009). On the longest common parameterized subsequence. *Theoretical Computer Science*, 410(51):5347–5353.

[Kolda and Bader (2009)] Kolda, T. G. and Bader, B. W. (2009). Tensor decompositions and applications. *SIAM Review*, 51(3):455–500.

Curriculum Vitae

Stamile Claudio

7 September 1989, Cosenza, Italy
Nationality: Italian
Adresse: 12 Bis rue Girié - 69003 Lyon

Mobile: +33 (0)6 21 45 07 61
E-Mail: claudiostamile@gmail.com

Formation:

PhD	**Biomedical Engineering** Université Claude Bernard Lyon 1 Ecole Doctorale Interdisciplinaire Sciences Santé CREATIS UMR 5220 CNRS U1206 INSERM	***2013-2017***
PhD	**Electronic Engineering** Katholieke Universiteit Leuven Arenberg Doctoral School ESAT, STADIUS	***2013-2017***
Master Degree	**Computer Science** University of Calabria, Arcavacata, Italy	***2011-2013***
Bachelor Degree	**Computer Science** University of Calabria, Arcavacata, Italy	***2009-2011***

Research Experience:

- **PhD Thesis** ***2013-2017***
  "Development of Unsupervised Methods for Longitudinal Fiber-Bundle Analysis"
  Ecole Doctorale Interdisciplinaire Sciences Santé - Arenberg Doctoral School
  Laboratoire CREATIS UMR CNRS 5220 INSERM 1206 - ESAT, STADIUS
  Université Claude Bernard Lyon 1 - Katholieke Universiteit Leuven
  Supervisors: Dr. D. Sappey-Marinier, Prof. Dr. Ir. S. Van Huffel

- **Stage Master Degree** ***08/2012 - 10/2012***
  Definition of innovative models for the prediction of cerebral edema evolution
  Center of Neurological Imaging (CNI) and Multiple Sclerosis Center at Brigham and Women's Hospital, Harvard Medical School, Boston USA - University of Calabria
  Supervisors: Prof. F. Calimeri, Dr. C. Guttmann

- **Stage Bachelor Degree** ***2011 - 2012***
  Design and implementation of a system for automatic analysis of biomedical images
  University of Calabria - "Sapienza" University of Rome
  Supervisors : Prof. F. Calimeri, Dr. R. Umeton, M. Cavallari, F. Orzi

Teaching Experience:
University of Calabria, Italy A.Y. 2011/2012
Academic Tutor, supporting TAs and Lecturers for classes in the following areas:

- Algorithms and data structures
- Formal languages and compilers
- Introduction to Databases
- Object oriented programming
- Advance computer programming
- Computational Logic

Patents:

- C. Stamile, M. Cavallari, R. Umeton, F. Calimeri, F. Orzi, **"Smart Fundus Camera - a portable medical device for quantitative retinal image analysis through a smartphone,"** U.S. Patent 9,468,377. October 18, 2016.
- C. Stamile, M. Cavallari, R. Umeton, F. Calimeri, F. Orzi, **"Dispositivo medico portatile e metodo per l'acquszizione di immagini della retina e esecuzione di analisi quantitative sull'immagine della retina,"** IT Patent 1423628. August 22, 2016.

Skills:
Theoretical:

- **Computer science:** Complexity theory, Data Structures, Parallel Programming, Graph Theory, Algorithm Design, Logic Programming, OLAP
- **Machine Learning:**
  - **Supervised**: Random Three, Random Forest, SVM, Bayesian classifier, Convolutional Neural Network, Evolutive Neural Network (NEAT, HyperNeat), Transfer Learning
  - **Unsupervised**: K-Means/Medoids, Tensor and Matrix Factorization, Hierarchical clustering, Density based clustering
- **Optimization:** Linear and Non Linear Least Square, Convex Optimization, Genetic Algorithm, Gradient Descending, Trust-Regions, Newton and Quasi Newton Methods

Practical:

