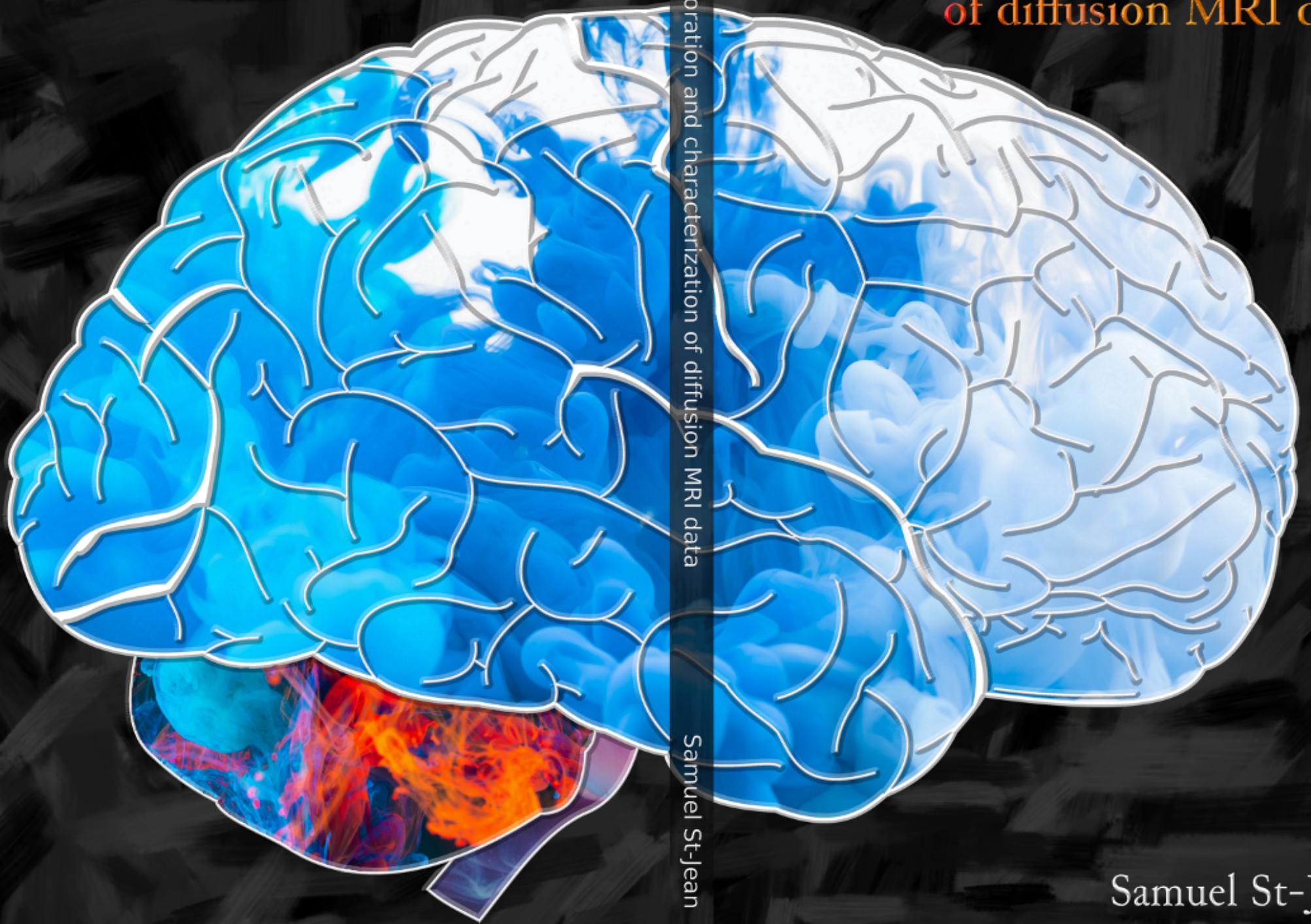


# Restoration and characterization of diffusion MRI data



Restoration and characterization of diffusion MRI data

Samuel St-Jean

Samuel St-Jean

# Restoration and characterization of diffusion MRI data

Samuel St-Jean

**Colophon:** A pictorial representation of the human brain with various diffusivities. In the cerebellum, diffusion is represented by paint as it is denser and contains deeply packed gray matter. In the cerebrum, ink is freely diffusing in water. The frontal lobe is not filled while the region of the corticospinal tract has been explored by the diffusion process. In the medulla, the diffusion properties are represented by smoke as they are yet different from the other tissues.

**Licensing:** Original figure outline by Hugh Guiney

<https://commons.wikimedia.org/wiki/File:Human-brain.SVG>, Human-brain

Color and overlaying by Samuel St-Jean, available under the CC BY-SA 4.0 license

<https://creativecommons.org/licenses/by-sa/4.0/legalcode>

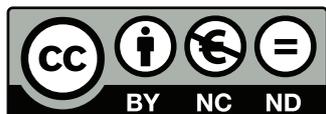
## Restoration and characterization of diffusion MRI data

©2020—Samuel St-Jean

PhD thesis, Utrecht University, the Netherlands

ISBN: 978-90-393-7228-9

Printed by: Ridderprint BV, the Netherlands



This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. To view a copy of this license, visit <https://creativecommons.org/licenses/by-nc-nd/4.0/>

The work presented in this thesis was supported by a scholarship from the Fonds de recherche du Québec - Nature et technologies (FRQNT) (Dossier 192865) and by the VIDI Grant 639.072.411 from the Netherlands Organization for Scientific Research (NWO).

# **Restoration and characterization of diffusion MRI data**

**Restauratie en karakterisering van diffusie MRI data**  
(met een samenvatting in het Nederlands)

**Restauration et caractérisation des données d'IRM de diffusion**  
(avec un résumé en français)

## **Proefschrift**

ter verkrijging van de graad van doctor aan de  
Universiteit Utrecht  
op gezag van de  
rector magnificus, prof.dr. H.R.B.M. Kummeling,  
ingevolge het besluit van het college voor promoties  
in het openbaar te verdedigen op

donderdag 9 januari 2020 des middags te 2.30 uur

door

**Samuel St-Jean**

geboren op 15 juli 1988  
te Sherbrooke, Canada

**Promotor:**

Prof. dr. ir. M.A. Viergever

**Copromotor:**

Dr. dr.h.c.mult. A.L.G. Leemans

# Contents

	Page
<b>Chapter 1</b>	
<b>Introduction</b>	<b>1</b>
1.1 Magnetic resonance imaging	2
1.2 Diffusion MRI: a brief introduction to theory and concepts	4
1.3 Outline of this thesis	9
1.4 Software implementations and datasets	11
<b>Chapter 2</b>	
<b>Non Local Spatial and Angular Matching: Enabling higher spatial resolution diffusion MRI datasets through adaptive denoising</b>	<b>17</b>
2.1 Introduction	19
2.2 Theory	21
2.3 Method	24
2.4 Results	31
2.5 Discussion	39
2.6 Conclusion	45
2.7 Appendix: The NLSAM Algorithm	46
<b>Chapter 3</b>	
<b>Obtaining representative core streamlines for white matter tractometry of the human brain</b>	<b>51</b>
3.1 Introduction	53
3.2 Theory and Methods	54
3.3 Results	56
3.4 Discussion and Conclusion	58

<b>Chapter 4</b>	
<b>Reducing variability in along-tract analysis with diffusion profile realignment</b>	<b>62</b>
4.1 Introduction	64
4.2 Theory	66
4.3 Materials and methods	68
4.4 Results	77
4.5 Discussion	89
4.6 Conclusion	94
4.7 Appendix: The diffusion profile realignment algorithm	94
4.8 Supplementary materials	96
<b>Chapter 5</b>	
<b>Automated characterization of noise distributions in diffusion MRI data</b>	<b>110</b>
5.1 Introduction	112
5.2 Theory	113
5.3 Material and Methods	116
5.4 Results	121
5.5 Discussion	133
5.6 Conclusions	136
5.7 Appendix	137
<b>Chapter 6</b>	
<b>Harmonization of diffusion MRI datasets with adaptive dictionary learning</b>	<b>146</b>
6.1 Introduction	148
6.2 Theory	149
6.3 Methods	151
6.4 Results	158
6.5 Discussion	166
6.6 Conclusions	170
6.7 Appendix: The harmonization algorithm	171
<b>Chapter 7</b>	
<b>Summary and discussion</b>	<b>178</b>
7.1 Summary	178

7.2 Discussion	180
<b>Samenvatting</b>	<b>188</b>
<b>Résumé</b>	<b>191</b>
<b>Publications</b>	<b>194</b>
<b>Acknowledgments</b>	<b>198</b>
<b>About the author</b>	<b>199</b>

*If it's on the Internet, then it must be true.*

George Washington

# 1

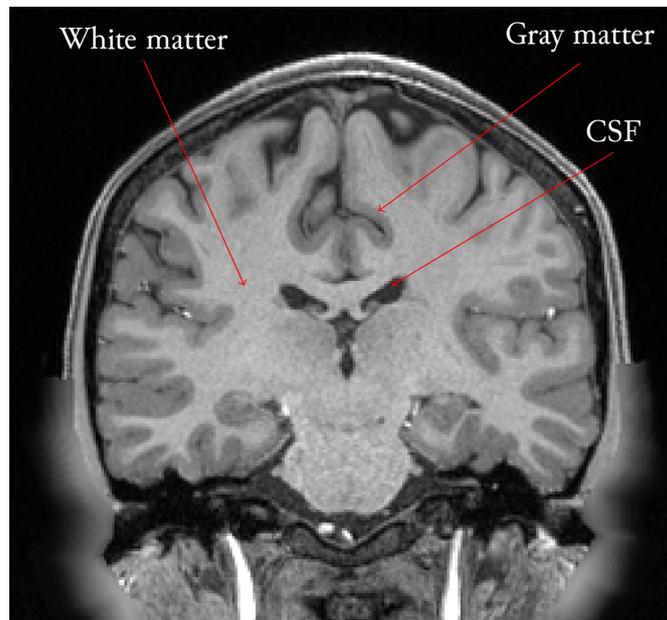
## Introduction

The turn of the new millennium saw the completion of the human genome characterization. One of the major challenges of the next century is to map the human connectome—a map of the intricate complexity of the white matter circuitry comprising the human brain. This would be a central achievement to understand and study noninvasively the healthy white matter, but will also open new doors to characterize neurodegenerative diseases such as Alzheimer or Parkinson. In order to do so, diffusion weighted magnetic resonance imaging (MRI) is one of the powerful, noninvasive tools at our disposal. However, diffusion MRI only offers an *indirect* way to probe the white matter microstructure at the macroscopic scale, but allows one to infer on the microscopic scale of the white matter. This is done by carefully influencing the random displacement of water molecules (the Brownian motion) in an experimentally controlled way using magnetic fields. The displacement of these molecules can then be used to infer information about the microstructure they encountered during this controlled displacement. As diffusion MRI is only a coarse and indirect view of the microstructure, it can be prone to overinterpretation of its findings—designing diffusion MRI experiments, processing the collected data, analyzing the results and understanding their limitations is not always straightforward. The chapters contained in this thesis present some recent advances in diffusion MRI to enhance data analysis and

subsequent studies of the human brain.

## 1.1 Magnetic resonance imaging

The everyday clinical MRI acquisition relies on T1-weighted (T1w) imaging (Brant-Zawadzki et al., 1992), where the white matter will typically be white, the gray matter gray and the cerebrospinal fluid (CSF) black as shown by Fig. 1.1. T1w images are a quick way to get an overview of the brain—the sequence is relatively fast to acquire (around 5 minutes on a clinical scanner), is easy to understand and provides a high spatial resolution of around 1 mm isotropic. However, understanding the underlying cause-to-effect phenomena affecting the contrast may be less straightforward as multiple competing processes can cause a change in the T1 relaxation time and therefore a different contrast.



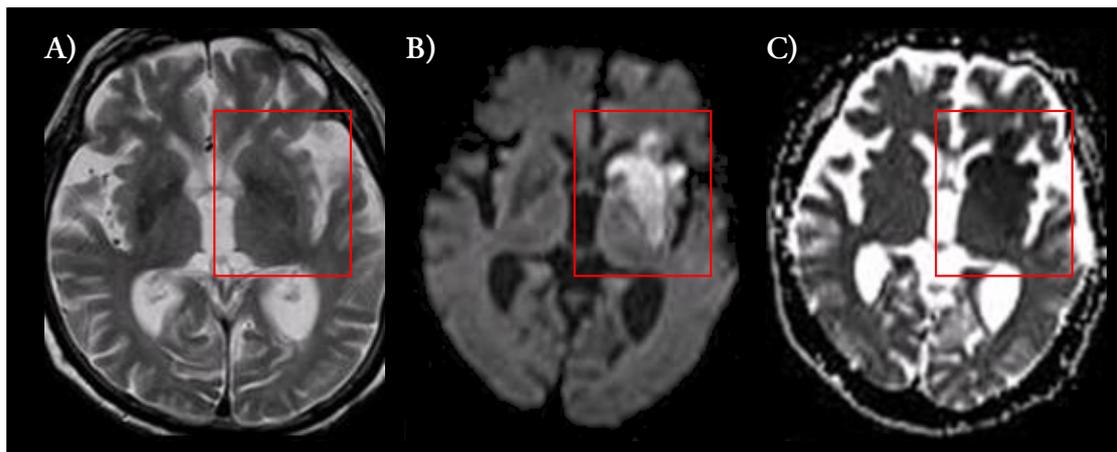
**Figure 1.1:** A structural T1w image showing the white matter (in white), the gray matter (in gray) and the CSF (in black). It is not possible to distinguish the organization of the underlying white matter on a T1w image, but this is possible using diffusion MRI. The figure is adapted from St-Jean (2015).

On the other hand, diffusion MRI can yield quantitative information about both the rate of diffusivity and direction of displacement of water molecules (Le Bihan et al., 1986; Le Bihan, 2014). These two complementary measures are affected by pathology through a

## 1.1. Magnetic resonance imaging

---

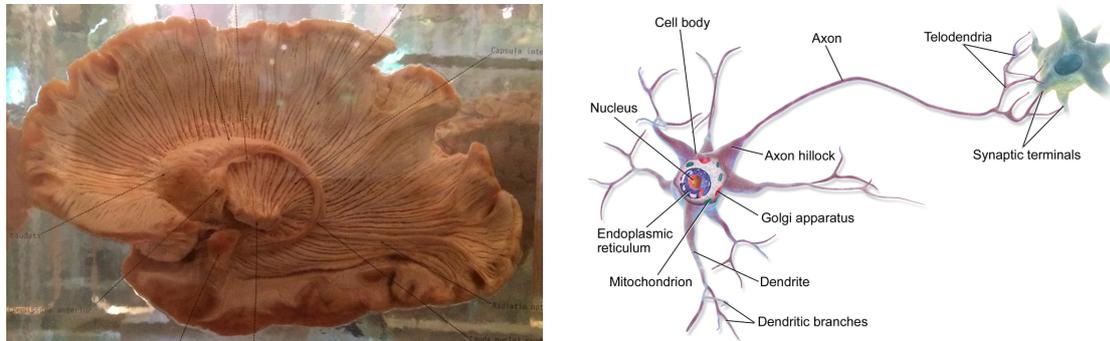
different mechanism than T1 relaxation, which makes it possible to disentangle the cause in some specific cases. One of the earliest applications was in stroke imaging, where the apparent diffusion coefficient (ADC) is lower and the diffusion weighted image is hyperintense in the affected region, which may be difficult to identify (or even invisible) in a classical T1w image (Baird and Warach, 1998). This difference in the diffusion weighted images can even be identified minutes after the incident (Birenbaum et al., 2011) as shown in Fig. 1.2.



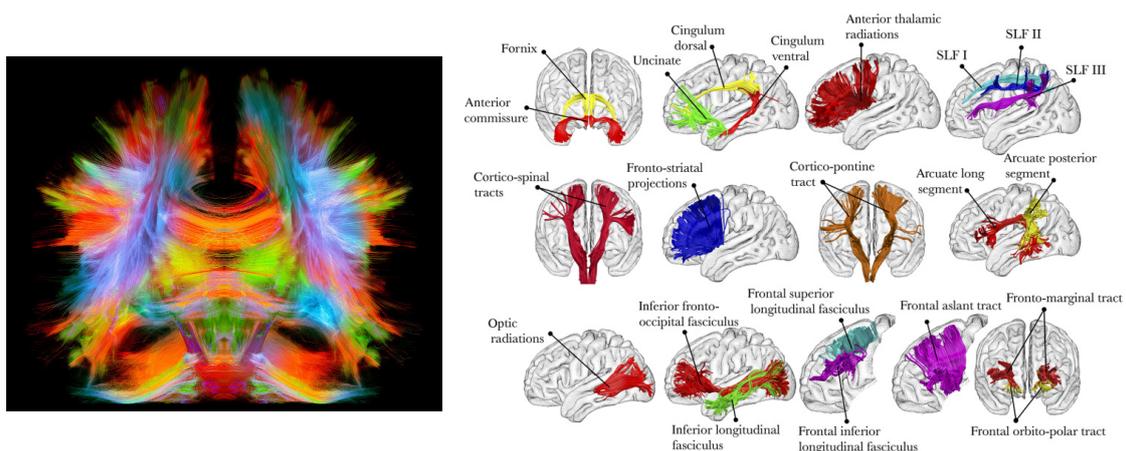
**Figure 1.2:** Example of a brain region affected by stroke 4 hours after the incident. In **A**), the T2w image does not show any change in contrast while there is a higher signal in the affected region on the diffusion weighted image in **B**). This is seen as a lower mean diffusivity in the ADC map as shown by the darkened region in **C**). The figure is adapted from Shen et al. (2011), available under the CC BY 2.0 license.

While diffusion MRI can provide information about the tissue microstructure through the change in contrast alone, it is also additionally possible to infer the directional information about these changes. Under an oriented magnetic field, such as used in an MRI scanner with a diffusion sensitizing gradient, the water molecules tend to diffuse parallel to the white matter fibers rather than perpendicularly. This can be used to infer the underlying structure and organization of the brain by changing the orientation of the diffusion gradient and taking multiple images subject to various orientations. The same idea can even be applied to other tissues of interest, either in vivo or ex vivo. Fig. 1.3 shows a schematic of a single neuron and a sagittal cut of an ex vivo brain, for which the white matter pathways can be reconstructed in vivo and noninvasively using diffusion MRI, a process known as tractography (Basser, Pajevic, et al., 2000; Mori and Van Zijl, 2002). An

example of a whole brain reconstruction using tractography and with a virtual dissection of some common fiber pathways are shown in Fig. 1.4.



**Figure 1.3:** Left: Histological cut in a sagittal view of the brain. Diffusion MRI enables noninvasive imaging of the white matter. Photograph courtesy of Maxime Chamberland. Right: Schematic representation of a neuron. The axon is enveloped by a myelin sheath, which constitutes the white matter in the brain. Image taken from Wikipedia.



**Figure 1.4:** Left: Tractography of the white matter reconstructed using diffusion MRI. This is an in vivo reconstruction, allowing to visualize the major pathways and white matter structures of the brain. Image courtesy of Maxime Chamberland. Right: Virtual anatomical dissection for some of the white matter pathways as obtained from whole brain tractography. The figure is adapted from Thiebaut de Schotten et al. (2015), available under the CC BY 4.0 license.

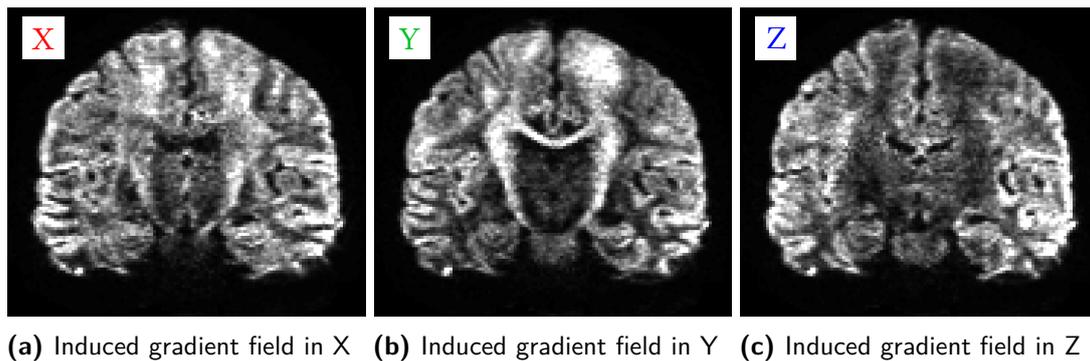
## 1.2 Diffusion MRI: a brief introduction to theory and concepts

As we have mentioned previously, one advantage of diffusion MRI is its ability to infer the directional organization of the imaged tissue. For a given diffusion weighting, different

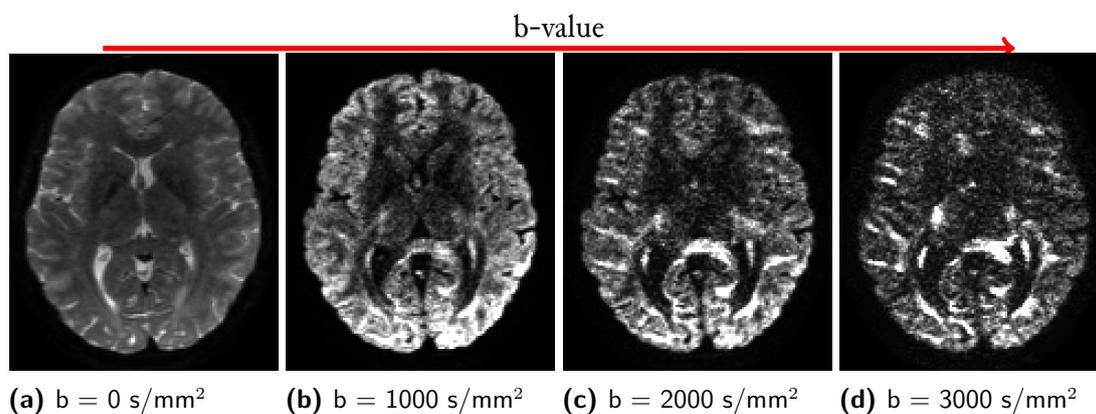
## 1.2. Diffusion MRI: a brief introduction to theory and concepts

---

orientation of the diffusion sensitizing gradients will give different contrasts as shown in [Fig. 1.5](#). This difference of angular contrast is one of the key experimental conditions which is used to infer properties of the tissue. The other factor available is the amount of diffusion weighting applied, usually referred to as the b-value. For a given angular orientation, an increase in diffusion weighting generally translates to a lower measured signal as shown in [Fig. 1.6](#).



**Figure 1.5:** The signal measured in diffusion MRI is orientation-dependent. When the structure is aligned with the applied gradient, the signal loss is accentuated due to the preference of water molecules to go parallel to the structure, rather than perpendicularly. As the CSF is an isotropic medium with a high diffusivity value, the signal loss is equal in all directions and generally speaking the region of highest diffusion in the healthy human brain. The figure is adapted from St-Jean (2015).



**Figure 1.6:** Example of increasing diffusion weighting using datasets from the human connectome project (HCP). For a fixed orientation of the diffusion gradient, an increase in b-value translates to a lower measured signal due to longer diffusion time. The behavior of this signal loss is at the foundation of diffusion MRI and is used to infer indirectly the white matter architecture. The figure is adapted from St-Jean (2015).

To summarize the information from multiple diffusion weighted images, various mathematical models offer a compact representation of the diffusion process, allowing the extraction of scalar values to simplify interpretation and visualization of the diffusion features. The most well-known signal representation is the diffusion tensor imaging (DTI) model (Basser, Mattiello, et al., 1994; Basser and Pierpaoli, 1996), which is valid in the case of free diffusion or in the presence of a single, coherently oriented fiber population. In DTI, the diffusion equation is written as a 3D symmetric ellipsoid, whose largest eigenvalue and eigenvector indicate the main axis of diffusion as shown by Eq. (1.1).

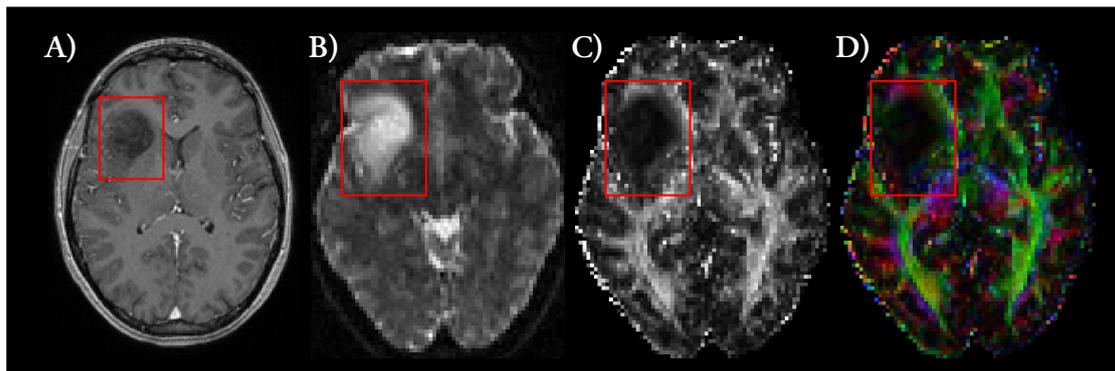
$$S(b, \mathbf{g}) = S_0 e^{-b\mathbf{g}^T \mathbf{D} \mathbf{g}}, \quad (1.1)$$

where  $\mathbf{D}$  is the diffusion tensor,  $b$  the b-value and  $\mathbf{g}$  the diffusion sensitizing gradient orientation. Solving the equations for  $\mathbf{D}$  requires at least six diffusion weighted images as the diffusion process is usually assumed to be symmetric as shown by Eq. (1.2), but collecting additional measurements is generally encouraged to obtain a stable solution (Jones et al., 2013; Tournier, Mori, et al., 2011).

$$\mathbf{D} = \begin{pmatrix} D_{xx} & - & - \\ D_{xy} & D_{yy} & - \\ D_{xz} & D_{yz} & D_{zz} \end{pmatrix} \quad (1.2)$$

Once the diffusion tensor is known, scalar metrics such as the ADC (which is the mean of the eigenvalues) and the fractional anisotropy (FA), a normalized measure of dispersion obtained by the ratio of the standard deviation over the mean of the eigenvalues, can be computed from  $\mathbf{D}$ . These measures can be useful to visually identify abnormal diffusion as shown by Fig. 1.7.

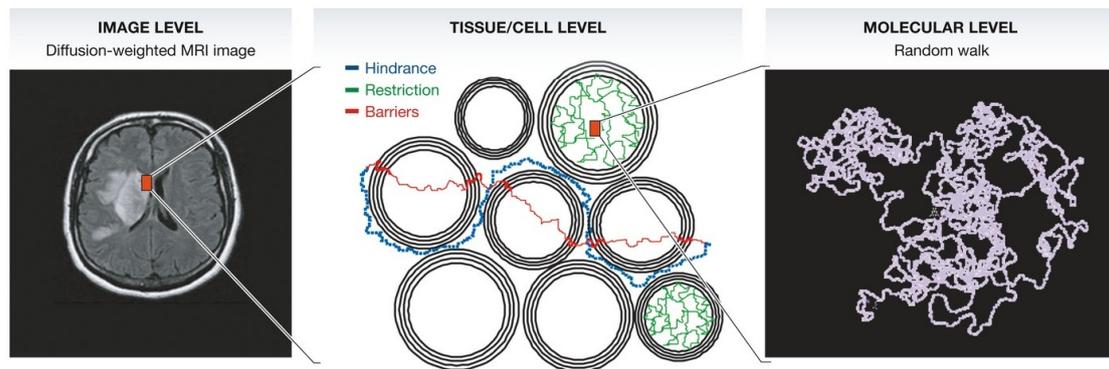
However, DTI is inadequate for long diffusion times where the water molecules may hit cell boundaries and become subject to different regimes of diffusion such as hindered and restricted diffusion as shown in Fig. 1.8. DTI is also unable to resolve the presence of crossing fibers, which are prevalent throughout the human brain (Jeurissen, Leemans, et al., 2013). Advanced models of diffusion, such as diffusion kurtosis imaging (DKI) (Jensen and Helpert, 2010; Jensen, Helpert, et al., 2005) or spherical deconvolution techniques (Dell'Acqua et al., 2007; Descoteaux, Deriche, et al., 2009; Jeurissen, Tournier, et al., 2014; Tournier, Calamante, et al., 2007) amongst others, are valid for longer diffusion



**Figure 1.7:** A tumor in a **A)** T1w image and a **B)** diffusion weighted image. While the tumor can be seen in the T1w image, the affected microstructure, presence of edema and displacement of the normal white matter is easily seen in the diffusion derived **C)** FA map and **D)** direction color coded FA map. The figure is adapted from St-Jean (2015).

times or in the presence of crossing fiber geometries respectively. This comes at the cost of increased mathematical complexity and longer acquisition protocols than classical DTI, but offer complementary information to DTI. As the diffusion MRI literature itself is quite vast, the reader who would like to broaden his knowledge can find several reviews for each major topic in diffusion MRI, such as artifacts correction (Tournier, Mori, et al., 2011), diffusion modeling (Descoteaux and Poupon, 2014) or tractography (Jeurissen, Descoteaux, et al., 2017) to name but a few. Recent reviews and special issues to the multiple topics in diffusion MRI include for example Leemans (2019) and Tournier (2019) while books are also dedicated to the subject such as Johansen-Berg and Behrens (2009) and Jones (2011).

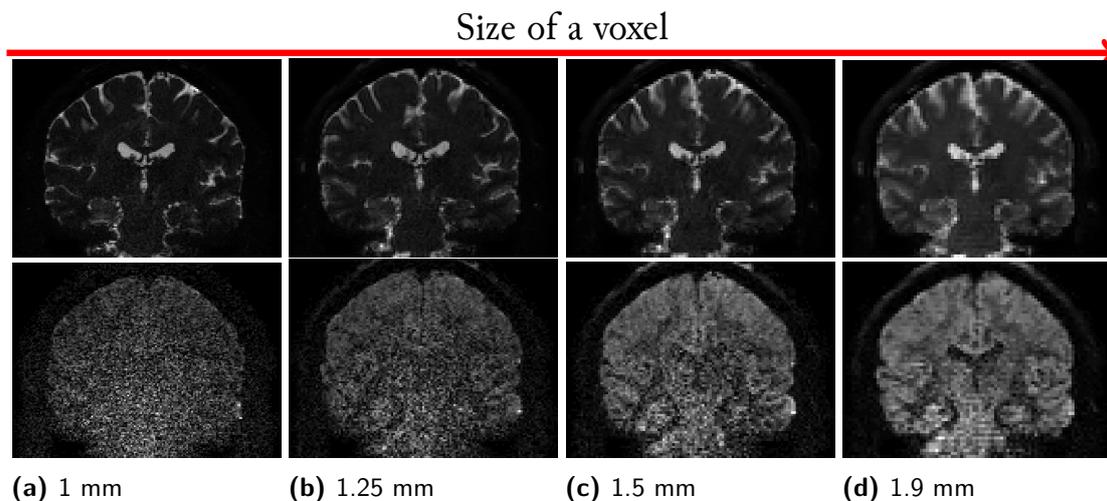
Since diffusion MRI can be used to provide directional information not available to conventional T1w imaging, one can wonder what might be its limitations besides longer scan times. Unfortunately, the classical version of diffusion MRI uses a spin-echo sequence (Stejskal and Tanner, 1965) and is based on T2 relaxation effects and relatively long echo times. This means that in clinical practice the resolution, shorter scan time and signal-to-noise ratio (SNR) that can be achieved is severely limited by the hardware available. A T1w image can be acquired in approximately 5 minutes at a spatial resolution of 1 mm isotropic while an equivalent diffusion MRI acquisition would be at a spatial resolution of 2 mm isotropic with 30 DWIs, which includes the use of parallel imaging acceleration (Griswold et al., 2002; Pruessmann et al., 1999) and echo planar imaging (EPI) (Mansfield, 1977; Rzedzian et al., 1983). In the research world, this is not so much an issue as several



**Figure 1.8:** Macroscopic and microscopic views of water diffusion. Given sufficient time, the water molecules hit the cell membranes, which may reduce their rate of diffusion (hindered diffusion, in blue) or may even be trapped inside the cell (restricted diffusion, in green) as opposed to free diffusion happening outside of the cellular environment. These different regimes of diffusion results in different contrasts as observed on the diffusion weighted images, provided the acquisition protocol allows sufficient diffusion time to explore these effects. Image adapted from Le Bihan (2014), available under the CC BY 4.0 license.

techniques have been designed to minimize scanning time (Lustig et al., 2007; Ning et al., 2016; Paquette et al., 2015; Scherrer et al., 2011) and may be available as off-the-shelves sequences or upon request from the authors. However, the decrease in SNR associated with a higher spatial resolution is oftentimes unavoidable.

An analogy to explain why this effect is hard to counterbalance can be to think of a 1 meter squared sandbox filled with a fixed amount of sand. Say that we divide this sandbox by using a grid of 100 boxes of size 10 cm by 10 cm. Now, if we were to subdivide it again by another factor 10, there would be 10 000 boxes of size 1 cm by 1 cm. Each box would contain fewer grains of sand than if we only used 100 boxes of size 10 cm by 10 cm, even though the total number of sand in the whole sandbox is always the same no matter the subdivision. The same idea applies when we use smaller voxels in diffusion MRI; the number of water molecules present in each voxel contributing to the signal is less than if coarser voxels, which would each contain more water molecules, would be instead imaged. As more slices also need to be acquired for an equivalent coverage, this also lengthens the acquisition time due to an increase in echo time (TE) and repetition time (TR). Fig. 1.9 shows the reduced signal in a DWI if only the voxel size is increased at each subsequent step. This tradeoff between SNR, spatial resolution and acquisition time is a careful balance that needs to be optimized for every diffusion experiment.



**Figure 1.9:** Datasets of the same subject at various spatial resolutions for a fixed b-value. At the top, a set of  $b = 0 \text{ s/mm}^2$  images and at the bottom a diffusion weighting of  $b = 1000 \text{ s/mm}^2$ . As the voxel resolution increases, the images contain less water molecules per volume and therefore less signal is measured. This increase in spatial resolution is at the cost of a lower SNR and increased scan time relative to their coarser counterpart, limiting the achievable resolution in vivo. The figure is adapted from St-Jean (2015).

### 1.3 Outline of this thesis

As we have seen so far, diffusion MRI can help to identify (and even quantify) abnormalities in the white matter by providing contrasts not available from classical MRI. This is however subject to limitations in terms of both acquisition time and SNR, which depends on the scanner and gradient hardware available. Fortunately, theory and methods taking root in image and signal processing, statistics and machine learning can be employed to enhance and facilitate the analysis of diffusion MRI datasets as will be showcased in this thesis.

As shown in Fig. 1.9, the increase in spatial resolution in diffusion MRI comes at the cost of a progressively lower SNR. The Non Local Spatial and Angular Matching (NL-SAM) algorithm for denoising diffusion MRI is presented in Chapter 2, which makes use of dictionary learning to iteratively construct an adaptive basis to represent the data at hand. Important features of the signal across diffusion weighted images are automatically identified and used for an efficient reconstruction, discarding artifacts associated with lower SNR datasets in the process. Synthetic simulations and comparisons with three other algorithms show that the method improves the estimation of diffusion metrics when compared to the original, unprocessed data. Experiments on a 1.2 mm isotropic dataset show qualitative

improvements in restoring coherence in crossing fiber configurations and subsequent reconstruction of fiber bundles, showing more anatomical details than a comparative 1.8 mm isotropic dataset of the same subject for matched acquisition times.

Chapters 3 and 4 go hand in hand, allowing quantitative analysis based on tractography as presented in the previous chapter. Since tractography is a virtual reconstruction of fiber bundles, it is possible to study the change in diffusion metrics along those fiber pathways. This provides an alternative analysis method to the classical region of interest and voxelwise analysis popular in structural MRI by projecting metrics along-tract in a new 1D space following the 3D fiber bundles. Chapter 3 first shows an optimal assignment strategy to create this 1D representation of the metric of interest. As tractography is different for every subject due to anatomical variations, this results in potential differences in reconstructed pathways. Delineation of those pathways therefore results in slightly different bundles of various length for each subject across spatial coordinates. The projection from a 3D space to this new 1D space is not necessarily straightforward since matching points in 3D do not always correspond to the same anatomical locations across subjects depending on the assignment strategy that is used.

Chapter 4 then expands upon Chapter 3 and presents the Diffusion Profile Realignment (DPR) algorithm, which is designed to realign the extracted 1D profiles just before statistical analysis. Using the Fourier transform, the 1D coordinates are realigned toward a common template subject that is automatically chosen from the set of subjects currently analyzed. Only the 1D segments that are sufficiently overlapping are kept and realigned altogether, ensuring that the pointwise coordinates are in fact matching across all spatial locations of every subject before analysis. This is demonstrated on synthetic experiments and on in vivo data, where the coefficient of variation is lower after realignment for the studied diffusion metrics. Using 100 in vivo subjects, additional experiments are performed by locally altering the shapes of the 1D profiles. After realignment, the affected regions are easier to identify than before applying the realignment algorithm while preserving the effect of interest.

Chapter 5 takes us back to the acquisition of diffusion datasets by proposing an automated method to estimate noise distributions in diffusion MRI, but without requiring *a priori* knowledge of the acquisition process itself. As the statistical distribution of the signal in MRI depends on the reconstruction algorithm and the type of parallel acceleration used for the acquisition, an automated method that can identify these characteristics without

user interaction may help to inform subsequent steps of the processing pipeline requiring such information. Using a transformation to a gamma distribution, voxels are automatically identified as belonging to the noise distribution or rejected as containing tissues or artifacts using equations based on the moments and maximum likelihood equations of the gamma distribution. The proposed algorithms are compared against three other methods using numerical simulations on phantoms, simulations with parallel acceleration and acquired datasets of a water bottle. Two in vivo datasets from different hardware manufacturers are also analyzed in addition to a bias correction and a denoising task. As the signal measured in MRI dictates the extracted scalar values from diffusion MRI modeling, different statistical properties could be mistakenly interpreted as genuine biological differences if not taken into account during analysis. This is even more important for multicenter studies that pool data from various acquisition protocols and hardware manufacturers as datasets exhibit small signal variations even between scans of the same subject.

[Chapter 6](#) follows on these previous ideas by proposing a new method to harmonize diffusion datasets acquired on different scanners. Based once again on dictionary learning just like [Chapter 2](#), the datasets from three different scanners are harmonized either towards a common space or from one scanner to the other in a fully automated manner. This can even be done if the datasets are acquired at different spatial or angular resolutions through subsampling and matching of the learned dictionaries, which do not require matched pairs of samples for the training phase. Experiments on simulated alteration of the datasets show that the algorithm preserves the induced effects while reducing variability between scanners on the studied diffusion MRI metrics.

Finally, [Chapter 7](#) summarizes the results presented throughout this thesis and presents new promising directions for diffusion MRI which could be combined with the ideas in this thesis.

## 1.4 Software implementations and datasets

The development of this thesis also lead to new algorithms, their implementation and to the acquisition of diffusion MRI datasets to support the experiments and results presented. To make the algorithms useful to the community in general, implementations are made available with examples and easy to use installation instructions and accompanying documentation. Moreover, most of the datasets specifically acquired for the work presented in

these chapters are made freely available so they may be of use to researchers doing similar experiments, expand upon the presented methods or compare fairly their algorithms with the same datasets that we used previously. The implementation of the NLSAM algorithm from [Chapter 2](#) is available online at <https://github.com/samuelstjean/nlsam> and the datasets used in the manuscript are also available at [https://github.com/samuelstjean/nlsam\\_data](https://github.com/samuelstjean/nlsam_data). The diffusion profile realignment algorithm from [Chapter 4](#) is available online at <https://github.com/samuelstjean/dpr>. An archival copy of each version of the code is also available on Zenodo (St-Jean, 2019) and the synthetic datasets and extracted metrics for the in vivo datasets are also available on Zenodo (St-Jean, Chamberland, et al., 2018). The algorithm for automatically estimating the noise distribution from [Chapter 5](#), which is an improved version of the implementation previously described in St-Jean, De Luca, Viergever, et al. (2018), can be obtained at <https://github.com/samuelstjean/autodmri>. An online archived version is also available on Zenodo (St-Jean, De Luca, Tax, et al., 2019) along with the synthetic datasets and the phantom datasets acquired for the experiments (St-Jean, De Luca, Tax, et al., 2018). Finally, the harmonization algorithm presented in [Chapter 6](#) is available online at <https://github.com/samuelstjean/harmonization>. The usual archived copy on Zenodo is also available (St-Jean, Viergever, et al., 2019).

# Bibliography

- [1] A. E. Baird and S. Warach. “Magnetic Resonance Imaging of Acute Stroke”. In: *Journal of Cerebral Blood Flow & Metabolism* 18.6 (June 1998), pp. 583–609 (cit. on p. 3).
- [2] P. J. Basser, J. Mattiello, and D. LeBihan. “MR diffusion tensor spectroscopy and imaging.” In: *Biophysical journal* 66.1 (Jan. 1994), pp. 259–67 (cit. on p. 6).
- [3] P. J. Basser, S. Pajevic, C. Pierpaoli, J. Duda, and A. Aldroubi. “In vivo fiber tractography using DT-MRI data.” In: *Magnetic Resonance in Medicine* 44.4 (Oct. 2000), pp. 625–32 (cit. on p. 3).
- [4] P. J. Basser and C. Pierpaoli. “Microstructural and Physiological Features of Tissues Elucidated by Quantitative-Diffusion-Tensor MRI”. In: *Journal of Magnetic Resonance, Series B* 111.3 (June 1996), pp. 209–219 (cit. on p. 6).
- [5] D. Birenbaum, L. W. Bancroft, and G. J. Felsberg. “Imaging in acute stroke.” In: *The western journal of emergency medicine* 12.1 (2011), pp. 67–76 (cit. on p. 3).
- [6] M. Brant-Zawadzki, G. D. Gillan, and W. R. Nitz. “MP RAGE: a three-dimensional, T1-weighted, gradient-echo sequence—initial experience in the brain.” In: *Radiology* 182.3 (Mar. 1992), pp. 769–775 (cit. on p. 2).
- [7] F. Dell’Acqua, G. Rizzo, P. Scifo, R. A. Clarke, G. Scotti, and F. Fazio. “A Model-Based Deconvolution Approach to Solve Fiber Crossing in Diffusion-Weighted MR Imaging”. In: *IEEE Transactions on Biomedical Engineering* 54.3 (Mar. 2007), pp. 462–472 (cit. on p. 6).
- [8] M. Descoteaux and C. Poupon. “Diffusion-Weighted MRI”. In: *Comprehensive Biomedical Physics*. Elsevier, 2014, pp. 81–97 (cit. on p. 7).
- [9] M. Descoteaux, R. Deriche, T. Knosche, and A. Anwander. “Deterministic and Probabilistic Tractography Based on Complex Fibre Orientation Distributions”. In: *IEEE Transactions on Medical Imaging* 28.2 (Feb. 2009), pp. 269–286 (cit. on p. 6).
- [10] M. A. Griswold et al. “Generalized Autocalibrating Partially Parallel Acquisitions (GRAPPA)”. In: *Magnetic Resonance in Medicine* 47.6 (2002), pp. 1202–1210 (cit. on p. 7).
- [11] J. H. Jensen and J. A. Helpert. “MRI quantification of non-Gaussian water diffusion by kurtosis analysis.” In: *NMR in biomedicine* 23.7 (Aug. 2010), pp. 698–710 (cit. on p. 6).
- [12] J. H. Jensen, J. A. Helpert, A. Ramani, H. Lu, and K. Kaczynski. “Diffusional kurtosis imaging: The quantification of non-gaussian water diffusion by means of magnetic resonance imaging”. In: *Magnetic Resonance in Medicine* 53.6 (June 2005), pp. 1432–1440 (cit. on p. 6).

- 
- [13] B. Jeurissen, M. Descoteaux, S. Mori, and A. Leemans. “Diffusion MRI fiber tractography of the brain”. In: *NMR in Biomedicine* (Sept. 2017), e3785 (cit. on p. 7).
- [14] B. Jeurissen, A. Leemans, J.-D. Tournier, D. K. Jones, and J. Sijbers. “Investigating the prevalence of complex fiber configurations in white matter tissue with diffusion magnetic resonance imaging”. In: *Human Brain Mapping* 34.11 (Nov. 2013), pp. 2747–2766 (cit. on p. 6).
- [15] B. Jeurissen, J.-D. Tournier, T. Dhollander, A. Connelly, and J. Sijbers. “Multi-tissue constrained spherical deconvolution for improved analysis of multi-shell diffusion MRI data”. In: *NeuroImage* 103 (Dec. 2014), pp. 411–426 (cit. on p. 6).
- [16] H. Johansen-Berg and T. E. J. Behrens. *Diffusion MRI*. Elsevier, 2009, p. 502 (cit. on p. 7).
- [17] D. K. Jones. *Diffusion MRI: Theory, Methods, and Applications*. Cambridge: Cambridge University Press, 2011, p. 784 (cit. on p. 7).
- [18] D. K. Jones, T. R. Knösche, and R. Turner. “White matter integrity, fiber count, and other fallacies: The do’s and don’ts of diffusion MRI”. In: *NeuroImage* 73 (June 2013), pp. 239–254 (cit. on p. 6).
- [19] D. Le Bihan, E. Breton, D. Lallemand, P. Grenier, E. Cabanis, and M. Laval-Jeantet. “MR imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders.” In: *Radiology* 161.2 (1986), pp. 401–407 (cit. on p. 2).
- [20] D. Le Bihan. “Diffusion MRI: what water tells us about the brain”. In: *EMBO Molecular Medicine* 6.5 (Apr. 2014), pp. 569–573 (cit. on pp. 2, 8).
- [21] A. Leemans. “Diffusion MRI of the brain: The naked truth”. In: *NMR in Biomedicine* 32.4 (Apr. 2019), e4084 (cit. on p. 7).
- [22] M. Lustig, D. Donoho, and J. M. Pauly. “Sparse MRI: The application of compressed sensing for rapid MR imaging”. In: *Magnetic Resonance in Medicine* 58.6 (Dec. 2007), pp. 1182–1195 (cit. on p. 8).
- [23] P. Mansfield. “Multi-planar image formation using NMR spin echoes”. In: *Journal of Physics C: Solid State Physics* 10.3 (1977), pp. 55–58 (cit. on p. 7).
- [24] S. Mori and P. C. M. Van Zijl. “Fiber tracking: Principles and strategies - A technical review”. In: *NMR in Biomedicine* 15.7-8 (2002), pp. 468–480 (cit. on p. 3).
- [25] L. Ning et al. “A joint compressed-sensing and super-resolution approach for very high-resolution diffusion imaging”. In: *NeuroImage* 125 (Jan. 2016), pp. 386–400 (cit. on p. 8).
- [26] M. Paquette, S. Merlet, G. Gilbert, R. Deriche, and M. Descoteaux. “Comparison of sampling strategies and sparsifying transforms to improve compressed sensing diffusion spectrum imaging”. In: *Magnetic Resonance in Medicine* 73.1 (Jan. 2015), pp. 401–416 (cit. on p. 8).
- [27] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger. “SENSE: Sensitivity encoding for fast MRI”. In: *Magnetic Resonance in Medicine* 42.5 (Nov. 1999), pp. 952–962 (cit. on p. 7).

## Bibliography

---

- [28] R. Rzedzian et al. “Real-Time Nuclear Magnetic Resonance Clinical Imaging in Paediatrics”. In: *The Lancet* 322.8362 (Dec. 1983), pp. 1281–1282 (cit. on p. 7).
- [29] B. Scherrer, A. Gholipour, and S. K. Warfield. “Super-Resolution in Diffusion-Weighted Imaging”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Vol. 14. Pt 2. Jan. 2011, pp. 124–132 (cit. on p. 8).
- [30] J.-M. Shen, X.-W. Xia, W.-G. Kang, J.-J. Yuan, and L. Sheng. “The use of MRI apparent diffusion coefficient (ADC) in monitoring the development of brain infarction”. In: *BMC Medical Imaging* 11.1 (Dec. 2011), p. 2 (cit. on p. 3).
- [31] S. St-Jean. “Acquisitions d’IRM de diffusion à haute résolution spatiale : nouvelles perspectives grâce au débruitage spatialement adaptatif et angulaire”. In: *MSc Thesis Université de Sherbrooke* (2015), p. 149 (cit. on pp. 2, 5, 7, 9).
- [32] S. St-Jean. *samuelstjean/dpr: [v0.1.1b] - 2019-03-13*. Mar. 2019.  
URL: <https://doi.org/10.5281/zenodo.2592014> (cit. on p. 12).
- [33] S. St-Jean, M. Chamberland, M. A. Viergever, and A. Leemans. *Datasets for ‘Reducing variability in along-tract analysis with diffusion profile realignment’*. 2018.  
URL: <http://dx.doi.org/10.5281/zenodo.2483169> (cit. on p. 12).
- [34] S. St-Jean, A. De Luca, C. M. W. Tax, M. A. Viergever, and A. Leemans. *Datasets for ‘Automated characterization of noise distributions in diffusion MRI data’*. 2018.  
URL: <https://zenodo.org/record/2483105> (cit. on p. 12).
- [35] S. St-Jean, A. De Luca, C. M. W. Tax, M. A. Viergever, and A. Leemans. *samuelstjean/autodmri: First release - 2019-07-17*. July 2019.  
URL: <https://doi.org/10.5281/zenodo.3339158> (cit. on p. 12).
- [36] S. St-Jean, A. De Luca, M. A. Viergever, and A. Leemans. “Automatic, Fast and Robust Characterization of Noise Distributions for Diffusion MRI”. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*. Ed. by A. F. Frangi, J. A. Schnabel, C. Davatzikos, C. Alberola-López, and G. Fichtinger. Springer International Publishing, May 2018, pp. 304–312 (cit. on p. 12).
- [37] S. St-Jean, M. A. Viergever, and A. Leemans. *samuelstjean/harmonization: Harmonization of diffusion MRI datasets with adaptive dictionary learning*. Sept. 2019.  
URL: <https://doi.org/10.5281/zenodo.3385925> (cit. on p. 12).
- [38] E. O. Stejskal and J. E. Tanner. “Spin Diffusion Measurements: Spin Echoes in the Presence of a Time-Dependent Field Gradient”. In: *The Journal of Chemical Physics* 42.1 (Jan. 1965), pp. 288–292 (cit. on p. 7).
- [39] M. Thiebaut de Schotten et al. “From Phineas Gage and Monsieur Leborgne to H.M.: Revisiting Disconnection Syndromes”. In: *Cerebral Cortex* 25.12 (Dec. 2015), pp. 4812–4827 (cit. on p. 4).
- [40] J.-D. Tournier. “Diffusion MRI in the brain – Theory and concepts”. In: *Progress in Nuclear Magnetic Resonance Spectroscopy* 112–113 (June 2019), pp. 1–16 (cit. on p. 7).

- [41] J.-D. Tournier, F. Calamante, and A. Connelly. “Robust determination of the fibre orientation distribution in diffusion MRI: non-negativity constrained super-resolved spherical deconvolution.” In: *NeuroImage* 35.4 (May 2007), pp. 1459–72 (cit. on p. 6).
- [42] J.-D. Tournier, S. Mori, and A. Leemans. “Diffusion tensor imaging and beyond.” In: *Magnetic Resonance in Medicine* 65.6 (2011), pp. 1532–1556 (cit. on pp. 6, 7).

### Additional attributions

- [1] BruceBlas. *Neuron*. <https://creativecommons.org/licenses/by-sa/3.0/legalcode>. 2019.  
URL: [https://commons.wikimedia.org/wiki/File:Blausen\\_0657\\_MultipolarNeuron.png](https://commons.wikimedia.org/wiki/File:Blausen_0657_MultipolarNeuron.png).

*In God we trust; all others must bring data.*

William Edwards Deming

# 2

## Non Local Spatial and Angular Matching: Enabling higher spatial resolution diffusion MRI datasets through adaptive denoising

Based on

---

Samuel St-Jean, Pierrick Coupé and Maxime Descoteaux  
Non Local Spatial and Angular Matching: Enabling higher spatial resolution diffusion  
MRI datasets through adaptive denoising  
Medical Image Analysis, Volume 32, 2016, Pages 115-130

## Abstract

Diffusion magnetic resonance imaging (MRI) datasets suffer from low Signal-to-Noise Ratio (SNR), especially at high b-values. Acquiring data at high b-values contains relevant information and is now of great interest for microstructural and connectomics studies. High noise levels bias the measurements due to the non-Gaussian nature of the noise, which in turn can lead to a false and biased estimation of the diffusion parameters. Additionally, the usage of in-plane acceleration techniques during the acquisition leads to a spatially varying noise distribution, which depends on the parallel acceleration method implemented on the scanner. This paper proposes a novel diffusion MRI denoising technique that can be used on all existing data, without adding to the scanning time. We first apply a statistical framework to convert both stationary and non stationary Rician and non central Chi distributed noise to Gaussian distributed noise, effectively removing the bias. We then introduce a spatially and angular adaptive denoising technique, the Non Local Spatial and Angular Matching (NLSAM) algorithm. Each volume is first decomposed in small 4D overlapping patches, thus capturing the spatial and angular structure of the diffusion data, and a dictionary of atoms is learned on those patches. A local sparse decomposition is then found by bounding the reconstruction error with the local noise variance. We compare against three other state-of-the-art denoising methods and show quantitative local and connectivity results on a synthetic phantom and on an in vivo high resolution dataset. Overall, our method restores perceptual information, removes the noise bias in common diffusion metrics, restores the extracted peaks coherence and improves reproducibility of tractography on the synthetic dataset. On the 1.2 mm high resolution in vivo dataset, our denoising improves the visual quality of the data and reduces the number of spurious tracts when compared to the noisy acquisition. Our work paves the way for higher spatial resolution acquisition of diffusion MRI datasets, which could in turn reveal new anatomical details that are not discernible at the spatial resolution currently used by the diffusion MRI community.

**Keywords:** Diffusion MRI, Denoising, Block Matching, Noise bias, Dictionary learning

## 2.1 Introduction

Diffusion magnetic resonance imaging (MRI) is an imaging technique that allows probing microstructural features of the white matter architecture of the brain. Due to the imaging sequence used, the acquired images have an inherently low signal-to-noise ratio (SNR), especially at high b-values. Acquiring data at high b-values contains relevant information and is now of great interest for connectomics (Van Essen et al., 2013) and microstructure (Alexander et al., 2010) studies. High noise levels bias the measurements because of the non-Gaussian nature of the noise, which in turn prohibits high resolution acquisition if no further processing is done. This can also lead to a false and biased estimation of the diffusion parameters, which impacts on the scalar metrics (e.g. fractional anisotropy (FA)), or in the fitting of various diffusion models (e.g. diffusion tensor imaging (DTI) and high angular resolution diffusion imaging (HARDI) models). This can further impact subsequent tractography and connectivity analysis if the spatially variable noise bias is not taken into account. Therefore, high SNR diffusion weighted images (DWIs) are crucial to draw meaningful conclusions in subsequent data or group analyses (Jones et al., 2013).

This paper focuses on denoising techniques since they can be used on all existing data, without adding to the scanning time. They also can be readily applied to any already acquired dataset just like motion and eddy current corrections that are commonly applied on acquired datasets. One possible way to acquire higher quality data is to use better hardware, but this is costly and not realistic in a clinical setting. One can also use a bigger voxel size to keep the relative SNR at the same level, but at the expense of a lower spatial resolution or acquiring fewer directions to keep an acceptable acquisition time (Descoteaux and Poupon, 2014). Averaging multiple acquisitions also increases the SNR, but this should be done either using Gaussian distributed noisy data (Eichner et al., 2015) or in the complex domain to avoid the increased noise bias (Jones et al., 2013).