- **Programming languages:** C++, Java, Python, MATLAB, Bash, Datalog, DLV

- **Technologies:**
  - **GUI Libraries:** PyQt, Qt, Electron
  - **Data Mining:** WEKA, scikit-learn, MATLAB
  - **Data Visualization:** gnuplot, Tableau, Gephi
  - **Natural Language Processing:** NLTK
  - **Numerical Optimization:** CPLEX, Ceres, numpy, scipy, Theano
  - **Parallel Programming:** Open MPI, OpenMP, CUDA
  - **Framework:** Hadoop, Spring, Apache Tomcat
- **Operative systems:** Mac OS X, Linux, Windows
- **Office Tools:** Ms Excel, Ms Power Point, Ms Word, Open Office, LaTeX, reveal.js

Languages:

- **Mother tongue:** Italian
- **Professional Level:** English, French
- **Entry Level:** Spanish

Awards:

- Marie Curie Fellowship. 2013
- Best Student Presentation Prize, 1st place. Nijmegen, The Netherlands. 2nd March 2016
- ERC Fellowship. 2016

Supervisor Activities

- Thesis: *Definition of a white matter genetic atlas based based on Diffusion Tensor Imaging.* Student: Ferdinando Primerano. Supervisors: Francesco Calimeri, Charles Guttman, Claudio Stamile
- Thesis: *A web application for advanced image analysis.* Student: Pierpaolo Rametta. Supervisors: Francesco Calimeri, Charles Guttman, Claudio Stamile
- Thesis: *A web-based framework for the design of novel tools for the analysis of biomedical images.* Student: Aldo Marzullo. Supervisors: Francesco Calimeri, Claudio Stamile
- Thesis: *Advanced techniques for biomedical images elaboration.* Student: Salvatore Isabella. Supervisors: Francesco Calimeri, Claudio Stamile

- Thesis: *Mobile porting of ImageJ.* Student: Pierangela Bruno. Supervisors: Francesco Calimeri, Claudio Stamile
- Thesis: *A new method for MRI defacing.* Student: Daniele Pezzolla. Supervisors: Francesco Calimeri, Claudio Stamile
- Thesis: *Reconstruction and analysis of retinal blood vessels using S-Rep.* Student: Luca Surace. Supervisors: Francesco Calimeri, Claudio Stamile
- Thesis: *Segmentation of images using neural networks.* Student: Aldo Marzullo. Supervisors: Francesco Calimeri, Claudio Stamile

List of Publications

Journal Papers

1. C. Stamile, G. Kocevar, F. Cotton, F. Durand-Dubief, S. Hannoun, C. Frindel, C. R.G. Guttmann, D. Rousseau, D. Sappey-Marinier, **"Detection of Longitudinal DTI Changes in Multiple Sclerosis Patients Based on Sensitive WM Fiber Modeling,"** PLoS ONE 11(5): e0156405. doi:10.1371/journal.pone.0156405, 2016.

2. C. Stamile, G. Kocevar, F. Cotton, F. Maes, D. Sappey-Marinier, S. Van Huffel **"Multiparametric Nonnegative Matrix Factorization for Longitudinal Variations Detection in White-Matter Fiber Bundles,"** *IEEE Journal of Biomedical and Health Informatics, 2016, in press.*

3. C. Stamile, G. Kocevar, F. Cotton, D. Sappey-Marinier, **"A Genetic Algorithm-Based Model for Longitudinal Changes Detection in White Matter Fiber-Bundles of Patient with Multiple Sclerosis,"** *Computers in Biology and Medicine, vol. 84, pp. 182-188, 2017.*

4. F. Cauteruccio*, C. Stamile*, G. Terracina, D. Ursino, D. Sappey-Marinier, **"A Model-Guided String-Based Approach to White Matter Fiber-Bundles Extraction,"** *Computers in Biology and Medicine, vol. 77, pp. 64-75, 2016.*

5. G. Kocevar*, C. Stamile*, S. Hannoun, F. Cotton, S. Vukusic, F. Durand-Dubief, D. Sappey-Marinier, **"Graph Theory-Based Brain Connectivity for Automatic Classification of Multiple Sclerosis Clinical Courses,"** Front. Neurosci. 10:478. doi: 10.3389/fnins.2016.00478, 2016.