With the advance of parallel imaging and acceleration techniques such as the generalized autocalibrating partially parallel acquisitions (GRAPPA) or the sensitivity encoding for fast MRI (SENSE), taking into account the modified noise distribution is the next step (Aja-Fernández et al., 2014; Dietrich et al., 2008). The noise is usually modeled with a Rician distribution when SENSE is used and a non central Chi ( $nc-\chi$ ) distribution with  $2N$  degrees of freedom (with  $N$  the number of receiver coils) when a Sum of Squares (SoS) reconstruction is used. If GRAPPA acceleration is also used with a SoS reconstruction, the degrees of freedom of the  $nc-\chi$  distribution will vary between 1 and  $2N$  (Aja-Fernández et al., 2014). Some techniques have been specifically adapted by the medical imaging community to take into account the Rician nature of the noise such as non local means algorithms (Coupe et al., 2008; Manjón, Coupé, Martí-Bonmatí, et al., 2010; Tristán-Vega and Aja-Fernández, 2010), Linear Minimum Mean Square Error estimator (Aja-Fernandez et al., 2008; Brion et al., 2013), generalized total variation (Liu et al., 2014), a majorize-minimize framework with total variation denoising (Varadarajan and Haldar, 2015), maximum likelihood (Rajan et al., 2012) or block matching (Maggioni et al., 2013). Some methods (Bao

et al., 2013; Becker et al., 2014; Brion et al., 2013; Gramfort et al., 2014; Lam et al., 2014; Manjón, Coupé, Concha, et al., 2013; St-Jean et al., 2014; Tristán-Vega and Aja-Fernández, 2010) have also been specifically designed to take advantage of the properties of the diffusion MRI signal such as symmetry, positivity or angular redundancy. Since the data acquired in diffusion MRI depicts the same structural information, but under different sensitizing gradients and noise realization, these ideas take advantage of the information contained in the multiple acquired diffusion MRI datasets.

We thus propose to exploit the structural redundancy across DWIs through a common sparse representation using dictionary learning and sparse coding to reduce the noise level and achieve a higher SNR. Our method can be thought of a Non Local Spatial and Angular Matching (NLSAM) with dictionary learning. To the best of our knowledge, most recent state-of-the-art denoising algorithms either concentrate on modeling the  $nc-\chi$  noise bias or the spatially varying nature of the noise in a Rician setting. Our method thus fills the gap by being robust to both of these aspects at the same time, as seen in Table 2.1. We will compare our method against one structural MRI method and two other publicly available algorithms: the Adaptive Optimized Non Local Means (AONLM) (Manjón, Coupé, Martí-Bonmati, et al., 2010), which is designed for 3D structural MRI, the Local Principal Component Analysis (LPCA) (Manjón, Coupé, Concha, et al., 2013) and the multi-shell Position-Orientation Adaptive Smoothing (msPOAS) algorithm (Becker et al., 2014), both designed for processing diffusion MRI datasets. More information on each method features and parameters will be detailed later.

Noise type		AONLM	LPCA	msPOAS	NLSAM
Stationary	Rician	✓	✓	✓	✓
	$nc-\chi$	X	X	✓	✓
Variable	Rician	✓	✓	X	✓
	$nc-\chi$	X	X	X	✓
Use 4D angular information		X	✓	✓	✓

**Table 2.1:** Features of the compared denoising algorithm, see Section 2.3.3 for an in-depth review of each method. The NLSAM algorithm is the only technique robust to both the spatially varying nature of the noise and the  $nc-\chi$  bias at the same time.

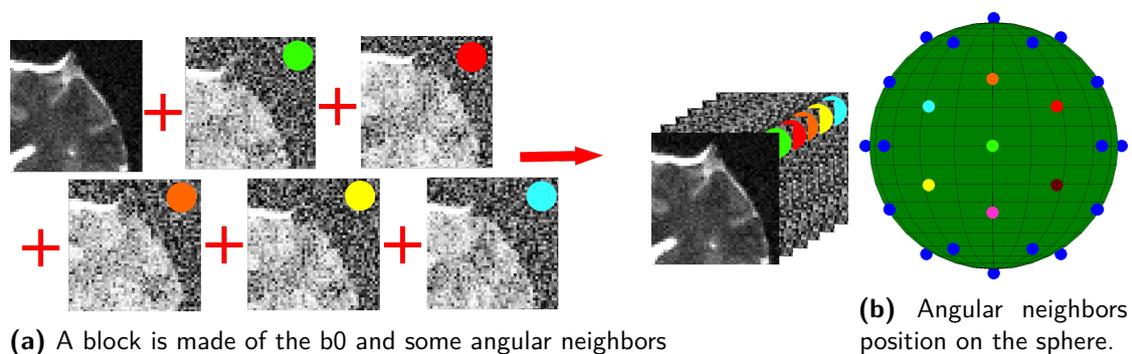
The contributions of our work are:

- i) Developing a novel denoising technique specifically tailored for diffusion MRI, which takes into account spatially varying Rician and  $nc-\chi$  noise.
- ii) Quantitatively comparing all methods on common diffusion MRI metrics.
- iii) Quantifying the impact of denoising on local reconstruction models.

- iv) Analyzing the impact of denoising on tractography with a synthetic phantom and a high spatial resolution dataset.

## 2.2 Theory

We now define two important terms used throughout the present work. Firstly, a patch is defined as a 3D region of neighboring spatial voxels, i.e. a small local region of a single 3D DWI. Secondly, a block is defined as a collection of patches taken at the *same* spatial position, but in different DWIs, i.e. a block is a 4D stack of patches that are similar in the angular domain. The reader is referred to Fig. 2.1 for a visual representation of the process.



**Figure 2.1:** **a)** A 3D block is made by stacking along the 4th dimension the  $b_0$ , a DWI and its angular neighbors, which share similar structure, but under a slightly different noise realization. **b)** Disposition of equidistant angular neighbors on the sphere.

**The Block Matching Algorithm** Reusing the key ideas from the non local means, the block matching algorithm (Dabov et al., 2007) further exploits image self-similarity. Similar 2D patches found inside a local neighborhood are stacked into a 3D transform domain and jointly filtered via wavelet hard-thresholding and Wiener filtering. Combining these filtered estimates using a weighted average based on their sparsity leads to superior denoising performance than the non local means filter. The idea has been extended in 3D for MRI image denoising in (Maggioni et al., 2013) and an adaptive patch size version for cardiac diffusion MRI image denoising was successfully employed by (Bao et al., 2013).

**The Dictionary Learning Algorithm** Dictionary learning has been used in the machine learning community to find data driven sparse representations (Elad and Aharon, 2006; Mairal et al., 2009). Typically, a set of atoms (called the dictionary) is learned over the data, providing a way to represent it with a basis tailored to the signal at hand (Olshausen and Field, 1996). This is analogous to using an off-the-shelf basis like the discrete cosine transform or wavelets, but in a data-driven manner, which gives better results than using

a fixed, general-purpose basis since it can also be overcomplete, i.e. it can have more atoms than coefficients. Given a set of input data  $\mathbf{X} = [x_1, \dots, x_n] \in \mathbb{R}^{m \times n}$  organized as column vectors, the process is expressed as

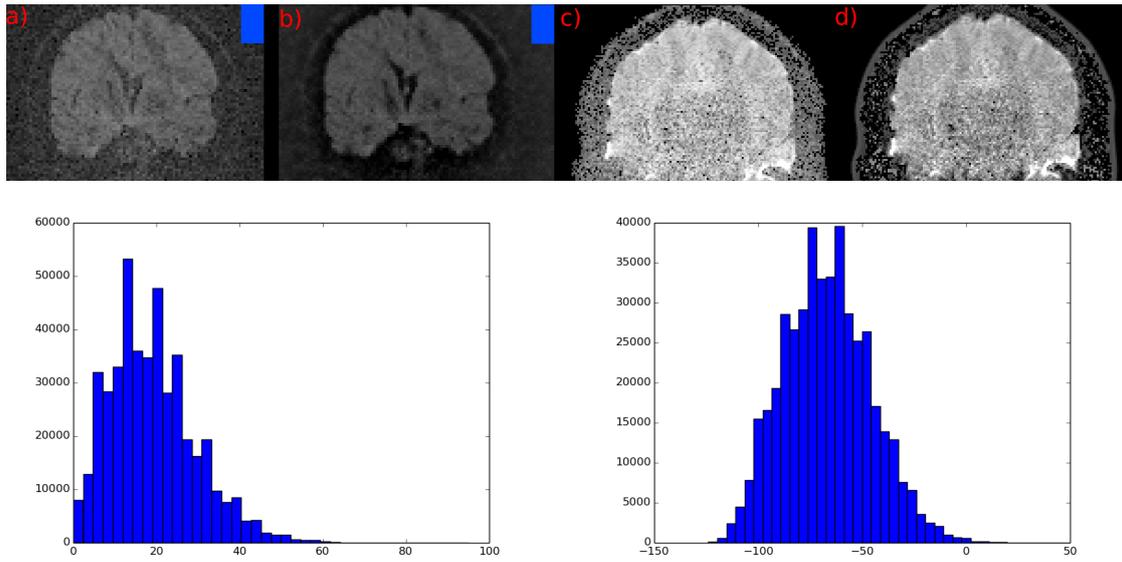
$$\min_{\mathbf{D}, \alpha} \frac{1}{n} \sum_{i=1}^n \left( \frac{1}{2} \|x_i - \mathbf{D}\alpha_i\|_2^2 + \lambda \|\alpha_i\|_1 \right) \quad \text{s.t.} \quad \|\mathbf{D}\|_2^2 = 1, \quad (2.1)$$

where  $\mathbf{D} \in \mathbb{R}^{m \times p}$  is the learned dictionary,  $\lambda$  is a trade-off parameter between the data fidelity term and the penalization on the coefficients  $\alpha = [\alpha_1, \dots, \alpha_n] \in \mathbb{R}^{p \times n}$ . A higher value of  $\lambda$  promotes sparsity at the expense of the similarity with the original data. The columns of  $\mathbf{D}$  are also constrained to be of unit  $\ell_2$  norm in order to avoid degenerated solutions (Elad and Aharon, 2006; Gramfort et al., 2014; Mairal et al., 2009). The key is to devise a sparse representation to reconstruct structural information and discard noise, since the latter does not typically allow a sparse representation in any basis. Using a penalization on the  $\ell_1$  norm of the coefficients promotes sparsity, hence providing denoising through the regularized reconstruction. This idea has led to inpainting and denoising applications from the machine learning community (Elad and Aharon, 2006; Mairal et al., 2009) or even to accelerated acquisition process in the diffusion MRI community for diffusion spectrum imaging (DSI) (Gramfort et al., 2014).

**Adjusting for various noise types** Although the original formulations of Eqs. (2.1) and (2.4) assume stationary, white additive Gaussian noise, this is usually not true in diffusion MRI data, especially at high b-values and low SNR. The noise is usually modeled as following a Rician distribution or a nc- $\chi$  distribution when used with parallel imaging depending on the reconstruction algorithm and the number of coils  $N$  used during the acquisition (Aja-Fernández et al., 2014; Dietrich et al., 2008). This introduces a bias, which depends on the intensity of the signal that must be taken into account to recover the expected value of the original signal as shown in Fig. 2.2. Note, though, that other common preprocessing corrections, such as motion correction or eddy current correction, require interpolation and could thus change the theoretical noise distribution (Veraart et al., 2013).

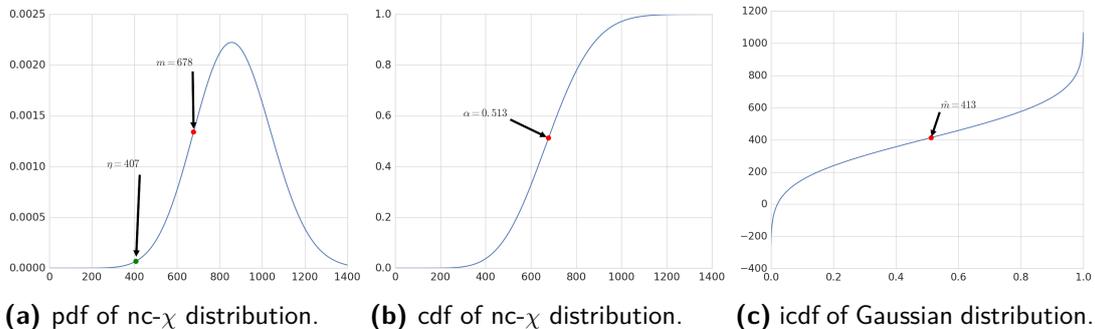
The key idea lies in the fact that the nc- $\chi$  distribution is actually made from a sum of Gaussians, from which the Rician distribution is a special case with  $N = 1$ . By making the hypothesis that each of the  $2N$  Gaussian distributions shares the same standard deviation  $\sigma_G$  (Koay, Özarslan, and Basser, 2009), one can map a value  $m$  from a nc- $\chi$  distribution to a equivalent value  $\hat{m}$  from a Gaussian distribution. We first compute estimates for  $\sigma_G$  and  $\eta$  (which is an estimate of the signal value in a Gaussian setting). If  $\eta$  is below the noise floor due to a low local SNR, that is when  $\eta < \sigma_G \sqrt{\pi/2}$ , we set  $\eta = 0$  instead of being negative as suggested by (Bai et al., 2014). The next step uses the cumulative distribution function (cdf) of a nc- $\chi$  distribution and the inverse cumulative distribution function (icdf) of a Gaussian distribution to find the equivalent value  $\hat{m}$  between the two distributions. This effectively maps a *noisy* nc- $\chi$  distributed signal  $m$  to a equivalent *noisy*

## 2.2. Theory



**Figure 2.2:** **Top:** **a)** A noisy acquisition with slowly varying  $nc\text{-}\chi$  noise and **b)** the resulting stabilized, Gaussian distributed noisy DWI. **c)** A noisy acquisition with fast varying Rician noise where the background was masked by the scanner with **d)** its stabilized counterpart. **Bottom:** Histogram of the  $nc\text{-}\chi$  noise distribution in the selected background region of **a)** before stabilization and **b)** after stabilization. Note the non-Gaussianity of the noise in **a)** versus **b)**.

Gaussian distributed signal  $\hat{m}$ . See Fig. 2.3 for a synthetic example with a visual depiction of the process for mapping  $nc\text{-}\chi$  signals to Gaussian distributed signals and (Koay, Özarlan, and Basser, 2009) for the original in-depth details.



**Figure 2.3:** A synthetic example of the stabilization algorithm. **a)** Given a noisy value  $m = 678$  observed in a  $nc\text{-}\chi$  distribution with  $N = 4$  and  $\sigma_G = 200$ , the underlying value is estimated as  $\eta = 407$ . **b)** The associated probability in the  $nc\text{-}\chi$  cdf with  $\eta$  is  $\alpha = 0.513$ , **c)** thus giving from the inverse cdf of a Gaussian distribution with mean  $\mu = 407$  and standard deviation  $\sigma_G = 200$  a new noisy value  $\hat{m} = 413$ .

Using a variance stabilization means considering the noise as additive white Gaussian

noise, which allows any already designed technique for Gaussian noise to be used without any modification. The author of (Foi, 2011) has shown that techniques with a Rician noise adaptation performed equally well as their Gaussian noise version through the use of a noise stabilization approach. The same idea has been directly applied with block matching (Dabov et al., 2007) for structural MRI in (Maggioni et al., 2013). The classical solution to remove the noise bias is to include the noise model into the denoising algorithm itself, as for example done in (Aja-Fernandez et al., 2008; Becker et al., 2014; Lam et al., 2014; Manjón, Coupé, Martí-Bonmatí, et al., 2010). The drawback with this solution is that each method has to be rethought to account for any other noise type not considered in its original formulation.

## 2.3 Method

**Adjusting for various noise types** In this paper, we will deal with both the Rician and nc- $\chi$  noise model on a voxelwise basis through the noise stabilization technique of (Koay, Özarslan, and Basser, 2009). This indeed makes our algorithm easily adaptable for any noise type by simply changing the pre-applied transformation as needed. We will use the Probabilistic Identification and Estimation of Noise (PIESNO) (Koay, Özarslan, and Pierpaoli, 2009) to estimate the stationary noise standard deviation. PIESNO works on a slice by slice basis and assumes the background noise as stationary along the selected slice, and is designed to find the underlying standard deviation of the Gaussian noise given its Rician or nc- $\chi$  nature. Voxels that are considered as pure background noise are found automatically by the method, using the fact that the squared mean of those voxels follows a Gamma distribution. Once automatically identified, the standard deviation  $\sigma_G$  of those voxels can be computed and a new estimation of the Gamma distribution is made with the updated  $\sigma_G$  until convergence. In the case of spatially varying noise, we will use a method similar to (Manjón, Coupé, Martí-Bonmatí, et al., 2010), where the noise is estimated locally as

$$\sigma_i^2 = \min \|u_i - u_j\|_2^2, \forall i \neq j, \quad (2.2)$$

with  $u_i$  a noisy patch computed by subtracting a patch to a low-pass filtered version of itself and applying the local Rician correction factor of (Koay and Basser, 2006). If the background was masked automatically by the scanner or is unreliable due to the scanner preprocessing for statistical estimation, we use a similar idea by computing the local standard deviation of the noise field as

$$\sigma_i = \text{std}(u_i - \text{low\_pass}(u_i)) \quad (2.3)$$

If a noise map was acquired during the scanning session, it can be sampled directly to estimate the parameters of the noise distribution. In the event that such a map is unavailable, a synthetic one can be constructed by subtracting the image from its low-pass filtered coun-

terpart (see Eq. (2.3)). Since the noise is assumed as independent and identically distributed across DWIs, we apply a median filter on the 4D dataset to get a 3D noise field. Finally, a Gaussian filter with a full-width at half maximum of 10 mm is applied to regularize the noise field, which is then corrected for the more general nc- $\chi$  bias with the correction factor of (Koay and Basser, 2006). A similar approach based on extracting the noise field with a principal component analysis was used by (Manjón, Coupé, Concha, et al., 2013).

**Locally Adapting the Dictionary Learning** To locally adapt the method to spatially varying noise, we add some more constraints to the classical formulation of Eq. (2.1). Firstly, since the measured signal in diffusion MRI is always positive, we use this assumption to constrain the positivity of the global dictionary  $\mathbf{D}$  and the coefficients  $\alpha$ , i.e.  $\mathbf{D} \geq 0, \alpha \geq 0$  as done in (Gramfort et al., 2014). We fixed the regularization parameter  $\lambda$  for Eq. (2.1) in the same fashion as (Mairal et al., 2009), that is  $\lambda = 1.2/\sqrt{m}$ , with  $m = ps^3 \times an$ ,  $ps$  is the patch size and  $an$  the number of angular neighbors. Secondly, once  $\mathbf{D}$  is known, we use Eq. (2.4) (see the next paragraph) iteratively until convergence with the constraint  $\alpha \geq 0$  and  $\lambda_i = \sigma_i^2(m + 3\sqrt{2m})$ , where  $\sigma_i^2$  is the local noise variance found either with PIESNO or Eq. (2.3). In accordance with (Candès et al., 2008),  $\lambda_i$  is an upper bound on the  $\ell_2$  norm of the noise. We set the convergence as reached for  $\alpha_i$  at iteration  $j$  when  $\max |\alpha_{i,j} - \alpha_{i,j-1}| < 10^{-5}$  or until a maximum of 40 iterations is realized.

**Adaptive and Iterative  $\ell_1$  Minimization** While Eq. (2.1) will both construct the dictionary  $\mathbf{D}$  and find the coefficients  $\alpha$ , there are specialized iterative algorithms for solving  $\ell_1$  problems to yield sparser solutions (Candès et al., 2008; Daubechies et al., 2010). An equivalent constrained formulation for solving each column  $i$  of  $\alpha$  is

$$\min_{\alpha_i} \|w_j \alpha_i\|_1 \quad \text{s.t.} \quad \frac{1}{2} \|x_i - \mathbf{D} \alpha_i\|_2^2 \leq \lambda_i, \quad (2.4)$$

where  $w_j$  is a weighting vector penalizing the coefficients of  $\alpha_i$  at iteration  $j$ . Eq. (2.4) can thus be iterated to further identify non zero coefficients in  $\alpha_i$  by setting  $w_{j+1} = \frac{1}{|\alpha_i| + \epsilon}$  for the next iteration. The algorithm is then started with  $w_0 = 1$  and  $\epsilon = \max |\mathbf{D}^T \xi|$ . As suggested by (Candès et al., 2008),  $\xi \sim \mathcal{N}(0, \sigma^2)$  is set as a random Gaussian vector, which gives a baseline where significant signal components might be recovered. While similar in spirit to Eq. (2.1), Eq. (2.4) provides a way to find the sparser representation for  $\alpha_i$  while bounding the  $\ell_2$  reconstruction error.

To the best of our knowledge, our paper is also the first to use the noise variance as an explicit bound on the  $\ell_2$  reconstruction error. This yields a sparse representation while controlling at the same time the fidelity with respect to the original data, while the classical way is to use the variance as a regularized penalization factor.

### 2.3.1 The proposed algorithm

Our new NLSAM algorithm combines ideas from block matching and sparse coding. We will use the same kind of framework, but by replacing the thresholding part in the block matching with a step of dictionary denoising instead, allowing the penalization on the sparsity of the signal to regularize the noisy blocks. We also take explicit advantage of the fact that diffusion MRI data is composed of multiples volumes of the *same* structure, albeit with different noise realizations and contrasts across DWIs. This allows sparser estimates to be found, further enhancing the separation of the data from the noise (Olshausen and Field, 1996). Our method is thus composed of three steps:

1. Correct the noise bias if needed.
2. Find angular neighbors on the sphere for each DWI.
3. Apply iterative local dictionary denoising on each subset of neighbors.

**Step 1.** In case the noise is not Gaussian distributed, we first correct for the noise bias by finding the Gaussian noise standard deviation with PIESNO (Koay, Özarlan, and Pierpaoli, 2009). If the background is masked, we instead use Eq. (2.3). We then transform the DWIs into Gaussian distributed, noisy signals using the correction scheme of (Koay, Özarlan, and Basser, 2009).

**Step 2.** We find the angular neighbors for each of the DWIs. In this step, the local angular information is encoded in a 4D block of similar angular data, as seen in Fig. 2.1. The gradients are symmetrized to account for opposite polarity DWIs, which share similar structure to their symmetrized counterpart. The search is also made along all the shell at the same time, since structural information (such as sharp edges) is encoded along the axial part of the data. This encodes the similar angular structure of the data along the 4th dimension in a single vector.

**Step 3.** The dictionary  $\mathbf{D}$  is constructed with Eq. (2.1) and the blocks are then denoised with Eq. (2.4). This step can be thought of finding a linear combination with the smallest number of atoms to represent a block. To adapt to spatially varying noise, each block is penalized differently based on the local variance of the noise. This enables the regularization to adapt to the amount of noise in the block, which is usually stronger as the acquired signal is farther from the receiver coils. Since each overlapping block is extracted, each voxel is represented many times and they are recombined using a weighted average based on their sparsity as in (Maggioni et al., 2013; Manjón, Coupé, Concha, et al., 2013). For each voxel  $i$  with intensity  $v_i$  contained in multiple overlapping blocks  $V_j$  in neighborhood  $V$ , we set the final value of  $v_i$  as

$$v_i = \frac{\sum_{j \in V} v_j (1 + \|V_j\|_0)}{\sum_{j \in V} 1 + \|V_j\|_0}, \quad (2.5)$$

where  $V$  is the same spatial position for voxel  $i$  across multiple blocks  $V_j$ . This assumes that more coefficients in block  $V_i$  also mean more noise in the reconstruction. The  $\ell_0$  norm thus penalizes reconstructions with more coefficients and assigns a lower weight in that case for the overlapping weighted average.

This third step is then repeated for all the DWIs. Since each DWI will be processed more than once with a different set of neighbors each time (see Fig. 2.1 for the block formation process), we obtain multiple denoised volumes of *exactly* the same data, but denoised in a different angular context. Once all the DWIs have been processed, we average the multiple denoised versions obtained previously to further reduce any residual noise. See Section 2.7 for an outline of the NLSAM algorithm as pseudocode. The result will be a denoised version of the input, through both dictionary learning and spatial and angular block matching.

### 2.3.2 Datasets and acquisition parameters

**Synthetic phantom datasets** The synthetic data simulations were based on the ISBI 2013 HARDI challenge phantom<sup>1</sup> and were made with phantomas<sup>2</sup>. We used the given 64 gradients set from the challenge at b-values of  $b = 1000$  s/mm<sup>2</sup> and  $b = 3000$  s/mm<sup>2</sup>. For simplicity, we will now refer to these datasets as the  $b = 1000$  s/mm<sup>2</sup> and the  $b = 3000$  s/mm<sup>2</sup> datasets. The datasets were generated with Rician and nc- $\chi$  noise profile, both stationary and spatially varying, at two different signal-to-noise ratios (SNR) for each case. In total, we thus have 8 different noise profiles for each b-value. The stationary noise was generated with SNR 10 and 20 and the spatially varying noise was generated with SNR varying linearly from 5 to 15 and from 7 to 20. The noise distributions were generated for each SNR by setting  $N = 1$  for the Rician noise and  $N = 12$  for the nc- $\chi$  noise. The noisy data was generated according to

$$\hat{I} = \sqrt{\sum_{i=0, j=0}^N \left( \frac{I}{\sqrt{N}} + \beta \epsilon_i \right)^2 + \beta \epsilon_j^2}, \quad \text{where } \epsilon_i, \epsilon_j \sim \mathcal{N}(0, \sigma^2), \quad (2.6)$$

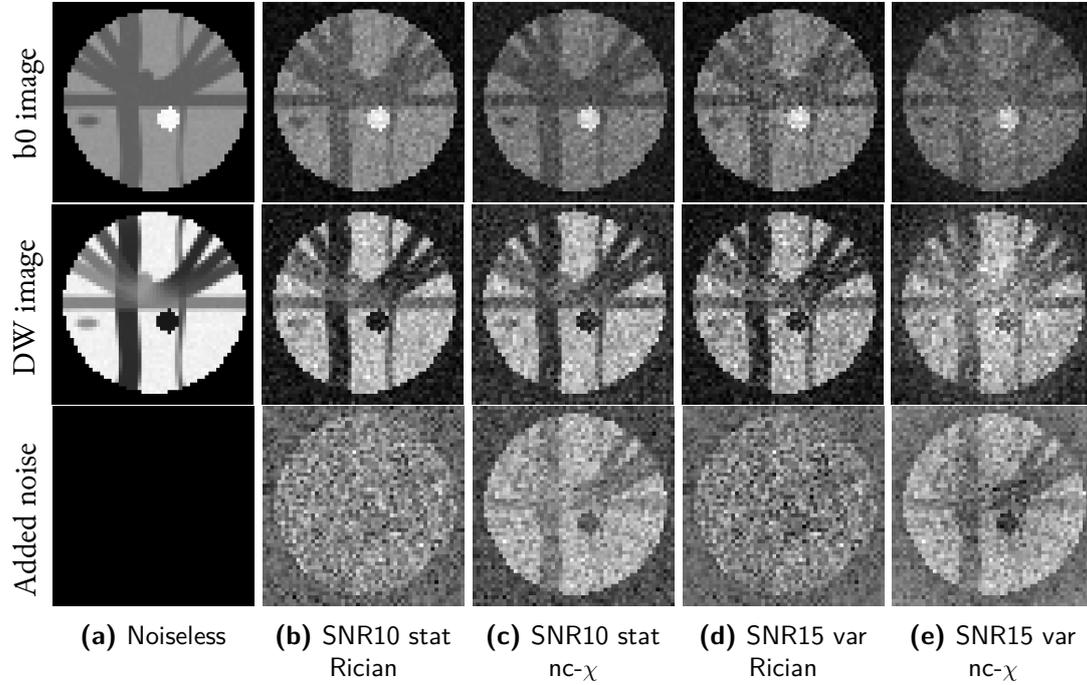
where  $\hat{I}$  is the resulting noisy volume,  $\mathcal{N}(0, \sigma^2)$  is a Gaussian distribution of mean 0 and variance  $\sigma^2$  with  $\sigma = \text{mean}(b0) / \text{SNR}$  and  $\text{mean}(b0)$  is the mean signal value of the  $b = 0$  s/mm<sup>2</sup> image.  $\beta$  is a mask to create the noise distribution set to 1 in the constant noise case and as a sphere for the spatially varying noise case. For the spatially varying noise experiments,  $\beta$  has a value of 1 on the borders up to a value of 3 at the middle of the mask, thus generating a stronger noise profile near the middle of the phantom than for the stationary (constant) noise case. As shown on Fig. 2.4, this results in a variable SNR ranging from approximately SNR 5 and SNR 7 in the middle of the phantom up to SNR

---

<sup>1</sup>[http://hardi.epfl.ch/static/events/2013\\_ISBI/](http://hardi.epfl.ch/static/events/2013_ISBI/)

<sup>2</sup><http://www.emmanuelcaruyer.com/phantomas.php>

15 and SNR 20 for the spatially varying noise case. This noise mimics a homogeneous noise reconstruction as implemented by some scanners while still having a spatially varying noise map.



**Figure 2.4:** Synthetic  $b = 1000 \text{ s/mm}^2$  datasets with various noise profiles used in the experiments. The top row shows the  $b_0$  image, the middle row shows the same DWI across noise types and the bottom row shows the various noise distribution which generated the middle row. From left to right: the noiseless data, SNR 10 with stationary Rician noise, SNR 10 with stationary  $nc-\chi$  noise, SNR 15 with spatially variable Rician noise, SNR 15 with spatially variable  $nc-\chi$  noise.

**Real datasets** To compare our NLSAM method on a real dataset, we acquired a full brain in vivo dataset consisting of 40 DWIs at  $b = 1000 \text{ s/mm}^2$  and one  $b = 0 \text{ s/mm}^2$ . The acquisition spatial resolution was  $1.2 \times 1.2 \times 1.2 \text{ mm}^3$ ,  $TR/TE = 18.9 \text{ s} / 104 \text{ ms}$ , gradient strength of  $45 \text{ mT/m}$  on a 3T Philips Ingenua scanner with a 32 channels head coil for a total acquisition time of 13 minutes. An in-plane parallel imaging factor of  $R = 2$  was used with the SENSE reconstruction algorithm, thus giving a fast spatially varying Rician noise distribution (hence, the denoising algorithms will be set with  $N = 1$ ) even if multiple coils are used by the reconstruction algorithm for producing the final image (see Fig. 2.2). No correction was applied to the dataset, as we wanted to show the effectiveness of denoising without any other preprocessing step such as eddy current or motion correction, which could introduce blurring caused by interpolation. To obtain a comparable clinical-like baseline dataset and show the advantage of acquiring directly high resolution DWIs, we

## 2.3. Method

---

also obtained a 64 DWIs dataset at  $b = 1000 \text{ s/mm}^2$  and one  $b = 0 \text{ s/mm}^2$  of the same subject. The spatial resolution was  $1.8 \times 1.8 \times 1.8 \text{ mm}^3$ ,  $TR/TE = 11.1 \text{ s} / 63 \text{ ms}$ , for a total acquisition time of 12 mins. The acquisition was made on the same scanner, but during another scanning session. No further processing nor denoising was done on this dataset for the reasons mentioned above. This can be thought of having a higher angular resolution at the cost of a lower spatial resolution for a comparable acquisition time.

### 2.3.3 Other denoising algorithms for comparison

We now present the various features and cases covered by the denoising algorithms studied in this paper. The Adaptive Optimized Nonlocal Means (AONLM) method (Manjón, Coupé, Martí-Bonmatí, et al., 2010) is designed for Rician noise removal in a 3D fashion and works separately on each DWIs volume. It also includes a Rician bias removal step and is able to spatially adapt to a varying noise profile automatically. We used the recommended default parameter of a 3D patch size of  $3 \times 3 \times 3$  voxels with the Rician bias correction in all cases. The Local Principal Component Analysis (LPCA) method (Manjón, Coupé, Concha, et al., 2013) is also made to take into account the Rician noise bias and is spatially adaptive, but also uses the information from all the DWIs in the denoising process. We used the automatic threshold set by the method with the Rician noise correction for all experiments. Both AONLM and LPCA can be downloaded from the author's website<sup>3</sup>. The multi-shell Position-Orientation Adaptive Smoothing (msPOAS) algorithm (Becker et al., 2014) was designed for both Rician and  $nc\text{-}\chi$  noise, while also taking into account the angular structure of the data for adaptive smoothing. We discussed with the authors of msPOAS<sup>4</sup> for their recommendations and using their suggestion, we set  $k^* = 12$  and  $\lambda = 18$ . We also supplied the correct value for  $N$  and used the implemented automatic detection of the noise standard deviation from msPOAS. For the NLSAM algorithm, we used a patchsize of  $3 \times 3 \times 3$  voxels with 4 angular neighbors, which correspond to the number of angular neighbors at the same distance on the sphere for each selected DWI. The value of  $N$  was given to the algorithm and the number of atoms was set to two times the number of voxels in a block for the dictionary learning part, which was repeated for 150 iterations. The other parameters were set as described in Section 2.3. As shown in Table 2.1, our method is designed to work on both stationary and spatially variable Rician and  $nc\text{-}\chi$  noise. The NLSAM algorithm is implemented in python and is freely available.<sup>5</sup>

Finally, we quantitatively assess the performance of each method by comparing them against the noiseless synthetic data using

- i) The peak signal-to-noise ratio (PSNR) in dB and the structural similarity index (SSIM) on the raw data intensities (Wang et al., 2004).

---

<sup>3</sup><http://personales.upv.es/jmanjon/denoising/index.htm>

<sup>4</sup><http://cran.r-project.org/web/packages/dti/index.html>

<sup>5</sup><https://github.com/samuelstjean/nlsam>

- ii) The dispersion of the FA error, computed from a weighted least-square diffusion tensor model.
- iii) The mean angular error (AE) in degrees and the discrete number of compartments (DNC) error for a region of crossings (Daducci et al., 2014; Paquette et al., 2015).
- iv) The Tractometer (Côté, Girard, et al., 2013) ranking platform on deterministic tractography algorithms for the synthetic datasets. This platform computes global connectivity metrics, giving an insight on the global coherence of the denoised datasets in a tractography setting.
- v) Tracking some known bundles on the high resolution in vivo dataset and qualitatively comparing them to their lower spatial resolution counterpart.

#### 2.3.4 Local models reconstruction and fiber tractography

The weighted least-square diffusion tensors were reconstructed using the default parameters of Dipy (Garyfallidis et al., 2014) to compute the FA values. We used the Constrained Spherical Deconvolution (CSD) (Tournier et al., 2007) with a spherical harmonics of order 8 to reconstruct the fODFs and extract the peaks subsequently used for the deterministic tracking. To compute the fiber response function (frf), we used all the voxels in the white matter that had an FA superior to 0.7. If less than 300 voxels meeting this criterion were found, the FA threshold was lowered by 0.05 until the criterion was met. See Sections 2.4.2 and 2.5.3 for more information about the bias introduced in the FA. For the synthetic datasets, the tracking was done inside the white matter mask and the seeding was done from the bundles extremities to mimic seeding from the white-gray matter interface (Girard et al., 2014). We used 100 seeds per voxels to allow sufficient bundle coverage, a stepsize of 0.2 mm and a maximum angle deviation of 60 degrees. The other parameters used were the defaults supplied by the tractometer pipeline (Côté, Girard, et al., 2013). The in vivo datasets deterministic tracking was made with the technique of (Girard et al., 2014) by seeding from the white matter and gray matter interface with the particle filtering and generating approximately 1 million of streamlines. White matter masks were created by segmenting a T1-weighted image with FSL FAST<sup>6</sup> from the same subject and then registered with ANTs<sup>7</sup> to each in vivo dataset. The bundles were finally automatically segmented using the White Matter Query Language (WMQL) (Wassermann et al., 2016) Tract Querier tool with regions obtained from a T1-weighted white matter and gray matter parcellation. This atlas-based automatic dissection method extracts fiber bundles automatically using anatomical definitions in a reproducible manner for all methods, as opposed

<sup>6</sup><http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FAST>

<sup>7</sup><http://picsl.upenn.edu/software/ants/>

to the traditional way of manually defining including and excluding ROIs to define bundles. Visualization of fODFs, peaks and tractography was made using the fibernavigator<sup>8</sup> (Chamberland et al., 2014).

## 2.4 Results

### 2.4.1 Preserving the raw DWI data

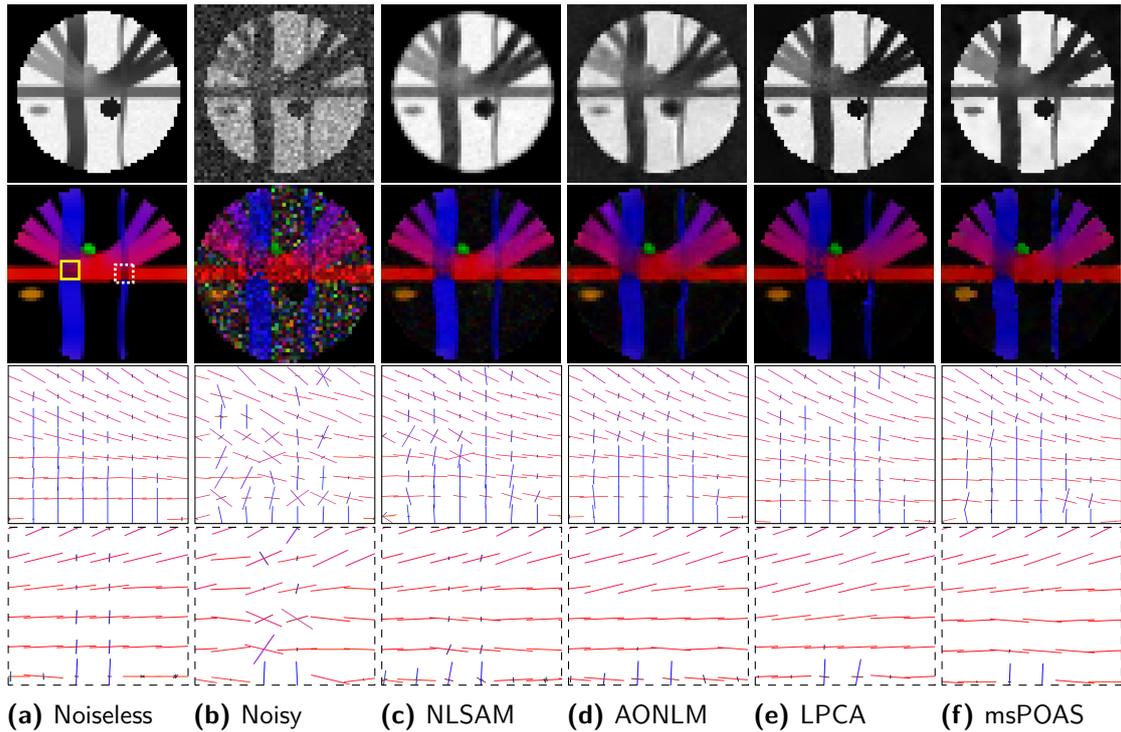
Fig. 2.5 shows the  $b = 1000 \text{ s/mm}^2$  noiseless data, the noisy input data at SNR 10 for  $nc\text{-}\chi$  stationary noise and the results of the denoising on the synthetic phantom for all compared methods. This is the noise case theoretically covered by msPOAS and our NLSAM algorithm. We also show two zoomed regions of crossings with the reconstructed peaks extracted from fODFs. All perceptual and FA metrics were computed on the slice shown while angular metrics were computed in the zoomed region depicted by the yellow box. Note how the small blue bundle and its crossings are preserved on the NLSAM denoised dataset, while other denoising methods tend to introduce blurring.

Fig. 2.6 shows the noisy high resolution in vivo dataset, the denoised version obtained for each algorithm and the low spatial resolution acquisition of the same subject without any denoising. Since our scanner uses a 32 channels head-coil but implements the SENSE reconstruction algorithm, the resulting spatially varying Rician noise distribution is the case covered by AONLM, LPCA and our NLSAM algorithm. We show a coronal slice for the gradient direction closest to  $(0, 1, 0)$ , the colored FA map and a zoom on two regions of crossings. The yellow region shows the junction of the corticospinal tract (CST) and superior longitudinal fasciculus (SLF) while the white region shows the junction of the corpus callosum (CC) and the CST. While the high resolution dataset is noisier than its lower resolution counterpart is, the highlighted crossings regions are well recovered by the denoising algorithms and thus offer an improvement in anatomical details over the lower spatial resolution dataset. We also see in the yellow box that the NLSAM denoised dataset recovers crossings extending from the CC which are almost absent in the compared datasets.

Fig. 2.7 shows the PSNR and SSIM for the SNR 10 (stationary noise) and SNR 15 (spatially varying noise) synthetic datasets. The LPCA algorithm performs best in term of PSNR on the Rician noise case, but attains a lower score for  $nc\text{-}\chi$  noisy datasets. The same trend is seen for AONLM and msPOAS algorithms, where the SNR 15  $nc\text{-}\chi$  case is the hardest test case. In contrast, our NLSAM technique is above 30 dB for the PSNR and 0.9 for the SSIM in most cases, with a relatively stable performance amongst the majority of tested cases. We also note that even though msPOAS is made to adjust itself to  $nc\text{-}\chi$  noise, the fact that the algorithm does not account for the intensity bias makes the perceptual metrics drop for the  $nc\text{-}\chi$  noise cases.

---

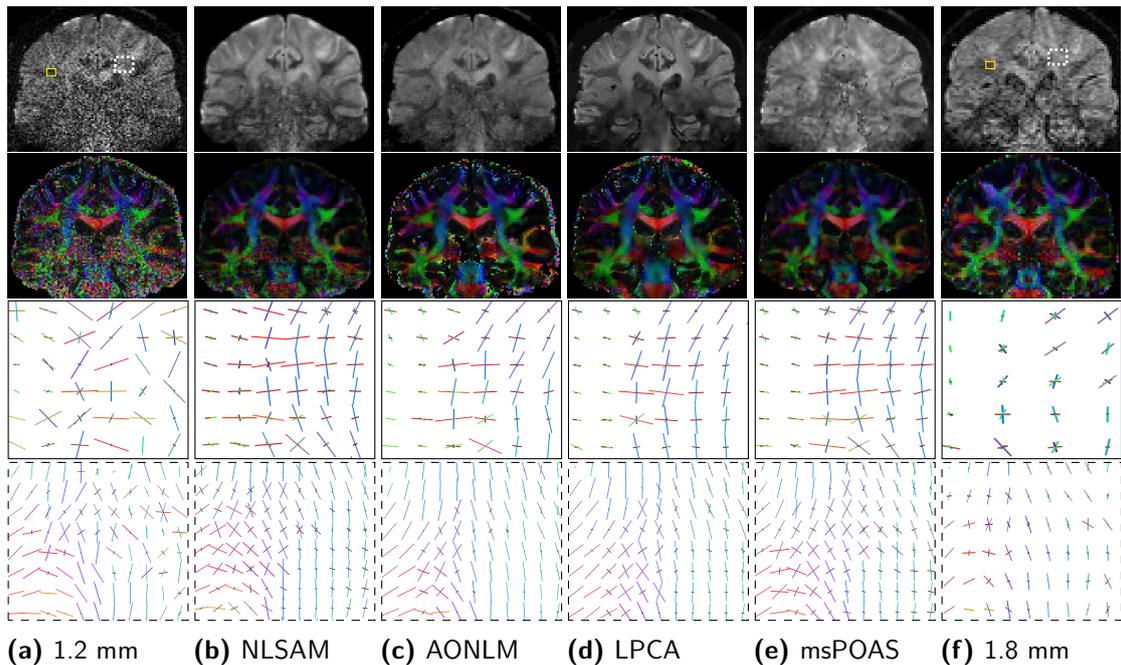
<sup>8</sup><https://github.com/scilus/fibernavigator>



**Figure 2.5:** Phantoms  $b = 1000 \text{ s/mm}^2$  synthetic dataset at SNR 10 for stationary  $nc\text{-}\chi$  noise on the  $y = 24$  slice. From top to bottom: Raw diffusion MRI, colored FA map, zoom on extracted peaks from fODF of order 8. Note how NLSAM restores the structure without blurring on the colored FA map and is the only method to restore the peaks from the noisy dataset in the zoomed white box region.

#### 2.4.2 Bias introduced in the FA

As shown by the FA difference map on Fig. 2.8, our NLSAM method commits a small FA error locally with a smaller maximum error than the compared methods. Voxelwise underestimation is denoted in blue and overestimation in red, where white means the computed value is close to the expected value. The noisy data largely overestimates the FA values for the synthetic datasets, while other denoising methods underestimate the real FA value most of the time. On the  $b = 1000 \text{ s/mm}^2$  datasets, NLSAM has the smallest spread of FA error. The effect of stabilizing the data prior to denoising can also be seen by the stable FA median error committed by NLSAM across all noise types. For the  $b = 3000 \text{ s/mm}^2$  datasets, the need to correct the intensity bias caused by the noise becomes more important, as seen by the increased error in underestimating the correct FA value for most methods. For the spatially varying Rician noise case, our method commits the lowest overestimation, as opposed to AONLM and LPCA, which are developed for this particular noise case. It is also important to note that in contrast to the other methods, msPOAS does not explicitly correct for the intensity bias by design, but rather leaves this correction

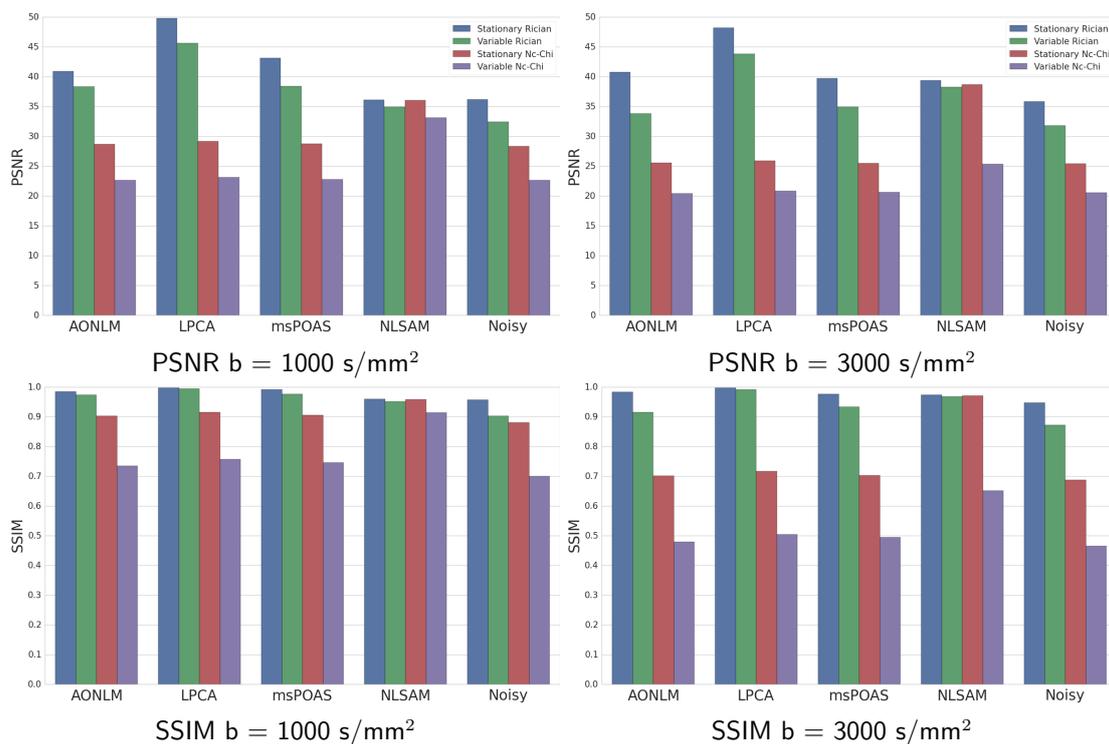


**Figure 2.6:** From top to bottom, the raw high resolution in vivo data corrupted with spatially varying Rician distributed intensities, the colored FA map and a zoom on two regions of crossings. All denoised methods were applied on the high spatial resolution 1.2 mm dataset. We also show an acquisition of the same subject at 1.8 mm for visual comparison. Our NLSAM algorithm is able to recover more crossings from the 3 way junction of the SLF, the CST and the CC as shown in the yellow and white boxes. While the 1.8 mm dataset is less noisy, its lower spatial resolution also means that each voxel contains more heterogeneous tissues and mixed diffusion orientations. The 1.2 mm denoised dataset shows more crossings without the averaging effect of the larger voxel size. For a comparable acquisition time, the denoised high resolution dataset has more information than its lower resolution counterpart without processing.

to subsequent processing steps. The SNR 15  $nc-\chi$  noise case is where all the methods make the biggest error, as they reduce the variance but still suffer from a large bias in FA. Overall, our method restores the value of the FA for large bundles more accurately. We also see that most methods make their largest error near the partial volume ball mimicking cerebrospinal fluid (CSF).

### 2.4.3 Impact on angular and discrete number of compartments (DNC) error

We now study the angular error and the mean relative error in the discrete number of compartments (DNC) (Daducci et al., 2014; Paquette et al., 2015). The mean relative discrete number of compartments error is defined as  $DNC_i = 100 \times |P_{i_{true}} - P_{i_{est}}| / P_{i_{true}}$  for voxel  $i$ ,  $P_{i_{true}}$  and  $P_{i_{est}}$  is the number of crossings respectively found on the noiseless dataset and on the compared dataset. All metrics were computed on the voxels containing

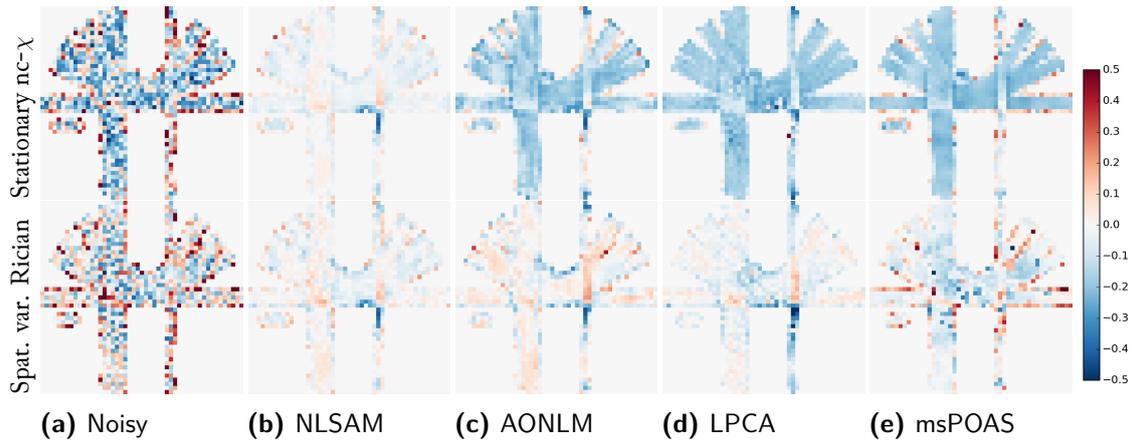


**Figure 2.7:** PSNR and SSIM metrics for the SNR 10 stationary and SNR 15 spatially variable noise cases datasets. All methods can correct the stationary and spatially varying Rician noise case to some extent while only our NLSAM algorithm has the best performance for the nc- $\chi$  noise case, especially for the spatially varying noise case.

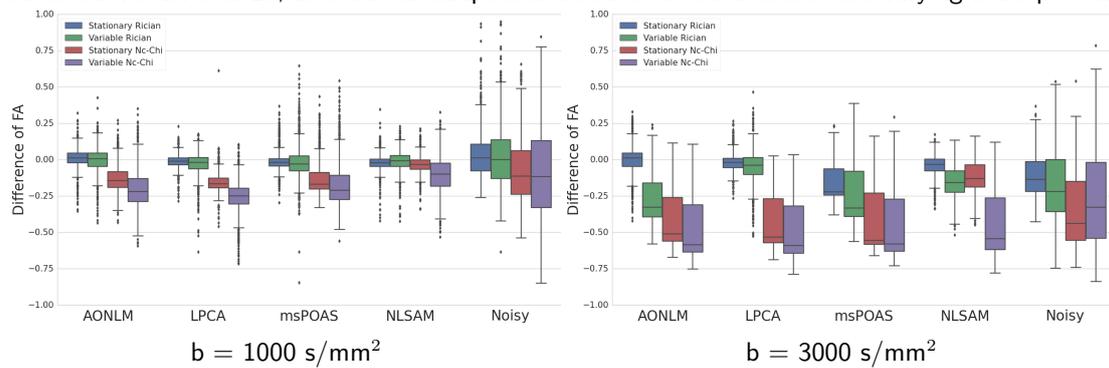
at least two crossings fibers on the noiseless dataset shown previously in Fig. 2.5.

Fig. 2.9 shows the distribution of the angular error and of the DNC error found in the region studied in addition to the mean angular error. All of the denoising algorithms have a lower median and mean angular error than the noisy datasets. In addition, the NLSAM denoised datasets have an almost equal or lower angular error than the other denoising methods, but with a lower maximum error most of the time as shown by the smaller whiskers. For the  $b = 1000$  s/mm<sup>2</sup> dataset DNC error, all three of AONLM, LPCA and NLSAM improve on the noisy dataset for the Rician noise case as they are devised for this kind of data. LPCA also has a better performance than the other two for the spatially varying Rician noise case, while NLSAM has a lower mean DNC error for both of the nc- $\chi$  noise case. The effect of the intensity bias is also seen on msPOAS, where the DNC error is always lower than the noisy dataset, but also higher than all the other methods that take into account the intensity bias. The  $b = 3000$  s/mm<sup>2</sup> dataset is much harder, where no method seems to have a clear advantage in all cases over the others. One interesting thing to note is that the noisy dataset has a low DNC error for both of the Rician noise case, but the confidence interval indicates it is in the same range as the denoised datasets.

## 2.4. Results



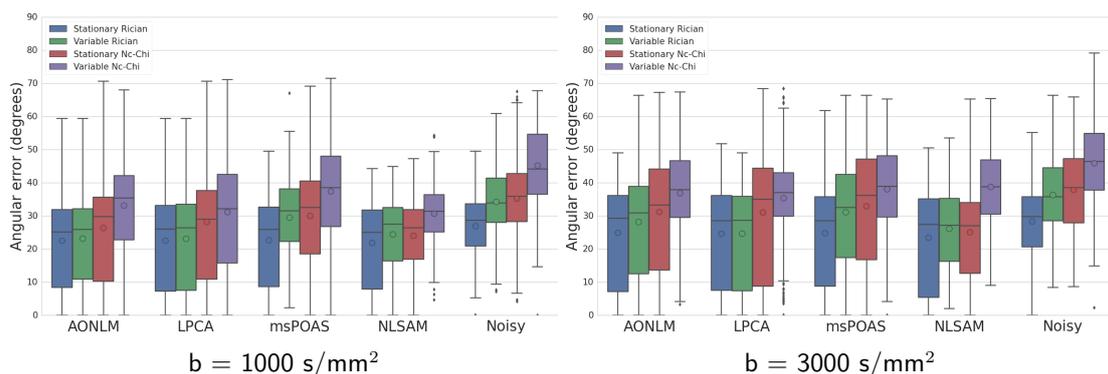
FA difference for the phantoms stationary  $nc-\chi$  SNR 10 and spatially varying Rician SNR 15  $b = 1000 \text{ s/mm}^2$  datasets. Blue values denote underestimation while red values show overestimation of the FA. **Top:** Stationary  $nc-\chi$  noise. NLSAM is less biased than the other methods in large, homogeneous regions, while the compared methods produces more underestimation for the  $nc-\chi$  case. **Bottom:** Spatially variable Rician noise. While being a harder case than the SNR 10 dataset since it varies from SNR 5 to 15, all methods adapt themselves to some extent to the varying noise profile.