6. G. Kocevar, C. Stamile, S. Hannoun, J.-A. Roch, F. Durand-Dubief, S. Vukusic, F. Cotton, D. Sappey-Marinier, **"Weekly Follow up of Acute Lesions in Three Early Multiple Sclerosis Patients using MR Spectroscopy and Diffusion,"** Journal of Neuroradiology, 2017, *in press.*

7. A. Ion-Margineanu, G. Kocevar, C. Stamile, D. M. Sima, F. Durand-Dubief, S. Van Huffel, D. Sappey-Marinier, **"Machine learning approach for classifying Multiple Sclerosis courses by combining clinical data with lesion loads**

and Magnetic Resonance metabolic features," Front. Neurosci. 11:398. doi: 10.3389/fnins.2017.00398, 2017.

8. F. Nusbaum, S. Hannoun, G. Kocevar, C. Stamile, P. Fourneret, O. Revol, D. Sappey-Marinier, **"Hemispheric Differences in White Matter Microstructure between two Profiles of Children with High Intelligence Quotient versus Controls: A Tract-Based Spatial Statistics Study,"** Front. Neurosci. 11:173. doi: 10.3389/fnins.2017.00173, 2017.
9. J. Yankam Njiwa, N. Costes, C. Bouillot, S. Bouvard, S. Fieux, G Becker, E. Levigoureux, G. Kocevar, C. Stamile, J.B. Langlois, R. Bolbos, C. Bonnet, L. Bezin, L. Zimmer, A. Hammers, **"Quantitative longitudinal imaging of activated microglia as a marker of inflammation in the pilocarpine rat model of epilepsy using [11C]-(R)-PK11195 PET and MRI,"** Journal of Cerebral Blood Flow & Metabolism, doi: 10.1177/0271678X16653615, 2016.

*These authors contributed equally to this work

Peer Reviewed Conference Papers

1. C. Stamile, G. Kocevar, S. Hannoun, F. Durand-Dubief, D. Sappey-Marinier, **"A Graph Based Classification Method for Multiple Sclerosis Clinical Form Using Support Vector Machine,"** *in: Machine Learning Meets Medical Imaging*, LNCS vol. 9487, pp. 57-64, 2015.
2. C. Stamile*, F. Cauteruccio*, G. Terracina, D. Ursino, G. Kocevar, D. Sappey-Marinier, **"A Model-Guided String-Based Approach to White Matter Fiber-Bundles Extraction,"** *in: Brain Informatics and Health*, LNCS vol. 9250, pp.135-144, 2015.
3. C. Stamile, G. Kocevar, S. Hannoun, C. Frindel, F. Cotton, F. Durand-Dubief, D. Rousseau, D. Sappey-Marinier, **"A Longitudinal Model for Variations Detection in White Matter Fiber-Bundles,"** *in: International Conference on Systems, Signals and Image Processing (IWSSIP)*, pp. 57-60, 2015.
4. C. Stamile, F. Cotton, F. Maes, D. Sappey-Marinier, S. Van Huffel, **"White Matter Fiber-Bundle Analysis Using Non-Negative Tensor Factorization,"** *in: Image Analysis and Recognition (ICIAR)*, LNCS vol. 9730, pp. 650-657, 2016.
5. C. Stamile, G. Kocevar, S. Hannoun, C. Frindel, F. Cotton, F. Durand-Dubief, D. Rousseau, D. Sappey-Marinier, **"A Random Forest Based Method for classification of White Matter Fiber-Bundles in Multiple Sclerosis,"** *in: 25^{e} Groupe d'Etudes du Traitement du Signal et des Images (GRETSI), 2015.*

6. C. Stamile, F. Cotton, D. Sappey-Marinier, S. Van Huffel, **“Longitudinal Neuroimaging Analysis Using Non-Negative Matrix Factorization,”** *in: The 12th International Conference on Signal Image Technology & Internet Systems (SITIS)*, pp. 55-61, 2016.

7. F. Cauteruccio*, C. Stamile*, G. Terracina, D. Ursino, D. Sappey-Marinier, **“Integrating QuickBundle with a string-based fiber representation for extracting anatomically coherent WM fiber-bundles,”** *in: Image Analysis and Recognition (ICIAR)*, LNCS vol. 9730, pp. 633-641, 2016.

8. F. Cauteruccio*, C. Stamile*, G. Terracina, D. Ursino, D. Sappey-Marinier, **“An Automated String-Based Approach to White Matter Fiber-Bundles Clustering,”** *in: 2015 International Joint Conference on Neural Networks (IJCNN)*, pp. 1-8, 2015.

9. F. Cauteruccio*, C. Stamile*, G. Terracina, D. Ursino, D. Sappey-Marinier, **“Integrating QuickBundles into a model-guided approach for extracting “anatomically-coherent” and “symmetry-aware” White Matter fiber-bundles,”** *in: 26th Italian Workshop on Neural Networks (WIRN), 2016*, to appear.

10. A. Marzullo, C. Stamile, G. Terracina, F. Calimeri, S. Van Huffel **“A Tensor-based Mutation Operator for Neuroevolution of Augmenting Topologies (NEAT),”** *in: IEEE Congress on Evolutionary Computation (CEC)*, pp. 681-687, 2017.

11. G. Kocevar, C. Stamile, S. Hannoun, F. Cotton, F. Durand-Dubief, D. Rousseau, D. Sappey-Marinier, **“Connectivité Structurelle par Graphes : Application á la Sclérose en Plaques,”** *in: 25e Groupe d’Etudes du Traitement du Signal et des Images (GRETSI), 2015.*

12. F. Calimeri, A. Marzullo, C. Stamile, G. Terracina, **“Biomedical Data Augmentation using Generative Adversarial Neural Networks,”** *in: 26th International Conference on Artificial Neural Networks (ICANN), 2017*, to appear.

13. A. Ion-Margineanu, G. Kocevar, C. Stamile, D. M. Sima, F. Durand-Dubief, S. Van Huffel, D. Sappey-Marinier, **“A comparison of Machine Learning approaches for classifying Multiple Sclerosis courses using MRSI and brain segmentations,”** *in: 26th International Conference on Artificial Neural Networks (ICANN), 2017*, to appear.

14. F. Calimeri, A. Marzullo, C. Stamile, G. Terracina, **“Blood Vessel Segmentation using Hypercube NeuroEvolution of Augmenting Topologies (HyperNeat),”** *in: 27th Italian Workshop on Neural Networks (WIRN), 2017*, to appear.

15. F. Calimeri, A. Marzullo, C. Stamile, G. Terracina, **"Optic Disc Detection using Fine Tuned Convolutional Neural Networks,"** *in: The 12th International Conference on Signal Image Technology & Internet Systems (SITIS)*, pp. 69-75, 2016.

16. F. Calimeri, A. Marzullo, M. Caracciolo, C. Stamile, **"BioHIPI: Biomedical Hadoop Image Processing Interface,"** *in: The Third International Conference on Machine Learning, Optimization and Big Data (MOD), 2017*, to appear.

*These authors contributed equally to this work

Peer Reviewed Conference Abstracts

1. C. Stamile, G. Kocevar, F. Cotton, F. Maes, D. Sappey-Marinier, S. Van Huffel, **"Non-Negative Matrix Factorization for White-Matter Fiber-Bundles Longitudinal Analysis,"** ESMRMB, Vienna, 2016.

2. C. Stamile, G. Kocevar, F. Durand-Dubief, F. Cotton, C. Frindel, S. Hannoun, D. Sappey-Marinier, **"Classification of Multiple Sclerosis Clinical Forms Using DTI Fiber-Bundles Information,"** ESMRMB, Edinburgh, 2015.

3. C. Stamile, G. Kocevar, F. Cotton, F. Durand-Dubief, S. Hannoun, C. Frindel, D. Rousseau, D. Sappey-Marinier, **"Detection of Longitudinal DTI Changes in Multiple Sclerosis Patients Based on Sensitive WM Fiber Modeling,"** ISMRM, Toronto, 2015.

4. C. Stamile, G. Kocevar, F. Durand-Dubief, F. Cotton, C. Frindel, S. Hannoun, D. Sappey-Marinier, **"Multiple Sclerosis Clinical Classification Based on DTI Fiber Analysis,"** ISMRM, Toronto, 2015.

5. C. Stamile, G. Kocevar, F. Cotton, F. Durand-Dubief, S. Hannoun, C. Frindel, D. Rousseau, D. Sappey-Marinier, **"Détection des changements longitudinaux rapides chez les patients SEP par une modélisation des faisceaux de substance blanche,"** SFRMBM, Grenoble, 2015.