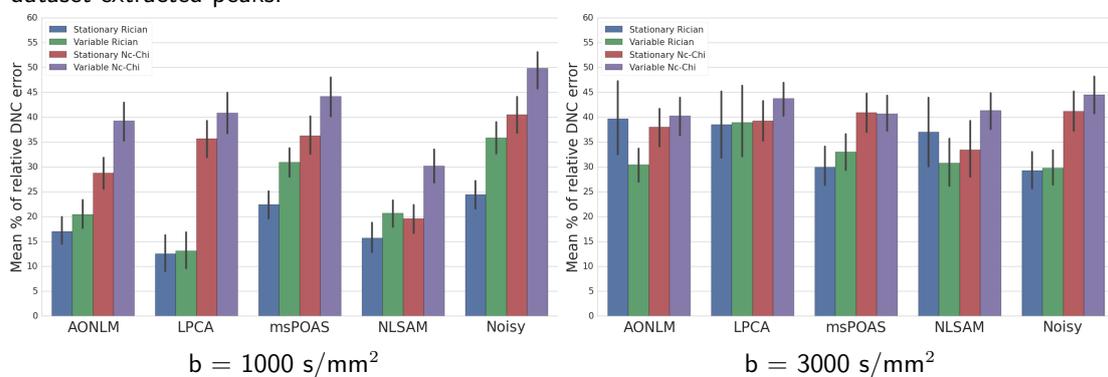
**Figure 2.8:** Boxplot of the difference in FA for the synthetic datasets at  $b = 1000 \text{ s/mm}^2$  (left) and  $b = 3000 \text{ s/mm}^2$  (right). The whiskers show 1.5 times the interquartile range ( $1.5 \times \text{IQR}$ ), where outliers are plotted individually. The bars represent the first quartile, the median and the third quartile. No method performs well on the  $nc-\chi$   $b = 3000 \text{ s/mm}^2$  spatially varying noise case, which is the hardest test case. NLSAM overall produces less error or is equal to the other methods, but has a lower bias in the FA error along noise type.

### 2.4.4 Impact on tractography

We now show how denoising techniques impact tractography by evaluating the number of valid bundles (VB), invalid bundles (IB) (Côté, Girard, et al., 2013) and the valid connection to connection ratio (VCCR) (Girard et al., 2014) found by the tracking algorithm. A valid bundle is defined as connecting two ROIs in the ground truth data while an invalid bundle is a connection made between two ROIs which is not supported by the ground truth data. The valid connection to connection ratio is the total of valid connections (VC) over the



Boxplot of the angular error in degrees on the synthetic datasets, where the dot represents the mean angular error. A low angular error means that the extracted fODFs peaks are aligned with the noiseless dataset extracted peaks.



**Figure 2.9:** The mean relative percentage of DNC error for the synthetic datasets. The bar represents the 95% confidence interval on the mean as computed by bootstrapping. The DNC error is the number of peaks found in excess or missing in each voxel with respect to the noiseless dataset.

sum of valid and invalid connections (IC), i.e.  $VCCR = VC / (VC + IC)$ . A good denoising algorithm should find a high number of valid bundles, a low number of invalid bundles and a high percentage of valid connection to connection ratio.

**Deterministic tractography on the synthetic phantom** Table 2.2 shows the results of deterministic tractography on the SNR 10, 15 and 20 synthetic datasets for both  $b = 1000 \text{ s/mm}^2$  and  $b = 3000 \text{ s/mm}^2$ . The noiseless  $b = 1000 \text{ s/mm}^2$  dataset had 25/27 valid bundles, 55 invalid bundles and a valid connection to connection ratio of 65% and the noiseless  $b = 3000 \text{ s/mm}^2$  dataset had 27/27 valid bundles, 40 invalid bundles and a valid connection to connection ratio of 68%. One of the first thing to note is that even though the noisy dataset always has a high number of valid bundles, it is at the price of a huge number of invalid bundles. Moreover, the valid connection to connection ratio is systematically lower for the SNR 10 datasets than any of the denoising methods. This indicates that only looking at the number of valid and invalid bundles does not show how many streamlines

## 2.4. Results

reached each region since only at least one streamline is required to make a connection, thus counting as a valid bundle. Another observation is that denoising helps controlling the number of invalid bundles and gives a better valid connection to connection ratio in most cases over the noisy data. For the SNR 15 cases, NLSAM has the highest number of valid bundles in almost all cases, but at the price of a larger number of invalid bundles at lower SNR. Another interesting trend is the tradeoff between valid bundles and invalid bundles: AONLM and LPCA both manage to get a lower number of invalid bundles, but also tend to have a lower number of valid bundles than msPOAS or NLSAM overall.

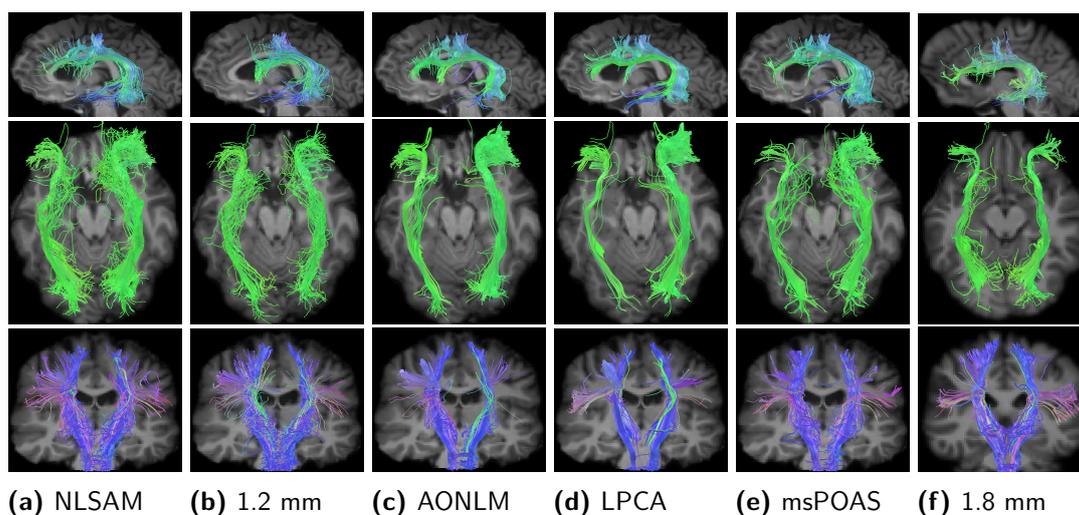
For the SNR 20 stationary noise cases, all methods are close in valid bundles with some difference in the number of invalid bundles. This shows that tractography could benefit from variable tracking parameters instead of fixed values depending on the preferred trade-off for the task at hand (Chamberland et al., 2014).

Method / Noise			Stationary noise						Spatially variable noise					
			SNR 10			SNR 20			SNR 15			SNR 20		
			VB	IB	VCCR	VB	IB	VCCR	VB	IB	VCCR	VB	IB	VCCR
AONLM	b = 1000 s/mm <sup>2</sup>	Rician	25	78	49%	25	75	51%	23	91	45%	25	89	50%
		nc- $\chi$	25	88	50%	26	88	52%	21	111	44%	23	93	47%
	b = 3000 s/mm <sup>2</sup>	Rician	25	69	52%	25	60	56%	24	85	50%	26	72	52%
		nc- $\chi$	25	78	55%	26	67	55%	20	95	48%	22	78	54%
LPCA	b = 1000 s/mm <sup>2</sup>	Rician	23	61	49%	25	64	54%	16	36	42%	18	38	45%
		nc- $\chi$	22	66	50%	24	70	54%	16	46	51%	20	56	52%
	b = 3000 s/mm <sup>2</sup>	Rician	23	44	47%	26	46	53%	17	37	42%	19	41	45%
		nc- $\chi$	20	42	58%	25	57	53%	18	40	55%	20	56	55%
msPOAS	b = 1000 s/mm <sup>2</sup>	Rician	25	101	49%	25	89	52%	23	129	44%	25	118	46%
		nc- $\chi$	23	121	40%	25	95	54%	20	131	35%	25	141	41%
	b = 3000 s/mm <sup>2</sup>	Rician	26	108	53%	26	74	58%	25	88	52%	25	93	49%
		nc- $\chi$	17	84	37%	25	84	57%	22	96	33%	23	94	47%
NLSAM	b = 1000 s/mm <sup>2</sup>	Rician	25	90	49%	26	96	54%	25	127	42%	25	114	45%
		nc- $\chi$	25	120	48%	25	90	54%	25	170	28%	26	144	43%
	b = 3000 s/mm <sup>2</sup>	Rician	25	92	50%	26	67	54%	25	108	43%	25	97	47%
		nc- $\chi$	23	100	45%	24	82	53%	23	173	29%	25	131	37%
Noisy	b = 1000 s/mm <sup>2</sup>	Rician	25	138	41%	25	107	53%	25	159	36%	25	134	42%
		nc- $\chi$	25	166	34%	26	119	49%	17	120	9%	25	209	24%
	b = 3000 s/mm <sup>2</sup>	Rician	25	116	46%	27	87	54%	25	160	36%	25	149	42%
		nc- $\chi$	25	182	36%	26	103	53%	18	124	9%	25	210	24%
Noiseless	b = 1000 s/mm <sup>2</sup>		VB	IC	VCCR									
			25	55	65%									
	b = 3000 s/mm <sup>2</sup>		27	40	68%									

**Table 2.2:** Tractometer results for the deterministic tracking.

**Tracking the real data** We now look at tractography on the in vivo high spatial resolution dataset and its clinical spatial resolution counterpart of the same subject previously shown on Fig. 2.6. The high spatial resolution dataset at 1.2 mm isotropic has 40 unique gradient directions while the lower spatial resolution dataset at 1.8 mm isotropic has 64 unique

gradient directions for a comparable acquisition time. The background is masked by the scanner and has a spatially varying Rician noise profile due to the SENSE reconstruction, which is the specific noise case covered by the AONLM and LPCA denoising algorithm. We use the deterministic tractography algorithm from (Girard et al., 2014), which considers anatomical constraints for more anatomically plausible tractography. Fig. 2.10 shows from top to bottom the left arcuate fasciculus (AF), the inferior fronto-occipital fasciculus (IFOF) and the corticospinal tract (CST) as dissected automatically by the Tract querier (Wassermann et al., 2016). The noisy 1.2 mm AF stops prematurely in the frontal part of the bundle, while the 1.8 mm noisy AF misses the temporal lobe. In contrast, the streamlines from the NLSAM denoised bundle go further into the temporal lobe. Also, note how the right IFOF has a better coverage for all the 1.2 mm datasets and more fanning near the front of the brain than the noisy 1.8 mm dataset. We also see that the left IFOF is thinner than its right counterpart is, but most of the bundles tracked from the denoised datasets produce less spurious tracks while keeping the anatomical details. The LPCA denoised IFOF stops prematurely for the left posterior part of the bundle, possibly because of a lost crossing region along the fibers during the denoising process. The CST does show some commissural fibers through the pons in the noisy 1.8 mm dataset, while they are present but look like spurious fibers on the noisy 1.2 mm dataset. AONLM can recover some of those commissural fibers, while NLSAM is the only algorithm that recovers them in addition to richer fanning near both sides of the motor cortex.



**Figure 2.10:** Deterministic tractography for selected bundles on the in vivo dataset. We also show a T1-weighted image aligned in the diffusion space for anatomical reference. **Top:** The left arcuate fasciculus. Note how the denoised NLSAM arcuate fasciculus goes further into the frontal and temporal region than both of its noisy 1.2 mm and 1.8 mm counterparts. **Middle:** The inferior fronto-occipital fasciculus. The AONLM denoised bundle has a denser part for the right IFOF while the LPCA bundle stops prematurely for the left IFOF, possibly due to a missing crossing along the bundle. **Bottom:** The corticospinal tract. We see that NLSAM recovers the commissural fibers in the pons from the noisy 1.2 mm dataset, which are not even present on the noisy 1.8 mm dataset nor on the other denoising algorithm’s bundles. NLSAM also recovers more fanning to both sides of the brain than all the compared methods.

## 2.5 Discussion

### 2.5.1 Enhancing the raw data

We quantitatively showed in [Fig. 2.7](#) that denoising restores perceptual information when compared to the unprocessed noisy data. Taking the spatially varying aspect and the particular nature of the noise into account is also important since modern scanners implement parallel imaging, which changes the nature of the noise ([Dietrich et al., 2008](#)), leading to a lower performance for denoising methods not fully taking into account the introduced bias. [Fig. 2.6](#) shows that this is also qualitatively true for in vivo data, where denoising visually restores information in regions heavily corrupted by noise. While perceptual metrics might indicate the performance of an algorithm, one must remember that the relative signal difference is of interest in diffusion MRI, which is not fully captured by perceptual metrics like the PSNR or the SSIM. One is also usually interested in diffusion MRI metrics as opposed to perceptual information brought by the raw diffusion MRI datasets. AONLM is able to remove most of the noise, but still shows some residuals near the inferior part of the brain, possibly due to only considering the 3D volumes separately, which means that

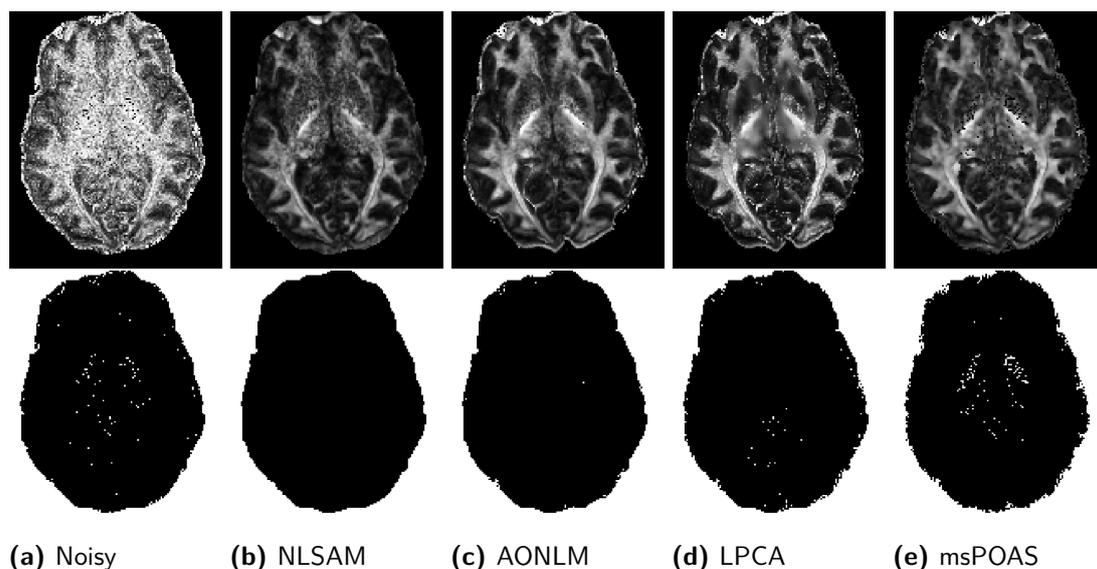
the algorithm can not benefit from the additional angular information brought by diffusion MRI. LPCA can restore visual information and sharp edges from the noisy dataset, but the region in the pons, where the noise level is higher and crossing fibers are more complex, also seems to be piecewise constant. This might arise from the fact that the algorithm uses all of the DWIs at once for its PCA decomposition step and treats all intensities at the same in the noise removal step. As for NLSAM, the algorithm only works in the local angular domain, thus exploiting similar contrast and redundant edge structure under different noise realization. msPOAS also uses a similar idea, where the angular similarity is weighted according to the Kullback-Leibler divergence to control the importance of dissimilar intensities in the denoising process. Nevertheless, these perceptual metrics show that denoising improves upon the noisy data, but one should also look at metrics derived from the studied object of interest i.e. tensor or fODF derived metrics, since high perceptual metrics might also reflect blurring of diffusion features, which is the main interest in this type of acquisition rather than the perceived quality.

### 2.5.2 Impact of the stabilization algorithm on the compared denoising methods

Fig. 2.11 shows the FA map when the compared denoising algorithm are applied on the stabilized data with the algorithm of (Koay, Özarlan, and Basser, 2009). For this experiment, we consider a voxel as being degenerated if its FA is exactly 0. The first thing to note is that the algorithm only reprojects the noisy data on plausible Gaussian distributed values and does not do any denoising. While we used here the algorithm of (Koay, Özarlan, and Basser, 2009) to correct the noise bias, another interesting approach consist of producing real-valued datasets as shown in (Eichner et al., 2015). This approach does not require estimation of  $\sigma^2$  or  $N$ , but instead use information contained in the phase of complex-valued acquisitions. Secondly, all of the other compared denoising algorithms produce some invalid voxels on the raw dataset, while having less degenerated voxels on the stabilized dataset as shown in Table 2.3. Nevertheless, only our NLSAM algorithm does not produce any degenerated FA voxel on the in vivo dataset. As tractography might rely on a thresholded FA mask (Chamberland et al., 2014), any missing white matter voxel will end the tractography early and produce anatomically invalid tractography. In the same way, computing FA based statistics in search of group differences inside a white matter mask might lead to erroneous conclusions when degenerated voxels are present. This undesirable side effect should be avoided when possible by choosing a method producing a low number of invalid voxels, such as NLSAM.

### 2.5.3 Reducing the diffusion metrics bias

Fig. 2.8 shows that knowing where errors are committed gives a better view of how denoising improves upon the noisy data. We see that our NLSAM algorithm actually has a smaller maximum error in underestimating the FA most of the time while other meth-



**Figure 2.11:** Effect of the stabilization algorithm on the compared methods. The top row shows an axial slice of the in vivo FA map computed on the stabilized dataset, where some voxels are degenerated. The bottom row shows their location on a binary brain mask. As shown in [Table 2.3](#), all methods produce degenerated FA voxels on both the regular and stabilized data, with the sole exception of NLSAM.

		AONLM	LPCA	msPOAS	NLSAM	Noisy	Mask
Built-in	Brain mask	83 314 (10.1%)	10 526 (1.3%)	84 319 (10.2%)	∅	5 994 (0.7%)	823 068 (100%)
	WM mask	29 664 (5.1%)	1 298 (0.2%)	16 665 (2.9%)	∅	1 769 (0.3%)	578 418 (100%)
Stabilization	Brain mask	10 052 (1.2%)	15 750 (1.9%)	29 377 (3.6%)	0 (0%)	9 395 (3.6%)	823 068 (100%)
	WM mask	404 (0.1%)	1 468 (0.3%)	4 267 (0.7%)	0 (0%)	2 850 (0.5%)	578 418 (100%)

**Table 2.3:** Number of degenerated FA voxels inside a brain mask and a white matter mask for the in vivo dataset. All methods were compared with their built-in noise estimation on the stabilized version, but without any additional noise correction factor. The percentage of degenerated voxels is indicated in parenthesis for each mask, where a voxel is considered degenerated if its FA value is exactly 0.

ods both over and underestimate the real FA value and make larger errors near CSF or at borders with the background. This could indicate that they are subject to problems with partial volume effect, which seems less important for NLSAM.

While stabilizing the data alleviates the FA underestimation problem in most cases, it also helps to reduce the number of degenerated voxels in the in vivo data as shown in [Fig. 2.11](#). Both AONLM and msPOAS produce less degenerated FA voxels on the stabilized dataset as shown in [Table 2.3](#), while NLSAM does not produce any degenerated voxel. In contrast, the noisy data and LPCA have an increased number of degenerated FA voxels, which might be caused by the diffusion signal being near the noise floor, thus producing a flat profile that is not properly recovered in this case. Reducing the FA bias

and avoiding degenerated voxels by including denoising in the processing pipeline could improve the statistical analysis in along-tract metrics (Colby et al., 2012) when looking for group differences.

#### 2.5.4 Restoring the coherence of local models

The CSD algorithm relies on the estimation of the fiber response function (frf), which in turn relies on the diffusion tensor. To estimate the frf, one must select voxels containing only a single fiber population. One way to do this is to estimate it from voxels with a high FA, usually with  $FA > 0.7$  (Descoteaux, Deriche, et al., 2009; Tournier et al., 2007). We observed that for the SNR 10 dataset with  $nc-\chi$  noise, the noisy dataset, AONLM and LPCA could not find as many single fiber voxels based on the FA threshold method as msPOAS or NLSAM since their reconstructed tensors have an inherently lower FA. This in turn impacts deconvolution since the estimates used for the deconvolution kernel are less stable, and lowering the FA threshold too much might violate the single fiber assumption, which is crucial for the CSD algorithm. One way to circumvent this could be by using the method of (Tax et al., 2014), which is based on a peak amplitude criterion instead of an FA threshold to identify single fiber voxels.

Fig. 2.9 shows that msPOAS and NLSAM have larger angular error than AONLM or LPCA, but this does not seem to impact much the number of valid bundles found by deterministic tractography. Indeed, the noisy data has the largest angular error in all cases, but still has a high number of valid bundles in most cases. This also suggests that a large overestimation or underestimation of fiber crossings (as reflected by the DNC error) has a higher impact on tractography. Both LPCA and msPOAS have a lower number of valid bundles than AONLM or NLSAM, which both have a rather symmetric under and over estimation of the number of peaks. This means that an overall estimation or underestimation of the number of crossings bias tractography, as it tends to follow false crossings or stops prematurely due to a lack of crossings, while a distributed error is not skewed toward these effects.

In Fig. 2.6, we see that denoising restores coherence in regions of crossing fibers that were lost on the noisy dataset or not even present in the lower spatial resolution 1.8 mm dataset due to a smaller voxel size. We also see that NLSAM restores more coherent crossings than the other denoising methods in the junction of the SLF and the CST, whereas the noisy dataset only shows incoherent crossings. This actually enables tractography algorithms to reconstruct tracts that are in agreement with the expected anatomy. In the same amount of acquisition time, one can thus acquire higher spatial resolution DWIs and get better angular information by post-processing the acquired data with denoising.

### 2.5.5 Limiting spurious fibers from tractography

We studied the impact of denoising techniques on deterministic tractography on a synthetic dataset in [Section 2.4.4](#). One often has to choose between finding a high number of valid bundles and invalid bundles, and finding less valid bundles and at the same time reducing the amount of invalid bundles. The noisy dataset always reaches a high number of valid bundles, but also at the price of having the highest number of invalid bundles most of the time. Our NLSAM algorithm shows a good balance between the number of valid and invalid bundles at low SNR, especially for the spatially varying noise case. This is always a tradeoff as seen in the ISMRM 2015 tractography challenge<sup>9</sup>.

For example, the LPCA algorithm has always a low number of invalid bundles, but also the lowest number of valid bundles for the spatially varying noise case. In opposition, NLSAM has a high number of valid bundles, but also a high number of invalid bundles most of the time.

Regarding the deterministic tractography, changing the stepsize or the maximum curving angle would give different results in terms of connectivity metrics, indicating that the tractography algorithm and chosen tractography parameters have a non-negligible influence on the results (Chamberland et al., 2014; Girard et al., 2014). We also used a seeding strategy of 100 seeds per voxel from the ROIs at each bundles endpoints to ensure a maximal number of valid bundles, which promotes a high number of valid bundles for each dataset. This shows that the missed bundles are hard to recover or not supported by the data itself, as opposed to being missed because of inadequate seeding (Côté, Girard, et al., 2013). On the other hand, this can artificially increase the number of invalid bundles, which could be reduced by reducing the number of seeds per voxel. Since automatic tractography pruning techniques such as (Côté, Garyfallidis, et al., 2015) might help reduce the number of spurious tracks, this would indicate that having a higher number of valid bundles would be preferable since invalid bundles could be potentially removed afterward. In contrast, a low number of valid bundles cannot be circumvented with further post-processing. Nevertheless, denoising increases the valid connection to connection ratio and reduces the number of invalid bundles, thus bringing confidence in the validity of the tractography results when compared to the noisy datasets.

For the in vivo dataset tracking shown in [Fig. 2.10](#), we see that tractography benefits from higher spatial resolution acquisitions, but that the produced tracts are slightly noisier. Combining the high spatial resolution, highly noisy dataset with a denoising algorithm at the beginning of the processing pipeline gives more anatomically plausible tracts in the end. The AF and CST produced by the NLSAM denoised dataset are both more anatomically plausible than their noisy or lower spatial resolution counterpart, which have less fanning fibers in the case of the CST. This shows that high resolution DWIs exhibits additional anatomical information due to the smaller voxel size, which might not be easily

---

<sup>9</sup>[http://www.tractometer.org/ismrm\\_2015\\_challenge/](http://www.tractometer.org/ismrm_2015_challenge/)

discernible at a lower spatial resolution (Sotiropoulos et al., 2013). Acquiring at higher spatial resolution could also help resolve complicated fiber configurations such as crossings fibers from fanning fibers or disentangle small structures like the optic chiasm (Roebroek et al., 2008), which is not possible at lower spatial resolution (Calabrese et al., 2014; Jones et al., 2013).

### 2.5.6 Other methods for high spatial resolution acquisitions

We have shown in Fig. 2.6 that high spatial resolution acquisitions which are noisy at first can reveal improved anatomical details when they are subsequently denoised. This indeed suggests that high resolution acquisition can now be acquired on clinical scanners. Recently, other algorithms enabling a high spatial resolution at the acquisition level have been suggested (Ning et al., 2016; Scherrer et al., 2015). These methods both rely on smartly fusing the (complementary) data of multiple acquisitions acquired at a lower spatial resolution to obtain a single high resolution volume. While the approach we suggest here is to acquire a single volume using a standard sequence, both techniques are fundamentally exploiting different aspects to increase the available spatial resolution. As such, it would be possible to combine our denoising algorithm with the reconstruction algorithms presented in (Ning et al., 2016; Scherrer et al., 2015).

### 2.5.7 Current limitations and possible improvements

Although most models assume a Rician or  $nc\text{-}\chi$  noise distribution, this does not take into account the noise correlation between each coils or the effect of acceleration techniques that subsample the k-space (Aja-Fernández et al., 2014). The development of correction factors for existing algorithms relies on computing the *effective* values for the noise standard deviation  $\sigma$  and the number of degrees of freedom of the  $nc\text{-}\chi$  distribution, which is expected to be smaller than  $2N$ . These values can be used to take into account the correlation introduced between the coils in parallel imaging acquisitions (Brion et al., 2013). To consider the fact that the noise distribution nature might vary spatially in addition to the noise variance, one can use *a priori* information obtained from the scanner through the SENSE sensitivity maps or the GRAPPA weights and need to estimate the correlation between each of the receiver coils. We could explicitly add such a correction to our algorithm since we work locally with the stabilization algorithm, Eqs. (2.1) and (2.4). Using multiband acceleration also modifies the noise properties due to the introduced aliasing, which further strengthen the idea that spatially adaptive denoising algorithms should be used on modern scanners and sequences (Uğurbil et al., 2013). Nevertheless, obtaining the needed map for a SENSE reconstruction or the required GRAPPA weights might not be easily feasible in a clinical setting. We also intend to revisit the order in which pre-processing algorithms (motion correction, eddy currents correction, distortions correction) should be applied since these steps require interpolation, which could also introduce spatial

## 2.6. Conclusion

---

correlation in the noise profile. This also makes the noise distribution deviate slightly from its theoretical distribution, where parameters vary spatially instead of being fixed constants for the whole volume (Aja-Fernández et al., 2014). Nevertheless, we have observed experimentally that our NLSAM algorithm is robust to small subject motion thanks to the local neighborhood processing. In cases where artifacts (such as EPI distortions) might undermine the denoising process, one can first correct for these artifacts using a nearest neighbor interpolation, which should not modify the noise distribution. Subsequent corrections can then be performed after denoising using other kinds of interpolation as needed.

While developing the NLSAM algorithm, we found that using a bigger 3D patchsize did not significantly improve the denoising quality, while augmenting both computing time and memory requirements. Our implementation also allows one create the smallest subset of angular neighbors covering all DWIs through a greedy set cover algorithm. This option (named “NLSAM fast” in Table 2.4) leads to a speedup of 3 to 4 times, but at the cost of slightly reducing the denoising performance since some DWIs might be denoised only once instead of multiple times. We used the fully covered version for our experiments, which were run on a machine running Ubuntu Linux 12.04 with a quad core Intel i7 930 at 2.8 GHz and 18 GB of RAM. Table 2.4 reports the runtime of the various algorithms in minutes and their RAM usage. While the computing time required by NLSAM is larger than the other methods, our Python implementation is fairly unoptimized and could be sped up to competitive runtimes by various code optimizations or lowering the maximum number of iterations in Eq. (2.4).

	AONLM	LPCA	msPOAS	NLSAM	NLSAM fast
Time (mins)	22.2	3.7	4.0	37.1	9.8
RAM usage (MB)	552	640	1543	606	412

**Table 2.4:** Required time and RAM usage for the compared denoising algorithms on the  $b = 1000$  s/mm<sup>2</sup> SNR 10 dataset with stationary Rician noise.

## 2.6 Conclusion

In this paper, we introduced a new denoising method, the Non Local Spatial and Angular Matching (NLSAM), which is specifically designed to take advantage of diffusion MRI data. Our method is based on 1) Correcting the spatially varying Rician and  $nc\text{-}\chi$  noise bias 2) Finding similar DWIs through angular neighbors to promote sparsity 3) Iteratively denoise similar patches and their neighbors locally with dictionary learning, where the local variance is used as an upper bound on the  $\ell_2$  reconstruction error. We extensively compared quantitatively our new method against three other state-of-the-art denoising methods on a synthetic phantom and qualitatively on an in vivo high resolution dataset.

We also showed that taking into account both the effect of spatially varying noise and non-Gaussian distributed noise is crucial to denoise effectively the DWIs. Our NLSAM algorithm is freely available<sup>10</sup>, restores perceptual information, removes the noise bias in common diffusion metrics and produces more anatomically plausible tractography on a high spatial resolution in vivo dataset when compared to a lower spatial resolution acquisition of the same subject.

Since our NLSAM algorithm can be used on any already acquired dataset and does not add any acquisition time, this shows that denoising the data should be a pre-processing part of every pipeline, just like any other correction method that is commonly applied for artifacts removal. With that in mind, the diffusion MRI community could aim for higher spatial resolution DWIs, without requiring the use of costly new hardware or complicated acquisition schemes. This could in turn reveal new anatomical details, which are not achievable at the spatial resolution currently used in diffusion MRI.

## 2.7 Appendix: The NLSAM Algorithm

This appendix outlines the NLSAM algorithm as pseudo code. The original implementation in Python is freely available at <https://github.com/samuelstjean/nlsam>.

---

**Algorithm 2.1:** The proposed NLSAM denoising algorithm.

---

**Data:** 4D dMRI data,  $an$  = number of angular neighbors,  $ps$  = spatial patch size,  $N$  = Number of coils,  $max\_iter = 40$

**Result:** Denoised data with NLSAM

**Step 1;**

Find  $\sigma_G$  with either PIESNO or Eq. (2.3);  
Apply noise stabilization with  $\sigma_G$  and  $N$  coils;

**foreach** *DWI in dMRI data* **do**

**Step 2;**

    Find the closest  $an$  angular neighbors;  
    Create 4D block with  $b_0$ , DWI and its  $an$  neighbors;  
    Extract all overlapping patches of size  $(ps, ps, ps)$ ;

**Step 3;**

    Apply Eq. (2.1) to find  $D$ ;  
    Iterate Eq. (2.4) to find  $\alpha$  until convergence or  $max\_iter$  is reached;  
    Average overlapping patches based on sparsity with Eq. (2.5);  
    Return to original shape;

**end**

**foreach** *Denoised DWI in dMRI data* **do**

  Average all Denoised DWI representations;

**end**

---

<sup>10</sup><https://github.com/samuelstjean/nlsam>

# Bibliography

- [1] S. Aja-Fernandez, C. Alberola-Lopez, and C.-F. Westin. “Noise and signal estimation in magnitude MRI and Rician distributed images: a LMMSE approach.” In: *IEEE transactions on image processing* 17.8 (Aug. 2008), pp. 1383–98 (cit. on pp. 19, 24).
- [2] S. Aja-Fernández, G. Vegas-Sánchez-Ferrero, and A. Tristán-Vega. “Noise estimation in parallel MRI: GRAPPA and SENSE.” In: *Magnetic resonance imaging* 32.3 (Apr. 2014), pp. 281–90 (cit. on pp. 19, 22, 44, 45).
- [3] D. C. Alexander et al. “Orientationally invariant indices of axon diameter and density from diffusion MRI”. In: *NeuroImage* 52.4 (2010), pp. 1374–1389 (cit. on p. 19).
- [4] R. Bai, C. G. Koay, E. Hutchinson, and P. J. Basser. “A framework for accurate determination of the T2 distribution from multiple echo magnitude MRI images”. In: *Journal of Magnetic Resonance* 244 (2014), pp. 53–63 (cit. on p. 22).
- [5] L. Bao, M. Robini, W. Liu, and Y. Zhu. “Structure-adaptive sparse denoising for diffusion-tensor MRI.” In: *Medical image analysis* 17.4 (May 2013), pp. 442–57 (cit. on pp. 19, 21).
- [6] S. Becker, K. Tabelow, S. Mohammadi, N. Weiskopf, and J. Polzehl. “Adaptive smoothing of multi-shell diffusion weighted magnetic resonance data by msPOAS”. In: *NeuroImage* 95 (July 2014), pp. 90–105 (cit. on pp. 20, 24, 29).
- [7] V. Brion et al. “Noise correction for HARDI and HYDI data obtained with multi-channel coils and sum of squares reconstruction: an anisotropic extension of the LMMSE.” In: *Magnetic resonance imaging* 31.8 (Oct. 2013), pp. 1360–71 (cit. on pp. 19, 20, 44).
- [8] E. Calabrese, A. Badea, C. L. Coe, G. R. Lubach, M. A. Styner, and G. A. Johnson. “Investigating the tradeoffs between spatial resolution and diffusion sampling for brain mapping with diffusion tractography: Time well spent?” In: *Human Brain Mapping* 35.11 (Nov. 2014), pp. 5667–5685 (cit. on p. 44).
- [9] E. J. Candès, M. B. Wakin, and S. P. Boyd. “Enhancing Sparsity by Reweighted l1 Minimization”. In: *Journal of Fourier Analysis and Applications* 14.5-6 (Oct. 2008), pp. 877–905 (cit. on p. 25).
- [10] M. Chamberland, K. Whittingstall, D. Fortin, D. Mathieu, and M. Descoteaux. “Real-time multi-peak tractography for instantaneous connectivity display.” English. In: *Frontiers in neuroinformatics* 8.May (Jan. 2014), p. 59 (cit. on pp. 31, 37, 40, 43).
- [11] J. B. Colby, L. Soderberg, C. Lebel, I. D. Dinov, P. M. Thompson, and E. R. Sowell. “Along-tract statistics allow for enhanced tractography analysis”. In: *NeuroImage* 59.4 (Feb. 2012), pp. 3227–3242 (cit. on p. 42).
- [12] M.-A. Côté, E. Garyfallidis, H. Larochelle, and M. Descoteaux. “Cleaning up the mess : tractography outlier removal using hierarchical QuickBundles clustering”. In: *International Symposium on Magnetic Resonance in Medicine (ISMRM’15)* 98 (2015), p. 6046 (cit. on p. 43).
- [13] M.-A. Côté, G. Girard, A. Boré, E. Garyfallidis, J.-c. Houde, and M. Descoteaux. “Tractometer: towards validation of tractography pipelines.” In: *Medical image analysis* 17.7 (Oct. 2013), pp. 844–57 (cit. on pp. 30, 35, 43).

- 
- [14] P. Coupe, P. Yger, S. Prima, P. Hellier, C. Kervrann, and C. Barillot. “An optimized blockwise nonlocal means denoising filter for 3-D magnetic resonance images.” In: *IEEE transactions on medical imaging* 27.4 (Apr. 2008), pp. 425–41 (cit. on p. 19).
- [15] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. “Image denoising by sparse 3-D transform-domain collaborative filtering.” In: *IEEE transactions on image processing* 16.8 (Aug. 2007), pp. 2080–95 (cit. on pp. 21, 24).
- [16] A. Daducci et al. “Quantitative Comparison of Reconstruction Methods for Intra-Voxel Fiber Recovery From Diffusion MRI.” In: *IEEE Transactions on Medical Imaging* 33.2 (Feb. 2014), pp. 384–399 (cit. on pp. 30, 33).
- [17] I. Daubechies, R. Devore, M. Fornasier, and C. S. Güntürk. “Iteratively reweighted least squares minimization for sparse recovery.” In: *Communications on Pure and Applied Mathematics* 63 (2010), pp. 1–38 (cit. on p. 25).
- [18] M. Descoteaux and C. Poupon. “Diffusion-Weighted MRI.” In: *Comprehensive Biomedical Physics*. Elsevier, 2014, pp. 81–97 (cit. on p. 19).
- [19] M. Descoteaux, R. Deriche, T. Knosche, and A. Anwender. “Deterministic and Probabilistic Tractography Based on Complex Fibre Orientation Distributions.” In: *IEEE Transactions on Medical Imaging* 28.2 (Feb. 2009), pp. 269–286 (cit. on p. 42).
- [20] O. Dietrich, J. G. Raya, S. B. Reeder, M. Ingrisch, M. F. Reiser, and S. O. Schoenberg. “Influence of multichannel combination, parallel imaging and other reconstruction techniques on MRI noise characteristics.” In: *Magnetic resonance imaging* 26.6 (July 2008), pp. 754–62 (cit. on pp. 19, 22, 39).
- [21] C. Eichner et al. “Real diffusion-weighted MRI enabling true signal averaging and increased diffusion contrast.” In: *NeuroImage* 122 (Nov. 2015), pp. 373–384 (cit. on pp. 19, 40).
- [22] M. Elad and M. Aharon. “Image Denoising Via Sparse and Redundant Representations Over Learned Dictionaries.” In: *IEEE Transactions on Image Processing* 15.12 (Dec. 2006), pp. 3736–3745 (cit. on pp. 21, 22).
- [23] A. Foi. “Noise estimation and removal in MR imaging: The variance-stabilization approach.” In: *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro 2* (Mar. 2011), pp. 1809–1814 (cit. on p. 24).
- [24] E. Garyfallidis et al. “Dipy, a library for the analysis of diffusion MRI data.” In: *Frontiers in Neuroinformatics* 8. February (2014), pp. 1–17 (cit. on p. 30).
- [25] G. Girard, K. Whittingstall, R. Deriche, and M. Descoteaux. “Towards quantitative connectivity analysis: reducing tractography biases.” In: *NeuroImage* 98 (Sept. 2014), pp. 266–78 (cit. on pp. 30, 35, 38, 43).
- [26] A. Gramfort, C. Poupon, and M. Descoteaux. “Denoising and fast diffusion imaging with physically constrained sparse dictionary learning.” In: *Medical image analysis* 18.1 (Jan. 2014), pp. 36–49 (cit. on pp. 20, 22, 25).
- [27] D. K. Jones, T. R. Knösche, and R. Turner. “White matter integrity, fiber count, and other fallacies: The do’s and don’ts of diffusion MRI.” In: *NeuroImage* 73 (June 2013), pp. 239–254 (cit. on pp. 19, 44).
- [28] C. G. Koay and P. J. Basser. “Analytically exact correction scheme for signal extraction from noisy magnitude MR signals.” In: *Journal of Magnetic Resonance* 179.2 (Apr. 2006), pp. 317–322 (cit. on pp. 24, 25).

## Bibliography

---

- [29] C. G. Koay, E. Özarslan, and P. J. Basser. “A signal transformational framework for breaking the noise floor and its applications in MRI”. In: *Journal of Magnetic Resonance* 197.2 (Apr. 2009), pp. 108–119 (cit. on pp. 22–24, 26, 40).
- [30] C. G. Koay, E. Özarslan, and C. Pierpaoli. “Probabilistic Identification and Estimation of Noise (PIESNO): A self-consistent approach and its applications in MRI”. In: *Journal of Magnetic Resonance* 199.1 (July 2009), pp. 94–103 (cit. on pp. 24, 26).
- [31] F. Lam, S. D. Babacan, J. P. Haldar, M. W. Weiner, N. Schuff, and Z. P. Liang. “Denoising diffusion-weighted magnitude MR images using rank and edge constraints”. In: *Magnetic Resonance in Medicine* 71.3 (Apr. 2014), pp. 1272–1284 (cit. on pp. 20, 24).
- [32] R. W. Liu, L. Shi, W. Huang, J. Xu, S. C. H. Yu, and D. Wang. “Generalized total variation-based MRI Rician denoising model with spatially adaptive regularization parameters.” In: *Magnetic resonance imaging* 32.6 (July 2014), pp. 702–20 (cit. on p. 19).
- [33] M. Maggioni, V. Katkovnik, K. Egiazarian, and A. Foi. “Nonlocal Transform-Domain Filter for Volumetric Data Denoising and Reconstruction”. In: *IEEE Transactions on Image Processing* 22.1 (Jan. 2013), pp. 119–133 (cit. on pp. 19, 21, 24, 26).
- [34] J. Mairal, F. Bach, J. Ponce, and G. Sapiro. “Online dictionary learning for sparse coding”. In: *Proceedings of the 26th Annual International Conference on Machine Learning - ICML '09*. Vol. 11. New York, New York, USA: ACM Press, Aug. 2009, pp. 1–8 (cit. on pp. 21, 22, 25).
- [35] J. V. Manjón, P. Coupé, L. Concha, A. Buades, D. L. Collins, and M. Robles. “Diffusion Weighted Image Denoising Using Overcomplete Local PCA”. In: *PLoS ONE* 8.9 (Sept. 2013). Ed. by G. Gong, e73021 (cit. on pp. 20, 25, 26, 29).
- [36] J. V. Manjón, P. Coupé, L. Martí-Bonmatí, D. L. Collins, and M. Robles. “Adaptive non-local means denoising of MR images with spatially varying noise levels.” In: *Journal of magnetic resonance imaging : JMRI* 31.1 (Jan. 2010), pp. 192–203 (cit. on pp. 19, 20, 24, 29).
- [37] L. Ning et al. “A joint compressed-sensing and super-resolution approach for very high-resolution diffusion imaging”. In: *NeuroImage* 125 (Jan. 2016), pp. 386–400 (cit. on p. 44).
- [38] B. A. Olshausen and D. J. Field. “Emergence of simple-cell receptive field properties by learning a sparse code for natural images.” In: *Nature* 381.6583 (June 1996), pp. 607–9 (cit. on pp. 21, 26).
- [39] M. Paquette, S. Merlet, G. Gilbert, R. Deriche, and M. Descoteaux. “Comparison of sampling strategies and sparsifying transforms to improve compressed sensing diffusion spectrum imaging”. In: *Magnetic Resonance in Medicine* 73.1 (Jan. 2015), pp. 401–416 (cit. on pp. 30, 33).
- [40] J. Rajan, J. Veraart, J. Van Audekerke, M. Verhoye, and J. Sijbers. “Nonlocal maximum likelihood estimation method for denoising multiple-coil magnetic resonance images.” In: *Magnetic Resonance Imaging* 30.10 (Dec. 2012), pp. 1512–1518 (cit. on p. 19).
- [41] A. Roebroeck et al. “High-resolution diffusion tensor imaging and tractography of the human optic chiasm at 9.4 T”. In: *NeuroImage* 39.1 (Jan. 2008), pp. 157–168 (cit. on p. 44).
- [42] B. Scherrer, O. Afacan, M. Taquet, S. P. Prabhu, A. Gholipour, and S. K. Warfield. *Accelerated High Spatial Resolution Diffusion-Weighted Imaging*. Ed. by S. Ourselin, D. C. Alexander, C.-F. Westin, and M. J. Cardoso. Vol. 9123. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2015, pp. 69–81 (cit. on p. 44).
- [43] S. N. Sotiropoulos et al. “Advances in diffusion MRI acquisition and processing in the Human Connectome Project”. In: *NeuroImage* 80 (Oct. 2013), pp. 125–143 (cit. on p. 44).

- 
- [44] S. St-Jean, P. Coupé, and M. Descoteaux. “Non Local Spatial and Angular Matching : a new denoising technique for diffusion MRI”. In: *International Symposium on Magnetic Resonance in Medicine (ISMRM’14)*. 2014 (cit. on p. 20).
- [45] C. M. Tax, B. Jeurissen, S. B. Vos, M. A. Viergever, and A. Leemans. “Recursive calibration of the fiber response function for spherical deconvolution of diffusion MRI data”. In: *NeuroImage* 86 (2014), pp. 67–80 (cit. on p. 42).
- [46] J.-D. Tournier, F. Calamante, and A. Connelly. “Robust determination of the fibre orientation distribution in diffusion MRI: non-negativity constrained super-resolved spherical deconvolution.” In: *NeuroImage* 35.4 (May 2007), pp. 1459–72 (cit. on pp. 30, 42).
- [47] A. Tristán-Vega and S. Aja-Fernández. “DWI filtering using joint information for DTI and HARDI.” In: *Medical image analysis* 14.2 (Apr. 2010), pp. 205–18 (cit. on pp. 19, 20).
- [48] K. Uğurbil et al. “Pushing spatial and temporal resolution for functional and diffusion MRI in the Human Connectome Project”. In: *NeuroImage* 80.21 (2013), pp. 80–104 (cit. on p. 44).
- [49] D. C. Van Essen, S. M. Smith, D. M. Barch, T. E. Behrens, E. Yacoub, and K. Ugurbil. “The WU-Minn Human Connectome Project: An overview”. In: *NeuroImage* 80 (Oct. 2013), pp. 62–79 (cit. on p. 19).
- [50] D. Varadarajan and J. P. Haldar. “A Majorize–Minimize Framework for Rician and Non–Central Chi MR Images”. In: *IEEE Transactions on Medical Imaging* 34.10 (Oct. 2015), pp. 2191–2202 (cit. on p. 19).
- [51] J. Veraart, J. Rajan, R. R. Peeters, A. Leemans, S. Sunaert, and J. Sijbers. “Comprehensive framework for accurate diffusion MRI parameter estimation.” In: *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine* 70.4 (Oct. 2013), pp. 972–84 (cit. on p. 22).
- [52] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. “Image quality assessment: from error visibility to structural similarity.” In: *IEEE transactions on image processing* 13.4 (Apr. 2004), pp. 600–12 (cit. on p. 29).
- [53] D. Wassermann et al. “The white matter query language: a novel approach for describing human white matter anatomy”. In: *Brain Structure and Function* 221.9 (Dec. 2016), pp. 4705–4721 (cit. on pp. 30, 38).

*There is no law except the law that there is no law.*

John Archibald Wheeler

# 3

## Obtaining representative core streamlines for white matter tractometry of the human brain

Based on

---

Maxime Chamberland, Samuel St-Jean, Chantal M. W. Tax and Derek K. Jones  
Obtaining representative core streamlines for white matter tractometry of the human brain  
Computational Diffusion MRI (CDMRI) workshop of MICCAI 2018, Granada. Springer  
International Publishing, Pages 3–19.

## Abstract

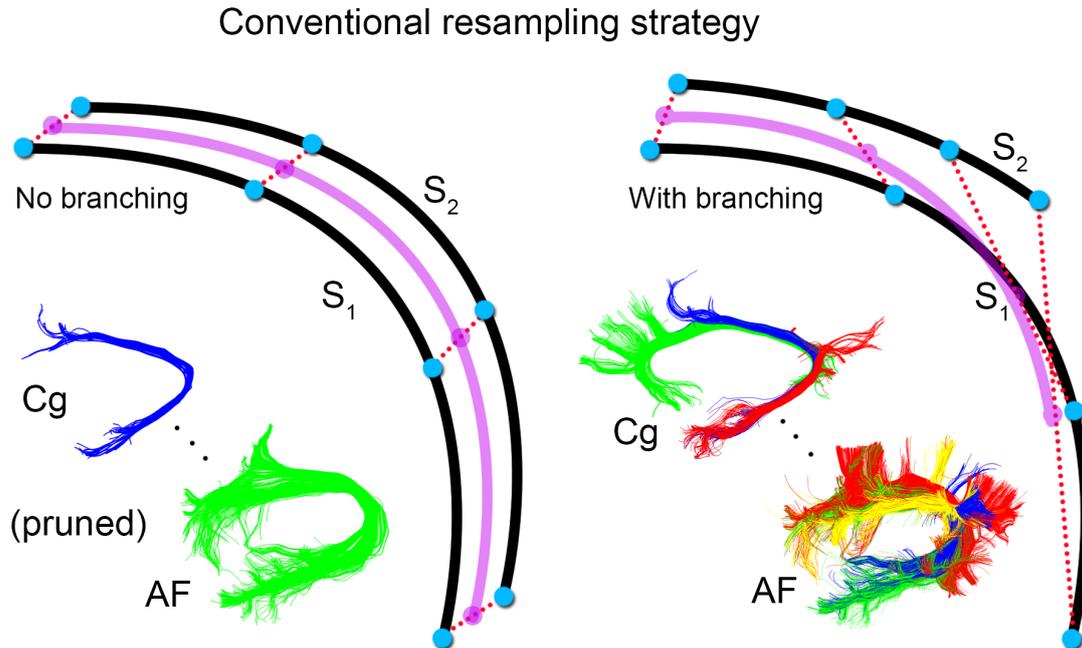
Diffusion MRI infers information about the micro-structural architecture of the brain by probing the diffusion of water molecules. The process of virtually reconstructing brain pathways based on these measurements is called tractography. Various metrics can be mapped onto pathways to study their micro-structural properties. Tractometry is an along-tract profiling technique that often requires the extraction of a representative streamline for a given bundle. This is traditionally computed by local averaging of the spatial coordinates of the vertices, and constructing a single streamline through those averages. However, the resulting streamline can end up being highly non-representative of the shape of the individual streamlines forming the bundle. In particular, this occurs when there is variation in the topology of streamlines within a bundle (e.g., differences in length, shape or branching). We propose an envelope-based method to compute a representative streamline that is robust to these individual differences. We demonstrate that this method produces a more representative core streamline, which in turn should lead to more reliable and interpretable tractometry analyses.

**Keywords:** Tractography, Tractometry, Bundle envelope, Core streamline, Diffusion MRI

### 3.1 Introduction

Tractography derived from diffusion MRI infers information about the structural architecture of the brain. In most studies, diffusion MRI metrics (e.g. fractional anisotropy (FA)) are often collapsed to a single scalar value per bundle (Jones, Catani, et al., 2005). Recently, a trend towards tract profiling (Colby et al., 2012; Yeatman et al., 2012) and direction-specific measurements within a voxel has emerged. Along-tract analysis is a technique that maps a given metric over the course of a bundle. The term *tractometry* was originally introduced by Bells et al. (2011) and the technique has been refined over the years by various groups (Colby et al., 2012; Corouge et al., 2006; De Santis et al., 2014; Jones, Travis, et al., 2005; Yeatman et al., 2012). It can be used to characterize areas of the brain with abnormal properties in patients (Cousineau et al., 2017; Dayan et al., 2016; Groeschel et al., 2014).

At the core of tractometry lies the concept of a *representative streamline* (O'Donnell et al., 2009) which is used to project metrics along the course of a given bundle. This is typically done by resampling all streamlines forming the bundle to  $n$  points and by averaging their spatial coordinates in a point-wise fashion (Colby et al., 2012; O'Donnell et al., 2009; Yeatman et al., 2012). This technique will produce a reasonable estimate of the average trajectory when there is very little branching and dispersion between the streamlines forming the pathway (Fig. 3.1, left). However, it is known that streamlines within a given bundle can vary in length and orientation, making direct averaging of their coordinates inappropriate (O'Donnell et al., 2009). Indeed, if the underlying streamlines are even slightly dispersed from each other, the resulting representative streamline obtained by simply averaging the coordinates can end up running outside of the shape of the bundle (Fig. 3.1, right). Not only does this representation become anatomically implausible, but also it can directly hamper further steps down the tractometry pipeline (e.g. when averaging metrics along different sections of the pathway). A common solution to overcome this problem is to perform tractometry only within the compact portion of the pathway by excluding data from the extremities, which tend to include fanning (Glozman et al., 2018; Yeatman et al., 2012). This approach greatly helps to 1) quickly obtain a representative streamline and 2) mitigate variability between subjects since bundles are essentially reduced to a simpler representation. However, cutting both extremities inherently limits the benefits conferred by state-of-the-art tractography techniques that can recover fanning and branching portions of white matter fasciculi. Here, we propose a technique to generate a representative streamline that is robust to multiple streamline lengths, arching configurations and orientations that naturally occur within a bundle.



**Figure 3.1:** Point-wise streamline averaging illustrated for two scenarios: pruned bundles (left) and fanning bundles (right). The mean streamline (purple) runs outside of the shape of the bundle when streamlines within a bundle slightly diverge from each other. Bundles: cingulum (Cg) and arcuate fasciculus (AF).

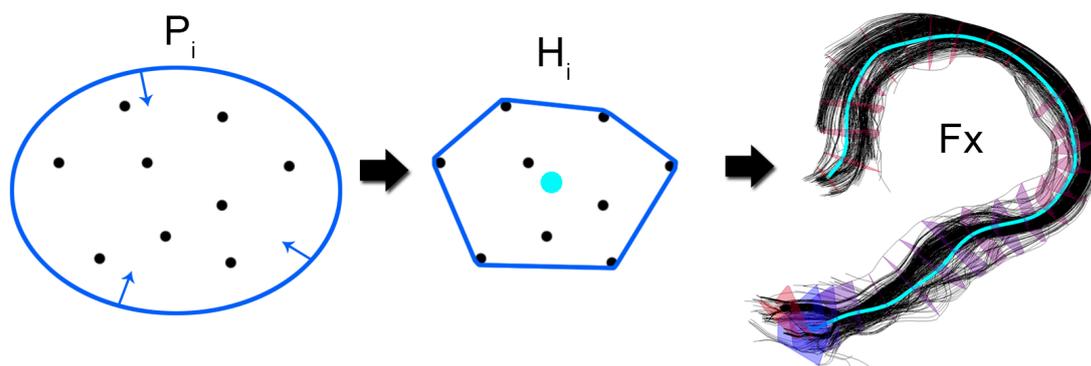
## 3.2 Theory and Methods

### 3.2.1 Acquisition and processing

Multi-shell high angular resolution diffusion MRI data were acquired on a Siemens Prisma scanner (TR = 4500 ms, TE = 80 ms, with b-values of 1200, 3000, 5000 s/mm<sup>2</sup>, 60 diffusion directions per shell and 15 non diffusion weighted images at a voxel size of 1.5 mm isotropic). Correction for subject motion and distortions caused by eddy currents were performed using *FSL eddy* and *topup* (Andersson and Sotiropoulos, 2016). Next, fiber orientation distributions functions (fODF) were computed using multi-shell multi-tissue constrained spherical deconvolution (Jeurissen et al., 2014). Tractography was performed using FiberNavigator (Chamberland et al., 2014), followed by manual bundle extraction of the corticospinal tract (CST), fornix (Fx), cingulum (Cg), arcuate fasciculus (AF) and inferior fronto-occipital fasciculus (iFOF). The main goal of the dissection plan was to preserve the characteristic anatomy, including fanning and branching, of each bundle (Rojkova et al., 2016). Tractography parameters were set as follows: min. fODF amplitude: 0.1, step size: 0.5 mm, max. angular threshold: 45°, min./max. streamline length: 30/200 mm with 1.5 million seeds covering the whole brain.

### 3.2.2 Proposed representative streamline extraction algorithm

The proposed method starts by averaging the top 5% longest streamlines from a bundle of interest to get a coarse approximation of the bundle’s core, defined by  $\mathbf{C} = \{p_i \in \mathbb{R}^3 | i = 1, \dots, n\}$  where  $p_i$  is a 3D point of the representative streamline. This core  $\mathbf{C}$  will serve as guidance for the propagation of a convex hull envelope along the entirety of the bundle. Next, we generate an orthogonal plane  $\mathbf{P}_i$  at each point  $p_i \in \mathbf{C}$  using the normal vector  $\vec{n}_i$  formed by  $p_i$  and  $p_{i+1}$ . Then, for each streamline from the bundle of interest, we find all line segments intersecting the current orthogonal plane  $\mathbf{P}_i$  up to a distance threshold of  $t$  mm (which is interactively defined in our implementation). If multiple points from a streamline intersect a plane  $\mathbf{P}_i$  (e.g., a spurious streamline doubling back on itself), only the closest point to  $\mathbf{C}$  is preserved so that the actual shape of the bundle is represented by its 3D envelope. We then compute a 2D convex hull  $\mathbf{H}_i \subset \mathbf{P}_i$  for each group of points found in the previous step as illustrated in Fig. 3.2. Finally, the centers of mass of each hull  $\mathbf{H}_i$  are linked together by fitting a 4<sup>th</sup> order b-spline curve comprising  $k$  knots, where  $k$  is a user-defined parameter. This step ensures that the representative streamline is located at the center since there is no guarantee that this is the case using the initial approximation of  $\mathbf{C}$  defined using the longest streamlines. The proposed framework is integrated within FiberNavigator, where all parameters are accessible to the user.



**Figure 3.2:** Example of a cross-section generated along the core of a bundle from a normal plane ( $\mathbf{P}_i$ ). The black dots represent in-plane streamlines with their convex hull  $\mathbf{H}_i$  in dark blue. On the right, the representative streamline (light blue) is obtained by linking the center of mass of each convex-hull (light blue dot).

### 3.2.3 Label maps generation

We compared our technique with conventional point-based resampling ( $n = 50$ ) computed using the mean distance flip algorithm, accounting for streamline direction (Cousineau et al., 2017). Distance maps were then generated by computing the minimum Euclidean distance between each point of the bundle and the core representative streamline. A transfer

function was used to visually map sections of the bundle assigned to respective locations along the core. A unique and smooth color grading from one end of the bundle to the other indicates a correct assignment along  $C$ , which in turn reflects how the diffusion metrics are averaged locally.

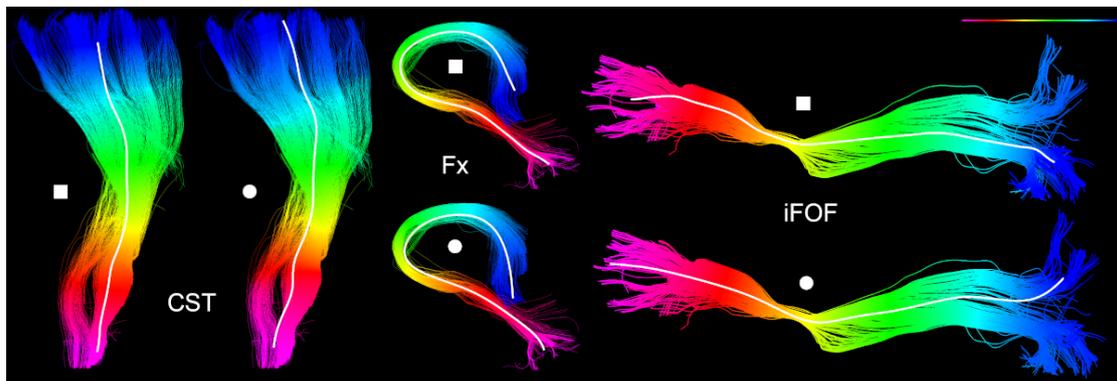
### 3.3 Results

#### 3.3.1 Bundles without branching and dispersion

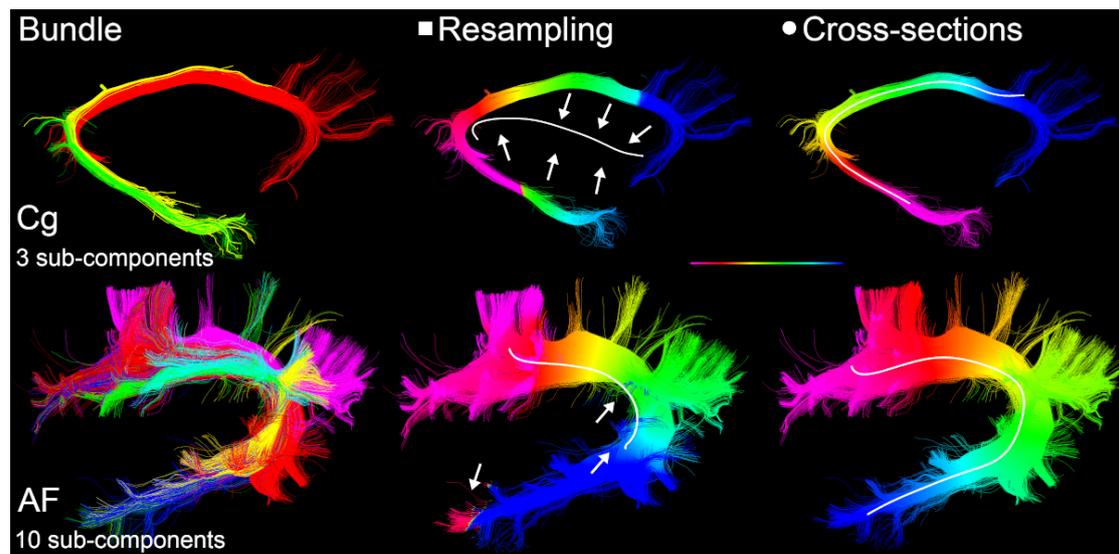
Fig. 3.3 shows results on tubular-shaped bundles (e.g. CST, Fx, iFOF) with little dispersion between the streamlines' starting and ending points. One can observe that the core streamline (i) has a length comparable to the rest of the streamlines within each bundle; (ii) lies inside the boundaries defined by all the streamlines. The color grading also shows a gradual labeling of the streamlines' points along the core.

#### 3.3.2 Bundles with branching and dispersion

Fig. 3.4 shows results on more complex bundles having different streamline lengths. In this example, the  $C_g$  consists of sub-components that connect the posterior cingulate cortex (PCC) to the medial prefrontal cortex (red streamlines) and the PCC to the parahippocampal gyrus (green streamlines). The complex configuration of the bundles inherently leads to an unrepresentative streamline when using the conventional point-based averaging (white arrows), as well as incorrect assignment of different sections along the bundle (indicated by the repeated green sections). The proposed cross-section method recovers an anatomically representative pathway that stays within the shape of the bundle.



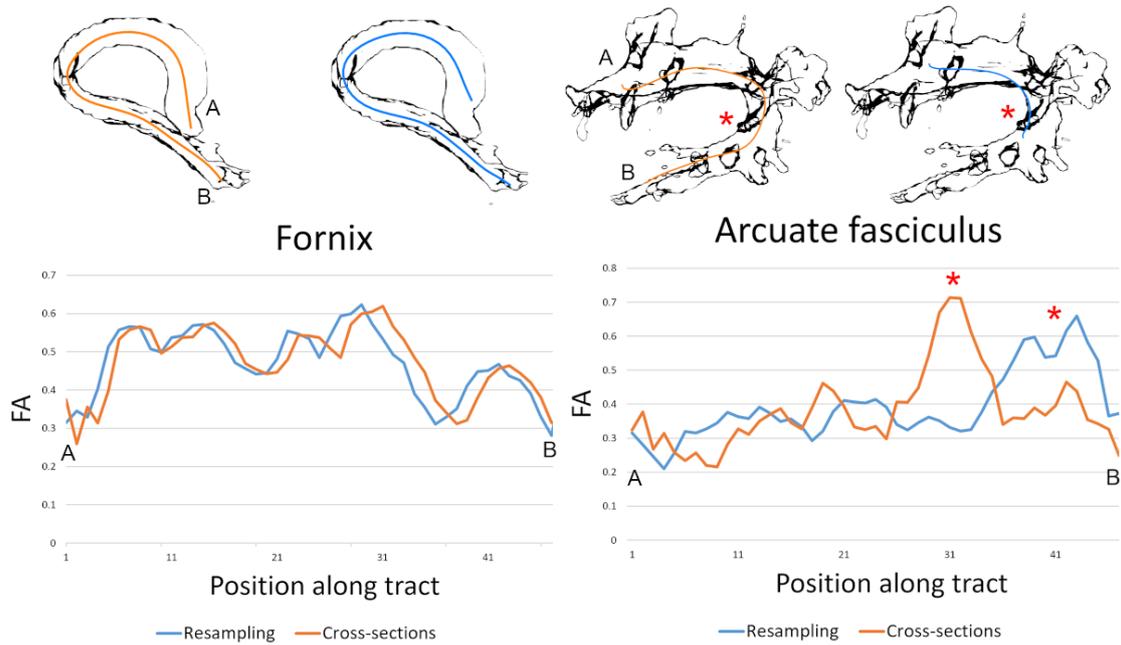
**Figure 3.3:** Label maps and representative streamlines illustrated for tubular-shaped bundles with little dispersion at the end points (e.g. directly connecting two brain regions) show agreement between the two methods (square: resampling method, circle: proposed method).



**Figure 3.4:** Complex fanning bundles with dispersed end-points reveal an astray mean streamline when using a conventional resampling approach to compute the core  $C$ , as well as incorrect label mapping (white arrows).

In the second example, the AF aggregates multiple sub-components with various arching streamlines that project to different areas of the lateral cortex. The representative streamline extracted from traditional averaging appears shorter than the full course of the bundle (white streamline). The inferior temporal aspect of the bundle (dark blue) is also incorrectly averaged and collapsed to the first point of the representative streamline. The last panel shows that the representative streamline traverses the full length of the bundle when using the proposed technique, which is also supported by the unique color grading of the label map.

Fig. 3.5 shows FA profiles computed along a tubular-shaped tract (Fx) and a fanning-shaped tract (AF). The Fx profiles appear similar in both resampling and cross-sections methods, except for a small shift induced by the larger anterior extent of the proposed method. The AF shows large differences between the two techniques in terms of profiling (red star), mostly due to the sub-optimality of the mean streamline as the representative streamline.



**Figure 3.5:** FA profiles illustrated for two representative streamline extraction algorithms. Left: Similar profiles are obtained for a tubular-shaped bundle (Fx). Right: Different profiles are obtained for a branching bundle (AF). The red star shows the location on the bundle where the maximum difference between the two methods occurs. Outlines of the bundles are shown for anatomical reference.

### 3.4 Discussion and Conclusion

We have shown that taking an average streamline as the representative pathway of a bundle can lead to non-representative results in the presence of tract branching and dispersion. To address this problem, we used orthogonal planes throughout the bundle to derive a representative core streamline that traverses its entire center of mass and therefore, allowing for a more interpretable tractometry. Generating core streamlines based on convex-hulls has been applied previously in the creation of 3D meshes for visualization (Enders et al., 2005), as well as for extracting skeleton streamlines for connectivity analysis using an atlas (Duda et al., 2010). Here, we additionally show that this approach produces more representative core streamlines for tractometry analyses in various bundle shapes. A drawback to the current approach is that it requires that at least some streamlines run from one end of the bundle to the other. Otherwise, the cross-section propagation may halt prematurely and thus affect the computation of the representative streamline. In addition, we assume that the input bundles are already pruned from streamline loops and undesired false positives.

Since tract morphology varies between subjects, truncation and resampling streamlines based on a length criterion (Colby et al., 2012) or number of points (Cousineau et al., 2017; Yeatman et al., 2012) may inadvertently discard important information that is specific to the pathway of interest. This loss of information is inherent to the truncation approach

and should be minimized when assigning diffusion metrics to a representative pathway to preserve the full extent of a bundle as much as possible (O'Donnell et al., 2009). We showed that this inevitably leads to discrepancies in tract profiling when assigning diffusion metrics to a representative streamline. An alternative approach to truncation and tract averaging could be to remove the need for a representative streamline by matching geometrical properties of streamlines between subjects (Glozman et al., 2018; Parker et al., 2016). Yet, the performance of those techniques still requires investigation for complex white matter configurations (e.g. fanning and branching). Nevertheless, truncation can also be useful to reduce potential issues associated with tractography and could still be applied in a post-processing step to our technique, once the representative streamline has been generated.

Finally, a potential improvement to the proposed method would be to recursively generate multiple convex hulls, allowing the algorithm to recover various sub-branches in fanning bundles. This could help in achieving a simplified—yet still anatomically accurate—representation of the core of a bundle, which is the centerpiece for the future design of tractometry pipelines.

# Bibliography

- [1] J. L. Andersson and S. N. Sotiropoulos. “An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging”. In: *Neuroimage* 125 (2016), pp. 1063–1078 (cit. on p. 54).
- [2] S. Bells et al. “Tractometry—comprehensive multi-modal quantitative assessment of white matter along specific tracts”. In: *Proc. ISMRM*. Vol. 678. 2011 (cit. on p. 53).
- [3] M. Chamberland, K. Whittingstall, D. Fortin, D. Mathieu, and M. Descoteaux. “Real-time multi-peak tractography for instantaneous connectivity display.” English. In: *Frontiers in neuroinformatics* 8.May (Jan. 2014), p. 59 (cit. on p. 54).
- [4] J. B. Colby, L. Soderberg, C. Lebel, I. D. Dinov, P. M. Thompson, and E. R. Sowell. “Along-tract statistics allow for enhanced tractography analysis”. In: *Neuroimage* 59.4 (2012), pp. 3227–3242 (cit. on pp. 53, 58).
- [5] I. Corouge, P. T. Fletcher, S. Joshi, S. Gouttard, and G. Gerig. “Fiber tract-oriented statistics for quantitative diffusion tensor MRI analysis”. In: *Medical image analysis* 10.5 (2006), pp. 786–798 (cit. on p. 53).
- [6] M. Cousineau et al. “A test-retest study on Parkinson’s PPMI dataset yields statistically significant white matter fascicles”. In: *NeuroImage: Clinical* 16 (2017), pp. 222–233 (cit. on pp. 53, 55, 58).
- [7] M. Dayan et al. “Profilometry: a new statistical framework for the characterization of white matter pathways, with application to multiple sclerosis”. In: *Human brain mapping* 37.3 (2016), pp. 989–1004 (cit. on p. 53).
- [8] S. De Santis, M. Drakesmith, S. Bells, Y. Assaf, and D. K. Jones. “Why diffusion tensor MRI does well only some of the time: variance and covariance of white matter tissue microstructure attributes in the living human brain”. In: *Neuroimage* 89 (2014), pp. 35–44 (cit. on p. 53).
- [9] J. T. Duda, C. McMillan, M. Grossman, and J. C. Gee. “Relating structural and functional connectivity to performance in a communication task.” In: *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)* 13.Pt 2 (2010), pp. 282–9 (cit. on p. 58).
- [10] F. Enders, N. Sauber, D. Merhof, P. Hastreiter, C. Nimsky, and M. Stamminger. “Visualization of White Matter Tracts with Wrapped Streamlines”. In: *IEEE Visualization 2005 - (VTS’05)*. IEEE, 2005, pp. 7–7 (cit. on p. 58).
- [11] T. Glozman, L. Bruckert, F. Pestilli, D. W. Yecies, L. J. Guibas, and K. W. Yeom. “Framework for shape analysis of white matter fiber bundles”. In: *Neuroimage* 167 (Feb. 2018), pp. 466–477 (cit. on pp. 53, 59).
- [12] S. Groeschel et al. “Identification and interpretation of microstructural abnormalities in motor pathways in adolescents born preterm”. In: *NeuroImage* 87 (2014), pp. 209–219 (cit. on p. 53).
- [13] B. Jeurissen, J.-D. Tournier, T. Dhollander, A. Connelly, and J. Sijbers. “Multi-tissue constrained spherical deconvolution for improved analysis of multi-shell diffusion MRI data”. In: *NeuroImage* 103 (2014), pp. 411–426 (cit. on p. 54).

## Bibliography

---

- [14] D. K. Jones, M. Catani, et al. “A diffusion tensor magnetic resonance imaging study of frontal cortex connections in very-late-onset schizophrenia-like psychosis”. In: *Am J Geriatr Psychiatry* 13.12 (Dec. 2005), pp. 1092–1099 (cit. on p. 53).
- [15] D. K. Jones, A. R. Travis, G. Eden, C. Pierpaoli, and P. J. Basser. “PASTA: pointwise assessment of streamline tractography attributes”. In: *Magnetic Resonance in Medicine* 53.6 (2005), pp. 1462–1467 (cit. on p. 53).
- [16] L. J. O’Donnell, C.-F. Westin, and A. J. Golby. “Tract-based morphometry for white matter group analysis”. In: *Neuroimage* 45.3 (2009), pp. 832–844 (cit. on pp. 53, 59).
- [17] G. D. Parker, D. Lloyd, and D. K. Jones. “The best of both worlds: Combining the strengths of TBSS and tract-specific measurements for group-wise comparison of white matter microstructure”. In: *In: International Symposium on Magnetic Resonance in Medicine (Singapore)*. 2036 (2016) (cit. on p. 59).
- [18] K. Rojkova, E. Volle, M. Urbanski, F. Humbert, F. Dell’Acqua, and M. Thiebaut de Schotten. “Atlasing the frontal lobe connections and their variability due to age and education: a spherical deconvolution tractography study”. In: *Brain Struct Funct* 221.3 (Apr. 2016), pp. 1751–1766 (cit. on p. 54).
- [19] J. D. Yeatman, R. F. Dougherty, N. J. Myall, B. A. Wandell, and H. M. Feldman. “Tract profiles of white matter properties: automating fiber-tract quantification”. In: *PloS one* 7.11 (2012), e49790 (cit. on pp. 53, 58).

*There's a lesson here, and I'm not the one that's gonna figure it out.*

Rick and Morty

# 4

## Reducing variability in along-tract analysis with diffusion profile realignment

Based on

---

Samuel St-Jean, Maxime Chamberland, Max A. Viergever and Alexander Leemans  
Reducing variability in along-tract analysis with diffusion profile realignment  
NeuroImage, Volume 199, 2019, Pages 663-679

---

## Abstract

Diffusion weighted magnetic resonance imaging (dMRI) provides a noninvasive virtual reconstruction of the brain's white matter structures through tractography. Analyzing dMRI measures along the trajectory of white matter bundles can provide a more specific investigation than considering a region of interest or tract-averaged measurements. However, performing group analyses with this along-tract strategy requires correspondence between points of tract pathways across subjects. This is usually achieved by creating a new common space where the representative streamlines from every subject are resampled to the same number of points. If the underlying anatomy of some subjects was altered due to, e.g. disease or developmental changes, such information might be lost by resampling to a fixed number of points. In this work, we propose to address the issue of possible misalignment, which might be present even after resampling, by realigning the representative streamline of each subject in this 1D space with a new method, coined diffusion profile realignment (DPR). Experiments on synthetic datasets show that DPR reduces the coefficient of variation for the mean diffusivity, fractional anisotropy and apparent fiber density when compared to the unaligned case. Using 100 in vivo datasets from the human connectome project, we simulated changes in mean diffusivity, fractional anisotropy and apparent fiber density. Independent Student's t-tests between these altered subjects and the original subjects indicate that regional changes are identified after realignment with the DPR algorithm, while preserving differences previously detected in the unaligned case. This new correction strategy contributes to revealing effects of interest which might be hidden by misalignment and has the potential to improve the specificity in longitudinal population studies beyond the traditional region of interest based analysis and along-tract analysis workflows.

**Keywords:** Diffusion profile realignment, Along-tract analysis, Tractometry, Tractography, Diffusion MRI, White matter

## 4.1 Introduction

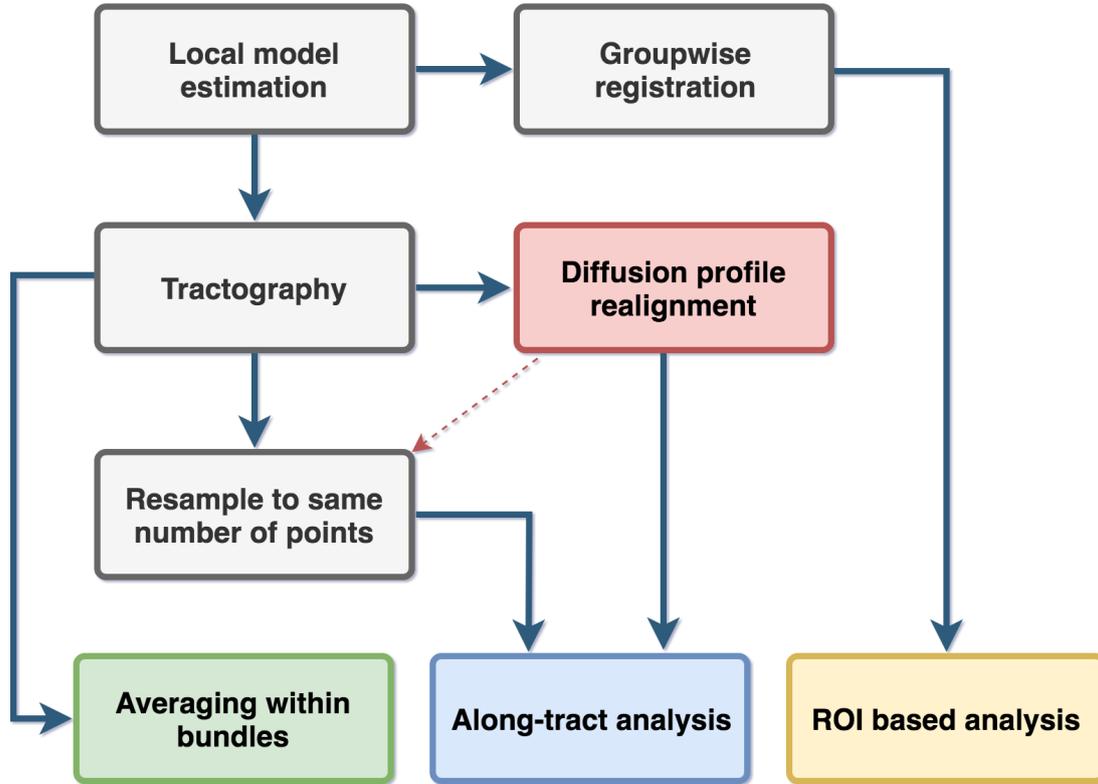
Diffusion weighted magnetic resonance imaging (dMRI) is a noninvasive technique that can be used to study microstructure in living tissues based on the displacement of water molecules. Since neurological diseases (e.g. multiple sclerosis (MS) (Cercignani and Gandini Wheeler-Kingshott, 2018), amyotrophic lateral sclerosis (ALS) (Haakma et al., 2017)) involve many processes that affect the density and properties of the underlying tissue, the corresponding changes are reflected on scalar values extracted from dMRI (Bodini and Ciccarelli, 2009). However, it remains challenging to pinpoint accurately the underlying cause as many of these changes (e.g. axonal damage, demyelination) may be reflected similarly by changes in measurements from dMRI (Beaulieu, 2002). Such changes could even be due to acquisition artifacts or from the use of a different processing method during data analysis (Jones and Cercignani, 2010), making dMRI sensitive, but not necessarily specific, to the various mechanisms involved in those changes (O'Donnell and Pasternak, 2015). Accurate characterization of the underlying processes affecting scalar metrics computed from dMRI still remains an open question.

A successful application of dMRI is to reconstruct the structure of the underlying tissues, a process known as tractography (see, e.g. Jeurissen et al. (2017) and Mori and Van Zijl (2002) for a review). Tractography enables a virtual reconstruction of the white matter bundles and pathways of the brain, which is central to preoperative neurosurgical planning (Nimsky et al., 2016) and at the heart of connectomics (Hagmann et al., 2007; Sporns et al., 2005). Over the last years, various analysis strategies have arisen to study scalar values computed from dMRI models. Two popular schools of techniques consist of using anatomical regions of interests (ROIs), either by manual or automatic delineation (Froeling et al., 2016; Smith et al., 2006), or using spatial information additionally brought by tractography to analyze scalar metrics along reconstructed bundles (Colby et al., 2012; Corouge et al., 2006; Cousineau et al., 2017; Jones, Travis, et al., 2005; Yeatman et al., 2012). Both approaches involve various user-defined settings and have their respective criticisms and drawbacks; ROI based analysis requires accurate groupwise registration (Bach et al., 2014), whereas tractography-based analysis needs to deal with false positives streamlines which can also look anatomically plausible (Maier-Hein et al., 2017). One key point shared between these methods is that they both require some form of correspondence between the studied structure of interest for each subjects, either by spatial registration to align the delineated ROIs (Froeling et al., 2016; Smith et al., 2006) or along the streamlines by resampling to a common number of points (Colby et al., 2012; Yeatman et al., 2012). Tractography-based approaches can analyze the voxels traversed by a specific white matter bundle in a data driven way and reveal subtle local changes inside a bundle, while ROI based analysis discards the 3D spatial information but reveal widespread changes in the bundle (O'Hanlon et al., 2015). For tractography-based analysis, metrics are either averaged by using all points forming a common bundle (Wakana et al., 2007) or collapsed as a representative pathway of the bundle (Colby et al., 2012; Cousineau et al., 2017; Yeatman et al., 2012) to study

changes in scalar values along its length. Once this per subject representative streamline has been defined, it is used to index scalar values along the length of this pathway (O’Hanlon et al., 2015; Szczepankiewicz et al., 2013). Recent applications include studying changes in diffusion metrics due to Alzheimer’s disease (AD) (Jin et al., 2017), which helped to uncover changes in mean diffusivity (MD) along the fornix for example. Studies in ALS patients also identified a diminution in fractional anisotropy (FA) along the corticospinal tract depending on the origin of the disease (Blain et al., 2011). Information from other MRI weighting such as myelin water fraction maps derived from T2 relaxometry have also been included to study changes due to MS (Dayan et al., 2016). As each subject respective morphology is different (i.e. reconstructed bundles from different subjects vary in shape and size) just as in ROIs based analysis, one needs to ensure correspondence between each segment of the studied bundle for all subjects. This correspondence is usually achieved by creating a new common space where all of the subjects representative streamlines are resampled to a common number of points. As noted by Colby et al. (2012), resampling to the same number of points makes the implicit assumption that the end points (and every point in between) are in correspondence across each subject. Yeatman et al. (2012) also mention that “it is important to recognize that the distal portions of the tract may not be in register across subjects”, even though the resampling step creates a new 1D space for point-by-point comparison. O’Donnell, Westin, et al. (2009) previously noticed the potential issue introduced by misalignment between subjects mentioning that “improved cross-subject alignment is of interest [...] as the high-frequency variations seen in individual subjects [...] are smoothed in the group average”. While many methods for registering dMRI volumes or streamlines were developed (see, e.g. O’Donnell, Daducci, et al. (2017) for a review), they do not directly address the issue of possible residual misalignment between the end points *after* extracting the representative streamlines of each subject. To ensure an adequate comparison between subjects, one must make sure that each streamline corresponds to the same underlying anatomical location.

In the present work, which extends our preliminary work presented at the ISMRM (St-Jean, Viergever, et al., 2016), we focus on the issue of possible misalignment between the final representative streamlines before conducting statistical analysis. To prevent this issue, we propose to realign the representative streamline of each subject while ensuring that the distance between each point is preserved, by resampling to a larger number of points than initially present. This strategy preserves the original 1D resolution of each subject, allowing a groupwise realignment based on maximizing the overall similarity by using the Fourier transform. After this realignment, points from individual streamlines that are identified as outliers can be discarded, as they would not overlap with the rest of the subjects. The representative, and now realigned, streamlines can be resampled to a lower number of points such as approximately one unit voxel size to facilitate group comparison and statistics (Colby et al., 2012). Fig. 4.1 shows an example of a typical workflow to analyze dMRI datasets and shows how the proposed diffusion profile realignment (DPR)

methodology can be used in preexisting pipelines.



**Figure 4.1:** Flowchart of current approaches and the proposed methodology. The diffusion profile realignment inserts itself in existing along-tract analysis workflows (red box) by combining a different resampling strategy with a realignment step. It is also possible to resample each streamlines to a smaller number of points (red arrow) if desired.

## 4.2 Theory

Each subject’s representative streamline is a 3D object, but the scalar metrics extracted along the tract can be viewed as a discrete 1D signal that may be non-stationary. In this work, we consider the 1D scalar metric profile to be a discrete signal equally sampled at each step of the tractography, which has a value of 0 outside the region delineated by the bundle it represents. We now present a realignment technique for 1D signals based on maximizing the cross-correlation function (CCF).

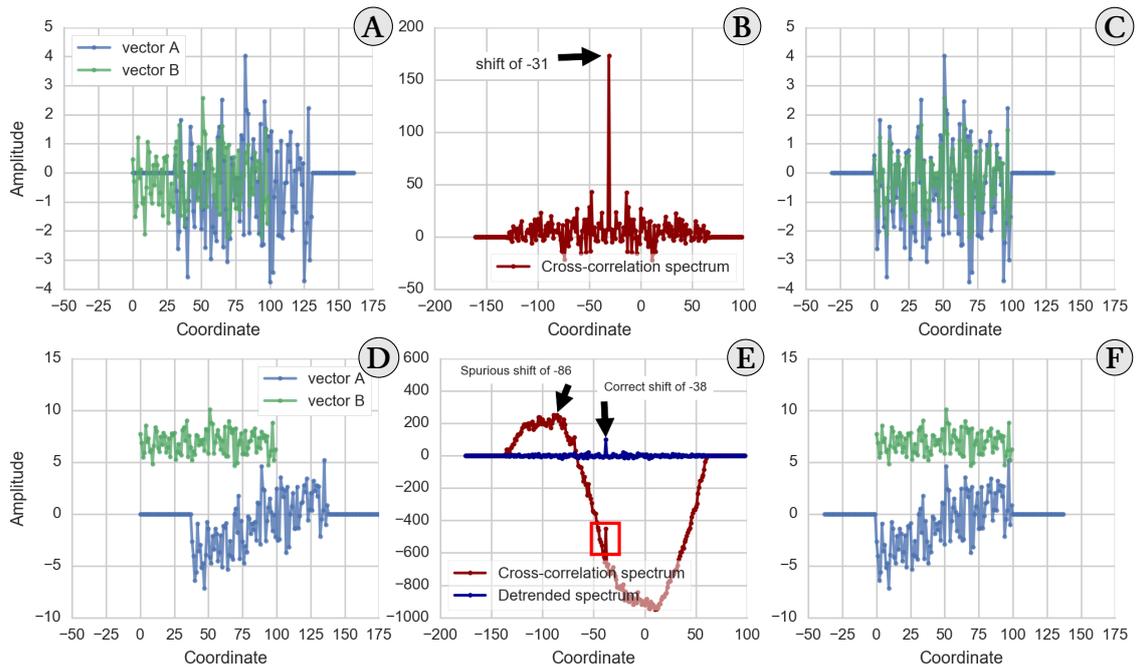
We can define the CCF using the fast Fourier transform (FFT) (Cooley and Tukey, 1965) as

$$\text{CCF}(x, y) = \mathcal{F}^{-1}(\mathcal{F}(x) \odot \mathcal{F}(y)^*), \quad (4.1)$$

## 4.2. Theory

---

where  $\mathcal{F}(x)$  and  $\mathcal{F}^{-1}(x)$  is the Fourier transform of  $x$  and its inverse,  $*$  is the complex conjugation and  $\odot$  the pointwise Hadamard product. The required shift to realign the vectors is given at the maximum coordinate of the CCF. The CCF measures the similarity between two vectors  $x$  and  $y$  assuming that the data is 1) stationary, 2) equally spaced between all points and 3) normally distributed (Denman, 1975; Platt and Denman, 1975). Stationarity can be achieved by fitting and subtracting a low degree polynomial from each vector before computing the cross-correlation, see, e.g. Box et al. (2008) and Stoica and Moses (2005) and references therein for more details. Equal spacing between each points can be obtained by resampling the data. The normality assumption seems less of an issue for large samples in practice (Platt and Denman, 1975). If the two vectors  $x$  and  $y$  have a different amplitude, the cross-correlation can be normalized by subtracting the mean and dividing by the standard deviation of each vector beforehand (Lewis, 1995). The shift computed at the maximum of the CCF is an integer displacement that can be refined by finding the maximum of the parabola around this point. Fig. 4.2 shows an example of the cross-correlation for both the stationary and non stationary case on two vectors. The first vector was randomly sampled from the standard normal distribution  $\mathcal{N}(0, 1)$ . The second vector was generated from the first vector by changing the offset and amplitude and then zero padding it at both ends to create an artificial displacement.



**Figure 4.2:** A synthetic example of the CCF between two randomly generated vectors. The top graphs showcase how the CCF spectrum can be used to find the displacement required to realign two different vectors by finding its maximum. **A)** Two vectors which are displaced with respect to each other, where vector B has a different amplitude from vector A. **B)** The cross-correlation spectrum, where the peak indicates the required shift to maximize the overlap between both vectors. **C)** The vectors after realignment, which is the exact displacement that had been applied. On the bottom graphs, removing the linear trend and normalizing the vectors satisfies the assumption of stationarity required by Eq. (4.1) and allows recovering the correct shift. **D)** Two unaligned vectors of different amplitude where vector B is also non stationary. **E)** The cross-correlation spectrum with detrending and normalization (in blue) and without these steps (in red). The detrended version recovers the correct shift, while the original CCF exhibits a variation in amplitude which hides the correct peak as a local (red box), but not global, maxima due to non stationarity. **F)** The vectors after realignment with the shift as computed by the detrended CCF. Both vectors are now realigned after shifting vector B with the shift computed in **E)**.

### 4.3 Materials and methods

To evaluate the proposed realignment procedure, we 1) generated synthetic datasets comprised of crossing bundles and 2) compared realignment on in vivo datasets with an altered version of their diffusion metrics. We now detail the various steps needed to perform an along-tract analysis and how the proposed realignment algorithm can be applied before performing a statistical analysis between subjects.

### 4.3.1 Resampling strategies for comparison between subjects

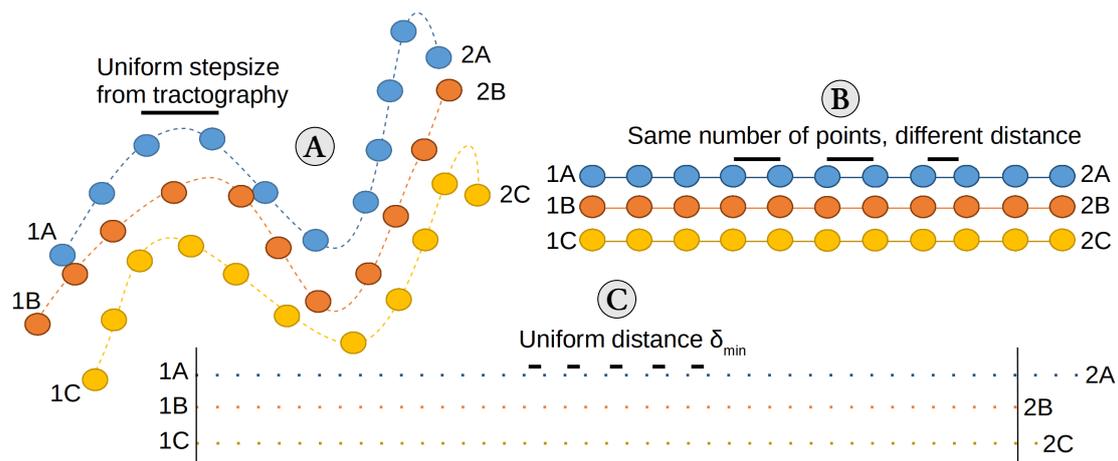
Various resampling strategies have been discussed in previous along-tract frameworks, with a common idea advocating resampling all representative streamlines to the same number of points. In Cousineau et al. (2017), the authors used a fixed number of points by resampling all of the studied bundles to 20 points while Yeatman et al. (2012) instead used 100 points. Colby et al. (2012) opted for resampling each bundle based on their average group length, ensuring that approximately one point per voxel was present. In this representation, each point of the streamlines is considered to correspond to the same anatomical location across subjects and is therefore blind to the intrinsic variance in shape or length between subjects. As each representative streamline most likely had a different length initially, the distance in millimeters between each sampled coordinate will be different for each subject. If the underlying anatomy of some subjects was altered due to, e.g. disease or developmental changes, such information might be lost by resampling to a fixed number of points as a first step. This can be prevented by ensuring that the new sampling resolution is at least equal or larger than the initial resolution used during tractography.

As a bundle is comprised of many individual streamlines, they are usually collapsed to a single representative pathway to facilitate subsequent analysis. This representative streamline is therefore an aggregation of many streamlines of various length and can be obtained either by averaging (Colby et al., 2012; Yeatman et al., 2012) or by finding representative clusters (Cousineau et al., 2017). Other assignment strategies towards a single representative pathway have been discussed in (Corouge et al., 2006; O'Donnell, Westin, et al., 2009). To ensure correspondence during this aggregation step, individual streamlines are usually resampled to a common number of points for all subjects. While this resampling is needed to obtain the representative streamline, it may also reduce the sampling resolution from the original streamlines given by the step size used for tractography if not enough points are kept. The representative streamline of each subject may also have a different orientation altogether and therefore might need to be flipped, ensuring that they share a common coordinate system (Colby et al., 2012).

In the present work, we instead advocate a novel two-step resampling strategy that builds upon the classical resampling strategy. After extracting the representative streamlines  $(S_1, \dots, S_n)$  for  $i = 1, \dots, n$  of each subject, each representative streamline  $S_i$  is defined by its number of points  $N_i$  and the distance between its points  $\delta_i$ . All streamlines are first resampled to  $M_i = N_i \times \delta_i / \delta_{\min}$  points, ensuring an equal distance between each point  $\delta_{\min} = \min(\delta_1, \dots, \delta_n)$ . In the end, the streamlines still have a different number of points  $M_i \geq \min(N_1, \dots, N_n)$  and points at the same coordinates across subjects do not implicitly assume to represent the same anatomical location. However, the distance  $\delta_i$  between each point  $M_i$  is now constant across subjects. While this idea may seem counterintuitive, the motivation behind this choice is due to Eq. (4.1), which relies on the FFT, and as such, needs equally sampled vectors and benefits from a high sampling resolution.

After the displacement has been applied, one can use the classical resampling strategies

presented by other authors, therefore making our approach fully compatible with already existing analysis techniques. We opted to use the methodology of Colby et al. (2012) since more than one point per unit voxel size would not carry additional information from the original data. This also alleviates further complications arising from multiple comparisons (Benjamini and Hochberg, 1995) for the subsequent statistical analysis one seeks to apply afterwards. Fig. 4.3 illustrates schematically the classical resampling versus our novel two-step resampling strategy.



**Figure 4.3:** An example of the classical and proposed resampling strategies on three representative streamlines. In **A**), three representative streamlines which have different shapes and lengths with their start (1A, 1B and 1C) and end points (2A, 2B and 2C) at different spatial locations. In **B**), the classical strategy of resampling to the same number of points (circles) introduces a common space to easily compare them. However, the end points of the underlying anatomies are artificially aligned when compared to their original representation and each point is at a different distance (black lines). In **C**), the proposed resampling strategy ensures that the distance  $\delta_{\min}$  (black lines) between every point is constant. Even though each streamline length is different as indicated by the location of the end points, they can now be realigned to identify the common anatomical positions between all subjects.

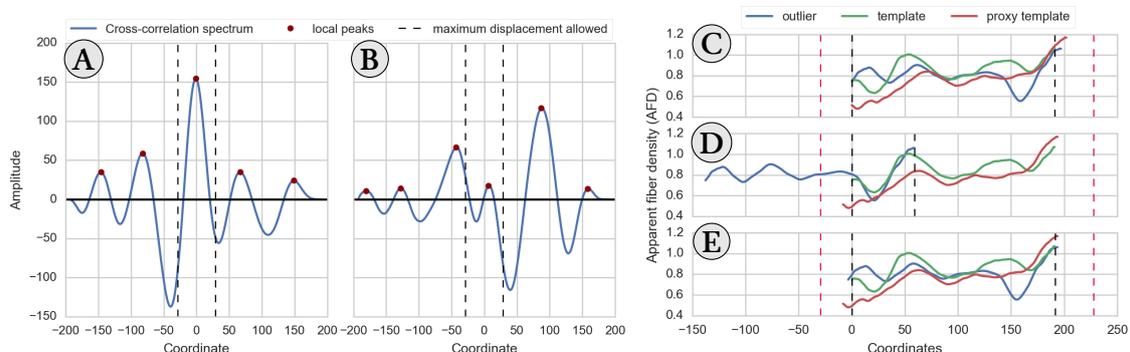
### 4.3.2 Proposed algorithm for diffusion profile realignment

DPR works in three steps once the 1D profiles have been resampled to an equal spacing as presented in Section 4.3.1. We also ensure stationarity of the data by fitting and subtracting a polynomial of degree one (i.e. a straight line) to each subject. It is important to mention here that this step is only to satisfy the stationarity assumption of Eq. (4.1) and does not modify the extracted diffusion profiles afterwards.

Firstly, a matrix of displacement is computed between every pairs of subjects and subsequently refined with parabola fitting as previously defined in Section 4.2. A maximum possible displacement in mm is then chosen. From the displacement matrix, the subject

### 4.3. Materials and methods

realigning the largest number of streamlines inside this maximum displacement is chosen automatically as the template subject. As Eq. (4.1) is symmetric, realigning subject  $A$  to  $B$  or subject  $B$  to  $A$  will have the same outcome in practice.



**Figure 4.4:** An example of a cross-correlation spectra (left) and finding a new template to realign outliers (right) using the HCP datasets. On the **left**, a threshold of 15% of the total streamline length is selected as the maximum allowed displacement (dashed vertical lines). **A)** A streamline with the global maximum of the CCF inside the chosen threshold. The maximum indicates the shift needed to realign it to the template. **B)** A streamline with a local maximum, but not the global maximum, of the CCF inside the chosen threshold. In this case, the two streamlines would not be realigned together as only small shifts should be needed for realignment. On the **right**, an example of realigning an outlier subject (in blue) to the original template (in green) via the closest matching new template (in red) using the AFD metric. The black dashed bars indicate the region where all three streamlines fully overlap and the red dashed bars shows the maximum allowed displacement of 15%. **C)** The three streamlines before realignment. **D)** Realigning the blue streamline with the template (in green) as given by the maximum of the CCF results in an outlier as in case **B)**. **E)** To circumvent the issue, a new template (in red) is found amongst the non-outlier subjects which minimizes the total displacement with the original template. The blue streamline is therefore not an outlier anymore as it now lies inside the displacement threshold as in case **A)**.

Secondly, all outliers with a displacement larger than the chosen threshold from the first step are realigned with the help of a new per-streamline template. For each outlier, a new template is selected amongst the remaining non-outlier subjects, which minimizes the total displacement between the original template from the first step and the current outlier. If the new minimum displacement is inside the chosen threshold, the subject that was previously an outlier is now registered through this new template. If no new template providing realignment inside the threshold can be found, then this subject is declared as an outlier and is not realigned at all. Fig. 4.4 shows the spectra of a normal subject and of an outlier for spectra computed with Eq. (4.1) from the HCP datasets. Even if the optimum displacement lies outside the chosen threshold, the outlier can still be realigned by finding a new template subject.

Finally, after realigning all the admissible streamlines to the template, there will be a different number of overlapping subjects for each coordinate. Just as ROIs were previously used to truncate the bundles' end points (recall Fig. 4.6), the resulting aligned streamlines

should be truncated once again to reduce their uncertainty since not all coordinates have the same number of overlapping streamlines anymore. A pseudocode version of the proposed algorithm is outlined in [Section 4.7](#). Our reference implementation is freely available as a standalone<sup>1</sup> (St-Jean, 2019), and will also be included in *ExploreDTI* (Leemans, Jeurissen, et al., 2009). We also make available the synthetic datasets and metrics extracted along the representative streamlines of the HCP datasets that are used in this manuscript (St-Jean, Chamberland, et al., 2018).

### 4.3.3 Datasets and acquisition parameters

**Synthetic datasets** A synthetic phantom consisting of 3 straight bundles crossing in the center at 60 degrees with a voxel size of 2 mm was created with phantomas (Caruyer et al., 2014). Each bundle has some partial voluming present on the outer edge to mimic the white matter / gray matter interface. We simulated 64 diffusion weighted images (DWIs) using gradient directions uniformly distributed on a half sphere and one  $b = 0$  s/mm<sup>2</sup> image with a signal-to-noise ratio (SNR) of 10, 20 and 30 with uniformly distributed Rician noise and a noiseless reference volume. Two distinct diffusion weightings of  $b = 1000$  s/mm<sup>2</sup> and  $b = 3000$  s/mm<sup>2</sup> were used, producing a total of 8 different synthetic datasets. The SNR was defined as  $\text{SNR} = S_0/\sigma$ , where  $S_0$  is the non-diffusion weighted signal and  $\sigma$  is the Gaussian noise standard deviation.

**HCP datasets** 100 subjects (50 males, 50 females) from the in vivo Human Connectome Project (HCP) database (Van Essen et al., 2012) aged between 26 and 30 years old were selected. All 18  $b = 0$  s/mm<sup>2</sup> volumes were kept along with the 90 volumes at  $b = 3000$  s/mm<sup>2</sup> in order to maximize the angular resolution (Tournier et al., 2013). The acquisition parameters were a voxel size of 1.25 mm isotropic, a gradient strength of 100 mT/m, a multiband acceleration factor of 3 and TR / TE = 5520 ms / 89.5 ms. We used the minimally preprocessed datasets which are already corrected for subject motion, EPI distortions and eddy currents induced distortions (Van Essen et al., 2012).

### 4.3.4 Local model reconstruction and fiber tractography

We used the constrained spherical deconvolution (CSD) algorithm (Tournier et al., 2007) with a recursive calibration of the response function (C. M. Tax et al., 2014) and spherical harmonics of order 8 to estimate the fiber orientation distribution functions (fODFs). We also computed the diffusion tensors using the REKINDLE approach (C. M. W. Tax et al., 2015) to exclude potential outliers from the data. We subsequently computed the apparent fiber density (AFD) maps (Dell’Acqua et al., 2013; Raffelt et al., 2012) from the fODFs and the FA and MD maps from the diffusion tensors (Basser and Pierpaoli, 1996) in all experiments. Whole-brain deterministic tractography was performed using the fODFs

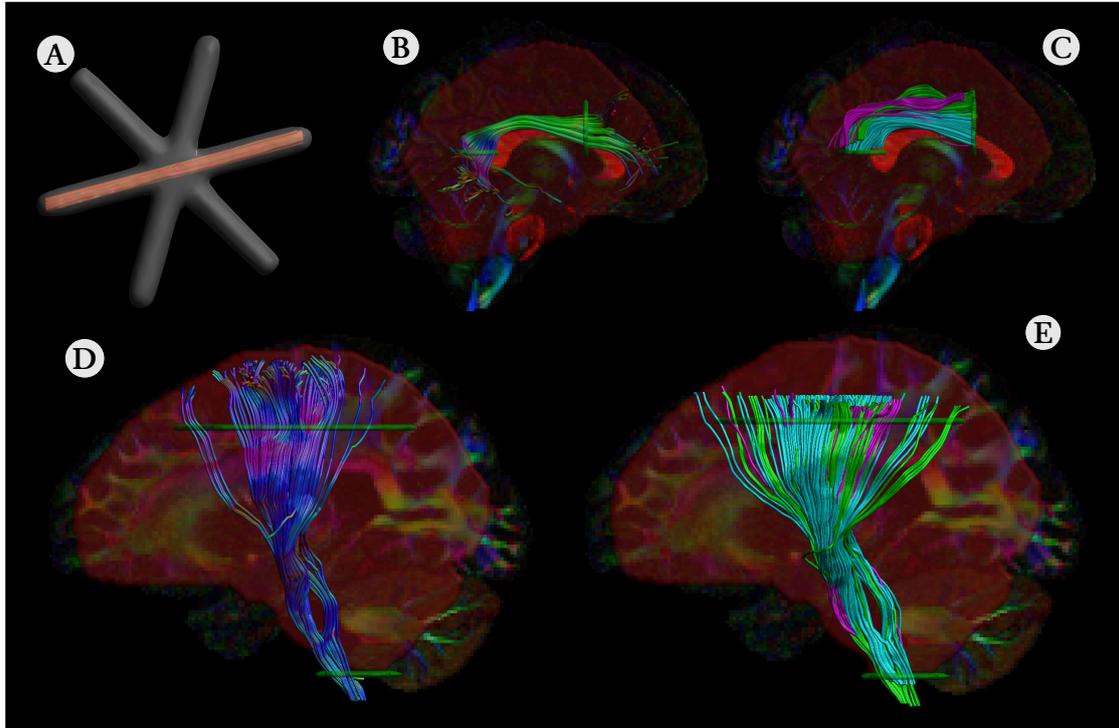
<sup>1</sup><https://github.com/samuelstjean/dpr>

with *ExploreDTI* (Leemans, Jeurissen, et al., 2009) with a step size of 0.5 mm, a fODFs threshold of 0.1 and an FA threshold of 0.2 for all datasets. The angle threshold, seeding grid resolution and streamlines length threshold used during tractography were different for the synthetic and HCP datasets as detailed below.

**Tractography parameters for the synthetic datasets** Tractography was performed with an angle threshold of 30 degrees and a seeding grid resolution of 0.5 mm on each axis to ensure a dense coverage of each bundle. Only the streamlines with a length of at least 10 mm and up to 150 mm were kept to prevent the presence of spurious streamlines. Two ROIs were manually drawn on one bundle to select only straight streamlines belonging to this bundle as shown in Fig. 4.5. The streamlines were kept to their full extent, including some small variations near the end points due to partial voluming, which ensures that the intersection of the three bundles is approximately at the center. To mimic similar representative streamlines extracted from various subjects, 150 streamlines were randomly selected and cut randomly from 1% up to 10% of their total length at both end points. Two sets of representative streamlines were created using classical resampling to the same number of points and our novel two-step resampling strategy, which is detailed in Section 4.3.1. In the first case, all streamlines were resampled to 50 points, which is approximately one unit point size per voxel. As each synthetic representative streamline had a different length after truncation, resampling to the same number of points allows a direct comparison between each coordinate, even if they do not match the same “anatomical” location by design of the experiment. No resampling was needed to simulate our proposed resampling strategy, as the distance between each point is already equal for this particular synthetic example.

**Tractography parameters for the HCP datasets** Whole-brain tractography was performed with an angle threshold of 45 degrees and a seeding grid resolution of 2 mm on each axis. Only the streamlines with a length of at least 10 mm and up to 300 mm were kept to limit the presence of spurious streamlines. ROIs were manually drawn to segment the left and right arcuate fasciculus (AF) and the left and right corticospinal tract (CST) on an exemplar subject (Wakana et al., 2007) as shown in Fig. 4.5. This exemplar subject FA map was used as a template and subsequently non linearly registered to each other subject respective FA map using Elastix (Klein et al., 2010). The obtained transformation was then applied on each ROIs drawn on the exemplar subject defining the four bundles, therefore warping the original ROIs unto each subject’s respective diffusion space as in Lebel et al. (2008). Only the segments between the ROIs were kept to only retain the straight sections and to remove spurious end points e.g. before the fanning in the CST. An alternative approach could be to extract the bundles automatically using a parcellation of the white matter obtained from each subject’s T1-weighted MR image (Cousineau et al., 2017; Wassermann et al., 2016). This would capture the full extent of the bundle instead of only keeping the sections between ROIs as done in the present work, but at the expense of possibly increas-

ing variability. Such an approach may be useful if important anatomical information is contained in these end regions.

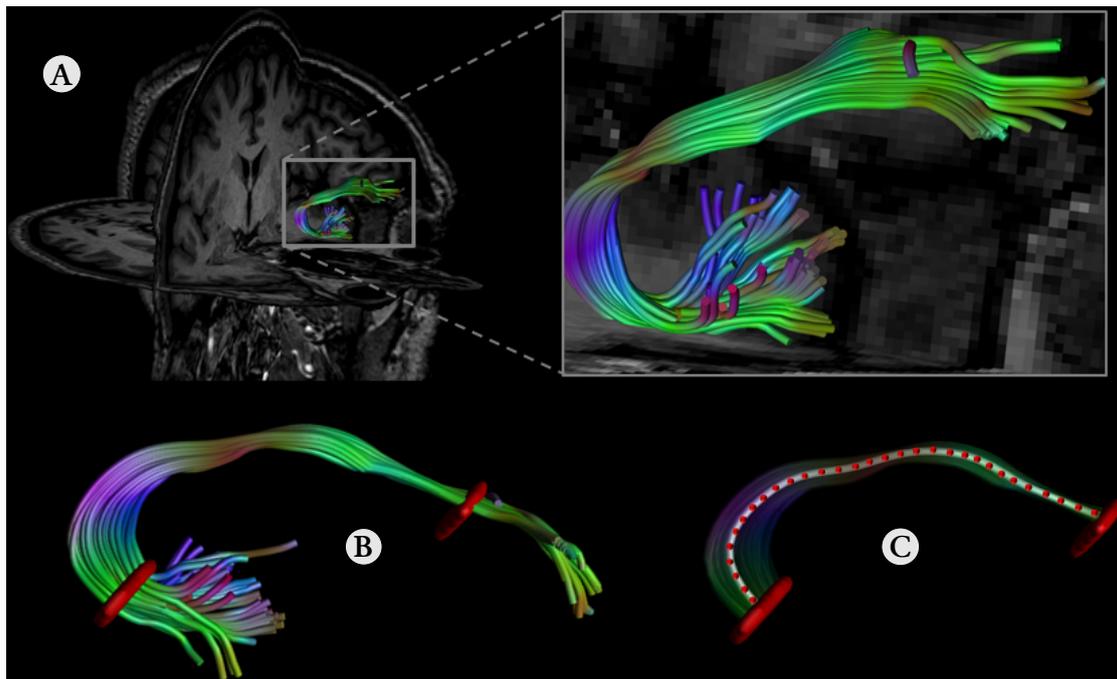


**Figure 4.5:** The synthetic bundles dataset and the locations of the ROIs used to segment some of the in vivo bundles on the exemplar subject with their automatically extracted counterpart for three subjects. In the top row, **A)** streamlines in a straight bundle of the synthetic datasets. Note that the streamlines are not truncated at the end points, but rather cover the full length of the red bundle so that they cross exactly at the center. The two inclusions (in green) and one exclusion (in red) ROIs segmenting **B)** the right AF on the exemplar subject and **C)** three automatically extracted right AF drawn in the exemplar subject native space (shown in green, cyan and magenta). On the bottom row, **D)** the left CST on the exemplar subject and **E)** three automatically extracted left CST bundles (shown in green, cyan and magenta) drawn in the exemplar subject native space. Note that each subject's bundle would correspond roughly to the same anatomical location in its own native space.

**Extracting representative streamlines for the HCP datasets** To extract the representative streamline of each subject, all streamlines forming a given bundle were linearly re-sampled to the same number of points, chosen as the number of points of the top 5% longest streamlines to reduce the effect of possible outliers. This choice is robust to possible outliers which might be longer (or much smaller) than the rest of the streamlines due to spurious results from tractography while also keeping a high sampling resolution, a desirable property for Eq. (4.1).

### 4.3. Materials and methods

In the present work, the mean streamline per bundle was extracted and finally resampled in two different ways: 1) using a fixed number of points for all subjects and 2) ensuring an equal distance between each point. For the classical resampling strategy, we resampled all subjects to 70 points for the arcuate fasciculi and 105 points for the corticospinal tracts. The second resampling strategy ensured that the distance  $\delta_{\min}$  (in mm) between each point is the same for all subjects. This also means that the representative streamlines of each subject do *not* have the same number of points and can not be compared directly at this stage when using this resampling strategy. Fig. 4.6 shows an example of selecting a representative segment between two ROIs as would be done for the uncinata fasciculus.



**Figure 4.6:** An example of along-tract analysis. **A)** The uncinata fasciculus is first segmented from a whole-brain tractography on an exemplar subject. **B)** The two ROIs (shown in red) that were defined to segment the uncinata fasciculus. Warping these ROIs to each subject provides an automatic dissection of the bundle. **C)** Only the portion of the mean streamline (shown in white) between the two ROIs is discretized (shown by the red dots), which allows mapping scalar metrics along the bundle itself.

#### 4.3.5 Extracting diffusion metrics for along-tract analysis

Once every representative streamline has been obtained, it can be used to collect diffusion derived metrics along the 3D pathway indexing a volume of interest. We collected the values of MD, FA and AFD for each subject along the streamline trajectory as in Colby et al. (2012). The resulting 1D segment is a vector of values varying along the length

of the representative streamline. This single representative pathway can now be realigned in a pointwise fashion to ensure correspondence between subjects before moving on to statistical analysis.