6. C. Stamile, G. Kocevar, F. Durand-Dubief, F. Cotton, C. Frindel, S. Hannoun, D. Sappey-Marinier, **"Classification des différentes formes cliniques de Sclérose en Plaques basée sur l'Analyse des Faisceaux de Substance Blanche,"** SFRMBM, Grenoble, 2015.

7. G. Kocevar, C. Stamile, I. Suprano, S. Hannoun, P. Fourneret, O. Revol, F. Nusbaum, D. Sappey-Marinier, **"Analyse de la Connectivité Structurelle Cérébrale en fonction du Quotient Intellectuel chez l'Adolescent,"** SFRMBM, Bordeaux, 2017.

8. G. Kocevar, C. Stamile, F. Cotton, F. Durand-Dubief, D. Sappey-Marinier, **"Characterization of Multiple Sclerosis Forms Through Fiber-Bundle Profile Analysis,"** ARSEP, Paris, 2017.
9. G. Kocevar, C. Stamile, S. Hannoun, F. Cotton, S. Vukusic, F. Durand-Dubief, D. Sappey-Marinier, **"Graph-Theory Based Classification of Multiple Sclerosis Clinical Courses,"** ARSEP MRI Workshop, Paris, 2017.
10. D. Pezzolla, C. Stamile, F. Durand-Dubief, D. Sappey-Marinier, F. Calimeri **"An Automatic Algorithm for MRI Anonymization Based on Face Features Detection,"** ESMRMB, Vienna, 2016.
11. G. Kocevar, C. Stamile, S. Hannoun, F. Cotton, F. Durand-Dubief, D. Sappey-Marinier, **"Characterization of Brain Structural Connectivity in Different Clinical Forms of Multiple Sclerosis Patients,"** ARSEP, Paris, 2016.
12. G. Kocevar, C. Stamile, S. Hannoun, F. Cotton, S. Vukusic, F. Durand-Dubief, D. Sappey-Marinier, **"Classification of Multiple Sclerosis Clinical Forms Using Structural Connectome,"** ESMRMB, Edinburgh, 2015.
13. G. Kocevar, C. Stamile, S. Hannoun, F. Cotton, F. Durand-Dubief, D. Sappey-Marinier, **"Characterization of DTI Brain Connectivity in Different Clinical Forms of Multiple Sclerosis Patients Based on Graph Theory,"** ISMRM, Toronto, 2015.
14. G. Kocevar, C. Stamile, S. Hannoun, F. Durand-Dubief, F. Cotton, D. Sappey-Marinier, **"Le Myo-Inositol, un marqueur métabolique de l'évolution rapide des lésions de Sclérose en Plaques,"** SFRMBM, Grenoble, 2015.
15. G. Kocevar, F. Durand-Dubief, C. Stamile, S. Hannoun, F. Cotton, D. Sappey-Marinier, **"Analyse de la connectivité structurelle cérébrale par la théorie des graphes : une nouvelle caractérisation des formes cliniques de sclérose en plaques,"** JNLF, Toulouse, 2017.
16. I. Suprano, C. Delon-Martin, G. Kocevar, C. Stamile, S. Hannoun, P. Fourneret, O. Revol, F. Nusbaum, D. Sappey-Marinier, **"Réorganisation Topologique des Réseaux Fonctionnels dans Deux Profils d'Enfants à Haut Potentiel Intellectuel,"** SFRMBM, Bordeaux, 2017.
17. G. Kocevar, C. Stamile, F. Cotton, F. Durand-Dubief, D. Sappey-Marinier, **"Characterization of multiple sclerosis forms through fiber-bundle profile analysis,"** ESMRMB, Barcelona, 2017.
18. I. Suprano, C. Delon-Martin, G. Kocevar, C. Stamile, O. Revol, F. Nusbaum, D. Sappey-Marinier, **"Differential Hemispheric Brain Network Reorganization**

in two profiles of High IQ Children: a resting-state fMRI study," ESM-RMB, Barcelona, 2017.