#### 4.3.6 Applying the diffusion profile realignment on representative streamlines

**Realignment of uniformly resampled and variable length streamlines** To evaluate the reduction in variability brought by our proposed DPR algorithm, we estimated the coefficient of variation (CV) at each coordinate along the streamlines before and after realignment using both resampling strategies. The CV, defined as  $CV = \sigma/\mu$  with  $\sigma$  the standard deviation and  $\mu$  the mean of each metric, is a unitless standardized measure of dispersion where a lower CV indicates a lower standard deviation around the mean value. For all experiments, we used a maximum displacement threshold of 15% to find the subject serving as a template during realignment. We computed the CV before and after the realignment of the representative streamlines using both resampling strategies. To compare the variability due to truncation of the end points, only the segments where 1% (at least one streamline present), 50%, 75% and 100% (all streamlines are fully overlapping) of the realigned streamlines were kept for computing the CV. In the synthetic datasets experiments, we weighted the CV by the number of points at each coordinate to account for the different number of points of the unaligned bundle. For experiments with the HCP datasets, we instead did a final resampling to the same number of points (if appropriate) after the realignment as previously used for the classical resampling strategy in order to ensure a fair comparison between both approaches.

**Simulating abnormal values of diffusion metrics in HCP subjects** An example application of the along-tract analysis framework could be to study neurological changes in a given population. These changes would presumably affect some specific white matter bundles and their underlying scalar values extracted from dMRI. Both the location and magnitude of these changes could reveal an effect of interest that might be hidden at first due to potential misalignment between subjects. To simulate a change in scalar metrics, 50 subjects were chosen randomly and had their representative streamlines profile modified while the other 50 subjects were left untouched. These 50 modified subjects are now classified as the “altered” subjects and the other untouched 50 subjects as the “controls” subjects in the subsequent experiments. For each altered subject, a location covering two times the affected length on both sides was chosen at random starting from the middle and the metrics were modified at this location. Two separate set of experiments were performed where the changes in metric was at first +10% and then -10% of its original value over 15% of the length. An additional set of experiments simulating highly focused damage of  $\pm 25\%$  and  $\pm 50\%$  of the metrics over 5% or 1% of the bundle length was performed. For the three cases, the randomly chosen location was at a position from 20% to 80%, 40% to 60% and 48% to 52% of the bundle length. This process is repeated for each metric

and each bundle, creating a different set of randomly modified subjects every time. The representative streamlines were finally realigned separately per group. As the control and altered subjects likely have different 1D profiles, realigning them separately makes it possible to select the best template for each group by itself. This strategy implicitly assumes that the neurological changes induce a similar increase or decrease in the diffusion metrics of each subject and that after realignment, each anatomical location is in correspondence between both groups. Correspondence between groups is also implied in classical along-tract analysis when resampling to the same number of points for comparison. Limiting the maximum displacement allowed also ensures that information carried by the diffusion metrics stays locally around the same position. The correspondence after separate realignment is assumed by resampling to the same number of points as the final step before analysis. In a clinical study setting, this could reflect neurological changes as induced by, e.g. a neurodegenerative disease or aging. The idea is to induce some changes in the extracted scalar values *only*, without modifying the underlying raw data or performing tractography and representative streamlines extraction once again. This choice of working in the extracted metric space only is to assess the changes in the metrics and realignment, in opposition to changes affecting the raw data itself. As the tractography process and extracted streamlines would most likely be slightly different due to the inherent challenges in reproducing tractography (Maier-Hein et al., 2017), the subsequent interpretation of the results could be confounded if tractography would be done anew.

**Statistical tests between HCP subjects** We conducted a Student’s t-test for independent samples between the controls and altered HCP subjects with a correction for the false discovery rate (FDR) of  $\alpha = 0.05$  (Benjamini and Hochberg, 1995) for one metric on each bundle. The t-test was realized on the datasets before and after the realignment of the representative streamlines metrics. However, the FDR correction only ensures an upper bound on the occurrence of false effects and do not indicate their location nor how many are present.

## 4.4 Results

### 4.4.1 Simulations with the synthetic datasets

We now present numerical simulations involving the synthetic datasets presented in [Section 4.3.3](#), comparing the two resampling strategies from [Section 4.3.1](#) before applying the DPR algorithm. [Fig. 4.7](#) shows the reduced CV for the realignment of the AFD metric on the SNR 20 dataset at  $b = 3000 \text{ s/mm}^2$  when compared to their non realigned counterpart. After realignment, the standard deviation at each coordinate is now generally lower, especially in the center portion where the three bundles are crossing. In the case of resampling to an equal distance  $\delta_{\min}$ , a few streamlines are overlapping at the end points, which might

reduce statistical power for these regions during subsequent analyses. As previously mentioned in [Section 4.3.2](#), portions where only a few streamlines are overlapping should be truncated accordingly to prevent these degenerate cases. [Fig. 4.8](#) shows summary boxplots of the CV in addition to the mean CV across all coordinates for the synthetic datasets for the MD, FA and AFD. In all cases, realignment provides a lower CV than the non realigned synthetic streamlines.

#### 4.4.2 Realignment of the in vivo HCP datasets

**Realignment of the arcuate fasciculi and corticospinal tracts** To quantify the improvements brought by the DPR algorithm for the in vivo datasets, we realigned the representative streamlines extracted from the 100 HCP datasets. [Fig. 4.9](#) shows the final outcome with the two previously discussed pipelines for producing along-tract averaged profiles: resampling to the same number of points as is conventionally done and after realignment with the DPR algorithm. For the realigned case, we kept only the segments where at least 75% of the subjects are overlapping and finally resampled all subjects to the same number of points. This last resampling step could be considered optional and is used to allow an easier visual comparison between the unaligned and realigned group profiles. While the overall shape of each profile is similar between the unaligned and realigned version, the end points and location of salient features are slightly different due to the realignment and the truncation threshold we used. As the maximum displacement threshold dictates which subject is used as a template for the realignment, average group profiles using a maximum displacement threshold of 5, 10 and 20% are shown in the supplementary materials [Section 4.8.1](#). To assess the effect of truncation on variance near the end points, we computed the CV for each metric at various truncation thresholds and for the unaligned metrics. [Fig. 4.10](#) shows the CV for the HCP datasets when the bundles are first resampled to the same number of points and after realignment (in brown). In all cases, the CV is approximately equal or lower after realignment with the DPR algorithm than when the representative streamlines are unaligned and resampled to the same number of points. We also show the CV in the unaligned case where all streamlines have an equal distance  $\delta_{\min}$  between points and for four truncation thresholds after applying the DPR algorithm (no truncation, 50%, 75% and 100% of overlap). In this particular case, the resampled and realigned bundles (light brown) and the realigned bundles with no truncation (green) are mostly equivalent as they are resampled to the same final number of points after realignment for comparison purposes. The main tendency shows a lower mean CV after realignment when compared to the non realigned cases. The CV values are also generally lower with increasing truncation thresholds as the number of overlapping points per coordinates is also increasing, contributing to a lower standard deviation of each metric.

**Robustness of the shapes of averaged profiles towards different metrics** When performing an along-tract analysis, tractography plays a key role as a spatial indexation method for

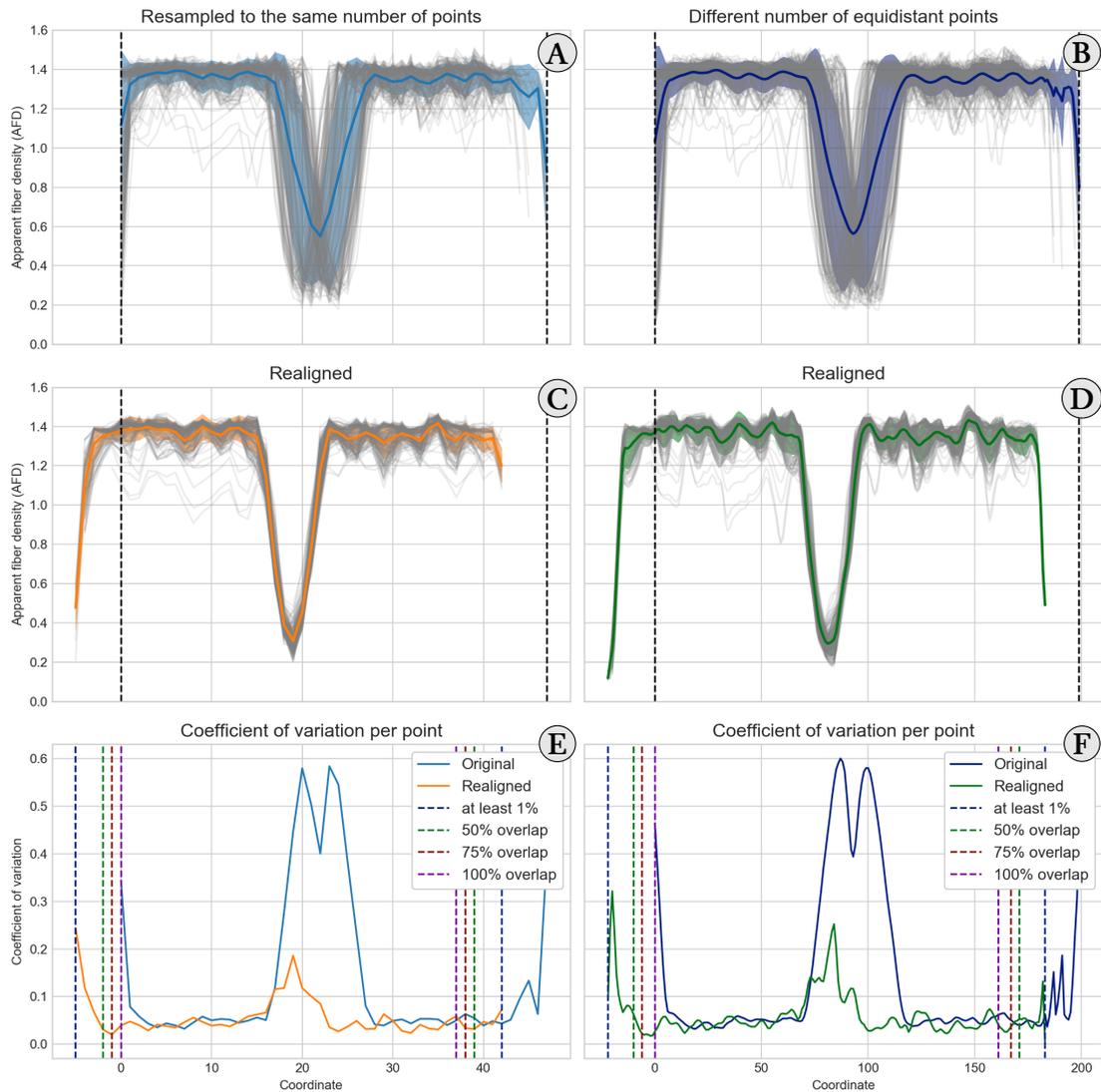
*extracting* the 1D metric profiles along the streamline. Given a particular subject representative streamline, the various scalar metrics that can be extracted each have their own distinct 1D profile along the streamline. In order to assess the robustness of our proposed DPR algorithm, we investigated whether for a given metric and template the resulting average group profile would be similar using the displacement computed from the other metrics. As the displacement depends on the spectrum of each 1D profile, each metric may use a different template and apply a different displacement for each subject. This may ultimately lead to a different group average profile due to our algorithm automatically choosing the template amongst the subjects. However, the *relative* displacement due to a change of template (and hence the resulting group average 1D profile) may be unaffected by this choice, leading to a similar group average profile. Fig. 4.11 shows the resulting average group profiles for each metric when using the original realignment and the realignment that would be applied from the two other remaining metrics with a maximum displacement threshold of 15%. As the AF is slowly varying in terms of diffusion metrics along its extracted path, the realignment of the MD metric is similar even when using the displacement computed from the FA or AFD metric. On the other hand, applying the realignment given by the MD to the FA and AFD profiles leads to different optimal realignments and a change in their overall profile. For the CST, as the representative streamline crosses other anatomical bundles along its path, the 1D profiles have more variation along coordinates than in the AF case. This is mostly notable in the MD metric profile, which is now similarly realigned when using either the FA or AFD. Due to these anatomical “landmarks”, the displacement given by the MD also yields similar profiles when applied to the FA and AFD metrics. Results for maximum displacement thresholds of 5, 10 and 20% produced similar trends, which are shown in the supplementary materials Section 4.8.2.

**Realignment with simulated diffusion abnormalities in HCP datasets** We first focus on the new strategy of resampling the representative streamlines, while ensuring that the distance between each point  $\delta_{\min}$  is the same. As one can always resample to a common number of points after realignment, this prevents a reduced sampling resolution when using Eq. (4.1). Automatically selecting a template from the subjects themselves allows the DPR algorithm to be as flexible as possible. The changes in scalar metrics (e.g. introduced by local alteration of tissue microstructure following disease) might not be obviously identified on the group average for the unaligned streamlines case, but the variations in shape of the realigned group average may be uncovered by selecting a new template. Fig. 4.12 shows four examples of the unaligned and realigned profiles of the scalar metrics for the datasets with and without simulated diffusion abnormalities for each white matter fiber bundle. Note how the original and altered unaligned streamlines have a similar profile for both metrics at first, but the realigned altered streamlines now have a different profile which was uncovered by realignment with the DPR algorithm (see the red boxes in Fig. 4.12). This is especially prevalent in the case of the MD metric where the unaligned profiles are similar for the

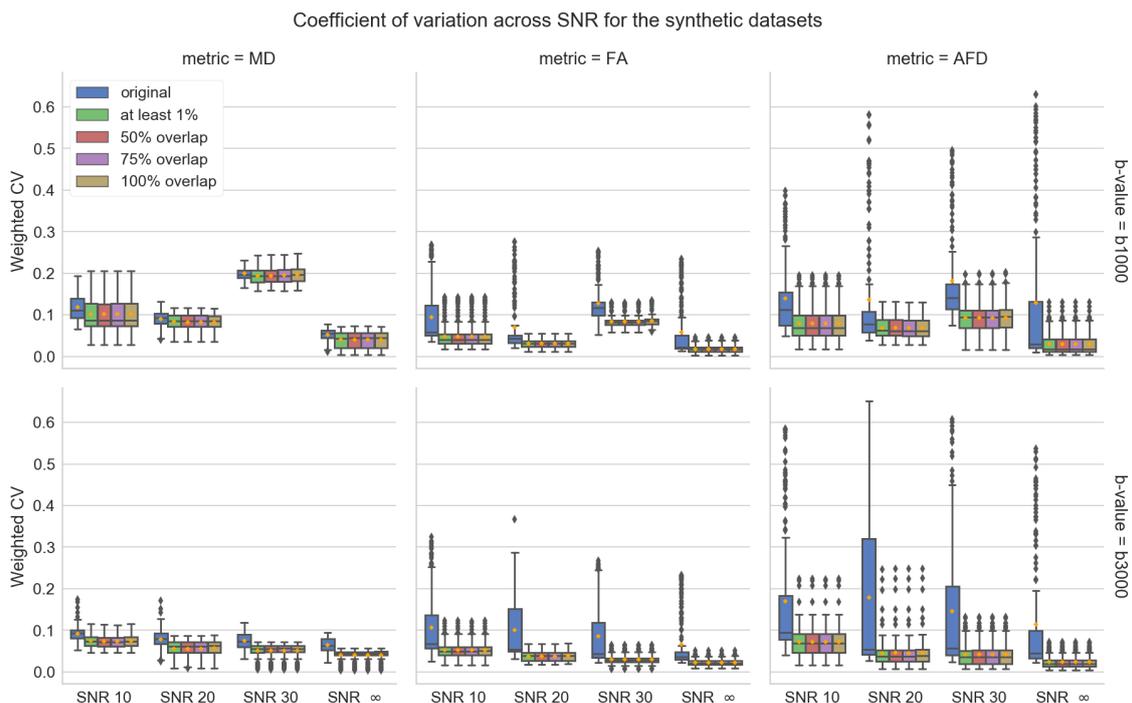
control and altered subject data, while realignment uncovers the higher MD values that were originally simulated.

**Statistical hypothesis testing** We now look at uncovering groupwise differences between the control and altered HCP subjects over the affected regions. Fig. 4.13 shows the results of the unpaired t-test for the HCP datasets before and after realignment for the A) AF left with the MD metric, B) AF right with the FA metric, C) CST left with the FA metric and D) CST right with the AFD metric, as previously shown in Fig. 4.12. All of the regions uncovered before using realignment are also identified as statistically significant at the level of  $p\text{-value} < 0.05$  after realignment. This indicates that findings for the unaligned case are preserved when using our proposed algorithm, with the addition of new affected regions, which might have been averaged out due to misalignment in the first place. For example, the left AF and left CST showcase an affected portion that is statistically significant only after realignment. However, using a lower statistical threshold or a higher level  $\alpha$  for the FDR might reveal additional affected regions at the cost of introducing potential false positives. Fig. 4.14 shows a second set of experiments on the four bundles realized with large alterations of the metrics which are spatially focused e.g. in the presence of tumors. Specifically, alterations in the metrics of 25% or 50% were induced over 1% or 5% of each bundle length and each group subsequently realigned with DPR. Unpaired t-test before and after realignment are conducted between the two groups at each location as in Fig. 4.13. Almost all affected regions are identified before and after realignment when the affected length is of 5%. For the CST left, the affected region is only identified after realignment when the alteration is of 25%. When only 1% of the bundle length is affected, no changes are identified before realignment, but are uncovered after realignment with the DPR algorithm in all cases. Results obtained with maximum displacement thresholds of 5%, 10% and 100% are shown in the supplementary materials Section 4.8.3.

#### 4.4. Results

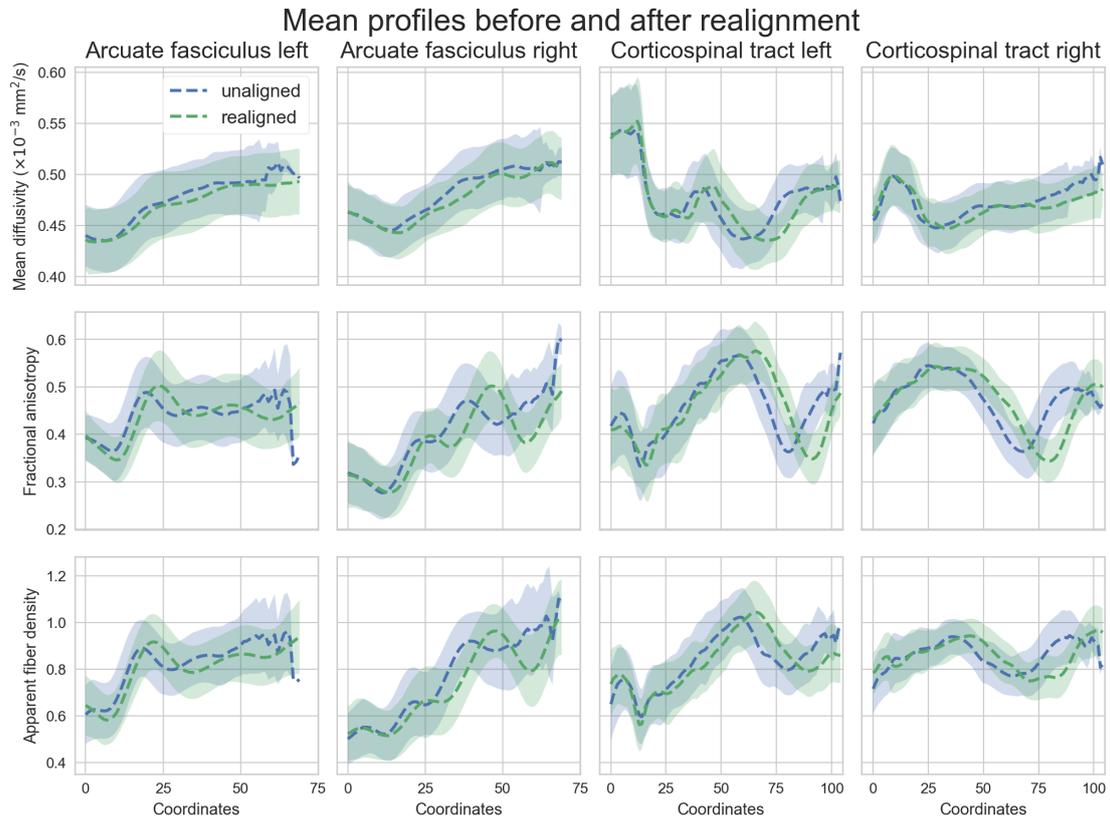


**Figure 4.7:** Realignment of representative streamlines resampled to 50 points (left column) and with an equal distance  $\delta_{\min}$  (right column) for the AFD case at SNR 20 and  $b = 3000 \text{ s/mm}^2$ . Each individual streamline is plotted in light gray, with the mean value in color and the standard deviation as the shaded area. The black vertical bars indicate the location of the original, non realigned streamlines. The colored vertical bars indicate the number of overlapping streamlines, ranging from at least 1 (all subjects, purple lines) to all of them (100%, red lines). Panels **A**) and **B**) show the streamlines before realignment. Note how individual streamlines are rather dispersed around the mean. Panels **C**) and **D**) show the streamlines after realignment, with the mean value being closer to all of the subjects and a smaller standard deviation than in panels **A**) and **B**). However, due to the realignment, the end points have less subjects contributing to the mean value and should be truncated according to the number of overlapping subjects. Panels **E**) and **F**) show the coefficient of variation (CV, where lower is better) for each point, which is in general lower or equal than the non realigned version in both cases. Note how the largest reduction in CV is in the crossing region, where the standard deviation is approximately three times smaller in the realigned case than for the unaligned case.

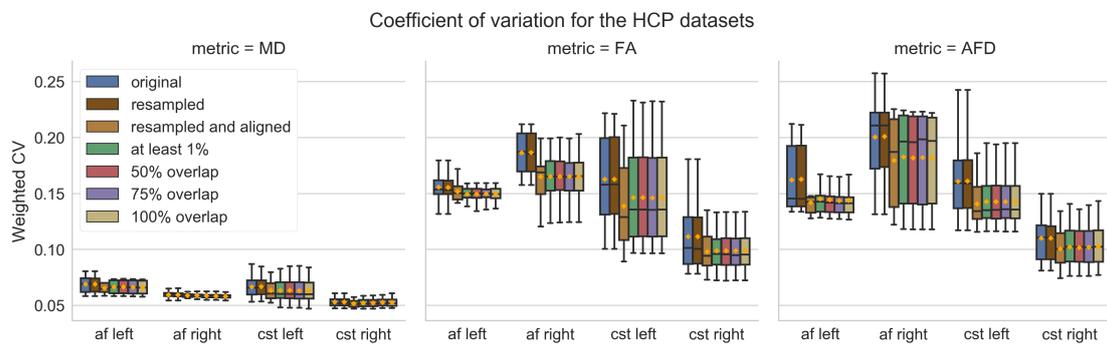


**Figure 4.8:** Boxplots of the CV for each point weighted by the number of overlapping subjects, for the MD (left), FA (center) and AFD (right) metrics and their respective mean value (in orange). The top row shows results for  $b = 1000 \text{ s/mm}^2$  on the synthetic datasets at SNR 10, 20, 30 and in the noiseless case while the bottom row shows results for  $b = 3000 \text{ s/mm}^2$ . In all cases, the realigned metrics (for any truncation percentage) have a lower or equal CV on average than the non realigned metrics (in blue). The FA and AFD metrics have a CV in the realigned case which is on average approximately two times smaller than the non realigned case across all SNRs and both b-values. This gain is smaller for the MD, which might be due to the relative homogeneity of the MD values.

## 4.4. Results

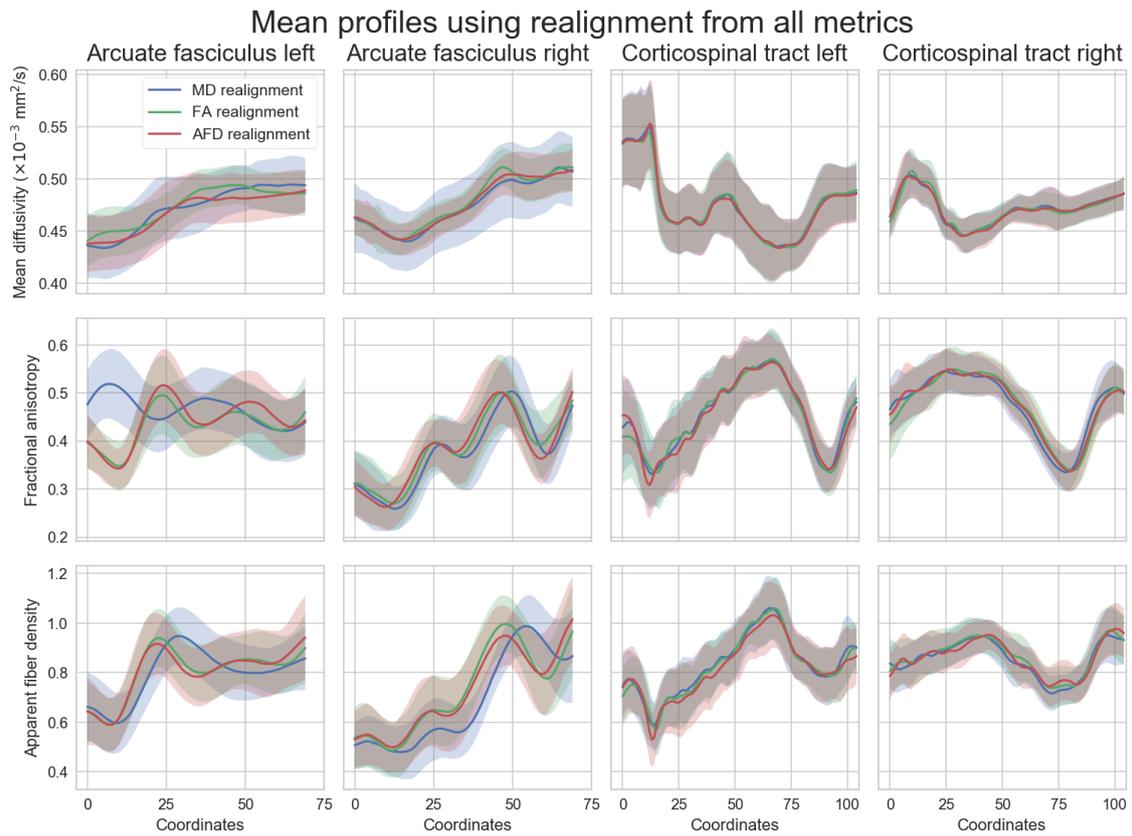


**Figure 4.9:** Along-tract averaged profiles (and standard deviation as the shaded area) of the unaligned (blue) and realigned (green) HCP subjects truncated to 75% of overlap with a final resampling to the same number of points. Each row shows the profile for one diffusion metric (MD, FA and AFD) while each column shows one of the studied bundles (AF left/right from anterior (coordinate 0) to posterior and CST left/right from inferior (coordinate 0) to superior). After realignment and truncation, the profiles are slightly different from their unaligned version at the end points while the center profile is similar. This is likely due to the misalignment mostly affecting the initial end points which are defined by the original truncation from the ROIs.

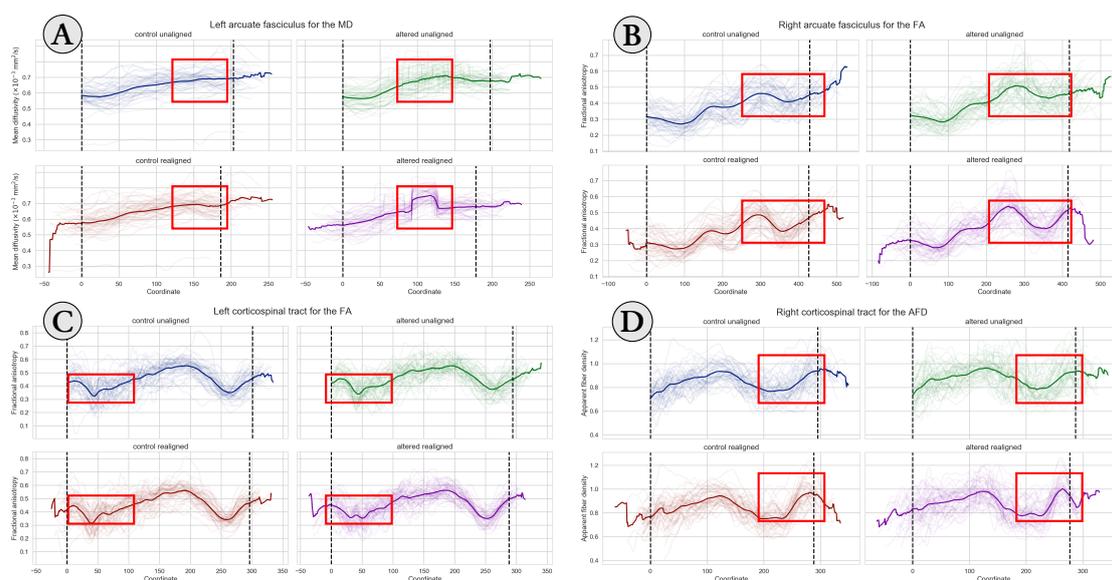


**Figure 4.10:** Boxplots of the CV for each point weighted by the number of overlapping subjects, for the MD (left), FA (center) and AFD (right) metrics and their respective mean value (in orange) for the four studied bundles. Similar to the synthetic datasets experiments, the in vivo datasets have a lower CV after realignment (green, red, purple and yellow boxplots) than when they are unaligned (brown boxplots). Even if the representative streamlines are truncated to the shortest number of points (yellow boxplot) or are resampled to the same length (light brown boxplot), the CV is smaller in the realigned cases than in the unaligned cases (brown and blue boxplots respectively). The gain in CV is once again smaller for the MD but larger for the FA and AFD in favor of the realigned cases, which is in line with the synthetic experiments.

## 4.4. Results

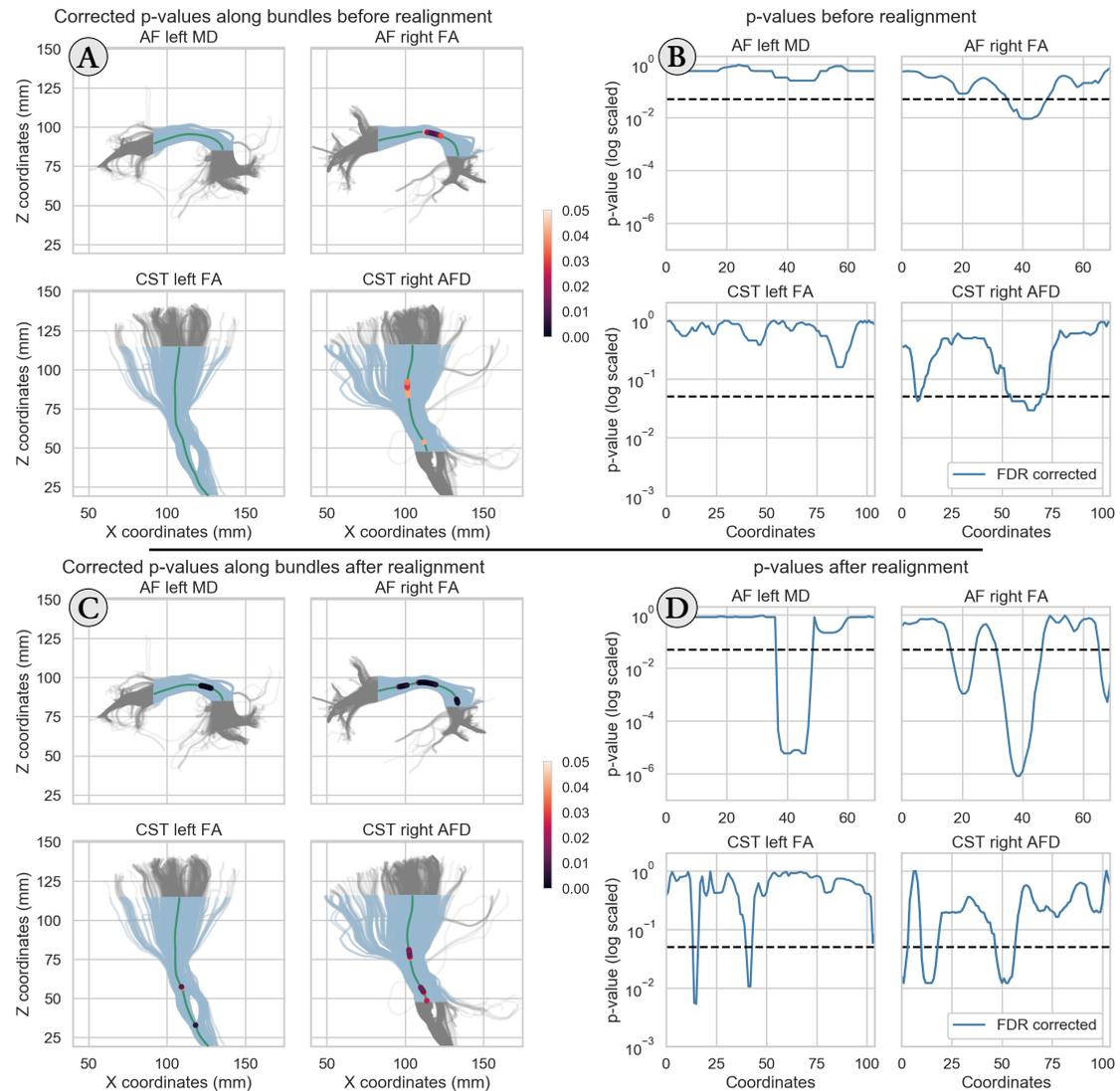


**Figure 4.11:** Along-tract averaged profiles (and standard deviation as the shaded area) of the white matter fiber bundles (columns) from the HCP datasets after realignment for each studied metric (rows). The metrics were truncated to 75% of overlap after realignment with a final resampling to the same number of points. On each row, the along-tract profile after realignment is shown for a given metric (MD on the first row, FA on the second row and AFD on the third row) using the displacement computed by the MD (blue), FA (green) and AFD (red). The AF are displayed from anterior (coordinate 0) to posterior and the CST from inferior (coordinate 0) to superior.



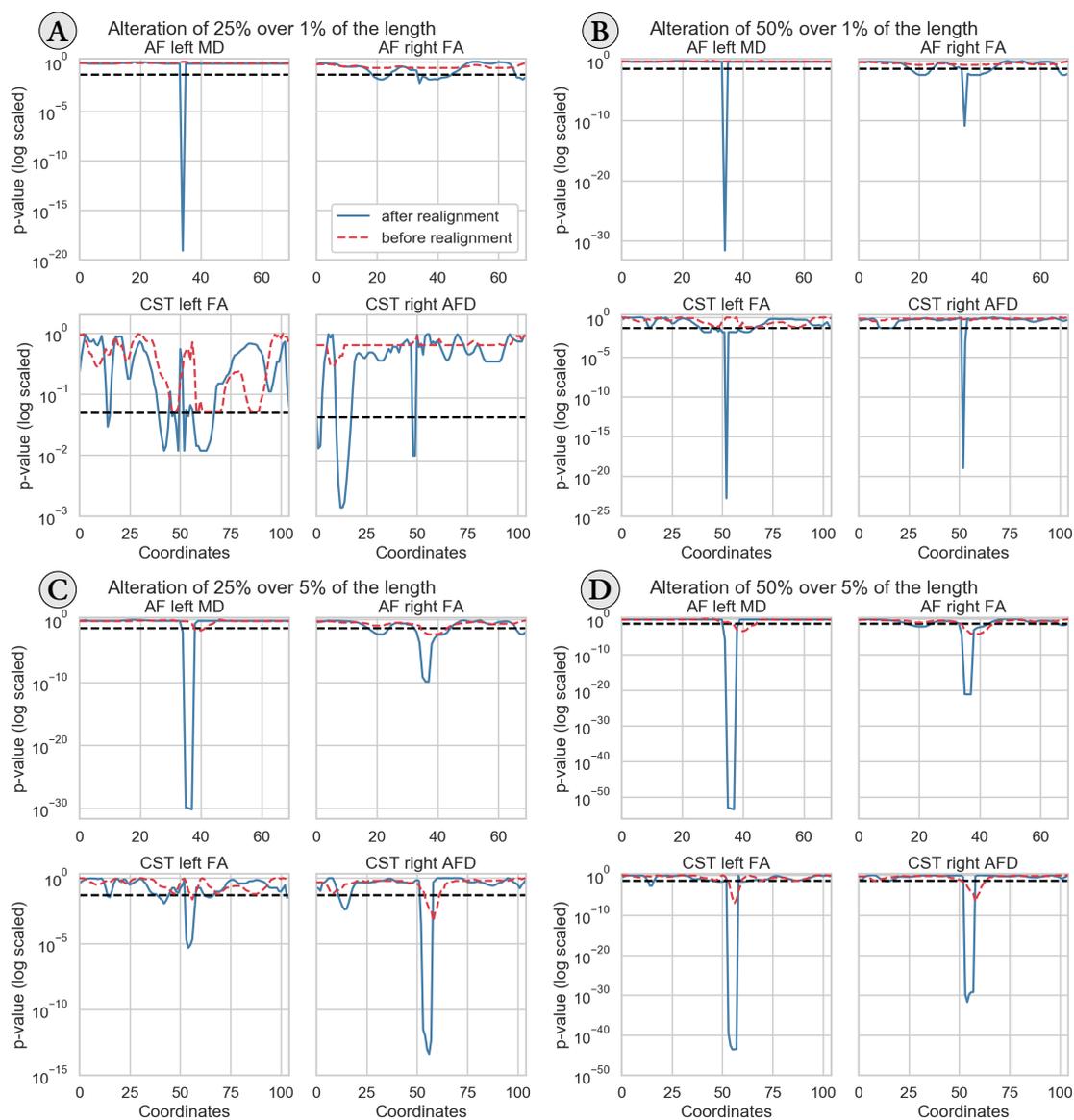
**Figure 4.12:** Comparisons between the unaligned and realigned profiles for the HCP datasets without (control column) and with (altered column) simulated diffusion alterations in the white matter fiber bundles. A different bundle for a specific metric is shown in each subfigure: **A)** AF left for the MD, **B)** AF right for the FA, **C)** CST left for the FA and **D)** CST right for the AFD. The AF are displayed from anterior (coordinate 0) to posterior and the CST from inferior (coordinate 0) to superior. Each subject representative streamline is rendered transparently and the group average representative streamline is represented by the solid line. The black bars indicate where at least 75% of the subjects are overlapping. Some key visual differences (red boxes) are hidden by misalignment between the control and altered subject data when they are unaligned, while realignment helps to uncover those hidden degeneracies. Note that the red boxes in the subgraphs have the same size and are aligned for easier visual comparison. The most striking example is in **A)** where the change in MD is easier to see after realignment as the control subjects are keeping their original shape while the altered datasets exhibit a drop in their scalar value around the same region. The unaligned group average streamline however makes this difference harder to uncover.

## 4.4. Results



**Figure 4.13:** Unpaired t-test corrected for false discovery rate (FDR) at  $\alpha = 0.05$  overlaid on the exemplar subject bundle for the same cases as in Fig. 4.12. On the left, fiber trajectories of the exemplar subject (in gray) and truncated portions of these pathways between the ROIs (in blue) expressed in world coordinates **A**) before realignment and **C**) after realignment with the DPR algorithm. The p-values at locations deemed statistically significant in the present work ( $p < 0.05$ ) are overlaid on the average streamline (in green). On the right, the p-values on a log scale after FDR correction along the average streamlines **B**) before realignment and **D**) after realignment with the DPR algorithm, but expressed as along-tract 1D point coordinates. The horizontal black bar is located at p-value = 0.05. In the realigned data case, the p-values are lower in the significant regions (corticospinal tract right) or even show affected regions which are not detected when the data is unaligned (arcuate fasciculi and corticospinal tract left). The most prominent case is for the left arcuate fasciculus, where the affected portion is not identified in the unaligned case (for our chosen significance threshold of 0.05), but has a corrected p-value of approximately  $10^{-5}$  after realignment.

## Focused alterations before and after realignment



**Figure 4.14:** Unpaired t-test (FDR corrected at  $\alpha = 0.05$ ) with focused alterations of the metrics for each bundle of **A)** 25% over 1% of the length, **B)** 50% over 1% of the length, **C)** 25% over 5% of the length and **D)** 50% over 5% of the length. The AF left/right are represented from anterior (coordinate 0) to posterior and the CST left/right from inferior (coordinate 0) to superior. The p-values are on a log scale along the average streamline before realignment (dashed red lines) and after realignment (solid blue lines) with the DPR algorithm. The horizontal dashed black lines indicate  $p\text{-value} = 0.05$ . When alterations cover 1% of the length, the affected profiles are identified only after realignment. At 5% of the length, the uncovered regions after realignment are concentrated around smaller sections than their counterpart before realignment.

## 4.5 Discussion

### 4.5.1 Reducing variability along bundles

Using simulations, we have shown how residual misalignment may hide the expected average profile of an along-tract analysis. [Fig. 4.7](#) shows this effect directly as the group mean profile from a set of streamlines only roughly corresponds to their individual, but in truth identical, shape as their spatial location differs due to small differences in their length. Realignment not only restores the expected group profile, but also reduces the pointwise variability of the metrics as the unequal streamlines are now aligned as reflected by the lower overall CV. Each individual subject is therefore participating to the group average instead of being spread out and biasing the estimated mean scalar value of the overall bundle in the crossing region. This is also true if the streamlines are first resampled to the same number of points. In this case, the variance at the end points is larger, possibly due to a loss in spectral resolution caused by resampling to a lower number of points than originally present. Resampling early in the along-tract analysis pipeline may not only inadvertently hide information for the realignment, but also hamper statistical testing by reducing the spatial specificity of the data (O'Donnell, Westin, et al., 2009).

For the realignment of the in vivo HCP datasets, [Fig. 4.9](#) shows that realignment alters the group profile at the end points while preserving the overall shape and the central portion of the bundle. This leads to a reduction of the CV, likely due to the reduction in variance at the end points while the overall mean profile is preserved as shown in [Fig. 4.10](#). As the realigned end points will also have less data from different subjects present at each coordinate, subsequent truncation further reduces the CV once again. The change of shape after realignment is possibly due to the difference in length between each subject and the subsequent mapping to their 1D metric profile. This 1D space hides the spatial 3D coordinates misalignment that may be present between subjects. However, this misalignment can still be mitigated afterwards. Even if the representative streamlines are shifted as a whole with the realignment, preservation of the overall shape and center portion might indicate that only the end points were dissimilar. The lower end point variance effect is also present when using the classical resampling strategy and subsequently realigning the representative streamlines. The misalignment at the end points between subjects is due in part to the truncation effect of the ROIs and to the nature of tractography itself and its many user-defined settings (Chamberland, Whittingstall, et al., 2014). The use of termination criteria (e.g. FA threshold, white matter mask, maximum curvature) or seeding strategy (e.g. white matter versus cortex seeding) (Girard et al., 2014) may prematurely terminate tractography in the middle of a white matter bundle, contributing in producing shorter streamlines which end before fully reaching the gray matter (Maier-Hein et al., 2017). New algorithms and seeding strategies are developed to enhance tractography end points near the cortex (St-Onge et al., 2018) and could help to reduce this truncation effect.

### 4.5.2 Effect of exchanging metrics for realignment

We have shown in [Fig. 4.11](#) the effect of applying the realignment computed from different metrics on the mean group tract profile. From these results, we can observe the different displacement values obtained from the dMRI metrics, even though the representative streamlines arise from the same anatomical location. This is due to the fact that our framework is fully driven by the 1D profiles of the studied metric, which all have different shapes and features, leading to slightly different realignment outcome depending on the bundle and the metric that is used. As the FA and AFD profiles are similar in the four studied bundles, exchanging their value still leads to the same overall profile in most cases. For the MD, results showed that the CST is also stable. This is most likely due to the complex 1D profile of the CST for the three metrics, as it defines unique landmarks that are picked by our algorithm for accurate realignment. Regarding the AF, exchanging the displacement from the FA or AFD yields similar profiles, an observation that does not hold for the MD metric. As the MD metric for the AF has a rather flat profile, the algorithm might pick up a spurious displacement due to the lack of well-defined features to exploit. Avants et al. (2011) also reached a similar conclusion in the context of 3D volume registration when using different metrics such as the mean square difference, cross-correlation or mutual information; using different metrics, type of registration or registering subject A onto subject B (and vice versa) leads to slightly different outcomes. We have fixed the maximally allowed displacement to 15% of the length of the bundle, but similar conclusions also applied for 10% and 20% of maximum displacement as shown in [Section 4.8.2](#). When the maximum displacement is only 5%, the AF show similar mean profiles for the three metrics, whereas the opposite is seen in the CST. This indicates that the maximally allowed displacement should be chosen per bundle and is data dependent. Short, straight and simpler bundles, such as the AF, might only need small realignment, whereas more complex structures with fanning, intersecting bundles and possibly large anatomical variations between subjects, such as the CST, likely benefit from larger maximum displacement thresholds to find their full overlap between subjects.

### 4.5.3 Identifying brain regions affected by abnormalities along-tract

One of the end goal of along-tract analysis is to uncover alteration of the white matter due to, e.g. disease at their specific locations. This is at the cost of trading the sensitivity of ROI averaging based analysis for additional specificity along the bundle, which also depends on the discretization of the points forming the streamlines (O'Donnell, Westin, et al., 2009). Using simulated changes in scalar metrics from the HCP subjects, we have shown in [Fig. 4.12](#) how misalignment can artificially reduce the specificity of along-tract analysis. As the affected portion of the bundle is usually unknown *a priori*, morphological differences between subjects might map the affected area to different points in their 1D profile during the representative streamlines extraction. The unaligned metrics might exhibit similar

mean profiles between the control and altered subjects in this case, as the affected portions would be originating from an adjacent anatomical location in each subject's original space, but would not be aligned in the 1D space. The mean representative streamline at the group level could therefore average out each subject's individual difference due to residual misalignment, hiding the effect of interest in the process. As we have previously mentioned in [Section 4.1](#), this effect of averaging out important information has also been theorized by O'Donnell, Westin, et al. (2009). However, the same effects can also become easier to detect after realignment since the control subjects mean profile will potentially be different from the altered subjects mean profile. This is thanks to the particular features of their 1D profile now being realigned instead of averaged out. In a similar way, if changes in the diffusion metrics are potentially present across the whole length of a white matter bundle, the maximum displacement threshold should be increased. This may reduce the number of subjects identified as outliers by using a smaller maximum displacement, which would not have been realigned in the first place. The tradeoff in allowing a larger maximum displacement is a potential reduction in statistical power or false discoveries as less subjects may be present at each along-tract location for statistical testing.

In our simulations, changes on the left AF and left CST are identifiable only after realignment whereas the original control and altered average profiles are mostly similar since each individual contribution is lost in the unaligned group averaging. After realignment, the altered region can be identified as each realigned subject now contributes to the group average at the same location. This effect is similar to what we observed in our simulations in [Section 4.4.1](#), where the CV is lower in the crossing-bundles region after realignment and how the mean group profile is also lower after realignment. It is also noticeable on the right AF bundle with the FA metric or on the CST bundles, but to a lesser extent, as the overall morphology of the CST bundles stays relatively similar even after altering the scalar metrics. Interestingly, the altered group profile seems to be subject to larger morphometric changes after realignment than the control group counterpart. This might indicate that sharp profile changes in each subject's shape due to disease are automatically picked up by our algorithm, providing realignment based on this change.

We also conducted unpaired Student's *t*-tests to statistically identify the altered regions on the same bundles and metrics as shown in [Fig. 4.13](#). While we used an FDR correction of  $\alpha = 0.05$ , different results could be obtained by choosing a different value of  $\alpha$ . However, the main conclusion should still be valid; statistical testing performed on the realigned datasets uncovered affected regions that were not identified in the unaligned case as shown from the global *p*-values plot. This difference could be partly due to the residual misalignment between subjects inadvertently canceling out the effect of interest, as coordinates are not overlapping. In this study, we considered statistical testing at the spatial resolution in the order of magnitude of one voxel size (1.25 mm in our case), but studying larger bundle segments could be used as a compromise between averaging data over the whole bundle in order to uncover effects of interest at the expense of spatial specificity (O'Donnell, Westin,

et al., 2009).

#### 4.5.4 Mapping to 1D space versus registration methods

In the present work, we concentrated on reducing the effect of residual misalignment between representative streamlines. As tractography is a mandatory step before using our approach, registration methods for raw dMRI datasets would likely not reduce the misalignment resulting from streamlines extraction. Some registration methods specifically work directly on the streamlines or bundles space (e.g. Leemans, Sijbers, et al. (2006)), but the same transformation should be applied to the underlying 3D volume containing the metric of interest. This is because we work on metrics extracted from representative streamlines, and not directly in the streamlines space itself, see e.g. Glzman et al. (2018) and O'Donnell, Daducci, et al. (2017) and references therein for a review of registration methods in dMRI.

O'Donnell, Westin, et al. (2009) state that “because within a bundle fibers have varying lengths and their point correspondence is not known a priori, it is not possible to directly average fiber coordinates to calculate a mean fiber”; care must be taken during the representative streamline extraction step that is at the core of the along-tract analysis framework. As such, the required step dictating this possible misalignment is the mapping strategy used to extract the representative streamline and how its end points are defined. Various schemes have been proposed such as assignment to perpendicular planes (Corouge et al., 2006), variants reducing the effect of outliers by additionally considering the spatial distance between streamlines (O'Donnell, Westin, et al., 2009), extracting representative core streamlines with splines (Chamberland, St-Jean, et al., 2018) or resampling to a common number of points (Colby et al., 2012). All these choices inevitably lead to differences and a mismatch across subjects after metric extraction, even if the original underlying anatomy would be perfectly aligned as we have shown in our synthetic experiments in Fig. 4.7. Assignment and truncation strategies between the common points of bundles have been explored in Colby et al. (2012) with the authors noting that all compared methods are generally successful in extracting a meaningful (but slightly different) representation as they use different strategies and parameters. Close similarities in the extracted metrics using the representative streamline could explain why 1D misalignment, while still present, had not been thoroughly investigated previously. Reliably extracting the information from fanning regions (e.g. CST towards the motor cortex) or from a splitting configuration (e.g. anterior pillars of the fornix) in a single representative streamline still remains an open problem (Chamberland, St-Jean, et al., 2018).

#### 4.5.5 Assumptions of the DPR algorithm and limitations of this study

In the present work, we exchanged the classical assumption of 1D *spatial correspondence* between points for the assumption of an equal 1D *spatial distance* between points. This

latter requirement is usually fulfilled with the use of a fixed step size during tractography, but might be void by the representative streamline extraction. Without loss of generality, we chose to resample each subjects' representative streamline a second time to ensure an equal distance  $\delta_{\min}$  between each point. We advocate resampling to a larger number of points than initially present to reduce possible complications due to aliasing or using windowing functions for filtering (Stoica and Moses, 2005). While this theoretically increases the computational complexity of the DPR algorithm, it also preserves the full spectra when applying Eq. (4.1). This is not a problem in practice owing to the existence of efficient FFT implementations; our algorithm can realign the 100 HCP subjects in less than 3 seconds on a standard desktop with a 3.5 GHz Intel processor. The resulting realigned metrics can then be resampled back to approximately one point per unit voxel size to minimize the effect of multiple comparisons during statistical testing. With the development of new methods that go beyond fixed step size tractography, such as the use of compressed streamlines (Rheault et al., 2017), it might be beneficial to avoid this resampling step for computational reasons after sampling metrics along non regularly spaced streamlines. Another approach to remove the need of resampling could be to use an FFT implementation dealing with non-equal sampling of the data (Dutt and Rokhlin, 1993; Scargle, 1989), but such implementations may not be as widely (and easily) available as the classical equispaced version of the FFT algorithm.

Due to the difficulty in reproducing tractography (Maier-Hein et al., 2017), our simulations on the in vivo datasets were designed around altered versions of already extracted scalar values. One would however expect true neurodegenerative changes to additionally influence the steps prior to tractography such as the main orientations extracted from tensors or fODF. The results we obtained should translate as long as a representative streamline for each bundle of interest can be reliably delineated for all subjects. Similar recommendations apply if the white matter bundle of interest is largely affected by disease or altered when compared to the expected overall shape from a healthy subject. Specific care should also be taken during the prior step of extracting the representative streamlines in these cases to ensure that relevant portions of the bundles of interest are present in all subjects (Parker et al., 2016).

Although not considered in the present work, any quantitative diffusion metric such as the diffusion kurtosis metrics (e.g. mean kurtosis (MK)) (Jensen and Helpert, 2010), the axon diameter (Assaf et al., 2008), or metrics provided by NODDI (Zhang et al., 2012) could be studied using our proposed framework. In cases of physical alterations of the white matter (e.g. tumors, lesions), the diffusion metrics themselves may not provide accurate landmarks for realignment due to differences in tractography when extracting the representative streamline of each subject. The use of shape descriptors, such as torsion or curvature of the bundles themselves (Leemans, Sijbers, et al., 2006), could also be employed with DPR instead of diffusion metrics as done in the present work. These descriptors may also be useful in cases where using a large maximum displacement threshold may yield false

positives detections if the effects are small, see the supplementary materials [Section 4.8.3](#) for examples. In a similar fashion, any other volume (e.g. T1 or T2 relaxometry values (Deoni et al., 2008)) providing anatomical information of interest can be used once co-registered to each subject’s native diffusion space. Combining the realignment information from multiple or complementary metrics (e.g. computing their average displacement) may improve the robustness of the DPR framework. When white matter alterations are affecting the diffusion metrics to an unacceptable extent, the average displacement from these independent anatomical features (which are presumably less affected by these effects) could be used to circumvent this issue.

We did not investigate realignment of lateralized bundles (e.g. realignment of the left and right AF together instead of separately) which can be useful for studying intra-hemispheric differences between subjects (Catani et al., 2007). Variations between left and right anatomical locations also implicitly assume that each coordinate in the 1D space is matched against its inter hemispheric counterpart. To facilitate this mapping between hemispheres, O’Donnell, Westin, et al. (2009) proposed to mirror all streamlines from one hemisphere to the other, allowing a direct correspondence between the subsequently extracted representative streamlines as they would be effectively identical. However, the 3D volume used to extract the scalar metrics of interest would possibly be different in each hemisphere. In this context, the realignment could be done separately for each side, providing different profiles reflecting lateralization.

## 4.6 Conclusion

In this paper, we developed a new correction strategy, the diffusion profile realignment (DPR), which is designed to address residual misalignments between subjects in along-tract analysis. Through simulations on synthetic and in vivo datasets, we have shown how realignment based on our novel approach can reduce variability at the group level between subjects. Furthermore, realignment of the in vivo datasets provided new insights and improved sensitivity about the location of the induced changes, which could not be completely identified at first when misalignment was present. The DPR algorithm can be integrated in preexisting along-tract analysis pipelines as it comes just before conducting statistical analysis. It can be used to reveal effects of interest, which may be hidden by misalignment and has the potential to improve the specificity in longitudinal population studies beyond the traditional ROI based analysis and along-tract analysis workflows.

## 4.7 Appendix: The diffusion profile realignment algorithm

This appendix outlines the diffusion profile realignment (DPR) algorithm. Our implementation is also freely available at <https://github.com/samuelstjean/dpr> (St-Jean, 2019) and will be a part of *ExploreDTI* (Leemans, Jeurissen, et al., 2009). The synthetic datasets

#### 4.7. Appendix: The diffusion profile realignment algorithm

---

and metrics extracted along the representative streamlines of the HCP datasets used in this manuscript are also available (St-Jean, Chamberland, et al., 2018).

To complement Eq. (4.1), the shift needed to maximize the overlap between the vector  $x$  and  $y$  is the maximum of the CCF, given by

$$\text{shift}(x, y) = \arg \max(\text{CCF}(x, y)). \quad (4.2)$$

In practice,  $x$  and  $y$  are discrete and must be both zero-padded sufficiently, that is, zeros are appended to each vector and make them artificially longer to prevent border effects when computing the linear cross-correlation (Stoica and Moses, 2005).

---

**Algorithm 4.1:** The proposed diffusion profile realignment (DPR) algorithm.

---

**Data:** Metrics extracted from streamlines discretized (with an equal distance  $\delta_{\min}$  and stationary metrics), displacement threshold  $\mathbf{t}$ , percentage of overlap  $\mathbf{p}\%$

**Result:** Realigned metrics

**Step 1 :** *Finding a common template;*

**foreach** *streamline* **do**

    | Compute the displacement  $\mathbf{d}$  with each other streamline using Eq. (4.2);

**end**

Define the template as the subject which realigns the most streamlines below the threshold  $\mathbf{t}$ ;

**foreach** *streamline* **do**

    | **if**  $|\mathbf{d}| \leq \mathbf{t}$  **then**

        | Realign the streamline unto the candidate template by its displacement  $\mathbf{d}$ ;

    | **else**

        | Do not touch the streamline and flag it as an outlier;

    | **end**

**end**

**Step 2 :** *Realigning outliers;*

**foreach** *outlier* **do**

    | Compute the new displacement  $\mathbf{nd}$  between the template, the outlier and each other non outlier;

    | **if**  $\min(|\mathbf{d} + \mathbf{nd}|) < \mathbf{t}$  **then**

        | Realign the streamline unto the template using the new displacement  $\mathbf{d} + \mathbf{nd}$  (see Fig. 4.4);

        | Add the streamline to the pool of non outliers candidates such that it can now be used;

    | **else**

        | Do not touch the streamline and flag it as an outlier;

    | **end**

**end**

**Step 3 :** *Truncating to overlapping coordinates;*

Truncate the realigned metrics to have at least  $\mathbf{p}\%$  of overlapping streamlines;

If outliers are still present from **Step 2**, (optionally) exclude them from further analysis as they can not be realigned inside the chosen displacement threshold  $\mathbf{t}$ ;

---

## 4.8 Supplementary materials

### 4.8.1 Realignment of the HCP datasets

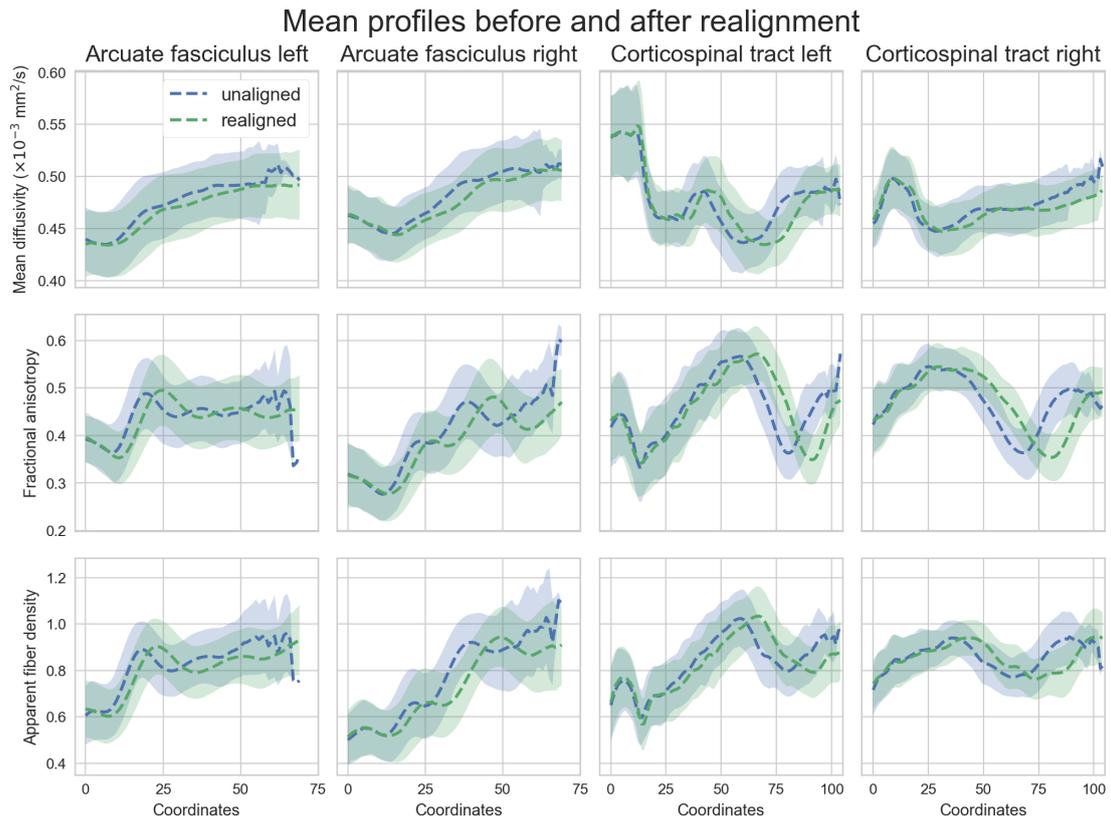
Section 4.8.1 presents counterpart results to Fig. 4.9, comparing along-tract averaged profiles before and after realignment, but instead using a maximally allowed displacement of 5%, 10% or 20%. Coordinates for the AF are from anterior (coordinate 0) to posterior and the CST are drawn from inferior (coordinate 0) to superior. In general, the overall mean profile is similar for every value of the maximally allowed displacement that were tested.

### 4.8.2 Displacement of the HCP datasets

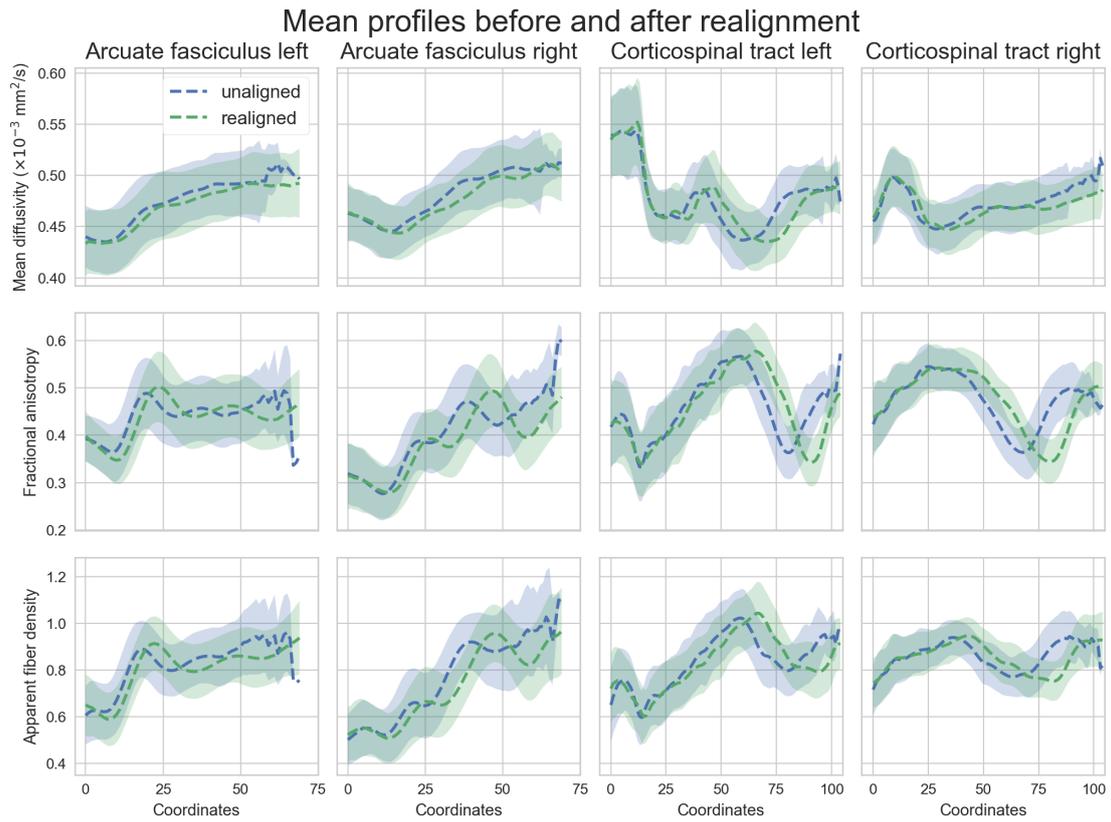
Section 4.8.2 presents counterpart results to Fig. 4.11 using realignment from other metrics, but instead using a maximally allowed displacement of 5%, 10% or 20%. Coordinates for the AF are from anterior (coordinate 0) to posterior and the CST are drawn from inferior (coordinate 0) to superior. Most of the trends observed previously when the displacement was of 15% are still valid.

### 4.8.3 Localized alterations of the HCP datasets

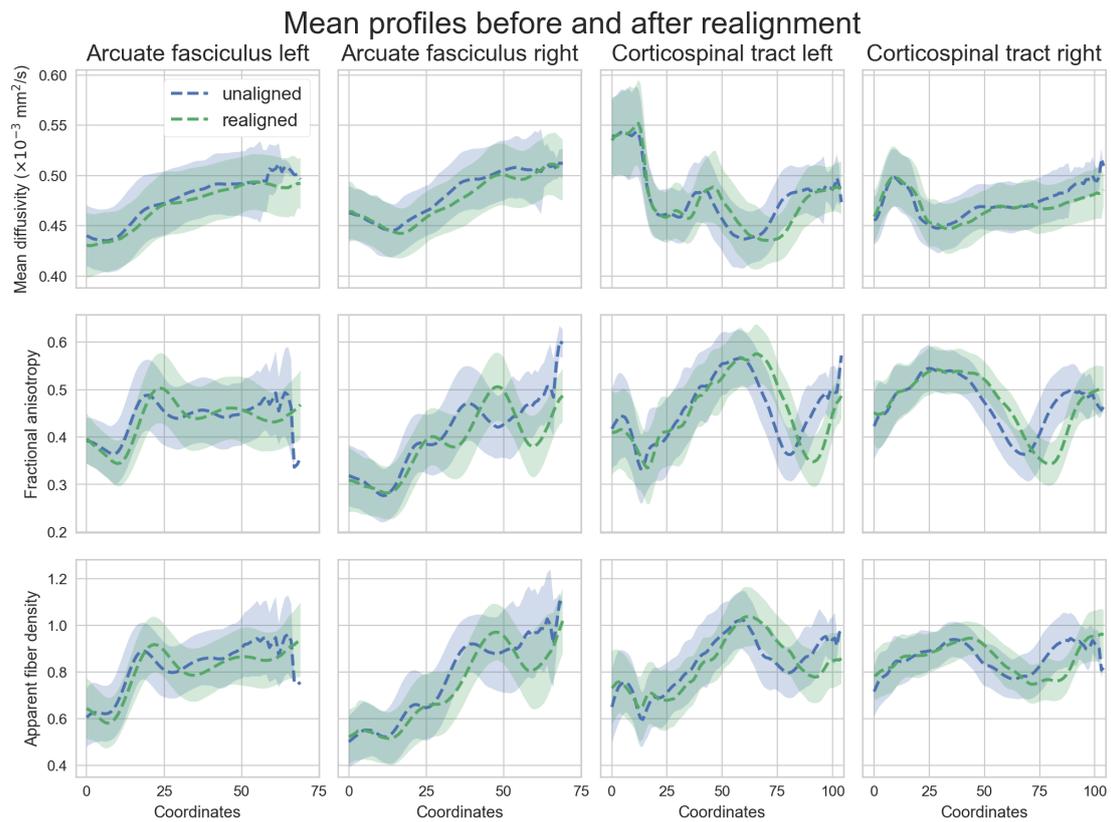
Section 4.8.3 presents counterpart results to Fig. 4.14, but instead using a maximally allowed displacement of 5%, 10% or 100% (no limit). Unpaired t-test (FDR corrected at  $\alpha = 0.05$ ) with focused alterations of the metrics for each bundle of **A**) 25% over 1% of the length, **B**) 50% over 1% of the length, **C**) 25% over 5% of the length and **D**) 50% over 5% of the length are shown. The AF left/right are represented from anterior (coordinate 0) to posterior and the CST left/right from inferior (coordinate 0) to superior. The p-values are on a log scale along the average streamline before realignment (dashed red lines) and after realignment (solid blue lines) with the DPR algorithm. The horizontal dashed black lines indicate p-value = 0.05. In general, alterations occurring over 5% of the length of the bundle can be detected, whereas small local changes over 1% of the length are detected only after realignment with the DPR algorithm.



**Figure 4.15:** Along-tract averaged profiles (and standard deviation as the shaded area) of the unaligned (blue) and realigned (green) HCP subjects truncated to 75% of overlap with a final resampling to the same number of points. These results are obtained by using a maximally allowed displacement of 5%.



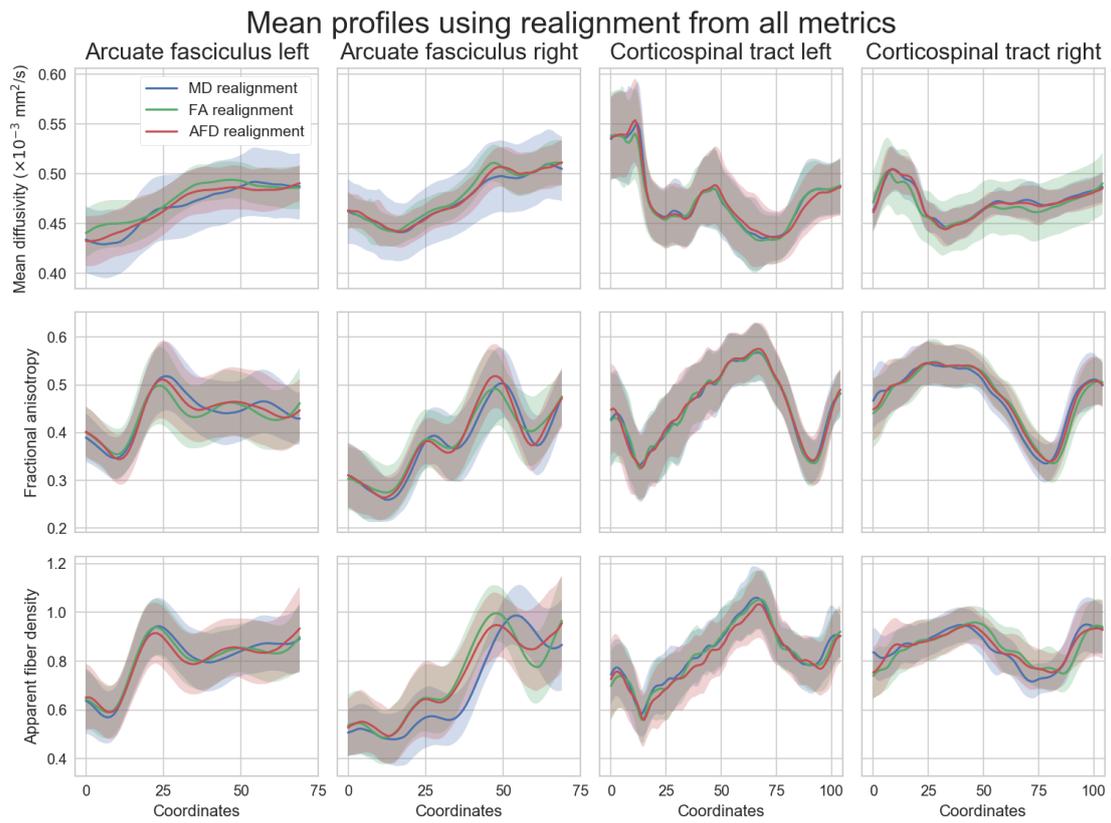
**Figure 4.16:** Along-tract averaged profiles (and standard deviation as the shaded area) of the unaligned (blue) and realigned (green) HCP subjects truncated to 75% of overlap with a final resampling to the same number of points. These results are obtained by using a maximally allowed displacement of 10%.



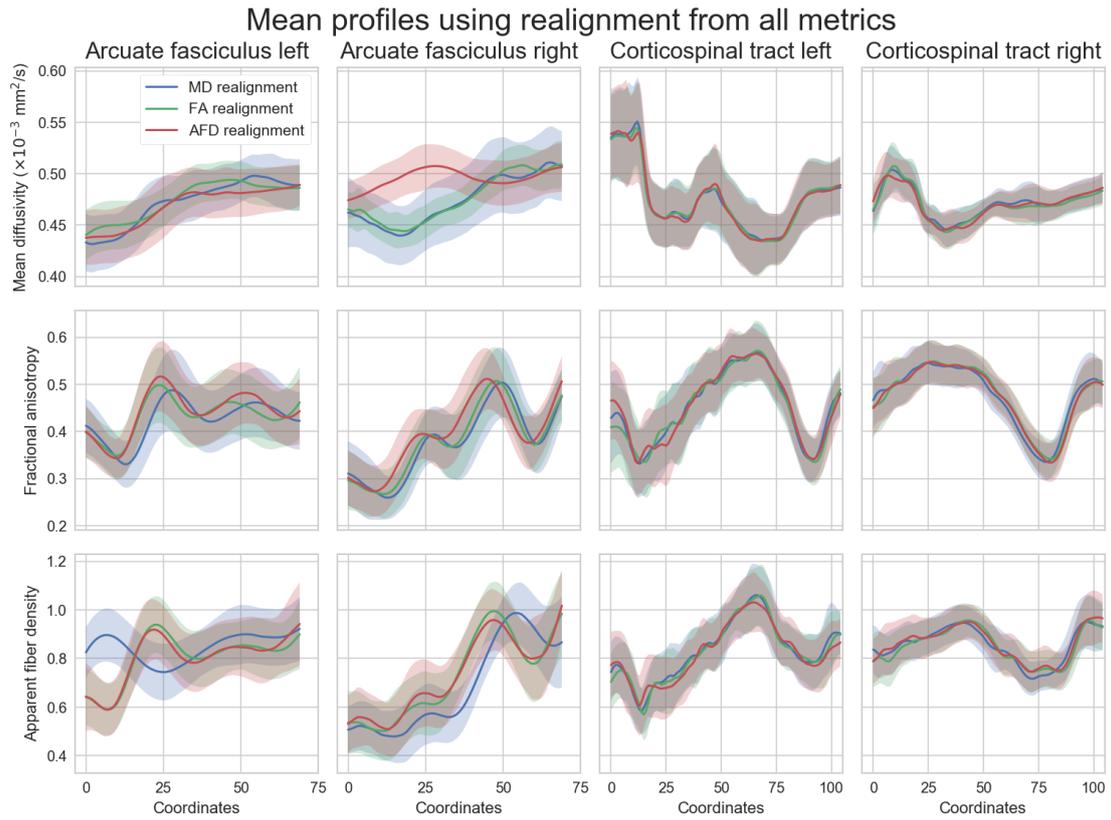
**Figure 4.17:** Along-tract averaged profiles (and standard deviation as the shaded area) of the unaligned (blue) and realigned (green) HCP subjects truncated to 75% of overlap with a final resampling to the same number of points. These results are obtained by using a maximally allowed displacement of 20%.



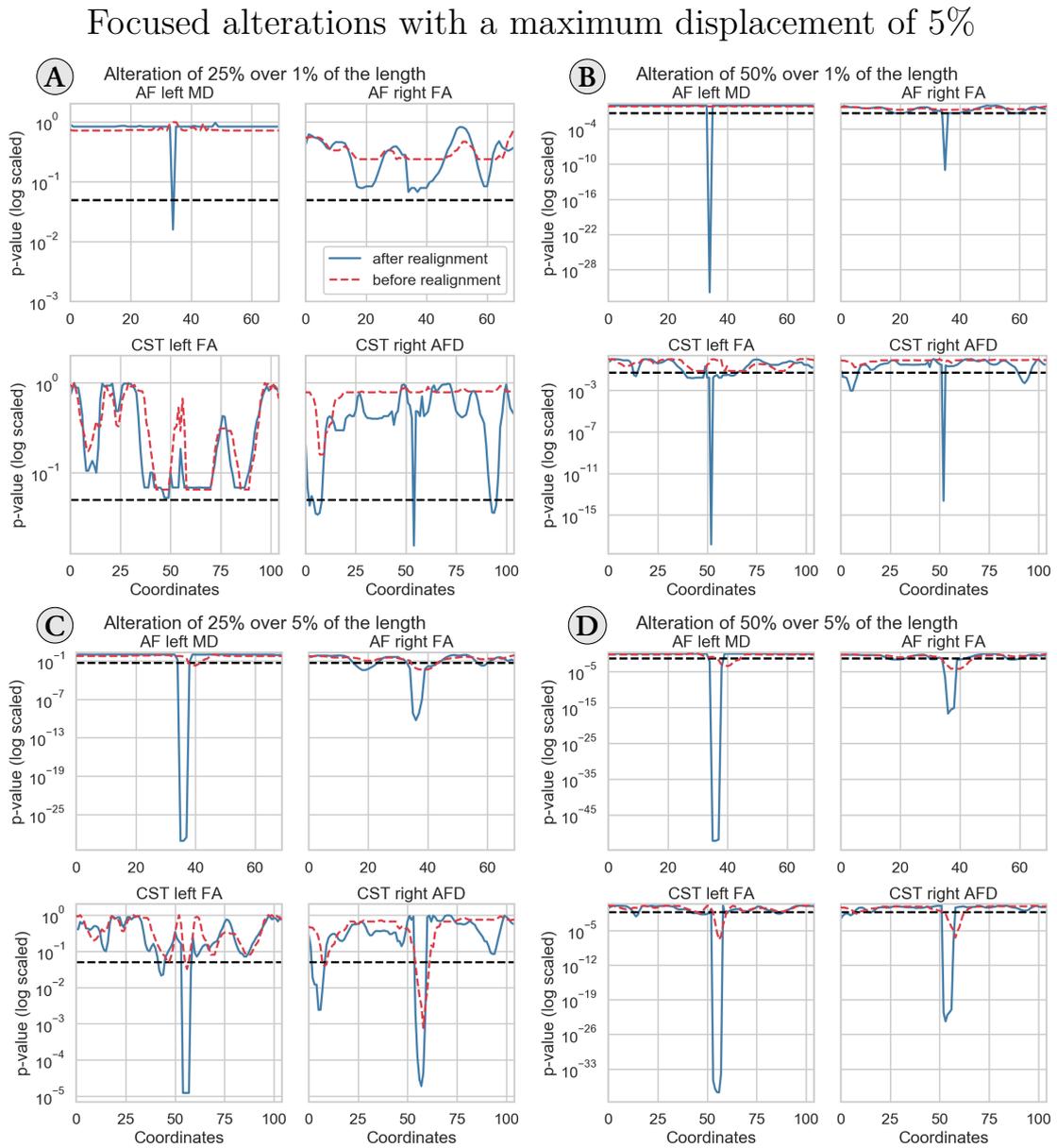
**Figure 4.18:** Along-tract averaged profiles (and standard deviation as the shaded area) of the white matter fiber bundles (columns) from the HCP datasets after realignment for each studied metric (rows). These results are obtained by using a maximally allowed displacement of 5%.



**Figure 4.19:** Along-tract averaged profiles (and standard deviation as the shaded area) of the white matter fiber bundles (columns) from the HCP datasets after realignment for each studied metric (rows). These results are obtained by using a maximally allowed displacement of 10%.

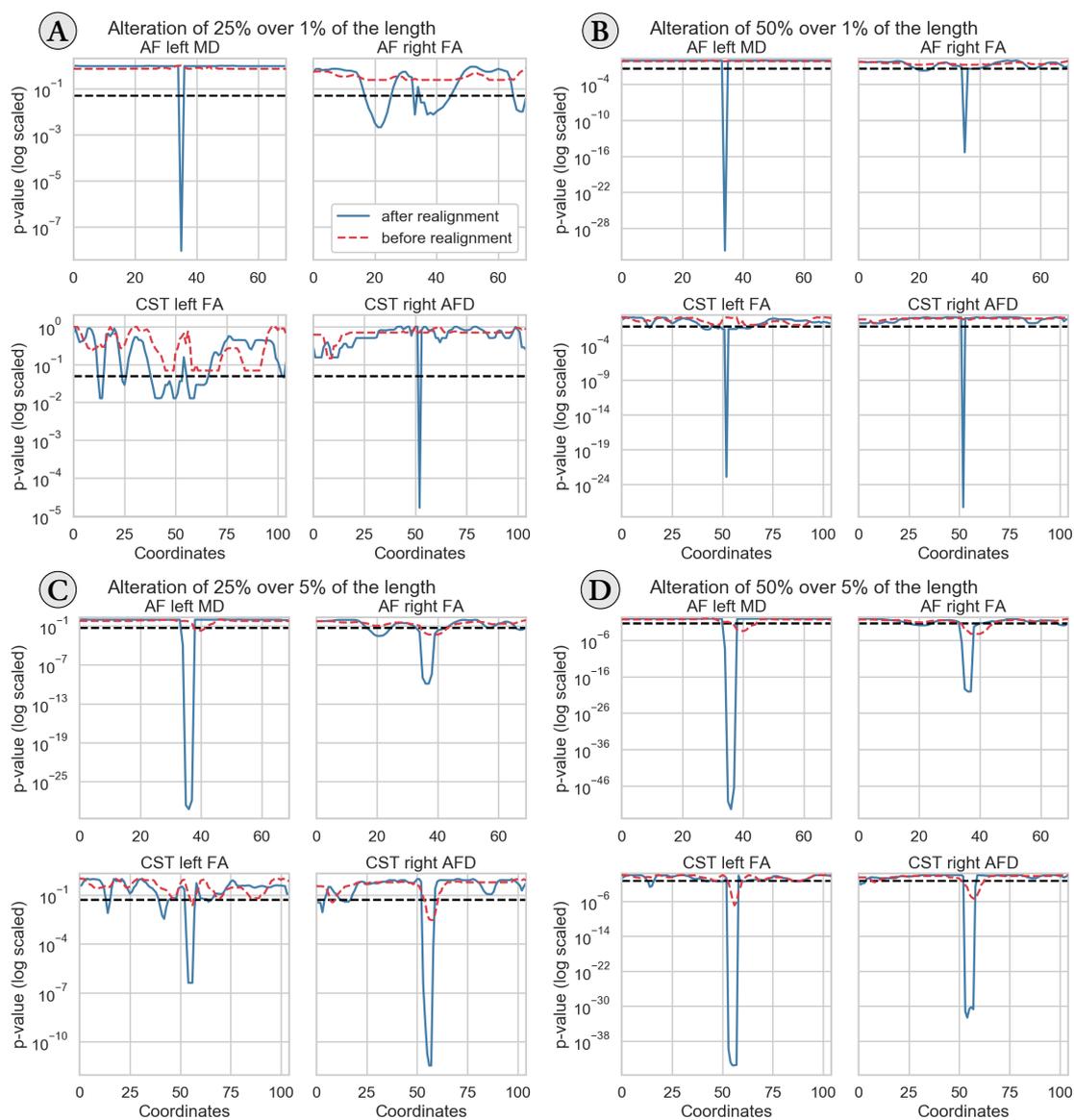


**Figure 4.20:** Along-tract averaged profiles (and standard deviation as the shaded area) of the white matter fiber bundles (columns) from the HCP datasets after realignment for each studied metric (rows). These results are obtained by using a maximally allowed displacement of 20%.



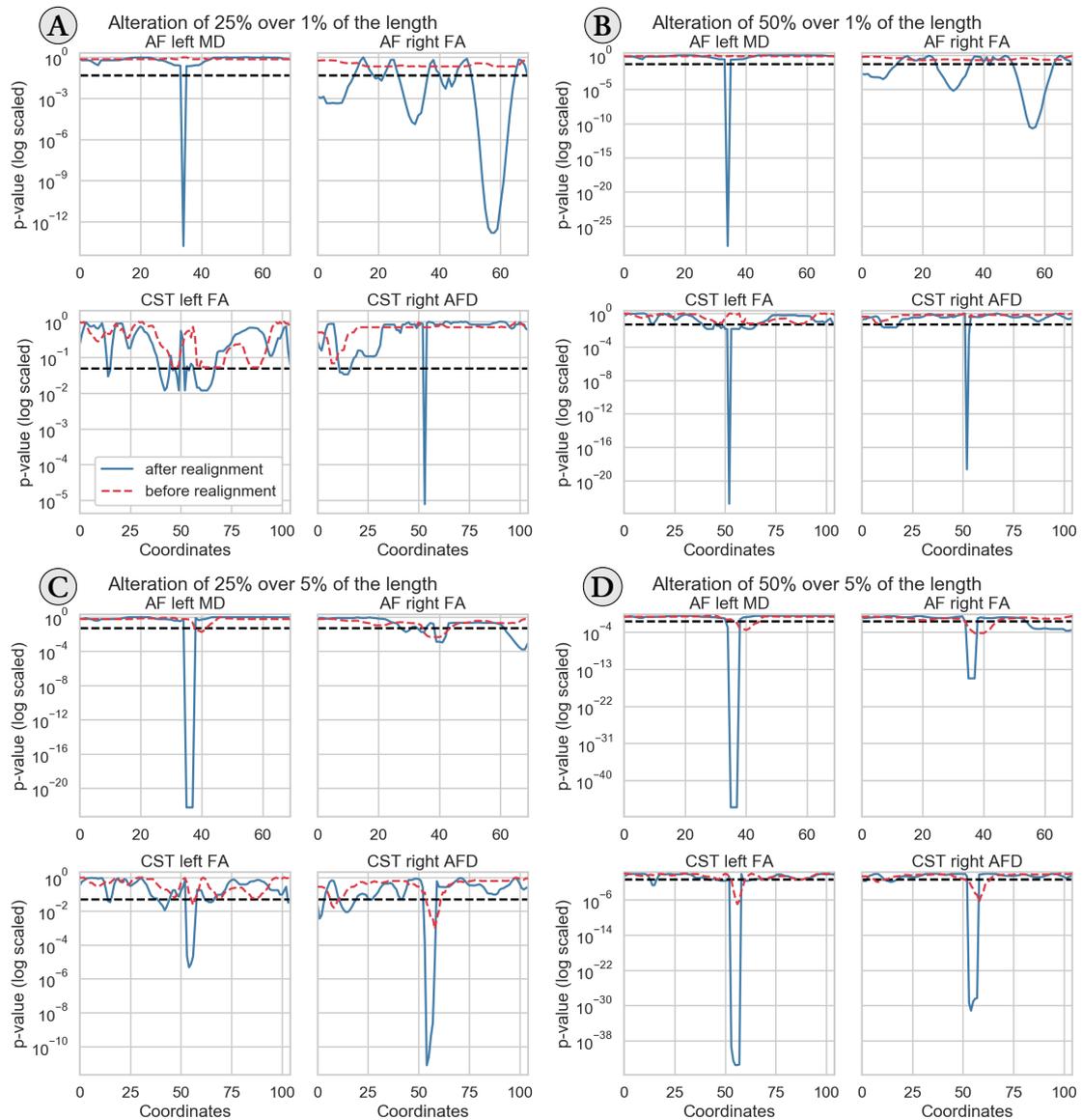
**Figure 4.21:** Unpaired t-test before and after realignment for the four bundles. These results are obtained by using a maximally allowed displacement of 5%.

### Focused alterations with a maximum displacement of 10%



**Figure 4.22:** Unpaired t-test before and after realignment for the four bundles. These results are obtained by using a maximally allowed displacement of 10%.

## Focused alterations without limiting the maximum displacement



**Figure 4.23:** Unpaired t-test before and after realignment for the four bundles. These results are obtained without limiting the allowed maximum displacement. This leads to false effects for the AF right bundles, presumably because structural differences, rather than local alterations, are driving the realignment process.

# Bibliography

- [1] Y. Assaf, T. Blumenfeld-Katzir, Y. Yovel, and P. J. Basser. “AxCaliber: a method for measuring axon diameter distribution from diffusion MRI.” In: *Magnetic resonance in medicine* 59.6 (June 2008), pp. 1347–54 (cit. on p. 93).
- [2] B. B. Avants, N. J. Tustison, G. Song, P. A. Cook, A. Klein, and J. C. Gee. “A reproducible evaluation of ANTs similarity metric performance in brain image registration.” In: *NeuroImage* 54.3 (Feb. 2011), pp. 2033–44 (cit. on p. 90).
- [3] M. Bach et al. “Methodological considerations on tract-based spatial statistics (TBSS)”. In: *NeuroImage* 100 (Oct. 2014), pp. 358–369 (cit. on p. 64).
- [4] P. J. Basser and C. Pierpaoli. “Microstructural and Physiological Features of Tissues Elucidated by Quantitative-Diffusion-Tensor MRI”. In: *Journal of Magnetic Resonance, Series B* 111.3 (June 1996), pp. 209–219 (cit. on p. 72).
- [5] C. Beaulieu. “The basis of anisotropic water diffusion in the nervous system - a technical review”. In: *NMR in Biomedicine* 15.7-8 (Nov. 2002), pp. 435–455 (cit. on p. 64).
- [6] Y. Benjamini and Y. Hochberg. “Controlling the false discovery rate: a practical and powerful approach to multiple testing”. In: *Journal of the Royal Statistical Society* 57.1 (Nov. 1995), pp. 289–300 (cit. on pp. 70, 77).
- [7] C. R. V. Blain et al. “Differential corticospinal tract degeneration in homozygous ‘D90A’ SOD-1 ALS and sporadic ALS.” In: *Journal of neurology, neurosurgery, and psychiatry* 82.8 (Aug. 2011), pp. 843–9 (cit. on p. 65).
- [8] B. Bodini and O. Ciccarelli. “Diffusion MRI in Neurological Disorders”. In: *Diffusion MRI*. Elsevier, 2009, pp. 175–203 (cit. on p. 64).
- [9] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel. *Time Series Analysis*. Hoboken, NJ: John Wiley & Sons, Inc., June 2008 (cit. on p. 67).
- [10] E. Caruyer, A. Daducci, M. Descoteaux, J.-C. Houde, J.-P. Thiran, and R. Verma. “Phantoms: a flexible software library to simulate diffusion MR phantoms”. In: *International Symposium on Magnetic Resonance in Medicine (ISMRM’14)*. Vol. 17. 2014, p. 6407 (cit. on p. 72).
- [11] M. Catani et al. “Symmetries in human brain language pathways correlate with verbal recall”. In: *Proceedings of the National Academy of Sciences* 104.43 (Oct. 2007), pp. 17163–17168 (cit. on p. 94).
- [12] M. Cercignani and C. Gandini Wheeler-Kingshott. “From micro- to macro-structures in multiple sclerosis: what is the added value of diffusion imaging”. In: *NMR in Biomedicine* October 2017 (Jan. 2018), e3888 (cit. on p. 64).
- [13] M. Chamberland, S. St-Jean, C. M. W. Tax, and D. K. Jones. “Obtaining representative core streamlines for white matter tractometry of the human brain”. In: *Computational Diffusion MRI*. Ed. by E. Bonet-Carne, F. Grussu, L. Ning, F. Seppehrband, and C. M. W. Tax. Granada: Springer International Publishing, 2018, pp. 3–19 (cit. on p. 92).

## Bibliography

---

- [14] M. Chamberland, K. Whittingstall, D. Fortin, D. Mathieu, and M. Descoteaux. “Real-time multi-peak tractography for instantaneous connectivity display.” English. In: *Frontiers in neuroinformatics* 8.May (Jan. 2014), p. 59 (cit. on p. 89).
- [15] J. B. Colby, L. Soderberg, C. Lebel, I. D. Dinov, P. M. Thompson, and E. R. Sowell. “Along-tract statistics allow for enhanced tractography analysis”. In: *NeuroImage* 59.4 (Feb. 2012), pp. 3227–3242 (cit. on pp. 64, 65, 69, 70, 75, 92).
- [16] J. W. Cooley and J. W. Tukey. “An algorithm for the machine calculation of complex Fourier series”. In: *Mathematics of Computation* 19.90 (May 1965), pp. 297–297 (cit. on p. 66).
- [17] I. Corouge, P. T. Fletcher, S. Joshi, S. Gouttard, and G. Gerig. “Fiber tract-oriented statistics for quantitative diffusion tensor MRI analysis”. In: *Medical Image Analysis* 10.5 (2006), pp. 786–798 (cit. on pp. 64, 69, 92).
- [18] M. Cousineau et al. “A test-retest study on Parkinson’s PPMI dataset yields statistically significant white matter fascicles.” In: *NeuroImage. Clinical* 16.July (2017), pp. 222–233 (cit. on pp. 64, 69, 73).
- [19] M. Dayan et al. “Profilometry: A new statistical framework for the characterization of white matter pathways, with application to multiple sclerosis”. In: *Human Brain Mapping* 37.3 (Mar. 2016), pp. 989–1004 (cit. on p. 65).
- [20] F. Dell’Acqua, A. Simmons, S. C. Williams, and M. Catani. “Can spherical deconvolution provide more information than fiber orientations? Hindrance modulated orientational anisotropy, a true-tract specific index to characterize white matter diffusion”. In: *Human Brain Mapping* 34.10 (Oct. 2013), pp. 2464–2483 (cit. on p. 72).
- [21] K. L. Denman. “Spectral Analysis: A summary of the theory and techniques”. In: *Fish. Mar. Servo Res. Dev. Tech. Rep.* 539 539 (1975), pp. 1–32 (cit. on p. 67).
- [22] S. C. L. Deoni, B. K. Rutt, T. Arun, C. Pierpaoli, and D. K. Jones. “Gleaning multicomponent T1 and T2 information from steady-state imaging data”. In: *Magnetic Resonance in Medicine* 60.6 (2008), pp. 1372–1387 (cit. on p. 94).
- [23] A. Dutt and V. Rokhlin. “Fast Fourier Transforms for Nonequispaced Data”. In: *SIAM Journal on Scientific Computing* 14.6 (Nov. 1993), pp. 1368–1393 (cit. on p. 93).
- [24] M. Froeling, P. Pullens, and A. Leemans. “DTI Analysis Methods: Region of Interest Analysis”. In: *Diffusion Tensor Imaging*. Ed. by W. Van Hecke, L. Emsell, and S. Sunaert. October 2017. New York, NY: Springer New York, 2016, pp. 175–182 (cit. on p. 64).
- [25] G. Girard, K. Whittingstall, R. Deriche, and M. Descoteaux. “Towards quantitative connectivity analysis: reducing tractography biases.” In: *NeuroImage* 98 (Sept. 2014), pp. 266–78 (cit. on p. 89).
- [26] T. Glozman, L. Bruckert, F. Pestilli, D. W. Yecies, L. J. Guibas, and K. W. Yeom. “Framework for shape analysis of white matter fiber bundles”. In: *NeuroImage* 167.May 2017 (Feb. 2018), pp. 466–477 (cit. on p. 92).
- [27] W. Haakma et al. “MRI shows thickening and altered diffusion in the median and ulnar nerves in multifocal motor neuropathy”. In: *European Radiology* 27.5 (May 2017), pp. 2216–2224 (cit. on p. 64).
- [28] P. Hagmann et al. “Mapping Human Whole-Brain Structural Networks with Diffusion MRI”. In: *PLoS ONE* 2.7 (July 2007). Ed. by O. Sporns, e597 (cit. on p. 64).
- [29] J. H. Jensen and J. A. Helpert. “MRI quantification of non-Gaussian water diffusion by kurtosis analysis.” In: *NMR in biomedicine* 23.7 (Aug. 2010), pp. 698–710 (cit. on p. 93).
- [30] B. Jeurissen, M. Descoteaux, S. Mori, and A. Leemans. “Diffusion MRI fiber tractography of the brain”. In: *NMR in Biomedicine* (Sept. 2017), e3785 (cit. on p. 64).

- [31] Y. Jin et al. “3D tract-specific local and global analysis of white matter integrity in Alzheimer’s disease.” In: *Human brain mapping* 38.3 (Mar. 2017), pp. 1191–1207 (cit. on p. 65).
- [32] D. K. Jones and M. Cercignani. “Twenty-five pitfalls in the analysis of diffusion MRI data”. In: *NMR in Biomedicine* 23.7 (Sept. 2010), pp. 803–820 (cit. on p. 64).
- [33] D. K. Jones, A. R. Travis, G. Eden, C. Pierpaoli, and P. J. Basser. “PASTA: Pointwise assessment of streamline tractography attributes”. In: *Magnetic Resonance in Medicine* 53.6 (June 2005), pp. 1462–1467 (cit. on p. 64).
- [34] S. Klein, M. Staring, K. Murphy, M. Viergever, and J. Pluim. “elastix: A Toolbox for Intensity-Based Medical Image Registration”. In: *IEEE Transactions on Medical Imaging* 29.1 (Jan. 2010), pp. 196–205 (cit. on p. 73).
- [35] C. Lebel, L. Walker, A. Leemans, L. Phillips, and C. Beaulieu. “Microstructural maturation of the human brain from childhood to adulthood”. In: *NeuroImage* 40.3 (Apr. 2008), pp. 1044–1055 (cit. on p. 73).
- [36] A. Leemans, B. Jeurissen, J. Sijbers, and D. Jones. “ExploreDTI: a graphical toolbox for processing, analyzing, and visualizing diffusion MR data”. In: *Proceedings 17th Scientific Meeting, International Society for Magnetic Resonance in Medicine* 17.2 (2009), p. 3537 (cit. on pp. 72, 73, 94).
- [37] A. Leemans, J. Sijbers, S. De Backer, E. Vandervliet, and P. Parizel. “Multiscale white matter fiber tract coregistration: A new feature-based approach to align diffusion tensor data”. In: *Magnetic Resonance in Medicine* 55.6 (June 2006), pp. 1414–1423 (cit. on pp. 92, 93).
- [38] J. P. Lewis. “Fast normalized cross-correlation”. In: *Vision interface* 10.1 (1995), pp. 120–123 (cit. on p. 67).
- [39] K. H. Maier-Hein et al. “The challenge of mapping the human connectome based on diffusion tractography”. In: *Nature Communications* 8.1 (Dec. 2017), p. 1349 (cit. on pp. 64, 77, 89, 93).
- [40] S. Mori and P. C. M. Van Zijl. “Fiber tracking: Principles and strategies - A technical review”. In: *NMR in Biomedicine* 15.7-8 (2002), pp. 468–480 (cit. on p. 64).
- [41] C. Nimsky, M. Bauer, and B. Carl. “Merits and Limits of Tractography Techniques for the Uninitiated”. In: *Advances and technical standards in neurosurgery*. 43. 2016, pp. 37–60 (cit. on p. 64).
- [42] L. J. O’Donnell, A. Daducci, D. Wassermann, and C. Lenglet. “Advances in computational and statistical diffusion MRI”. In: *NMR in Biomedicine* July (Nov. 2017), e3805 (cit. on pp. 65, 92).
- [43] L. J. O’Donnell and O. Pasternak. “Does diffusion MRI tell us anything about the white matter? An overview of methods and pitfalls”. In: *Schizophrenia Research* 161.1 (Jan. 2015), pp. 133–141 (cit. on p. 64).
- [44] L. J. O’Donnell, C.-F. Westin, and A. J. Golby. “Tract-based morphometry for white matter group analysis”. In: *NeuroImage* 45.3 (Apr. 2009), pp. 832–844 (cit. on pp. 65, 69, 89–92, 94).
- [45] E. O’Hanlon et al. “White matter differences among adolescents reporting psychotic experiences: A population-based diffusion magnetic resonance imaging study”. In: *JAMA Psychiatry* 72.7 (2015), pp. 668–677 (cit. on pp. 64, 65).
- [46] G. D. Parker, D. LLOYD, and D. K. Jones. “The best of both worlds: Combining the strengths of TBSS and tract-specific measurements for group-wise comparison of white matter microstructure”. In: *International Symposium on Magnetic Resonance in Medicine (ISMRM’16)*. 2016 (cit. on p. 93).
- [47] T. Platt and K. L. Denman. “Spectral Analysis in Ecology”. In: *Annual Review of Ecology and Systematics* 6.1 (Nov. 1975), pp. 189–210 (cit. on p. 67).

## Bibliography

---

- [48] D. Raffelt et al. “Apparent Fibre Density: A novel measure for the analysis of diffusion-weighted magnetic resonance images”. In: *NeuroImage* 59.4 (Feb. 2012), pp. 3976–3994 (cit. on p. 72).
- [49] F. Rheault, J.-C. Houde, and M. Descoteaux. “Visualization, Interaction and Tractometry: Dealing with Millions of Streamlines from Diffusion MRI Tractography”. In: *Frontiers in Neuroinformatics* 11.June (2017), p. 42 (cit. on p. 93).
- [50] J. D. Scargle. “Studies in astronomical time series analysis. III - Fourier transforms, autocorrelation functions, and cross-correlation functions of unevenly spaced data”. In: *The Astrophysical Journal* 343 (Aug. 1989), p. 874 (cit. on p. 93).
- [51] S. M. Smith et al. “Tract-based spatial statistics: Voxelwise analysis of multi-subject diffusion data”. In: *NeuroImage* 31.4 (July 2006), pp. 1487–1505 (cit. on p. 64).
- [52] O. Sporns, G. Tononi, and R. Kötter. “The human connectome: A structural description of the human brain”. In: *PLoS Computational Biology* 1.4 (2005), pp. 0245–0251 (cit. on p. 64).
- [53] S. St-Jean. *samuelstjean/dpr: [v0.1.1b] - 2019-03-13*. Mar. 2019.  
URL: <https://doi.org/10.5281/zenodo.2592014> (cit. on pp. 72, 94).
- [54] S. St-Jean, M. Chamberland, M. A. Viergever, and A. Leemans. *Datasets for 'Reducing variability in along-tract analysis with diffusion profile realignment'*. 2018.  
URL: <http://dx.doi.org/10.5281/zenodo.2483169> (cit. on pp. 72, 95).
- [55] S. St-Jean, M. A. Viergever, G. J. Biessels, and A. Leemans. “Correcting spatial misalignment between fiber bundles segments for along-tract group analysis”. In: *International Symposium on Magnetic Resonance in Medicine (ISMRM'16)*. 2016 (cit. on p. 65).
- [56] E. St-Onge, A. Daducci, G. Girard, and M. Descoteaux. “Surface-enhanced tractography (SET)”. In: *NeuroImage* 169.December 2017 (Apr. 2018), pp. 524–539 (cit. on p. 89).
- [57] P. Stoica and R. Moses. *Spectral Analysis of Signals*. Prentice Hall, 2005, p. 452 (cit. on pp. 67, 93, 95).
- [58] F. Szczepankiewicz et al. “Variability in diffusion kurtosis imaging: Impact on study design, statistical power and interpretation”. In: *NeuroImage* 76 (Aug. 2013), pp. 145–154 (cit. on p. 65).
- [59] C. M. W. Tax, W. M. Otte, M. A. Viergever, R. M. Dijkhuizen, and A. Leemans. “REKINDLE: Robust Extraction of Kurtosis INDices with Linear Estimation”. In: *Magnetic Resonance in Medicine* 73.2 (2015), pp. 794–808 (cit. on p. 72).
- [60] C. M. Tax, B. Jeurissen, S. B. Vos, M. A. Viergever, and A. Leemans. “Recursive calibration of the fiber response function for spherical deconvolution of diffusion MRI data”. In: *NeuroImage* 86 (2014), pp. 67–80 (cit. on p. 72).
- [61] J.-D. Tournier, F. Calamante, and A. Connelly. “Robust determination of the fibre orientation distribution in diffusion MRI: non-negativity constrained super-resolved spherical deconvolution.” In: *NeuroImage* 35.4 (May 2007), pp. 1459–72 (cit. on p. 72).
- [62] J.-D. Tournier, F. Calamante, and A. Connelly. “Determination of the appropriate b value and number of gradient directions for high-angular-resolution diffusion-weighted imaging.” In: *NMR in biomedicine* 26.12 (Dec. 2013), pp. 1775–86 (cit. on p. 72).
- [63] D. C. Van Essen et al. “The Human Connectome Project: A data acquisition perspective”. In: *NeuroImage* 62.4 (2012), pp. 2222–2231 (cit. on p. 72).
- [64] S. Wakana et al. “Reproducibility of quantitative tractography methods applied to cerebral white matter”. In: *NeuroImage* 36.3 (2007), pp. 630–644 (cit. on pp. 64, 73).

- [65] D. Wassermann et al. “The white matter query language: a novel approach for describing human white matter anatomy”. In: *Brain Structure and Function* 221.9 (Dec. 2016), pp. 4705–4721 (cit. on p. 73).
- [66] J. D. Yeatman, R. F. Dougherty, N. J. Myall, B. A. Wandell, and H. M. Feldman. “Tract Profiles of White Matter Properties: Automating Fiber-Tract Quantification”. In: *PLoS ONE* 7.11 (Nov. 2012). Ed. by C. Beaulieu, e49790 (cit. on pp. 64, 65, 69).
- [67] H. Zhang, T. Schneider, C. A. Wheeler-Kingshott, and D. C. Alexander. “NODDI: Practical in vivo neurite orientation dispersion and density imaging of the human brain”. In: *NeuroImage* 61.4 (July 2012), pp. 1000–1016 (cit. on p. 93).

*The good thing about science is that it's true whether or not you believe in it.*

Neil deGrasse Tyson

# 5

## Automated characterization of noise distributions in diffusion MRI data

Based on

---

Samuel St-Jean, Alberto De Luca, Chantal M. W. Tax, Max A. Viergever and Alexander Leemans

Automated characterization of noise distributions in diffusion MRI data

Submitted to Medical image analysis and available at <https://arxiv.org/abs/1906.12121>

## Abstract

Knowledge of the noise distribution in magnitude diffusion MRI images is the centerpiece to quantify uncertainties arising from the acquisition process. The use of parallel imaging methods, the number of receiver coils and imaging filters applied by the scanner, amongst other factors, dictate the resulting signal distribution. Accurate estimation beyond textbook Rician or noncentral chi distributions often requires information about the acquisition process (e.g. coils sensitivity maps or reconstruction coefficients), which is usually not available. We introduce two new automated methods using moments and maximum likelihood equations of the Gamma distribution to estimate noise distributions as they explicitly depend on the number of coils, making it possible to estimate all unknown parameters using only the magnitude data. A rejection step is used to make the framework automatic and robust to artifacts. Simulations using stationary and spatially varying noncentral chi noise distributions were created for two diffusion weightings with SENSE or GRAPPA reconstruction and 8, 12 or 32 receiver coils. Furthermore, MRI data of a water phantom with different combinations of parallel imaging were acquired on a 3T Philips scanner along with noise-only measurements. Finally, experiments on freely available datasets from a single subject acquired on a 3T GE scanner are used to assess reproducibility when limited information about the acquisition protocol is available. Additionally, we demonstrated the applicability of the proposed methods for a bias correction and denoising task on an in vivo dataset acquired on a 3T Siemens scanner. A generalized version of the bias correction framework for non integer values of  $N$  is also introduced. The proposed framework is compared with three other algorithms with datasets from three vendors, employing different reconstruction methods. Simulations showed that assuming a Rician distribution can lead to misestimation of the noise distribution in parallel imaging. Results on the acquired datasets showed that signal leakage in multiband can also lead to a misestimation of the noise distribution. Repeated acquisitions of in vivo datasets show that the estimated parameters are stable and have lower variability than compared methods. Results for the bias correction and denoising task show that the proposed methods reduce the appearance of noise at high b-value. The proposed algorithms herein can estimate both parameters of the noise distribution automatically, are robust to signal leakage artifacts and perform best when used on acquired noise maps.

**Keywords:** Diffusion MRI, Noise estimation, Parallel acceleration, Gamma distribution, GRAPPA, SENSE

## 5.1 Introduction

Diffusion magnetic resonance imaging (dMRI) is a noninvasive imaging technique that allows probing microstructural properties of living tissues. Advances in parallel imaging techniques (Griswold et al., 2002; Pruessmann et al., 1999), such as accelerated acquisitions (e.g. partial k-space (Storey et al., 2007), multiband imaging (Moeller et al., 2010; Nunes et al., 2006) and compressed sensing (Lustig et al., 2007; Paquette et al., 2015)), have greatly reduced the inherently long scan time in dMRI. New acquisition methods and pulse sequences in dMRI are also pushing the limits of spatial resolution while reducing scan time (Holdsworth et al., 2019), which also affects the signal distribution in ways that are challenging to model. Estimation of signal distributions deviating from theoretical cases is challenging and oftentimes requires information such as coil sensitivities or reconstruction matrices. This information may not be recorded at acquisition time or is even not available from the scanner, making techniques relying on these parameters difficult to apply in practice. Even though the magnitude signal model is still valid nowadays, the use of image filters (Dietrich et al., 2008), acceleration methods subsampling k-space (e.g. the SENSE (SENSitivity ENcoding) (Pruessmann et al., 1999), GRAPPA (GeneRALized Autocalibrating Partial Parallel Acquisition) (Griswold et al., 2002; Heidemann et al., 2012) or the homodyne detection methods (Noll et al., 1991)) and spatial correlation between coil elements (Aja-Fernández, Vegas-Sánchez-Ferrero, et al., 2014; Dietrich et al., 2008) influence, amongst other factors, the parameters of the resulting signal distribution.

With the recent trend towards open data sharing and large multicenter studies using standardized protocols (Duchesne et al., 2019; Emaus et al., 2015), differences in hardware, acquisition or reconstruction algorithms may inevitably lead to different signal distributions. This may affect large-scale longitudinal studies investigating neurological changes due to these “scanner effects” (Sakaie, Zhou, et al., 2018) as the acquired data may be fundamentally different across sites in terms of statistical properties of the signal. Algorithms have been developed to mitigate these potential differences (Mirzaalian et al., 2018; Tax et al., 2019), but characterization of the signal distribution from various scanners is challenging due to the black box nature of the acquisition process, especially in routine clinical settings. While some recent algorithms for dMRI are developed to include information about the noise distribution (Collier et al., 2018; Sakaie and Lowe, 2017), there is no method, to the best of our knowledge, providing a fully automated way to characterize the noise distribution using information from the magnitude data itself only. Due to this gap between the physical acquisition process and noise estimation theory, noise distributions are either assumed as Rician (with parameter  $\sigma_g$  related to the standard deviation) or noncentral chi (with fixed degrees of freedom  $N$ ) and concentrate in estimating the noise standard deviation  $\sigma_g$  (Koay, Özarlan, and Pierpaoli, 2009; Tabelow et al., 2015; Veraart et al., 2016). This assumption inevitably leads to misestimation of the true signal distribution as  $N$  and  $\sigma_g$  are interdependent for some reconstruction algorithms (Aja-Fernández, Brion, et al., 2013). Reconstruction filters preserving only the real part of the signal also cause  $N$  to

deviate from the Rician noise distribution, producing instead a half-Gaussian signal distribution (Dietrich et al., 2008). Misestimation of the appropriate signal distribution could impact subsequent processing steps such as bias correction (Koay, Özarslan, and Basser, 2009), denoising (St-Jean, Coupé, et al., 2016) or diffusion model estimation (Landman et al., 2007; Sakaie and Lowe, 2017; Zhang et al., 2012), therefore negating potential gains in statistical power from analyzing datasets acquired in different centers or from different vendors.

In this work, we propose to estimate the parameters  $\sigma_g$  and  $N$  from either the magnitude data or the acquired noise maps by using a change of variable to a Gamma distribution  $\text{Gamma}(N, 1)$  (Koay, Özarslan, and Pierpaoli, 2009), whose first moments and maximum likelihood equations directly depend on  $N$ . This makes the proposed method fast and easy to apply to existing data without additional information, while being robust to artifacts by rejecting outliers of the distribution. Preliminary results of this work have been presented at the annual meeting of the MICCAI (St-Jean, De Luca, Viergever, et al., 2018). This manuscript now contains additional theory, simulations including signal correlations and parallel acceleration, and experiments on phantoms and in vivo datasets acquired with parallel and multiband acceleration. As example applications, we perform bias correction and denoising on an in vivo dataset using the estimated distribution derived with each algorithm.