Modèles Non-Supervisés pour l'Analyse des Fibres de Substance Blanche dans la Sclérose en Plaques

L'exploitation de données longitudinales d'IRM de tenseur de diffusion (DTI) est un défi crucial pour mieux comprendre les mécanismes pathologiques de maladies neurologiques complexes comme la sclérose en plaques (SEP), dans laquelle les faisceaux de substance blanche (SB) sont différemment atteints par les processus inflammatoires. Dans ce travail, nous proposons une nouvelle famille d'algorithmes non-supervisés pour modéliser les altérations longitudinales des métriques de diffusion le long des faisceaux de SB. Dans la première partie de ce travail, nous dressons l'état de l'art des études déjà présentes dans la littérature. Cet état de l'art se focalisera sur les études montrant les effets de la SEP sur les faisceaux de SB grâce à l'emploi de la DTI. Dans la seconde partie de ce travail, nous introduisons deux nouvelles méthodes, "string-based", l'une semi-supervisée et l'autre non-supervisée, pour extraire les faisceaux de SB. Nous montrons comment ces algorithmes permettent d'améliorer l'extraction de faisceaux spécifiques comparés aux approches déjà présentes dans la littérature. De plus, dans un second chapitre, nous présentons une extension de la méthode proposée par le couplage du formalisme "string-based" aux informations spatiales des faisceaux de SB. Dans la troisième et dernière partie de ce travail, nous décrivons trois algorithmes automatiques permettant l'analyse des changements longitudinaux le long des faisceaux de SB chez des patients SEP. Ces méthodes sont basées respectivement sur un modèle de mélange Gaussien, la factorisation de matrices non-négatives et la factorisation de tenseurs non-négatifs. De plus, pour valider nos méthodes, nous introduisons un nouveau modèle pour simuler des changements longitudinaux réels, basé sur une fonction de probabilité Gaussienne généralisée. Des hautes performances ont été obtenues avec ces algorithmes dans la détection de changements longitudinaux fins le long des faisceaux de SB chez des patients atteints de SEP. En conclusion, nous avons proposé dans ce travail des nouveaux algorithmes non-supervisés pour une analyse précise des faisceaux de SB, permettant une meilleure caractérisation des altérations de SB survenant dans la SEP.

Mots clés : Analyse longitudinale; Imagerie par tenseur de diffusion; Imagerie par résonance magnétique; Sclérose en plaques; Extraction des faisceaux de SB; Factorisation de matrices non-négatives; Factorisation tensorielle.

Unsupervised Models for White Matter Fiber-Bundles Analysis in Multiple Sclerosis

Processing of longitudinal diffusion tensor imaging (DTI) data is a crucial challenge to better understand pathological mechanisms of complex brain diseases such as multiple sclerosis (MS) where white-matter (WM) fiber-bundles are variably altered by inflammatory and demyelinating/or events. In this work, we propose a new family of unsupervised algorithms to model longitudinal alterations in diffusivity metrics along WM fiber-bundles. In the first part, we give an overview of the studies already present in literature. We focus our analysis on studies showing the effects of MS in WM using DTI. In the second part, we introduce two new string-based methods, one semi-supervised and one unsupervised, to extract specific WM fiber-bundles. We show how these algorithms allow to improve the extraction of specific fiber-bundles compared to previous approaches already present in literature. Moreover, in the second chapter, we show an extension of the proposed method by coupling the string-based formalism with the spatial information of the fiber-tracks. In the third, and last part, we describe, in order of complexity, three different unsupervised algorithms to perform analysis of longitudinal changes visible along WM fiber-bundles in MS patients. These methods are based on Gaussian mixture model, non-negative matrix and non-negative tensor factorization respectively. Moreover, in order to validate our methods, we introduce a new model to simulate real longitudinal changes based on a generalized Gaussian probability density function. For those algorithms high level of performances were obtained in the detection of small longitudinal changes along the WM fiber-bundles of MS patients. In conclusion, we propose, in this work, a new set of unsupervised algorithms to perform a more sensitivity analysis of WM fiber-bundle useful for the characterization of pathological alterations occurring in MS patients.

Keywords: Longitudinal analysis; Diffusion tensor imaging; Magnetic resonance imaging; Multiple sclerosis; Fiber-bundle clustering; Non-negative matrix factorization; Tensor factorization.

INTITULE ET ADRESSE DE L'U.F.R. OU DU LABORATOIRE

CREATIS - CNRS UMR5220 - Inserm U1206 - Université de Lyon - Université Lyon1 - INSA-Lyon
7 Av. Jean Capelle, 69621 VILLLEURBANNE, France.