## 5.2 Theory

In this section, we introduce the necessary background on the Gamma distribution, its moments and maximum likelihood equations. Expressing the signal with a Gamma distribution highlights equations which can be solved to estimate parameters  $\sigma_g$  and  $N$ .

### 5.2.1 Probability distribution functions of MRI data

To account for uncertainty in the acquisition process, the complex signal measured in k-space by the receiver coil array can be modeled with a separate additive zero mean Gaussian noise for each channel with identical variance  $\sigma_g^2$  (Gudbjartsson and Patz, 1995). The signal acquired from the real and imaginary part of each coil in a reconstructed magnitude image can be expressed as (Constantinides et al., 1997)

$$m_N = \sqrt{\sum_{n=1}^N m_{Rn}^2 + m_{In}^2}, \quad (5.1)$$

where  $m_{Rn}$  and  $m_{In}$  are the real and imaginary parts of the signal, respectively, as measured by coil number  $n$ ,  $N$  is the number of degrees of freedom (which can be up to the number of coils in the absence of accelerated parallel imaging) and  $m_N$  is the resulting reconstructed signal value for a given voxel. The magnitude signal can therefore be approximated by a

noncentral chi distribution and has a probability density function (pdf) given by (Dietrich et al., 2008; Koay, Özarlan, and Basser, 2009)

$$pdf(m|\eta, \sigma_g, N) = \frac{m^N}{\sigma_g^2 \eta^{N-1}} \exp\left(\frac{-(m^2 + \eta^2)}{2\sigma_g^2}\right) I_{N-1}\left(\frac{m\eta}{\sigma_g^2}\right) dm, \quad (5.2)$$

where  $m$  is the noisy signal value for a given voxel,  $\eta$  is the (unknown) noiseless signal value,  $\sigma_g$  is the Gaussian noise standard deviation,  $N$  is the number of degrees of freedom and  $I_\nu(z)$  is the modified Bessel function of the first kind.

With the introduction of multiband imaging and other modern acquisition methods, parameters estimation of the magnitude data is not straightforward anymore. The number of degrees of freedom  $N$ , which is related to the number of receiver coils, likely deviates from heuristic estimation based on the actual number of coils as  $N$  also depends on the reconstruction technique employed (Sotiropoulos et al., 2013). The pdf of the magnitude data can be modeled by considering spatially varying degrees of freedom  $N_{eff}$  and standard deviation  $\sigma_{eff}$  (also called the *effective* values) and we generally have  $N_{eff} \leq N$ , (Aja-Fernández, Vegas-Sánchez-Ferrero, et al., 2014; Dietrich et al., 2008).

The noncentral chi distribution includes the Rician ( $N = 1$ ), the Rayleigh ( $N = 1, \eta = 0$ ) and the central chi distribution ( $\eta = 0$ ) as special cases (Dietrich et al., 2008). The probability density function (pdf) of the central chi distribution is given by

$$pdf(m|\eta = 0, \sigma_g, N) = \frac{m^{2N-1}}{2^{N-1} \sigma_g^{2N} \Gamma(N)} \exp\left(\frac{-m^2}{2\sigma_g^2}\right) dm, \quad (5.3)$$

where  $\Gamma(x)$  is the Gamma function. With a change of variable introduced by (Koay, Özarlan, and Pierpaoli, 2009), Eq. (5.3) can be rewritten as a Gamma distribution  $Gamma(N, 1)$  with  $t = m^2/2\sigma_g^2$ ,  $dt = m/\sigma_g^2 dm$  which has a pdf given by

$$pdf(t|N) = \frac{1}{\Gamma(N)} t^{N-1} \exp(-t) dt. \quad (5.4)$$

Eq. (5.4) only depends on  $N$ , which can be estimated from the sample values.

### 5.2.2 Parameter estimation using the method of moments and maximum likelihood

**The method of moments** The pdf of  $Gamma(\alpha, \beta)$  is defined as

$$pdf(x|\alpha, \beta) = \frac{x^{\alpha-1}}{\Gamma(\alpha)\beta^\alpha} \exp(-x/\beta) dx \quad (5.5)$$

and has mean  $\mu_{\text{gamma}}$  and variance  $\sigma_{\text{gamma}}^2$  given by

$$\mu_{\text{gamma}} = \alpha\beta, \quad \sigma_{\text{gamma}}^2 = \alpha\beta^2. \quad (5.6)$$

Another useful identity comes from the sum of Gamma distributions, which is also a Gamma distribution (Weisstein, 2017) such that if  $t_i \sim \text{Gamma}(\alpha_i, \beta)$ , then

$$\sum_{i=1}^K t_i \sim \text{Gamma}\left(\sum_{i=1}^K \alpha_i, \beta\right). \quad (5.7)$$

From Eq. (5.6), we obtain that the mean and the variance of the distribution  $\text{Gamma}(N, 1)$  are in fact equal and of value  $N$ . That is, we can estimate the Gaussian noise standard deviation  $\sigma_g$  and the number of coils  $N$  from the sample moments of the magnitude images themselves, provided we can select voxels without any signal contribution where  $\eta = 0$ . Firstly,  $\sigma_g$  can be estimated from Eq. (5.6) as

$$\sigma_g = \frac{1}{\sqrt{2}} \sqrt{\frac{\sum_{v=1}^V m_v^4}{\sum_{v=1}^V m_v^2} - \frac{1}{V} \sum_{v=1}^V m_v^2}, \quad (5.8)$$

where  $V$  is the number of identified noise only voxels and  $m_v$  the value of such a voxel, see Section 5.7.1 for the derivations. Once  $\sigma_g$  is known,  $N$  can be estimated from the sample mean of those previously identified voxels as

$$N = \frac{1}{V} \sum_{v=1}^V t_v = \frac{1}{2V\sigma_g^2} \sum_{v=1}^V m_v^2. \quad (5.9)$$

Derivations of Eqs. (5.8) and (5.9) are detailed in Section 5.7.1.

**Maximum likelihood equations for the Gamma distribution** Estimation based on the method of maximum likelihood yields two equations for estimating  $\alpha$  and  $\beta$ . Rearranging the equations for a Gamma distribution will give Eq. (5.9) and a second implicit equation for  $N$  that is given by (Thom, 1958)

$$\log(\beta) + \psi(\alpha) = \frac{1}{V} \sum_{v=1}^V \log t_v, \quad (5.10)$$

where  $\psi(x) = \frac{d}{dx} \log(\Gamma(x))$  is the digamma function. For the special case  $\text{Gamma}(N, 1)$ , we can rewrite Eq. (5.10) as

$$\psi(N) = \frac{1}{V} \sum_{v=1}^V \log(m_v^2/2\sigma_g^2). \quad (5.11)$$

Combining Eq. (5.9) and Eq. (5.11), we also have an implicit equation to find  $\sigma_g$

$$\psi\left(\frac{1}{2V\sigma_g^2}\sum_{v=1}^V m_v^2\right) = \frac{1}{V}\sum_{v=1}^V \log(m_v^2/2\sigma_g^2) = \frac{1}{V}\sum_{v=1}^V \log(m_v^2) - \log(2\sigma_g^2). \quad (5.12)$$

As Eqs. (5.11) and (5.12) have no closed form solution, they can be solved numerically e.g. using Newton's method. See Section 5.7.1 for practical implementation details.

## 5.3 Material and Methods

### 5.3.1 Automated and robust background separation

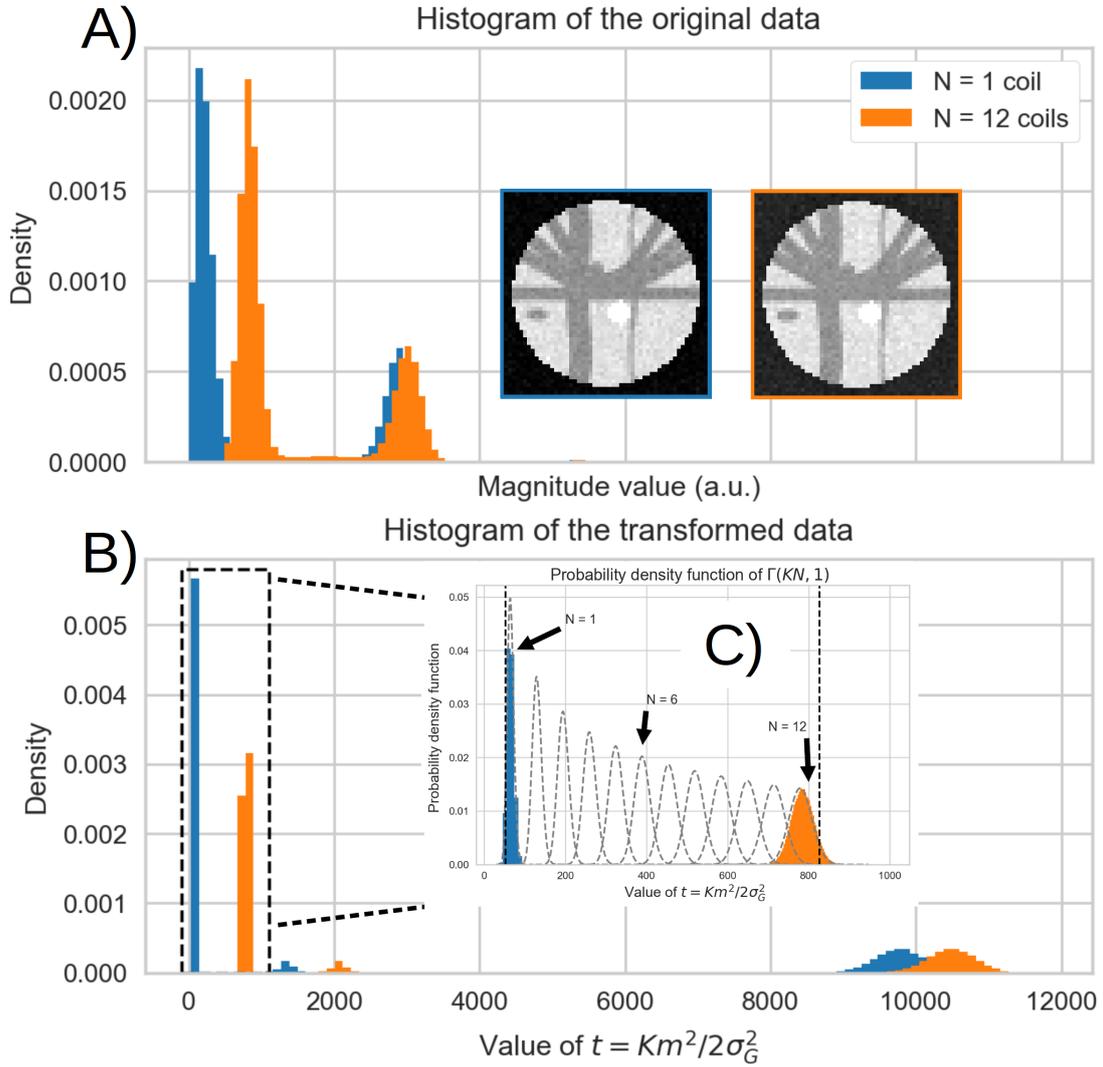
The equations we presented in Section 5.2.1 are only valid when  $\eta = 0$  by construction and assume that each selected voxel  $m_v$  belongs to the same Gamma distribution. Following a methodology similar to (Koay, Özarlan, and Pierpaoli, 2009), we assume that each 2D slice with the same spatial location belongs to the same statistical distribution throughout each 3D volume. This practical assumption allows selecting a large number of noise only voxels for computing statistics as well as identifying (and subsequently discarding) potential slice acquisition artifacts that may affect one volume, but not the rest of the acquisition. Using Eq. (5.7), the sum of all DWIs can be used to separate the voxels belonging to the Gamma distribution  $\text{Gamma}(KN, 1)$ , where  $K$  is the number of acquired DWIs, from the voxels not in that specific distribution with a rejection step using the inverse cumulative distribution function<sup>1</sup>(cdf). In the particular case  $\text{Gamma}(KN, 1)$  at a probability level  $p$ , the inverse cdf is  $\text{icdf}(\alpha, p) = P^{-1}(\alpha, p)$ , where  $P^{-1}$  is the inverse lower incomplete regularized gamma function<sup>2</sup>. This relationship can be used to identify potential outliers, such as voxels which contain non background signal, by excluding any voxel  $m_v$  whose value does not fall between  $\lambda_- = \text{icdf}(\alpha, p/2)$  and  $\lambda_+ = \text{icdf}(\alpha, 1 - p/2)$ , i.e.  $m_v$  is an outlier if  $m_v < \lambda_-$  or  $m_v > \lambda_+$ .

To provide a better understanding of the change of variable  $t = m^2/2\sigma_g^2$ , Fig. 5.1 shows the histogram for a synthetic dataset at  $b = 3000$  s/mm<sup>2</sup>, which will be detailed later in Section 5.3.2. Voxels belonging to the background are easily separated in terms of the Gamma distribution after transformation, thus allowing estimation of parameters from voxels truly belonging to the noise distribution, see Section 5.7.3 and (St-Jean, De Luca, Viergever, et al., 2018) for technical details. Our implementation of the proposed algorithm is freely available<sup>3</sup> (St-Jean, De Luca, Tax, et al., 2019).

<sup>1</sup>The inverse cdf is also known as the quantile function.

<sup>2</sup>As there is no analytical solution to the inverse cdf of a Gamma distribution, one can use the function `gaminv(p, alpha, beta = 1)` in Matlab or `InverseGammaRegularized(alpha, 1 - p)` in Mathematica to numerically estimate it.

<sup>3</sup><https://github.com/samuelstjean/autodmri>



**Figure 5.1:** Histogram of the synthetic data at  $b = 3000 \text{ s/mm}^2$  **A)** before the change of variable to a Gamma distribution and **B)** after the change of variable to a Gamma distribution for  $N = 1$  and  $N = 12$ . Summing all  $K$  DWIs together separates the background voxels from the rest of the data, which follows a Gamma distribution  $\text{Gamma}(KN, 1)$  by construction. In **C)**, a view of the left part from **B)** with the theoretical histograms of Gamma distributions from  $N = 1$  up to  $N = 12$ . The black dotted lines represent the lower bound  $\lambda_-$  to the upper bound  $\lambda_+$ , with  $p = 0.05$ ,  $N_{min} = 1$  and  $N_{max} = 12$ . This broad search covers the background voxels in both cases while excluding remaining voxels which do not belong to the distribution  $\text{Gamma}(KN, 1)$ .

### 5.3.2 Datasets and experiments

**Synthetic phantom datasets** Two synthetic phantom configurations from previous dMRI challenges were used. The first simulations were based on the ISBI 2013 HARDI chal-

lence<sup>4</sup> using phantoms (Caruyer et al., 2014). We used the given 64 gradient directions to generate two separate noiseless single-shell phantoms with either  $b = 1000$  s/mm<sup>2</sup> or  $b = 3000$  s/mm<sup>2</sup> and an additional  $b = 0$  s/mm<sup>2</sup> volume. The datasets were then corrupted with Rician ( $N = 1$ ) and noncentral chi noise profiles ( $N = 4, 8$  and  $12$ ), both stationary and spatially varying, at a signal-to-noise ratio (SNR) of 30 according to

$$\hat{I} = \sqrt{\sum_{i=0, j=0}^N \left( \frac{I}{\sqrt{N}} + \tau \epsilon_i \right)^2 + (\tau \epsilon_j)^2}, \quad \text{where } \epsilon_i, \epsilon_j \sim \mathbb{N}(0, \sigma_g^2), \quad (5.13)$$

where  $I$  is the noiseless volume,  $\hat{I}$  is the resulting noisy volume,  $\tau$  is a mask for the spatial noise pattern,  $\mathbb{N}(0, \sigma_g^2)$  is a Gaussian distribution of mean 0 and variance  $\sigma_g^2 = (\bar{m}/\text{SNR})^2$  and  $\bar{m} = (1/V \sum_{v=1}^V m_v)$  with  $m_v$  each voxel of the  $b = 0$  s/mm<sup>2</sup> image inside the white matter. In the stationary noise case,  $\tau$  is set to 1 so that the noise is uniform. For the spatially varying noise case,  $\tau$  is a sphere with a value of 1 in the center up to a value of 1.75 at the edges of the phantom, thus generating a stronger noise profile outside the phantom than for the stationary noise case. Since all datasets are generated at SNR 30, the noise standard deviation  $\sigma_g$  is the same even though the b-value or number of coils  $N$  is different, but the magnitude standard deviation  $\sigma_{m_N}$  is lower than  $\sigma_g$ .

The second set of synthetic experiments is based on the ISMRM 2015 tractography challenge (Maier-Hein et al., 2017), which consists of 25 manually delineated white matter bundles. Ground truth data consisting of 30 gradient directions at either  $b = 1000$  s/mm<sup>2</sup> or  $b = 3000$  s/mm<sup>2</sup> and 3  $b = 0$  s/mm<sup>2</sup> images at a resolution of 2 mm isotropic was generated using fiberfox (Neher et al., 2014) without artifacts or subject motion. Subsequent noisy datasets were created at SNR 20 by simulating an acquisition with 8, 12 and 32 coils using the parallel MRI simulation toolbox<sup>5</sup> with SENSE (Pruessmann et al., 1999) or GRAPPA (Griswold et al., 2002) reconstructions with an acceleration factor of  $R = 2$ . The SENSE simulated datasets also included spatial correlations between coils of  $\rho = 0.1$ , increasing the spatially varying effective noise standard deviation  $\sigma_g$  and keeping the signal Rician distributed ( $N = 1$ ). For the GRAPPA reconstructed datasets, 32 calibrating lines were sampled in the k-space center, neglecting spatial correlations ( $\rho = 0$ ) as it is a k-space method (Aja-Fernández and Tristán-Vega, 2015). The resulting effective values of  $N$  and  $\sigma_g$  will be both spatially varying. We additionally generated 33 synthetic noise maps per dataset by setting the underlying signal value to  $\eta = 0$  and performing the reconstruction using the same parameters as the DWIs. All generated datasets are available online (St-Jean, De Luca, Tax, et al., 2018).

**Acquired phantom datasets** We acquired phantom images of a bottle of liquid on a 3T Philips Ingenia scanner using a 32 channels head coil with a gradient strength of 45 mT/m.

<sup>4</sup>[http://hardi.epfl.ch/static/events/2013\\_ISBI/](http://hardi.epfl.ch/static/events/2013_ISBI/)

<sup>5</sup><https://mathworks.com/matlabcentral/fileexchange/36893-parallel-mri-noisy-phantom-simulator>

We varied the SENSE factor from  $R = 1, 2$  or  $3$  and multiband acceleration factors from no multiband ( $MB$ ),  $MB = 2$  or  $MB = 3$  while fixing remaining acquisition parameters to investigate their influence on the resulting signal distributions, resulting in 9 different acquisitions. The datasets consist of 5  $b = 0$  s/mm<sup>2</sup> volumes and 4 shells with 10 DWIs each at  $b = 500$  s/mm<sup>2</sup>,  $b = 1000$  s/mm<sup>2</sup>,  $b = 2000$  s/mm<sup>2</sup> and  $b = 3000$  s/mm<sup>2</sup> with a voxel size of 2 mm isotropic and  $TE / TR = 135$  ms / 5000 ms,  $\Delta/\delta = 66.5$  ms / 28.9 ms. Six noise maps were also acquired during each of the experiments by disabling the RF pulse and gradients of the sequence. The acquired phantom datasets are also available (St-Jean, De Luca, Tax, et al., 2018).

**In vivo datasets** A dataset consisting of four repetitions of a single subject<sup>6</sup> was also used to assess the reproducibility of noise estimation without *a priori* knowledge (Poldrack et al., 2015). This is the dataset we previously used in our MICCAI manuscript (St-Jean, De Luca, Viergever, et al., 2018). The acquisition was performed on a GE MR750 3T scanner at Stanford university, where a 3x slice acceleration with blipped-CAIPI shift of FOV/3 was used, partial Fourier 5/8 with a homodyne reconstruction and a minimum TE of 81 ms. Two acquisitions were made in the anterior-posterior phase encode direction and the two others in the posterior-anterior direction. The voxelsize was 1.7 mm isotropic with 7  $b = 0$  s/mm<sup>2</sup> images, 38 volumes at  $b = 1500$  s/mm<sup>2</sup> and 38 volumes at  $b = 3000$  s/mm<sup>2</sup>. As the acquisition used a homodyne filter to fill the missing k-space, this should lead in practice to a half Gaussian noise profile, a special case of the noncentral chi distribution with  $N = 0.5$ , due to using only the real part of the signal for the final reconstruction (Chap. 13 Bernstein et al., 2004; Dietrich et al., 2008; Noll et al., 1991).

In addition, one dataset acquired on a 3T Siemens Connectom scanner from the 2017 MICCAI harmonization challenge<sup>7</sup> consisting of 16  $b = 0$  s/mm<sup>2</sup> volumes and 3 shells with 60 DWIs each at  $b = 1200$  s/mm<sup>2</sup>,  $b = 3000$  s/mm<sup>2</sup> and  $b = 5000$  s/mm<sup>2</sup> was used (Tax et al., 2019). The voxel size was 1.2 mm isotropic with a pulsed-gradient spin-echo echo-planar imaging (PGSE-EPI) sequence and a gradient strength of 300 mT/m. Multiband acceleration  $MB = 2$  was used with GRAPPA parallel imaging with  $R = 2$  and an adaptive combine reconstruction employing a 32 channels head coil. Other imaging parameters were  $TE / TR = 68$  ms / 5400 ms,  $\Delta/\delta = 31.1$  ms / 8.5 ms, bandwidth of 1544 Hz/pixel and partial Fourier 6/8.

**Noise estimation algorithms for comparison** To assess the performance of the proposed methods, we used three other noise estimation algorithms previously used in the context of dMRI. Default parameters were used for all of the algorithms as done in St-Jean, De Luca, Viergever, et al. (2018). The local adaptive noise estimation (LANE) algorithm (Tabelow

<sup>6</sup><https://openfMRI.org/dataset/ds000031>

<sup>7</sup><https://www.cardiff.ac.uk/cardiff-university-brain-research-imaging-centre/research/projects/cross-scanner-and-cross-protocol-diffusion-MRI-data-harmonisation>

et al., 2015) is designed for noncentral chi signal estimation, but requires *a priori* knowledge of  $N$ . Since the method works on a single 3D volume, we only use the  $b = 0$  s/mm<sup>2</sup> image for all of the experiments to limit computations as the authors concluded that the estimates from a single DWI are close to the mean estimate. We also use the Marchenko-Pastur (MP) distribution fitting on the principal component analysis (PCA) decomposition of the diffusion data, which is termed MPPCA (Veraart et al., 2016). Finally, we also compare to the Probabilistic Identification and Estimation of Noise (PIESNO) (Koay, Özarlan, and Pierpaoli, 2009), which originally proposed the change of variable to the Gamma distribution that is at the core of our proposed method. PIESNO requires knowledge of  $N$  (which is kept fixed by the algorithm) to iteratively estimate  $\sigma_g$  until convergence by removing voxels which do not belong to the distribution  $\text{Gamma}(N, 1)$  for a given slice. We set  $p = 0.05$  and  $l = 50$  for the initial search of  $\sigma_g$  in PIESNO and our proposed method, with additional parameters set to  $N_{min} = 1$  and  $N_{max} = 12$  for all cases. When estimating distributions from noise maps, we compute values in small local windows of size  $3 \times 3 \times 3$ . To the best of our knowledge, ours is the first method which estimates both  $\sigma_g$  and  $N$  jointly without requiring any prior information about the reconstruction process of the MRI scanner. Because PIESNO and LANE both *require* knowledge of the value of  $N$ , we set the correct value of  $N$  for the spatially varying noise phantom experiments and  $N = 1$  for the remaining experiments as suggested by Tabelow et al. (2015) when  $N$  is unknown. We quantitatively assess the performance of each method on the synthetic datasets by measuring the standard deviation of the noise and the relative error inside the phantom against the known value of  $\sigma_g$ , computed as

$$\text{relative error} = 100 \times (\sigma_{g_{estimated}} - \sigma_{g_{true}}) / \sigma_{g_{true}}. \quad (5.14)$$

As PIESNO and our proposed methods estimate a single value per slice whereas MPPCA and LANE provide estimates from small spatial neighborhood, we report the mean value and the standard deviation estimated inside the synthetic phantoms on each slice. For the acquired phantom datasets, we report the estimated noise distributions using both the DWIs and the measured noise maps for all 9 combinations of parallel imaging parameters that were acquired. For the in vivo datasets, we report once again the noise distributions estimated by each method. The reproducibility of the estimated distributions is assessed on the four GE datasets while the Connectom dataset is used to evaluate the performance of each compared algorithm on a bias correction and denoising task. In addition, we report  $N$  as estimated by our proposed methods for all cases.

**Bias correction and denoising of the Connectom dataset** In a practical setting, small misestimation in the noise distribution (e.g. spatially varying distribution vs nature of the distribution) might not impact much the application of choice. We evaluate this effect of misestimation on the Connectom dataset with a bias correction and a denoising task. Specifically, we apply noncentral chi bias correction (Koay, Özarlan, and Bassler, 2009)

on the in vivo dataset from the CDMRI challenge using Eq. (5.37). The algorithm is initialized with a spherical harmonics decomposition of order 6 (Descoteaux et al., 2007) as done in St-Jean, Coupé, et al. (2016). The data is then denoised using the non local spatial and angular matching (NLSAM) algorithm with 5 angular neighbors where each b-value is treated separately (St-Jean, Coupé, et al., 2016). Default parameters of a spatial patch size of  $3 \times 3 \times 3$  were used and the estimation of  $\sigma_g$  as computed by each method was given to the NLSAM algorithm. For MPPCA, LANE and PIESNO, a default value of  $N = 1$  was used and the value of  $N$  as computed by the moments and maximum likelihood equations for the proposed methods. The bias correction algorithm was also generalized for non integer values of  $N$  as detailed in Section 5.7.2.

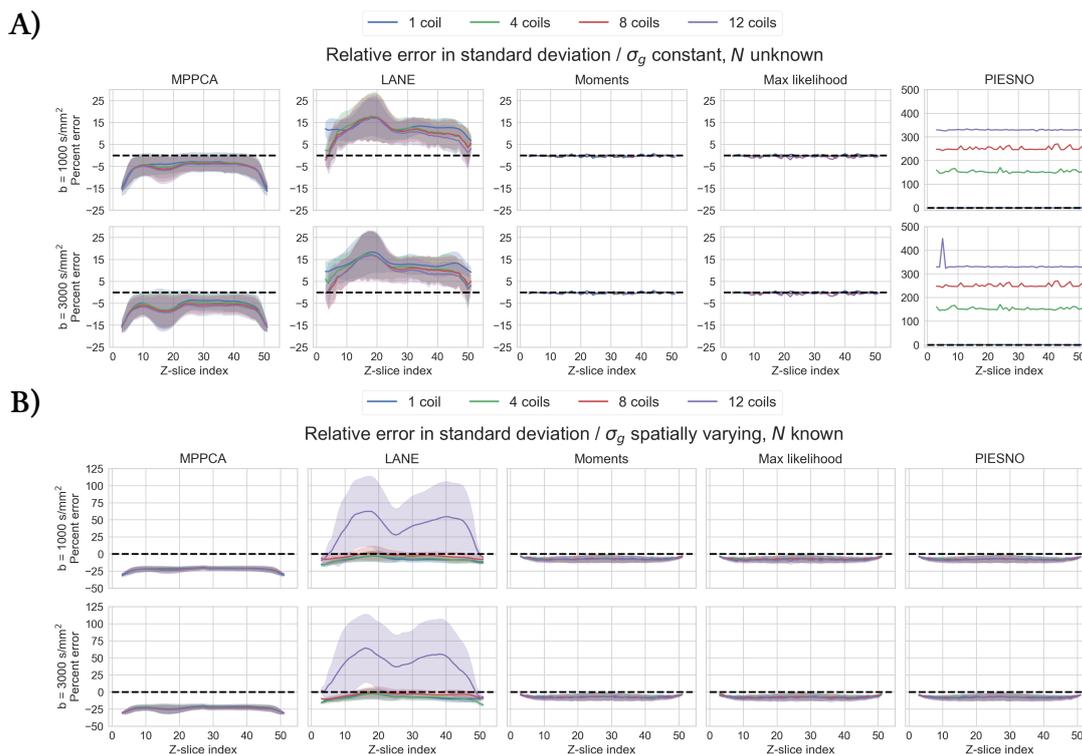
## 5.4 Results

We show here results obtained on the phantoms and in vivo datasets. The first set of simulations uses a sum of square reconstruction with stationary and spatially varying noise profiles. The second set of simulations includes SENSE and GRAPPA reconstructions, resulting in both spatially varying signal distribution profiles. Finally, the distributions estimated by each algorithm for the in vivo dataset are used for a bias correction and denoising task.

### 5.4.1 Synthetic phantom datasets

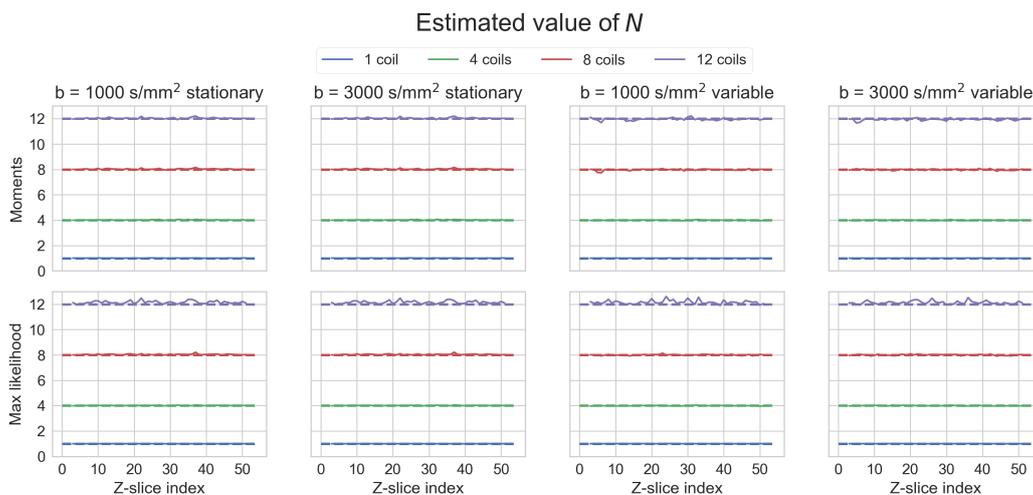
**Simulations with a sum of squares reconstruction** Fig. 5.2 shows results from simulations with stationary and spatially varying noise profiles for all datasets as estimated inside the phantom. For stationary noise profiles with  $N$  unknown, estimation of  $\sigma_g$  is the most accurate for the proposed methods with an error of about 1%, followed by MPPCA making an error of approximately 5% and LANE of 15%. The error of PIESNO increases with the value of  $N$ , presumably due to misspecification in the signal distribution, whereas MPPCA and LANE are both stable in their estimation with increasing values of  $N$ . The proposed methods using equations based on the moments and maximum likelihood recovers the correct value of  $\sigma_g$  in all cases with the lowest variance across slices, indicating that the estimated value of  $\sigma_g$  is similar in all slices as expected. The same behavior is observed for PIESNO when  $N = 1$ , but the estimated  $\sigma_g$  is larger than the correct value by two to three times when  $N$  is misspecified. In the spatially varying noise case where  $N$  is known, the moments, maximum likelihood equations and PIESNO all perform similarly with approximately 2% of error. LANE generally outperforms MPPCA except for the  $N = 12$  case, but still misestimates  $\sigma_g$  by approximately 15% and 25% respectively. Only the proposed methods and MPPCA are independent of correctly specifying  $N$ . Finally, Fig. 5.3 shows the estimated values of  $N$  by the proposed methods. Estimation generally follows the correct value, regardless of misestimation of  $\sigma_g$ .

## 5.4. Results



**Figure 5.2:** Percentage of error when the real value of  $N$  is known and  $\sigma_g$  is constant (in **A**) and  $N$  is known with  $\sigma_g$  spatially varying (in **B**) with the mean (solid line) and standard deviation (shaded area). All methods underestimate spatially varying  $\sigma_g$ , except for LANE with  $N = 12$  which overestimates it instead. On average, all methods are tied at around 5% of error with MPPCA reaching approximately 25% of error. Of interesting note, the proposed methods are tied with PIESNO when the correct value of  $N$  is given to the latter, but do not require an estimate of  $N$ , which is now an output instead of a prerequisite.

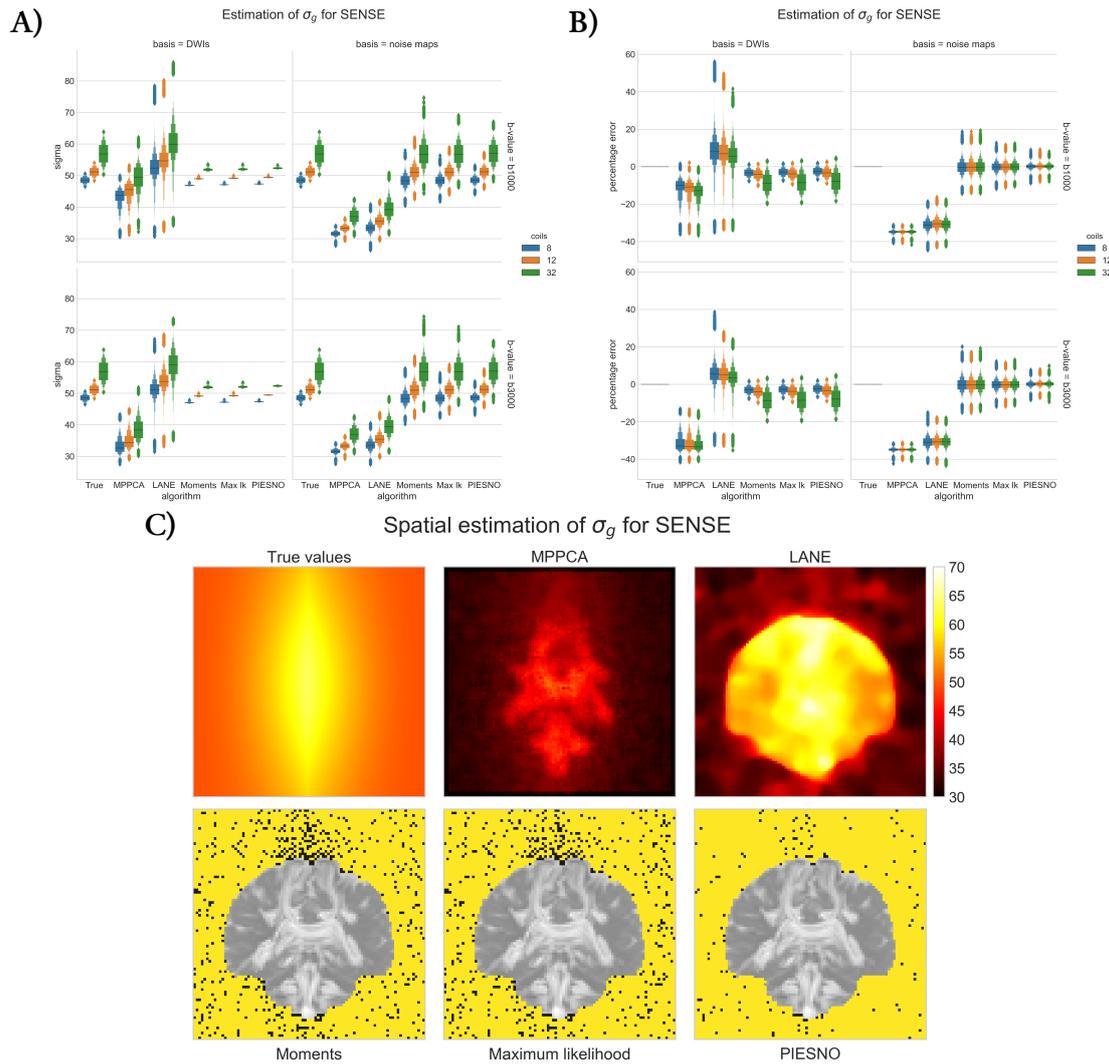
**Simulations with parallel imaging** Fig. 5.4 shows the estimated values of  $\sigma_g$  from a SENSE reconstruction and Fig. 5.5 shows the results for the GRAPPA reconstructed datasets. For SENSE, estimation using noise maps is the most precise for both proposed methods and PIESNO where the average error is around 0, followed by LANE when using DWIs as the input which results in 10% of overestimation. MPPCA generally underestimates  $\sigma_g$  by around 15% for data at  $b = 1000$  s/mm<sup>2</sup> and 30% for data at  $b = 3000$  s/mm<sup>2</sup>. LANE instead overestimates when using DWIs and underestimates  $\sigma_g$  when using noise maps and knowing the correct value of  $N = 1$ . The proposed methods (the moments and maximum likelihood equations) and PIESNO are performing similarly, but PIESNO requires knowledge of  $N = 1$ . Estimation is also more precise for the three methods using the Gamma distribution (moments, maximum likelihood and PIESNO) than those using local estimations (MPPCA and LANE) and closest to the true values when using noise maps. In the case of GRAPPA, results are similar to the SENSE experiments with the ex-



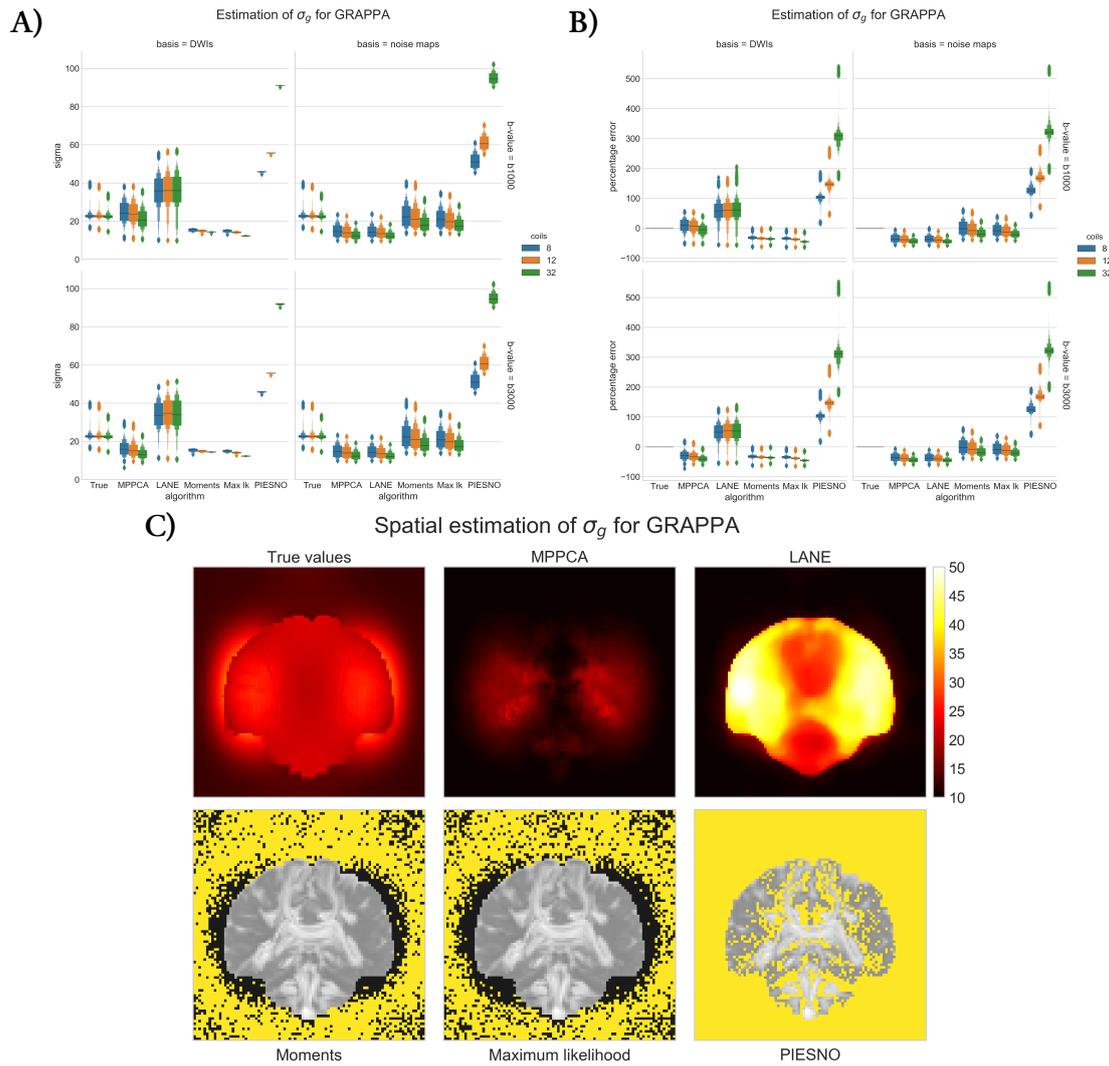
**Figure 5.3:** Estimated value of  $N$  using equations from the moments (top) and with maximum likelihood (bottom) for the proposed methods. Even for the spatially variable case where  $\sigma_g$  is slightly underestimated, the estimated values of  $N$  are stable and correspond to the real values used in the synthetic simulations in every case.

ception of MPPCA being more precise than the compared methods for the  $b = 1000 \text{ s/mm}^2$  case and performs equally well at  $b = 3000 \text{ s/mm}^2$  as the proposed methods with an average error of about 20%. Results using LANE are similar with increasing number of coils when assuming  $N = 1$ , while the estimated value from PIESNO also increases with the number of coils as previously seen in Fig. 5.2. In this case, LANE overestimates  $\sigma_g$  by around 50% when using DWIs, but performs similarly to MPPCA when estimating  $\sigma_g$  from the noise maps. Estimation from noise maps using the moments or maximum likelihood equations is the most precise in all cases. The error of PIESNO increases with  $N$  as seen in Fig. 5.5 panel C). This is caused by mistakenly including gray matter voxels of low intensity in the estimated distribution while they are correctly excluded automatically by the proposed methods. Finally, Fig. 5.6 shows the estimated values of  $N_{eff}$  using the datasets from Figs. 5.4 and 5.5 by the proposed methods. For the SENSE case, the true value is a constant  $N = 1$  by construction and the estimated values by both algorithms are on average correct with the maximum likelihood equations having the lowest variance. In the case of GRAPPA, values of  $N$  vary spatially inside the phantom and depend on the per voxel signal intensity, just as  $\sigma_g$  does in Fig. 5.5. This leads to some overestimation when only background voxels are considered, with the best estimation obtained when using the noise maps. For simulations using 8 and 12 coils, estimated values of  $N$  are, in general, following the expected values. However, the spatially varying pattern can not be fully recovered as the correct value of  $N$  depends on the true signal intensity  $\eta$  in each voxel, which is not present when collecting noise only measurements.

## 5.4. Results



**Figure 5.4:** Estimation of the noise standard deviation  $\sigma_g$  (in **A**) and the percentage error (in **B**) inside the phantom only for each method using a SENSE reconstruction with 8, 12 or 32 coils. The left columns (basis = DWIs) shows estimation using all of the DWIs, while the right column (basis = noise maps) shows the estimated values from synthetic noise maps. Results for  $b = 1000$  s/mm<sup>2</sup> are on the top row, while the bottom row shows results for the  $b = 3000$  s/mm<sup>2</sup> datasets. Figure **C**) shows the spatially estimated values of  $\sigma_g$  using the  $b = 3000$  s/mm<sup>2</sup> dataset with 32 coils for a single slice. The top row shows the results from the true distribution and local estimation as done by MPPCA and LANE. The general trend shows that even though MPPCA and LANE misestimate  $\sigma_g$ , they still follow the spatially varying pattern (lower at edges with the highest intensity near the middle) from the correct values. In the bottom row, voxels identified as belonging to the same distribution  $\Gamma(N, 1)$  are overlaid in yellow over the sum of all DWIs. Note how voxels containing signal from the DWIs are excluded by all three methods.

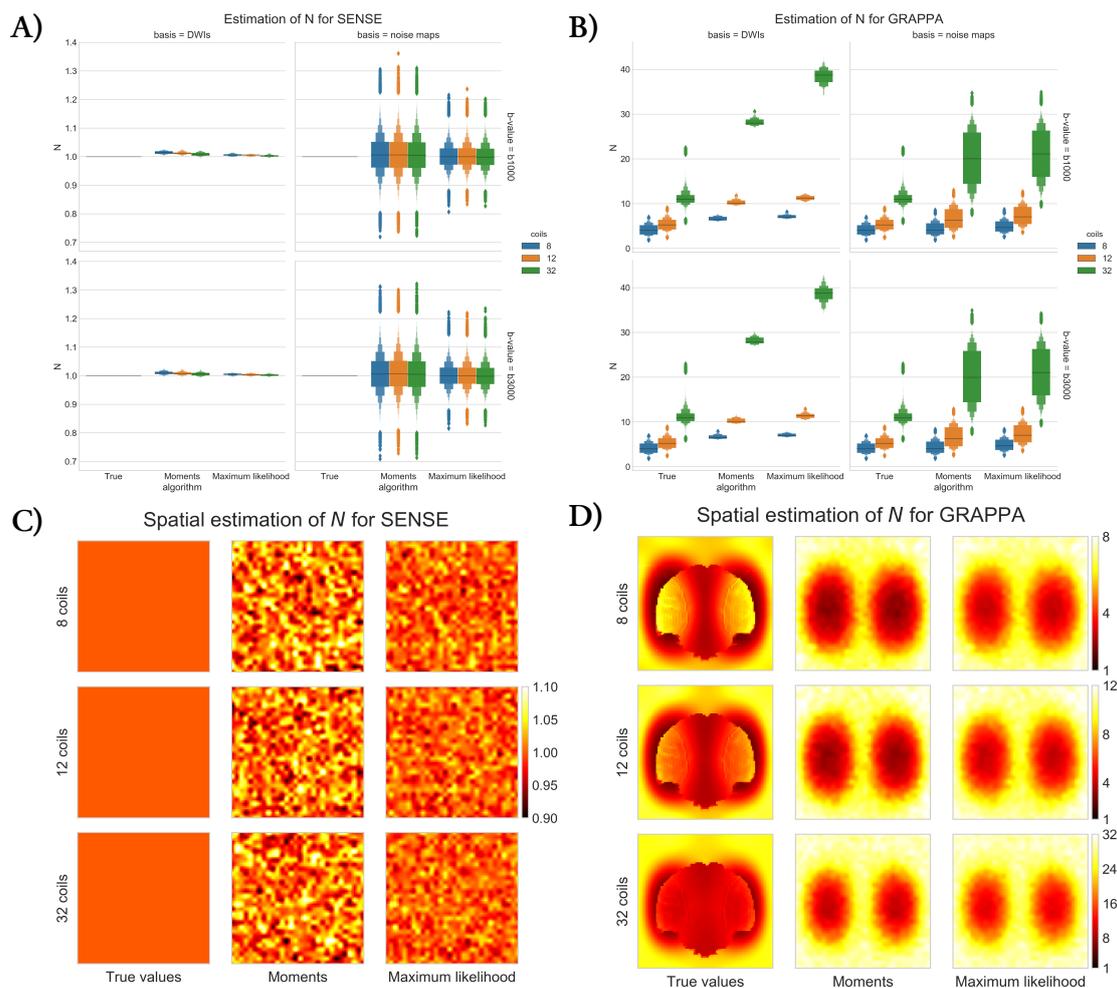


**Figure 5.5:** Estimation of the noise standard deviation  $\sigma_g$  (in **A**) and the percentage error (in **B**) inside the phantom only for each method using a GRAPPA reconstruction with 8, 12 or 32 coils, using the same conventions as Fig. 5.4. Figure **C**) shows the spatially estimated values of  $\sigma_g$  using the  $b = 3000$  s/mm<sup>2</sup> dataset with 32 coils for a single slice. The top row shows the true value of  $\sigma_g$  and the spatial estimation from MPPCA and LANE. There is once again a misestimation for both methods while following the correct spatially varying pattern. In the bottom row, voxels identified as belonging to the same distribution  $\text{Gamma}(N, 1)$  are overlaid in yellow over the sum of all DWIs. Note how PIESNO mistakenly selects some low intensity voxels belonging to the gray matter, in addition to all of the voxels in the background, which causes an overestimation of  $\sigma_g$  with a fixed value of  $N = 1$ . Both proposed methods instead select voxels with small variations in intensity as belonging to the same distribution without mistakenly selecting gray matter voxels.

#### 5.4.2 Acquired phantom datasets

Fig. 5.7 shows the estimated values of  $\sigma_g$  for all methods with a SENSE acceleration of rate  $R = 1, 2$  and 3 with multiband imaging at acceleration factors of  $MB = 2, MB = 3$  or 126

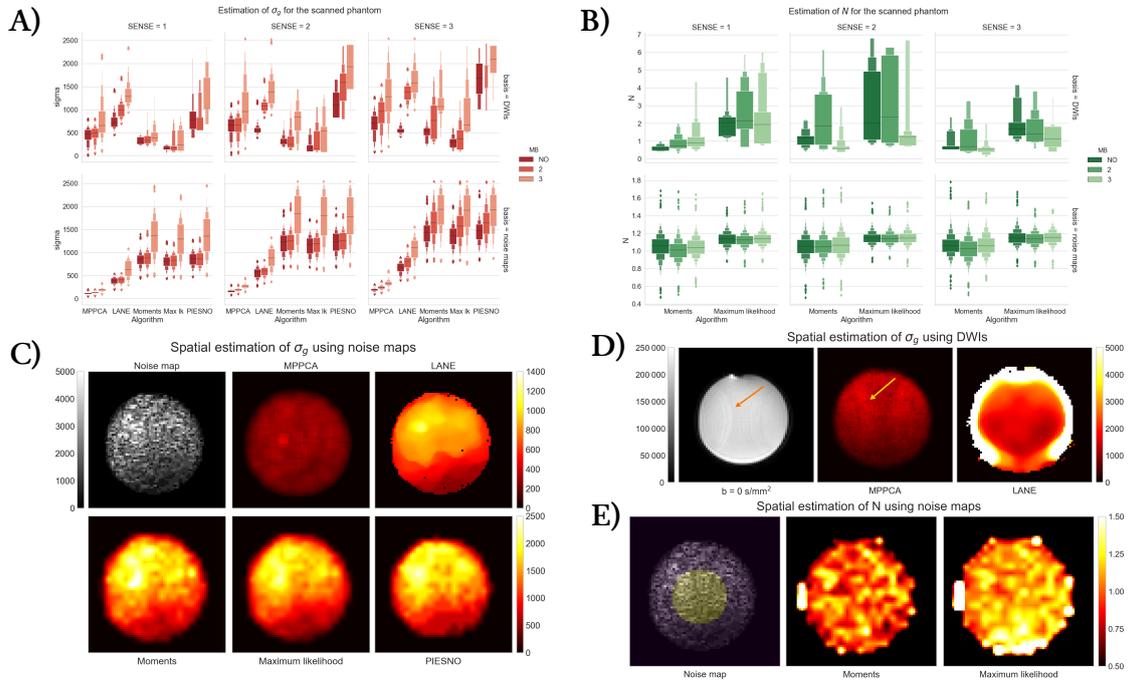
## 5.4. Results



**Figure 5.6:** The estimated values of  $N$  for SENSE (left column) and GRAPPA (right column) for the  $b = 1000$  s/mm<sup>2</sup> (first row of boxplots) and  $b = 3000$  s/mm<sup>2</sup> (second row of boxplots) datasets. The left column shows results computed from the automatically selected background voxels (basis = DWIs), while the right column shows local estimation using noise maps (basis = noise maps). In **A** and **B**), the boxplot of  $N$  inside the phantom for the SENSE/GRAPPA algorithm with a spatial map of  $N$  shown in **C**) and **D**) computed using the noise maps from the  $b = 3000$  s/mm<sup>2</sup> datasets.

deactivated in panel **A**). Results show that  $\sigma_g$  increases with  $R$  and is higher when  $MB = 3$  for  $R$  fixed, even if in theory  $\sigma_g$  should be similar for a given  $R$  and increasing  $MB$ . Panel **C**) shows the estimated values of  $\sigma_g$  when using noise maps as the input for  $R = 3$  and  $MB = 3$ . As in the synthetic experiments, MPPCA and LANE have the lowest estimates for  $\sigma_g$  with PIESNO and the proposed methods estimating higher values. Since the correct value is unknown, a reference sample slice of a noise map is also shown. When compared to values from the measured noise map, estimated values of  $\sigma_g$  are approximately fivefold lower for MPPCA, four times lower for LANE and around half for the other methods.

Estimation on the noise maps yields a value of around  $N = 1$  for both proposed methods as seen in panels **B)** and **E)**, irrespective of the acceleration used. In the case of estimation using the DWIs, the range of estimated values is larger and increases at acceleration factors of  $MB = 2$  and  $R = 2$  or  $3$ .



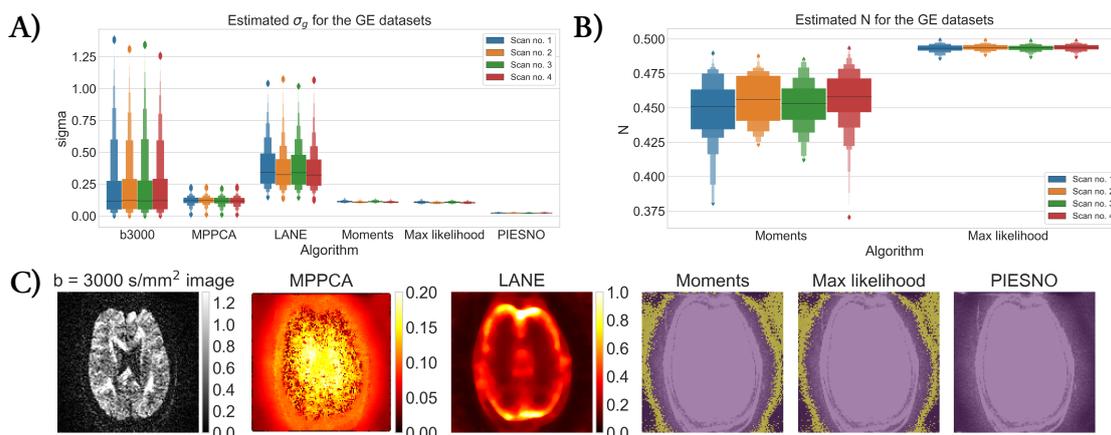
**Figure 5.7:** Estimation of noise distributions for the scanned phantom datasets inside a small ROI. Large outliers above the 95th percentile were removed to not skew the presented boxplots. In **A)**, the estimation of the noise standard deviation  $\sigma_g$  for each method using DWIs (top row) and using noise maps (bottom row). Each column shows an increasing SENSE factor, where  $\sigma_g$  increases (according to theory) with the square root of the SENSE factor. The different hues show an increasing multiband factor, which should not influence the estimation of  $\sigma_g$ . For the case  $MB = 3$ , there may be signal leakage to adjacent slices, which would increase the measured values of  $\sigma_g$  even when the estimation uses only noise maps. In **B)**, boxplots for the values of  $N$  estimated by both proposed methodologies for the experiments shown in **A)**. Estimated values using noise maps are always close to 1 on average while estimations using DWIs seems to be affected by the possible signal leakage inherent to the use of multiband imaging. In **C)**, an axial slice of a noise map and estimated values of  $\sigma_g$  by all methods for the case  $R = 3$  and  $MB = 3$ , which is the highest rate of acceleration from all of the investigated cases. Note the different scaling between the top and bottom row as MPPCA and LANE estimates of  $\sigma_g$  are two to three times lower than other methods. In **D)**, a  $b = 0$  s/mm<sup>2</sup> image of the phantom and spatially estimated values of  $\sigma_g$  for MPPCA and LANE. Note how some signal leakage (orange arrows) is affecting the  $b = 0$  s/mm<sup>2</sup> volume due to using  $MB = 3$ . In **E)**, location of the spherical ROI used for the boxplots overlaid on a noise map and spatially estimated values of  $N$  for both proposed methods. As less voxels are available near the borders of the phantom, estimating the noise distributions parameters results in lower precision.

### 5.4.3 In vivo datasets

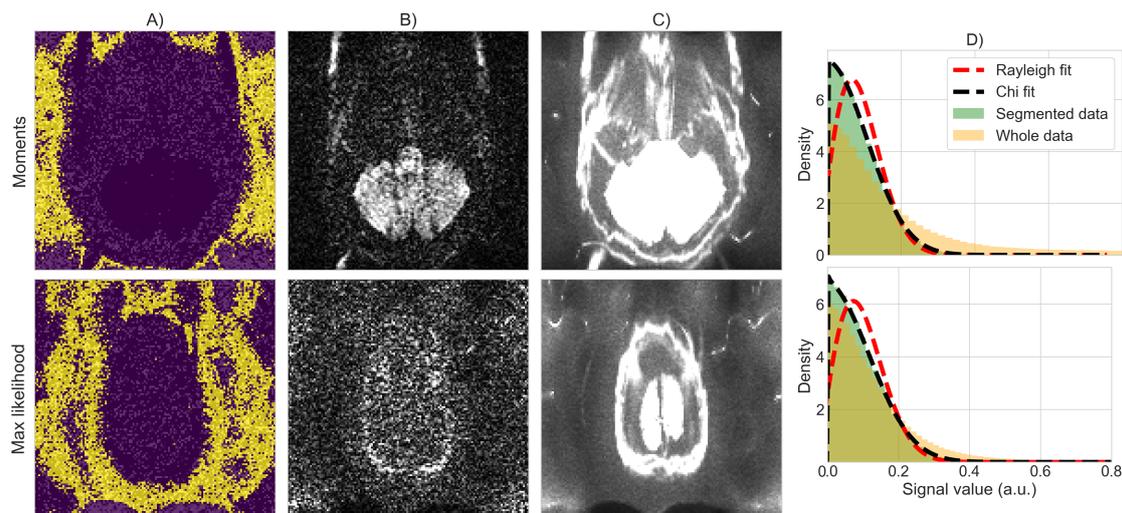
**Multiple datasets from a single subject** Fig. 5.8 shows the estimated value of  $\sigma_g$  on four repetitions of the GE datasets for each method as computed inside a brain mask. The values from a  $b = 3000 \text{ s/mm}^2$  volume (including background) is also shown as a reference for the values present at the highest diffusion weighting in the dataset. All methods show good reproducibility, as their estimates are stable across the data. The value of  $N$  as computed by our proposed methods is also similar for all datasets with the median at  $N = 0.45$  for the moments and  $N = 0.49$  for the maximum likelihood equations. This corresponds to a half Gaussian distribution as would be obtained by a real part magnitude reconstruction (Dietrich et al., 2008). However, LANE recovered the highest values of  $\sigma_g$  amongst all methods with a large variance and a median higher than the  $b = 3000 \text{ s/mm}^2$  values, which might indicate overestimation in some areas. The median of MPPCA and the proposed methods are similar, while PIESNO estimates of  $\sigma_g$  are approximately two times lower. This could indicate that specifying  $N = 1$  was incorrect for these datasets, as PIESNO identified about 10 noise only voxels.

Fig. 5.9 shows an axial slice around the cerebellum and the top of the head which are corrupted by acquisition artifacts likely due to parallel imaging. Voxels containing artifacts were automatically discarded by both methods, preventing misestimation of  $\sigma_g$  and  $N$ . The values computed from these voxels also offer a better qualitative fit than assuming a Rayleigh distribution or selecting non-brain data. We also timed each method to estimate  $\sigma_g$  on one of the GE datasets using a standard desktop computer with a 3.5 GHz Intel Xeon processor. The runtime to estimate  $\sigma_g$  (and  $N$ ) was around 5 seconds for the maximum likelihood equations, 9 seconds for the moments equations, 11 seconds for PIESNO, 3 minutes for MPPCA and 18 minutes for LANE.

**Estimation with a Connectom dataset** Fig. 5.10 shows in **A)** the estimated values of  $\sigma_g$  inside a brain mask and in **B)** the values of  $N$  computed by the proposed methods. Estimated values of  $\sigma_g$  vary by an order of magnitude between the different methods. In the case of MPPCA and LANE, the median of the estimates is higher than the reference  $b = 5000 \text{ s/mm}^2$  data, while PIESNO and the proposed methods estimate values lower than the reference and have lower variability in their estimated values. For the estimation of  $N$ , recovered values are distributed close to 1 as is expected from an adaptive combine reconstruction providing a Rician distribution. Values estimated with the maximum likelihood equations have a lower variability than with the moments equations. In **C)**, the top row shows the  $b = 5000 \text{ s/mm}^2$  volume and spatial maps of  $\sigma_g$  as estimated by MPPCA and LANE. The bottom row shows voxels identified as pure noise (in light purple) using the moments, the maximum likelihood equations and PIESNO. Ghosting artifacts are excluded, but presumably affect estimation using the entire set of DWIs shown in the top row. Fig. 5.11 shows in **A)** the signal intensity after applying bias correction (left column) and after denoising (right column) for each volume ordered by increasing b-value. The top



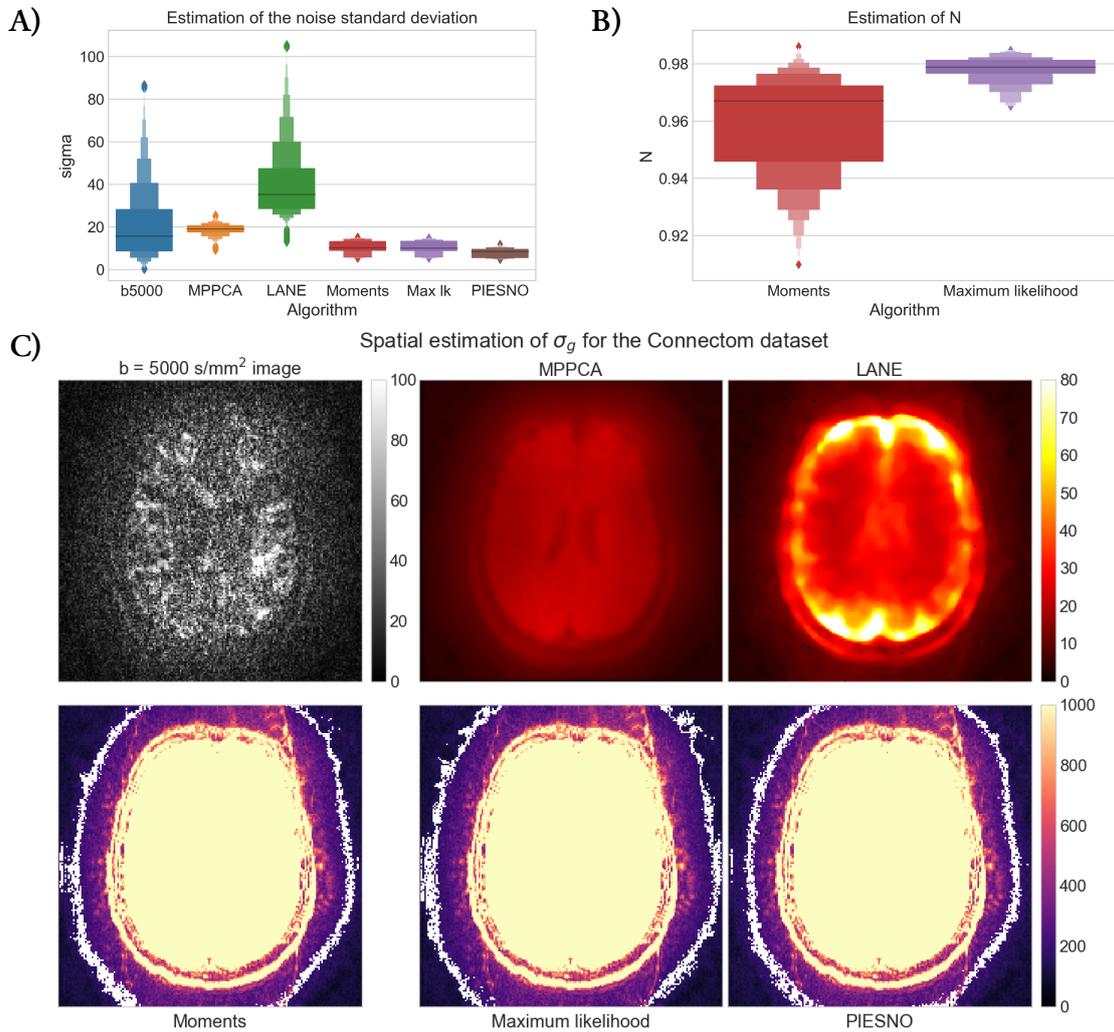
**Figure 5.8:** Estimation of the noise profiles on four repetitions of a single subject from a GE scanner. In **A)**, the baseline signal values of a  $b = 3000 \text{ s/mm}^2$  volume and estimated values of  $\sigma_g$  for all methods inside a brain mask and **B)** estimated values of  $N$  by the proposed methods are shown. Note that the values for LANE and the  $b = 3000 \text{ s/mm}^2$  volume were truncated at the 99 percentile to remove extreme outliers. In **C)**, an axial slice of a  $b = 3000 \text{ s/mm}^2$  image from one dataset and the estimated values of  $\sigma_g$  for MPPCA and LANE. For the proposed methods and PIESNO, a mask of the identified background voxels (in yellow) overlaid on the data.



**Figure 5.9:** An axial slice in the cerebellum from one of the GE datasets. Voxels identified in **A)** as noise only (yellow) are free of artifacts in a single slice in **B)** or along the sum of all volumes in **C)**. In **D)**, the normalized density histogram using the selected voxels from **A)** (green) fit well a chi distribution (black dashed lines), while assuming a Rayleigh distribution (red dashed lines) or using all non brain voxels (orange) leads to a worse visual fit.

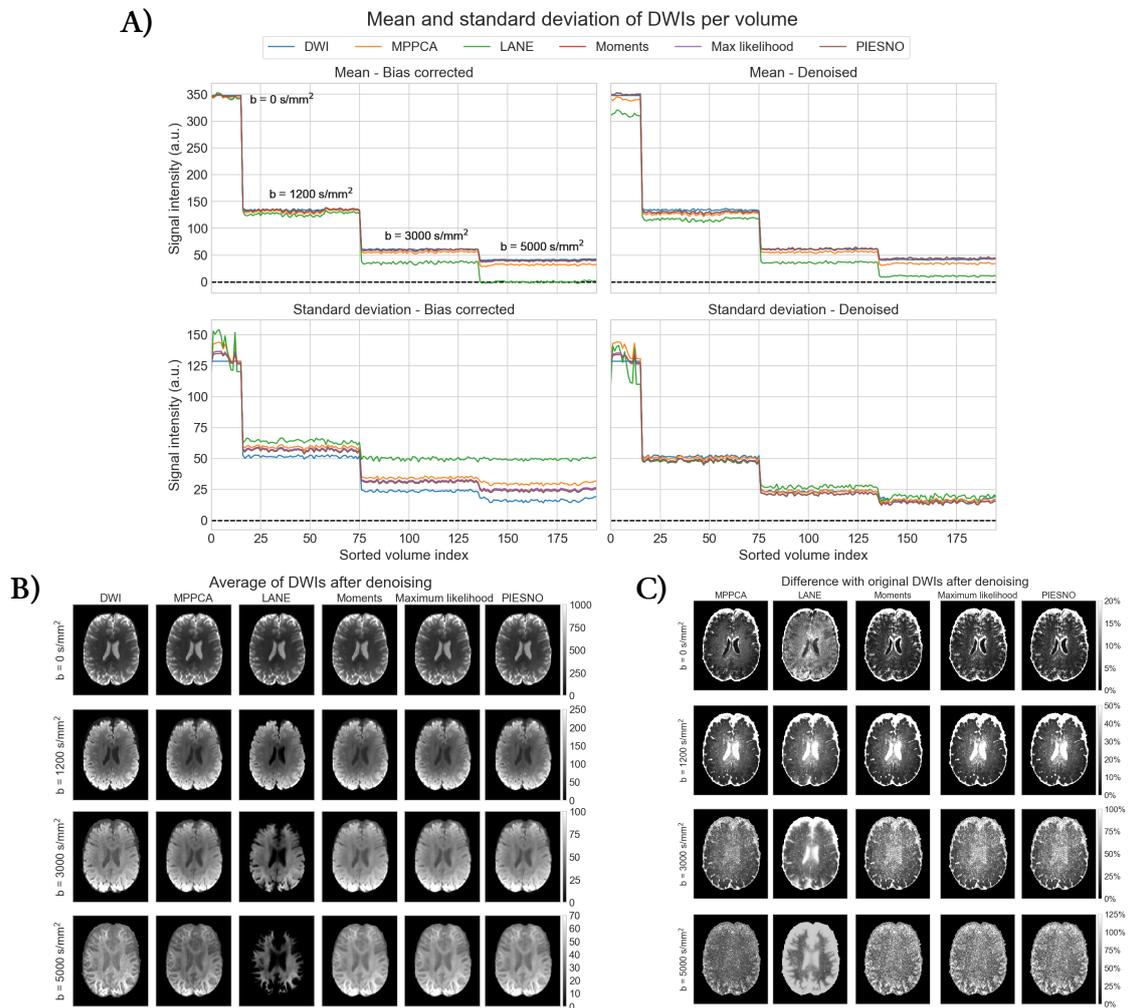
row (resp. bottom row) shows the mean (resp. standard deviation) as computed inside a white and gray matter mask. The mean signal decays with increasing b-value as expected,

but the standard deviation of the signal does not follow the same trend in the cases of LANE. After denoising, the mean signal and its standard deviation decays once again as for the original data. Panel **B)** shows the average DWI at a given b-value for the original dataset and after denoising using the noise distribution from each method. Results are similar for all methods for the  $b = 0$  s/mm<sup>2</sup> datasets, but the overestimation of  $\sigma_g$  by LANE produces missing values in the gray matter for  $b = 3000$  s/mm<sup>2</sup> and  $b = 5000$  s/mm<sup>2</sup>. In general, averaging reduces the noise present at  $b = 0$  s/mm<sup>2</sup> and  $b = 1200$  s/mm<sup>2</sup> while only denoising is effective at  $b = 3000$  s/mm<sup>2</sup> and  $b = 5000$  s/mm<sup>2</sup>. At  $b = 5000$  s/mm<sup>2</sup>, the MPPCA denoised volume is of lower intensity than when obtained by the moments, maximum likelihood equations or PIESNO. This is presumably due to LANE and MPPCA estimating higher values of  $\sigma_g$  than the three other methods. Finally, panel **C)** shows the absolute difference between the original and the denoised dataset obtained by each method. At  $b = 5000$  s/mm<sup>2</sup>, LANE removes most of the signal in the gray matter mistakenly due to overestimating  $\sigma_g$ . Other methods perform comparably well on the end result, despite estimates of  $\sigma_g$  of different magnitude.



**Figure 5.10:** Estimation of noise distributions for the Connectom dataset. In **A)**, signal distribution of the original data and noise standard deviation  $\sigma_g$  for all methods, where data above the 99th percentile for the  $b = 5000 \text{ s/mm}^2$  volume and LANE were discarded. In **B)**, values of  $N$  as estimated using the moments (in red) and by maximum likelihood (in purple). In **C)** on the top row, a  $b = 5000 \text{ s/mm}^2$  volume and spatial estimation of  $\sigma_g$  as measured by MPPCA and LANE. On the bottom row, voxels identified as containing only noise (in white) by the moments, maximum likelihood and PIESNO overlaid on top of the sum of the  $b = 0 \text{ s/mm}^2$  volumes. Note how each algorithm identifies different voxels, while automatically ignoring voxels belonging to the data or contaminated with signal leakage from multiband imaging.

## 5.4. Results



**Figure 5.11:** Bias correction and denoising of the Connectom dataset from the noise distributions estimated by each method. In **A)**, the left column (resp. right column) shows the result of noncentral chi bias correction (resp. denoising) on the signal value. The top row (resp. bottom row) shows the mean (resp. standard deviation) of the signal inside a white and gray matter mask for each volume. Note how the bias corrected value of LANE goes below 0 (dashed line) due to its high estimation of  $\sigma_g$ . After denoising, the standard deviation of the signal decreases as the b-value increases, an effect which is less noticeable for the bias corrected signal only. However, this effect is less pronounced for the bias corrected signal only in the case of LANE and MPPCA. In **B)**, spatial maps of the original data and after denoising (in each column) from averaged datasets at  $b = 0 \text{ s/mm}^2$ ,  $b = 1200 \text{ s/mm}^2$ ,  $b = 3000 \text{ s/mm}^2$  and  $b = 5000 \text{ s/mm}^2$  (in each row) for each method. Note how each b-value uses a different scale to enhance visualization even though the signal intensity is lower for increasing b-values. Panel **C)** shows the difference in percentage between the original data and after denoising using parameters as estimated by each algorithm.

## 5.5 Discussion

We have shown how a change of variable to a Gamma distribution  $\text{Gamma}(N, 1)$  can be used to robustly and automatically identify voxels belonging only to the noise distribution. At each iteration, the moments (Eqs. (5.8) and (5.9)) and maximum likelihood equations (Eqs. (5.11) and (5.12)) of the Gamma distribution can be used to compute the number of degrees of freedom  $N$  and the Gaussian noise standard deviation  $\sigma_g$  relating to the original noise distribution. Voxels not adhering to the distribution are discarded, therefore refining the estimated parameters until convergence. One of the advantage of our proposed methods is that no *a priori* knowledge is needed from the acquisition or the reconstruction process itself, which is usually not stored or hard to obtain in a clinical setting. Results from Section 5.4.1 show that we can reliably estimate parameters from the magnitude data itself in the case of stationary distributions. For spatially varying distributions without parallel acceleration, the proposed methods achieve an average relative error of approximately 10% when estimating  $\sigma_g$ , which is equal or better than the other methods compared in this work. Estimated values of  $N$  are around the true values, even when  $\sigma_g$  is misestimated. While these experiments may still be considered simplistic when compared to modern scanning protocols where parallel acceleration is ubiquitous, they highlight that even textbook cases can lead to misestimation if the correct signal distribution is not taken into account. Practical tasks taking advantage of the signal distribution such as bias correction (Peciak et al., 2018), noise floor removal (Sakaie and Lowe, 2017), deep learning reconstruction with various signal distributions (Lønning et al., 2019) or diffusion model estimation (Collier et al., 2018; Landman et al., 2007; Zhang et al., 2012) may be tolerant, but not perform optimally, to some misestimation of the noise distribution. See e.g. Hutchinson et al. (2017) for discussions on the impact of noise bias correction on diffusion metrics in an ex vivo rat brain dataset.

**Effects of misspecification of the noise distribution** Experiments with SENSE from Section 5.4.1 reveal that using a local estimation with noise maps provides the best estimates for the proposed methods and PIESNO. MPPCA and LANE perform better when using DWIs as the input rather than noise maps, but at the cost of a broader range of estimated values for  $\sigma_g$  and still underperform when compared to the three other methods. This is presumably because the signal diverges from a Gaussian distribution at low SNR (Gudbjartsson and Patz, 1995) and especially in noise maps, leading to a misspecification of its parameters when the assumed noise distribution is incorrect. Phantom experiments carried with GRAPPA show similar trends except for PIESNO, which overestimates  $\sigma_g$  as shown in Fig. 5.2. When erroneously fixing  $N = 1$ , low intensity voxels where  $\eta > 0$  (e.g. gray matter) may be mistakenly included in the distribution after the change of variables, leading to overestimation of  $\sigma_g$ .

The presence of tissue in voxels used for noise estimation might compromise the accuracy of the estimated distributions as shown in Section 5.4.1. This can be explained by

the lower number of noise only voxels available to the proposed methods and PIESNO and to difficulty in separating the signal from the noise for MPPCA and LANE at low SNR. Using measured noise maps is not a foolproof solution as by definition they set  $\eta = 0$ , while the (unknown) noiseless signal from tissues is  $\eta > 0$ . As the noise distribution may depend on  $\eta$  (Aja-Fernández, Vegas-Sánchez-Ferrero, et al., 2014), this means that its parameters (e.g. from a GRAPPA reconstruction) will be inherently different than the one estimated from noise maps. This effect can be seen in Fig. 5.6, where the estimated values of  $N$  from noise maps and DWIs are close to 1 for SENSE as expected in theory. For GRAPPA, they are either overestimated and underestimated in regions of the phantom and overestimated in background regions as  $N$  locally depends on  $\eta$ . Accurate estimation of  $\sigma_g$  and  $N$  over signal regions still remains an open challenge. Nevertheless, the median of the estimated distribution of  $\sigma_g$  is closer to the true distribution when using noise maps than when using DWIs for the proposed methods. Such noise map measurements could therefore provide improved signal distribution estimation for, e.g. body or cardiac imaging, where no intrinsic background measurements are available.

**Effects of parallel imaging and multiband in a phantom** Section 5.4.2 presented results from a scanned phantom using SENSE coupled with multiband imaging. While no ground truth is available, a SENSE acceleration should provide a Rician signal distribution ( $N = 1$ ) and  $\sigma_g$  should increase with  $\sqrt{R}$  (Aja-Fernández, Vegas-Sánchez-Ferrero, et al., 2014). Fig. 5.7 shows that for a common SENSE factor, all values of  $\sigma_g$  estimated with  $MB = 3$  are larger than at lower factors. The use of multiband imaging should not influence the estimation of  $\sigma_g$  as it only reduces the measured signal, and not the noise component unlike SENSE does. Indeed, estimated values of  $\sigma_g$  are stable until  $MB = 3$  or  $R = 2$  and  $MB = 2$  is used; this is possibly due to signal leakage and aliasing signal from multiband folding over from adjacent slices with higher factors (Barth et al., 2016; Todd et al., 2016). Noise maps are less affected by this artifact, which is already present when  $R = 2$  and  $MB = 2$ , as adjacent voxels have low values, similarly to unaffected voxels. However, leaking signal in DWIs might impact parameters estimation as it can be interpreted as an increase in SNR and therefore a lower noise contribution than expected. Estimation of  $\sigma_g$  is also increasing approximately with  $\sqrt{R}$  for all methods as expected (Aja-Fernández, Vegas-Sánchez-Ferrero, et al., 2014). While we can not quantify these results, this follows the synthetic experiments with SENSE shown in Fig. 5.4, where PIESNO and the proposed methods were more precise in estimating  $\sigma_g$  from noise maps.

In the case of estimation using DWIs as input, this expected increase in  $\sigma_g$  for increasing SENSE factor is less obvious e.g. LANE estimates of  $\sigma_g$  decrease from  $R = 2$  to  $R = 1$  for the no multiband case. As MPPCA and LANE also estimate  $\eta$ , it could explain the larger variance of  $\sigma_g$  as  $\eta$  fundamentally depends on the microstructural content of each voxel, which is complex and subject to large spatial variations, e.g. notably across DWIs. This also means that estimation over DWIs is susceptible to signal leakage, which would

explain the increased estimated values of  $\sigma_g$  for  $MB = 2$  and  $MB = 3$  for a given SENSE factor. In the noise maps, we have observed that MPPCA and LANE estimated  $\eta > \sigma_g$  in all cases (results not shown). Overestimating the true value of  $\eta = 0$ , which is an implicit assumption in PIESNO and the proposed methods, could explain underestimation of  $\sigma_g$  when using noise maps. This overestimation of  $\eta$  in turn leads to lower estimates of  $\sigma_g$ . The use of multiband and the inherent signal leakage at high factors could explain this overestimation of  $\eta$  and underestimation of  $\sigma_g$  for all tested cases. In the case of SENSE, the proposed methods estimated approximately  $N = 1$  in all cases, suggesting robustness to multiband artifacts.

**Estimation of noise distributions for in vivo datasets** To complement earlier sections, two datasets acquired on different scanners combining parallel and multiband imaging were analyzed in Section 5.4.3. Fig. 5.8 shows that assuming a Rician distribution with  $N = 1$  can prove inadequate in some situations. The four repetitions of a single subject acquired on a GE scanner point towards a half Gaussian distribution instead as evidenced by the computed values of  $N$  around 0.5. This is further evidenced by the low number of voxels (less than 10) detected by PIESNO while assuming  $N = 1$ . In the preliminary results of our MICCAI submission (St-Jean, De Luca, Viergever, et al., 2018), using  $N = 0.5$  for PIESNO gave similar results to the proposed methods, suggesting the departure of the data from a pure Rician distribution. Additionally, Fig. 5.9 shows that those voxels identified automatically as pure noise also adhere closer to a chi distribution than a Rayleigh distribution (where  $\eta = 0$  in both cases). Considering the whole distribution of the data, which is contaminated by artifacts, would also lead to a different distribution. Even if local methods can consider spatially varying noise profile, the local estimation of  $\sigma_g$  will invariably be affected whenever those same artifacts repeat over the data. This introduces a compromise between avoiding artifacts at the cost of reduced spatial specificity and local methods which may not be able to exclude artifacts, but provide local estimations of  $\sigma_g$ . Measurements from noise maps, if available, could therefore offer a middle ground if  $N$  is low or does not depend locally on the coil geometries (e.g. SENSE or homodyne reconstruction) as shown in Section 5.4.1.

Fig. 5.10 shows a large range of estimates for  $\sigma_g$  across methods. In particular, the moments and maximum likelihood equations estimate smaller values of  $\sigma_g$  than MPPCA and LANE, but larger than PIESNO, while still recovering values of  $N$  close to 1 and successfully discard voxels contaminated by multiband artifacts. The correct value of  $\sigma_g$  most likely sits between these two results as parallel MRI produces spatially varying noise profile, which is higher in the center and not fully captured by the background signal, but the local estimation methods also overestimated  $\sigma_g$  in our synthetic simulations. In panel A), MPPCA and LANE estimates of  $\sigma_g$  with DWIs are likely affected by multiband artifacts, as the median is larger than the signal level at  $b = 5000$  s/mm<sup>2</sup>. This indicates a possible overestimation as  $\sigma_g$  should be lower than the measured signal at the highest

b-value. For PIESNO and the proposed methods, the median  $\sigma_g$  is lower than the median of the reference  $b = 5000$  s/mm<sup>2</sup> data. An overestimation of  $N$  could explain the low values of  $\sigma_g$  estimated by the proposed methods just as misestimation of  $\eta$  by MPPCA and LANE could affect their respective estimate of  $\sigma_g$  by balancing out the misestimated values.

Fig. 5.11 shows the result of each method on a bias correction and denoising task on the Connectom dataset. In panel A), the standard deviation of the signal (bottom left panel) is increased after bias correction for LANE (green line) and decreased (around the same level) for the other methods when compared to the uncorrected data (blue line). The situation is similar after denoising, but to a lesser extent, while the moments, maximum likelihood equations and PIESNO follow the same signal level as the unprocessed data on average. Regarding the mean of the signal itself, LANE is on average lower or close to 0 after bias correction, indicating potential degeneracies due to overestimation of  $\sigma_g$ . From panels B) and C), the results of all methods are visually similar except for LANE (especially at  $b = 3000$  s/mm<sup>2</sup> and  $b = 5000$  s/mm<sup>2</sup>), indicating that the NLSAM denoising algorithm treated different values of  $\sigma_g$  in the same way. This is because the optimal regularized solution (which depends on  $\sigma_g$ ) is piecewise constant (St-Jean, Coupé, et al., 2016; Tibshirani and Taylor, 2011) and can tolerate small deviations in  $\sigma_g$ . Finally, MPPCA, the moments and maximum likelihood equations and PIESNO perform similarly, even if they estimated different values of  $\sigma_g$  and  $N$ , with MPPCA showing slightly lower signal intensity at  $b = 5000$  s/mm<sup>2</sup>. This could be due to the bias correction having a larger effect when  $\sigma_g$  is larger, increasing the standard deviation of the resulting signal. As shown in panel C), the difference with the original dataset for MPPCA is lower than the proposed methods or PIESNO, even though the estimated value of  $\sigma_g$  was larger.

## 5.6 Conclusions

We presented a new, fully automated framework for characterizing the noise distribution from a diffusion MRI dataset using the moments or maximum likelihood equations of the Gamma distribution. The estimated parameters can be subsequently used for e.g. bias correction and denoising as we have shown or diffusion models taking advantage of this information. This requires only magnitude data, without the use of dedicated maps or parameters intrinsic to the reconstruction process, which may be challenging to obtain in practice. The proposed framework is fast and robust to artifacts as voxels not adhering to the noise distribution can be automatically discarded using an outlier rejection step. This makes the proposed methods also applicable on previously acquired datasets, which may not carry the necessary information required by more advanced estimation methods. Experiments using parallel MRI and multiband imaging on simulations, an acquired phantom and in vivo datasets have shown how modern acquisition techniques complicate estimation of the signal distribution due to artifacts at high acceleration factor. This issue can be alleviated with the use of noise only measurements or by limiting the acceleration factor

to prevent signal leakage. Moreover, different vendors implement different default reconstruction algorithms which leads to different signal distributions, challenging the strategy of assuming a Rician distribution or approximations of  $N$  based on the physical amount of channels in the receiver coil. We also have shown how signal bias correction and denoising can tolerate some misestimation of the noise distribution using an in vivo dataset. Noteworthy is that the theory we presented also applies to any other MRI weighting using large samples of magnitude data (e.g. functional MRI, dynamic contrast enhanced MRI). This could help multicenter studies or data sharing initiatives to include knowledge of the noise distribution in their analysis in a fully automated way to better account for inter-scanner effects.

## 5.7 Appendix

### 5.7.1 Estimating parameters of the Gamma distribution

**Estimation using the method of moments** For any given distribution, we can estimate its parameters by relating the samples and the theoretical expression of its moments. The Gamma distribution is parametrized as  $Gamma(\alpha, \beta)$  and has a probability distribution function of

$$pdf(t|\alpha, \beta) = \frac{t^{\alpha-1}}{\Gamma(\alpha)\beta^\alpha} \exp(-t/\beta) dt \quad (5.15)$$

with  $t, \alpha, \beta > 0$  and  $\Gamma(x)$  the gamma function. The first moments are analytically given by (Chap. 5 Papoulis, 1991; Weisstein, 2017).

$$\mu_{gamma} = \alpha\beta, \sigma_{gamma}^2 = \alpha\beta^2, \quad (5.16)$$

In this paper, the Gamma distribution parameters are  $Gamma(\alpha = N, \beta = 1)$  after the change of variable  $t = m^2/(2\sigma_g^2)$  for our particular case. Since we have  $\beta = 1$ , this leads to a special case where the mean and variance are *equal* with a value of  $\alpha = N$  and can be expressed only in terms of the magnitude signal  $m$ . For simplicity, we will only use the mean  $\mu_{gamma}$  and variance  $\sigma_{gamma}^2$  to estimate the required parameters  $N$  and  $\sigma_g^2$ , but higher order moments could also be used. However, in practice, they might accumulate numerical errors due to the higher powers involved and are not used here since two equations are enough to estimate the two parameters. Starting from the analytical expression given by Eq. (5.16), we have for the special case  $Gamma(N, 1)$

$$\mu_{gamma} = \alpha, \sigma_{gamma}^2 = \alpha \quad (5.17)$$

Which we can compute using the sample mean and sample variance formulas such that

$$\alpha = \frac{1}{K} \sum_{k=1}^K t_k = \frac{1}{K} \sum_{k=1}^K t_k^2 - \left( \frac{1}{K} \sum_{k=1}^K t_k \right)^2 \quad (5.18)$$

Substituting the equation for the moments in terms of  $t = m^2/2\sigma_g^2$ , we obtain

$$\frac{1}{K} \sum_{k=1}^K \frac{m_k^2}{2\sigma_g^2} = \frac{1}{K} \sum_{k=1}^K \left( \frac{m_k^2}{2\sigma_g^2} \right)^2 - \left( \frac{1}{K} \sum_{k=1}^K \frac{m_k^2}{2\sigma_g^2} \right)^2 \quad (5.19)$$

$$\Rightarrow \frac{1}{2K\sigma_g^2} \sum_{k=1}^K m_k^2 = \frac{1}{4K\sigma_g^4} \sum_{k=1}^K m_k^4 - \frac{1}{4K^2\sigma_g^4} \left( \sum_{k=1}^K m_k^2 \right)^2 \quad (5.20)$$

$$\Rightarrow \sum_{k=1}^K m_k^2 = \frac{1}{2K\sigma_g^2} \left( K \sum_{k=1}^K m_k^4 - \left( \sum_{k=1}^K m_k^2 \right)^2 \right) \quad (5.21)$$

$$\Rightarrow 2K\sigma_g^2 = \frac{K \sum_{k=1}^K m_k^4 - \left( \sum_{k=1}^K m_k^2 \right)^2}{\sum_{k=1}^K m_k^2} \quad (5.22)$$

$$\Rightarrow \sigma_g = \frac{1}{\sqrt{2K}} \sqrt{\frac{K \sum_{k=1}^K m_k^4 - \left( \sum_{k=1}^K m_k^2 \right)^2}{\sum_{k=1}^K m_k^2}} \quad (5.23)$$

$$\Rightarrow \sigma_g = \frac{1}{\sqrt{2}} \sqrt{\frac{\sum_{k=1}^K m_k^4}{\sum_{k=1}^K m_k^2} - \frac{1}{K} \sum_{k=1}^K m_k^2} \quad (5.24)$$

Therefore, it is possible to estimate the Gaussian noise standard deviation using [Eq. \(5.24\)](#) and the values of magnitude data  $m_k$ , assuming that the voxels considered here do not contain any object signal. With the value of the noise variance  $\sigma_g^2$  now known, going back to the original Gamma distribution  $Gamma(\alpha = N, \beta = 1)$  yields the number of coils  $N$  as previously shown by [Eq. \(5.9\)](#)

$$N = \alpha = \mu_{gamma} = \frac{1}{2K\sigma_g^2} \sum_{k=1}^K m_k^2 \quad (5.25)$$

**Estimation using maximum likelihood equations** An alternative to the method of moments to estimate parameters from a given distribution is to solve the equations derived from its likelihood function for each unknown parameter. Given a set of observed data, maximizing the likelihood function from a known distribution (or equivalently, the log of the likelihood function) yields a set of equations to estimate its parameters. For the  $Gamma(\alpha, \beta)$  distribution, maximizing the log likelihood by equating the partial deriva-

tive to 0 for each parameter yields (Thom, 1958)

$$\frac{1}{K\beta} \sum_{k=1}^K t_k - \alpha = 0 \quad (5.26)$$

$$\log(\beta) + \frac{d}{d\alpha} \log(\Gamma(\alpha)) - \frac{1}{K} \sum_{k=1}^K \log(t_k) = 0 \quad (5.27)$$

Since we have  $\alpha = N$  and  $\beta = 1$ , in this special case Eq. (5.26) is the same as Eq. (5.25).

Combining Eqs. (5.26) and (5.27) yields an implicit equation to estimate  $\sigma_g$ , which can be written as

$$f(\sigma_g) = \psi \left( \frac{1}{2K\sigma_g^2} \sum_{k=1}^K m_k^2 \right) - \frac{1}{K} \sum_{k=1}^K \log(m_k^2) + \log(2\sigma_g^2) = 0 \quad (5.28)$$

$$f'(\sigma_g) = \frac{1}{K\sigma_g^3} \left[ \psi' \left( \frac{1}{2K\sigma_g^2} \sum_{k=1}^K m_k^2 \right) \sum_{k=1}^K m_k^2 - 2K\sigma_g^2 \right] = 0 \quad (5.29)$$

and Eq. (5.27) can be rewritten as an implicit equation of  $N$

$$f(N) = \psi(N) - \frac{1}{K} \sum_{k=1}^K \log(m_k^2/2\sigma_g^2) = 0 \quad (5.30)$$

$$f'(N) = \psi'(N) = 0 \quad (5.31)$$

where  $\psi(x) = \frac{d}{dx} \log(\Gamma(x))$  is the digamma function and  $\psi'$  is the derivative of  $\psi$ , called the polygamma function. Eqs. (5.29) and (5.31) can be solved numerically using Newton's method provided we have a starting estimate  $x_0$ . The update rule for Newton's method at iteration  $n$  is therefore

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \quad (5.32)$$

For the first iteration, a starting estimate  $x_0$  to approximate the solution is needed. For Eq. (5.29), we use  $x_0 = \sigma_{m_N}$  while a starting estimate for Eq. (5.31) is given by (Minka, 2012) considering  $y = \frac{1}{K} \sum_{k=1}^K \log(m_k^2/2\sigma_g^2)$ .

$$x_0 = \psi^{-1}(y) \approx \begin{cases} \exp(y) + 1/2 & \text{if } y \geq -2.22 \\ -1/(y + \psi(1)) & \text{if } y < -2.22 \end{cases} \quad (5.33)$$

In practice, we have observed that 5 iterations of Eq. (5.32) were sufficient to reach  $|x_n - x_{n-1}| < 10^{-13}$ .

### 5.7.2 Generalized bias correction

As an application that requires knowledge of both  $\sigma_g$  and  $N$ , we now present a general version for non integer values of  $N$  of the signal bias correction from Koay and Bassler (2006) and Koay, Özarlan, and Bassler (2009). The correction factor  $\xi(\eta|\sigma_g, N)$  can be used to obtain  $\eta$  from the magnitude measurement  $m_N$  given the values of  $\sigma_g$  and  $N$  such that

$$\xi(\eta|\sigma_g, N) = 2N + \frac{\eta^2}{\sigma_g^2} - \left( \beta_N {}_1F_1 \left( -1/2, N, \frac{-\eta^2}{2\sigma_g^2} \right) \right)^2 \quad (5.34)$$

where  ${}_1F_1$  is Kummer's function of the first kind. By defining

$$\beta_N = \sqrt{\pi/2} \binom{N-1/2}{1/2} \quad (5.35)$$

$$= \sqrt{\pi/2} \left( \frac{\Gamma(N+1/2)}{\Gamma(3/2)\Gamma(N)} \right) \quad (5.36)$$

$$= \sqrt{2} \left( \frac{\Gamma(N+1/2)}{\Gamma(N)} \right) \quad (5.37)$$

where  $\binom{n}{k}$  is a binomial coefficient, we obtain a generalized version of Eq. (5.34) which can now be applied for non integer values of  $N$ , such as in the case of a half Gaussian signal distribution ( $N = 0.5$ ) which occurs when employing half-Fourier reconstruction techniques (Dietrich et al., 2008). Estimation of  $\eta$  is finally done with

$$\eta = \sqrt{\hat{m}^2 + (\xi(\eta|\sigma_g, N) - 2N)\sigma_g^2} \quad (5.38)$$

where  $\hat{m}$  is an estimate of the first moment of a noncentral chi variable and is estimated from a spherical harmonics fit of order 6 in the present work. Eq. (5.38) can be solved iteratively w.r.t.  $\eta$  until convergence, see (Koay, Özarlan, and Bassler, 2009) for further implementation details.

### 5.7.3 Automated identification of noise only voxels

This appendix outlines the proposed algorithm and details for a practical implementation. Our implementation is also freely available at <https://github.com/samuelstjean/autodmri> (St-Jean, De Luca, Tax, et al., 2019) and will be a part of ExploreDTI (Leemans et al., 2009). The synthetic and acquired datasets used in this manuscript are also available (St-Jean, De Luca, Tax, et al., 2018).

---

**Algorithm 5.1:** Main algorithm to identify voxels belonging to the Gamma distribution
 

---

**Data:** 4D DWIs data, probability level  $p = 0.05$ , length of the search interval  $l = 50$ ,  $N_{min} = 1$ ,  
 $N_{max} = 12$

**Result:**  $\sigma_g$ ,  $N$ , mask of background only voxels

Compute the *median* of the whole dataset;

**foreach** 2D Slice of the 4D dataset **do**

    Compute the upper bound  $\sigma_{g_{max}} = \text{median} / \sqrt{2 \text{icdf}(N_{max}, 1/2)}$ ;

    Compute the search interval  $\Phi = [1\sigma_{g_{max}}/l, 2\sigma_{g_{max}}/l, \dots, l\sigma_{g_{max}}/l]$ ;

**while**  $\sigma_g, N$  not converged **do**

        Compute  $\lambda_- = \text{icdf}(\alpha, p/2)$  and  $\lambda_+ = \text{icdf}(\alpha, 1 - p/2)$ ;

**foreach**  $\sigma_{candidate} \in \Phi$  **do**

            Apply change of variable  $t = \frac{data^2}{2\sigma_{candidate}^2}$ ;

            Find voxels from the gamma distribution;

$\text{mask\_current} = \left( \lambda_- \leq \sum_{k=1}^K t_k \right) \cap \left( \sum_{k=1}^K t_k \leq \lambda_+ \right)$ ;

**if** number of voxels in  $\text{mask\_current} > \text{mask}$  **then**

                |  $\text{mask} = \text{mask\_current}$ ;

**end**

**end**

        Compute  $\sigma_g$  with the voxels inside the mask using Eq. (5.8) or Eq. (5.12);

        Compute  $N$  with the voxels inside the mask using Eq. (5.9) or Eq. (5.11);

        Set  $N_{min} = N_{max} = N$ ;

        Set  $\Phi = [0.95\sigma_g, 0.96\sigma_g, \dots, 1.05\sigma_g]$ ;

**end**

**end**

---

# Bibliography

- [1] S. Aja-Fernández, V. Brion, and A. Tristán-Vega. “Effective noise estimation and filtering from correlated multiple-coil MR data.” In: *Magnetic resonance imaging* 31.2 (Feb. 2013), pp. 272–85 (cit. on p. 112).
- [2] S. Aja-Fernández and A. Tristán-Vega. “A review on statistical noise models for Magnetic Resonance Imaging”. In: *LPI, ETSI Telecomunicacion, Universidad de Valladolid, Spain, Tech. Rep* TECH-LPI2013-01 (2015), pp. 1–30 (cit. on p. 118).
- [3] S. Aja-Fernández, G. Vegas-Sánchez-Ferrero, and A. Tristán-Vega. “Noise estimation in parallel MRI: GRAPPA and SENSE.” In: *Magnetic resonance imaging* 32.3 (Apr. 2014), pp. 281–90 (cit. on pp. 112, 114, 134).
- [4] M. Barth, F. Breuer, P. J. Koopmans, D. G. Norris, and B. A. Poser. “Simultaneous multislice (SMS) imaging techniques.” In: *Magnetic resonance in medicine* 75.1 (Jan. 2016), pp. 63–81 (cit. on p. 134).
- [5] M. A. Bernstein, K. F. King, and X. J. Zhou. *Handbook of MRI Pulse Sequences*. Elsevier Science, 2004, p. 1040 (cit. on p. 119).
- [6] E. Caruyer, A. Daducci, M. Descoteaux, J.-C. Houde, J.-P. Thiran, and R. Verma. “Phantoms: a flexible software library to simulate diffusion MR phantoms”. In: *International Symposium on Magnetic Resonance in Medicine (ISMRM’14)*. Vol. 17. 2014, p. 6407 (cit. on p. 118).
- [7] Q. Collier et al. “Diffusion kurtosis imaging with free water elimination: A bayesian estimation approach”. In: *Magnetic Resonance in Medicine* 80.2 (Aug. 2018), pp. 802–813 (cit. on pp. 112, 133).
- [8] C. D. Constantinides, E. Atalar, and E. R. McVeigh. “Signal-to-noise measurements in magnitude images from NMR phased arrays”. In: *Magnetic Resonance in Medicine* 38.5 (Nov. 1997), pp. 852–857 (cit. on p. 113).
- [9] M. Descoteaux, E. Angelino, S. Fitzgibbons, and R. Deriche. “Regularized, fast, and robust analytical Q-ball imaging”. In: *Magnetic Resonance in Medicine* 58.3 (Sept. 2007), pp. 497–510 (cit. on p. 121).
- [10] O. Dietrich, J. G. Raya, S. B. Reeder, M. Ingrisch, M. F. Reiser, and S. O. Schoenberg. “Influence of multichannel combination, parallel imaging and other reconstruction techniques on MRI noise characteristics.” In: *Magnetic resonance imaging* 26.6 (July 2008), pp. 754–62 (cit. on pp. 112–114, 119, 128, 140).
- [11] S. Duchesne et al. “The Canadian Dementia Imaging Protocol: Harmonizing National Cohorts”. In: *Journal of Magnetic Resonance Imaging* 49.2 (Feb. 2019), pp. 456–465 (cit. on p. 112).
- [12] M. J. Emaus et al. “MR Imaging as an Additional Screening Modality for the Detection of Breast Cancer in Women Aged 50–75 Years with Extremely Dense Breasts: The DENSE Trial Study Design”. In: *Radiology* 277.2 (Nov. 2015), pp. 527–537 (cit. on p. 112).
- [13] M. A. Griswold et al. “Generalized Autocalibrating Partially Parallel Acquisitions (GRAPPA)”. In: *Magnetic Resonance in Medicine* 47.6 (2002), pp. 1202–1210 (cit. on pp. 112, 118).
- [14] H. Gudbjartsson and S. Patz. “The rician distribution of noisy mri data”. In: *Magnetic Resonance in Medicine* 34.6 (Dec. 1995), pp. 910–914 (cit. on pp. 113, 133).

- 
- [15] R. M. Heidemann, A. Anwander, T. Feiweier, T. R. Knösche, and R. Turner. “k-space and q-space: Combining ultra-high spatial and angular resolution in diffusion imaging using ZOOPPA at 7T”. In: *NeuroImage* 60.2 (Apr. 2012), pp. 967–978 (cit. on p. 112).
- [16] S. J. Holdsworth, R. O’Halloran, and K. Setsompop. “The quest for high spatial resolution diffusion-weighted imaging of the human brain in vivo”. In: *NMR in Biomedicine* 32.4 (Apr. 2019), e4056 (cit. on p. 112).
- [17] E. B. Hutchinson et al. “Analysis of the effects of noise, DWI sampling, and value of assumed parameters in diffusion MRI models.” In: *Magnetic resonance in medicine* 78.5 (Nov. 2017), pp. 1767–1780 (cit. on p. 133).
- [18] C. G. Koay and P. J. Basser. “Analytically exact correction scheme for signal extraction from noisy magnitude MR signals”. In: *Journal of Magnetic Resonance* 179.2 (Apr. 2006), pp. 317–322 (cit. on p. 140).
- [19] C. G. Koay, E. Özarslan, and P. J. Basser. “A signal transformational framework for breaking the noise floor and its applications in MRI”. In: *Journal of Magnetic Resonance* 197.2 (Apr. 2009), pp. 108–119 (cit. on pp. 113, 114, 120, 140).
- [20] C. G. Koay, E. Özarslan, and C. Pierpaoli. “Probabilistic Identification and Estimation of Noise (PIESNO): A self-consistent approach and its applications in MRI”. In: *Journal of Magnetic Resonance* 199.1 (July 2009), pp. 94–103 (cit. on pp. 112–114, 116, 120).
- [21] B. Landman, P. L. Bazin, and J. Prince. “Diffusion tensor estimation by maximizing rician likelihood”. In: *Proceedings of the IEEE International Conference on Computer Vision* (2007), pp. 1–8 (cit. on pp. 113, 133).
- [22] A. Leemans, B. Jeurissen, J. Sijbers, and D. Jones. “ExploreDTI: a graphical toolbox for processing, analyzing, and visualizing diffusion MR data”. In: *Proceedings 17th Scientific Meeting, International Society for Magnetic Resonance in Medicine* 17.2 (2009), p. 3537 (cit. on p. 140).
- [23] K. Lønning, P. Putzky, J.-J. Sonke, L. Reneman, M. W. Caan, and M. Welling. “Recurrent inference machines for reconstructing heterogeneous MRI data”. In: *Medical Image Analysis* 53 (Apr. 2019), pp. 64–78 (cit. on p. 133).
- [24] M. Lustig, D. Donoho, and J. M. Pauly. “Sparse MRI: The application of compressed sensing for rapid MR imaging”. In: *Magnetic Resonance in Medicine* 58.6 (Dec. 2007), pp. 1182–1195 (cit. on p. 112).
- [25] K. H. Maier-Hein et al. “The challenge of mapping the human connectome based on diffusion tractography”. In: *Nature Communications* 8.1 (Dec. 2017), p. 1349 (cit. on p. 118).
- [26] T. P. Minka. “Estimating a Dirichlet distribution”. In: *Microsoft research* (2012) (cit. on p. 139).
- [27] H. Mirzaalian et al. “Multi-site harmonization of diffusion MRI data in a registration framework”. In: *Brain Imaging and Behavior* 12.1 (Feb. 2018), pp. 284–295 (cit. on p. 112).
- [28] S. Moeller et al. “Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI”. In: *Magnetic Resonance in Medicine* 63.5 (May 2010), pp. 1144–1153 (cit. on p. 112).
- [29] P. F. Neher, F. B. Laun, B. Stieltjes, and K. H. Maier-Hein. “Fiberfox: Facilitating the creation of realistic white matter software phantoms”. In: *Magnetic Resonance in Medicine* 72.5 (Nov. 2014), pp. 1460–1470 (cit. on p. 118).
- [30] D. C. Noll, D. G. Nishimura, and A. Macovski. “Homodyne Detection in Magnetic Resonance Imaging”. In: *IEEE Transactions on Medical Imaging* 10.2 (1991), pp. 154–163 (cit. on pp. 112, 119).

## Bibliography

---

- [31] R. G. Nunes, J. V. Hajnal, X. Golay, and D. J. Larkman. “Simultaneous slice excitation and reconstruction for single shot EPI”. In: *Proc. Intl. Soc. Mag. Reson. Med* 13.2 (2006), p. 293 (cit. on p. 112).
- [32] A. Papoulis. *Probability, random variables, and stochastic processes*. Tata McGraw-Hill Education, 1991 (cit. on p. 137).
- [33] M. Paquette, S. Merlet, G. Gilbert, R. Deriche, and M. Descoteaux. “Comparison of sampling strategies and sparsifying transforms to improve compressed sensing diffusion spectrum imaging”. In: *Magnetic Resonance in Medicine* 73.1 (Jan. 2015), pp. 401–416 (cit. on p. 112).
- [34] T. Pieciak, I. Rabanillo-Viloria, and S. Aja-Fernandez. “Bias correction for non-stationary noise filtering in MRI”. In: *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE, Apr. 2018, pp. 307–310 (cit. on p. 133).
- [35] R. A. Poldrack et al. “Long-term neural and physiological phenotyping of a single human”. In: *Nature Communications* 6.1 (Dec. 2015), p. 8885 (cit. on p. 119).
- [36] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger. “SENSE: Sensitivity encoding for fast MRI”. In: *Magnetic Resonance in Medicine* 42.5 (Nov. 1999), pp. 952–962 (cit. on pp. 112, 118).
- [37] K. Sakaie and M. Lowe. “Retrospective correction of bias in diffusion tensor imaging arising from coil combination mode”. In: *Magnetic Resonance Imaging* 37 (Apr. 2017), pp. 203–208 (cit. on pp. 112, 113, 133).
- [38] K. Sakaie, X. Zhou, J. Lin, J. Debbins, M. Lowe, and R. J. Fox. “Technical Note: Retrospective reduction in systematic differences across scanner changes by accounting for noise floor effects in diffusion tensor imaging”. In: *Medical Physics* 45.9 (Sept. 2018), pp. 4171–4178 (cit. on p. 112).
- [39] S. N. Sotiropoulos et al. “Effects of image reconstruction on fiber orientation mapping from multi-channel diffusion MRI: Reducing the noise floor using SENSE”. In: *Magnetic Resonance in Medicine* 70.6 (2013), pp. 1682–1689 (cit. on p. 114).
- [40] S. St-Jean, P. Coupé, and M. Descoteaux. “Non Local Spatial and Angular Matching: Enabling higher spatial resolution diffusion MRI datasets through adaptive denoising”. In: *Medical Image Analysis* 32.2016 (Aug. 2016), pp. 115–130 (cit. on pp. 113, 121, 136).
- [41] S. St-Jean, A. De Luca, C. M. W. Tax, M. A. Viergever, and A. Leemans. *Datasets for 'Automated characterization of noise distributions in diffusion MRI data'*. 2018.  
URL: <https://zenodo.org/record/2483105> (cit. on pp. 118, 119, 140).
- [42] S. St-Jean, A. De Luca, C. M. W. Tax, M. A. Viergever, and A. Leemans. *samuelstjean/autodmri: First release - 2019-07-17*. July 2019.  
URL: <https://doi.org/10.5281/zenodo.3339158> (cit. on pp. 116, 140).
- [43] S. St-Jean, A. De Luca, M. A. Viergever, and A. Leemans. “Automatic, Fast and Robust Characterization of Noise Distributions for Diffusion MRI”. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*. Ed. by A. F. Frangi, J. A. Schnabel, C. Davatzikos, C. Alberola-López, and G. Fichtinger. Springer International Publishing, May 2018, pp. 304–312 (cit. on pp. 113, 116, 119, 135).
- [44] P. Storey et al. “Partial k-space reconstruction in single-shot diffusion-weighted echo-planar imaging”. In: *Magnetic Resonance in Medicine* 57.3 (Mar. 2007), pp. 614–619 (cit. on p. 112).
- [45] K. Tabelow, H. U. Voss, and J. Polzehl. “Local estimation of the noise level in MRI using structural adaptation”. In: *Medical Image Analysis* 20.1 (Feb. 2015), pp. 76–86 (cit. on pp. 112, 119, 120).
- [46] C. M. W. Tax et al. “Cross-scanner and cross-protocol diffusion MRI data harmonisation: A benchmark database and evaluation of algorithms”. In: *NeuroImage* 195 (July 2019), pp. 285–299 (cit. on pp. 112, 119).

- 
- [47] H. C. S. Thom. “A Note on the Gamma Distribution”. In: *Monthly Weather Review* 86.4 (Apr. 1958), pp. 117–122 (cit. on pp. 115, 139).
- [48] R. J. Tibshirani and J. Taylor. “The solution path of the generalized lasso”. In: *The Annals of Statistics* 39.3 (June 2011), pp. 1335–1371 (cit. on p. 136).
- [49] N. Todd, S. Moeller, E. J. Auerbach, E. Yacoub, G. Flandin, and N. Weiskopf. “Evaluation of 2D multiband EPI imaging for high-resolution, whole-brain, task-based fMRI studies at 3T: Sensitivity and slice leakage artifacts”. In: *NeuroImage* 124.Pt A (Jan. 2016), pp. 32–42 (cit. on p. 134).
- [50] J. Veraart, E. Fieremans, and D. S. Novikov. “Diffusion MRI noise mapping using random matrix theory”. In: *Magnetic Resonance in Medicine* 76.5 (Nov. 2016), pp. 1582–1593 (cit. on pp. 112, 120).
- [51] E. W. Weisstein. *Gamma Distribution*. From *MathWorld—A Wolfram Web Resource*. 2017. URL: <http://mathworld.wolfram.com/GammaDistribution.html> (visited on [visited on 10/09/2017]) (cit. on pp. 115, 137).
- [52] H. Zhang, T. Schneider, C. A. Wheeler-Kingshott, and D. C. Alexander. “NODDI: Practical in vivo neurite orientation dispersion and density imaging of the human brain”. In: *NeuroImage* 61.4 (July 2012), pp. 1000–1016 (cit. on pp. 113, 133).

*This page intentionally left blank*

# 6

## Harmonization of diffusion MRI datasets with adaptive dictionary learning

Based on

---

Samuel St-Jean, Max A. Viergever and Alexander Leemans  
Harmonization of diffusion MRI datasets with adaptive dictionary learning  
Submitted to Neuroimage and available at <https://arxiv.org/abs/1910.00272>.

## Abstract

Diffusion weighted magnetic resonance imaging is a noninvasive imaging technique that can indirectly infer the microstructure of tissues and provide metrics which are subject to normal variability across subjects. Potentially abnormal values or features may yield essential information to support analysis of controls and patients cohorts, but subtle confounds affecting diffusion MRI, such as those due to difference in scanning protocols or hardware, can lead to systematic errors which could be mistaken for purely biologically driven variations amongst subjects. In this work, we propose a new harmonization algorithm based on adaptive dictionary learning to mitigate the unwanted variability caused by different scanner hardware while preserving the natural biological variability present in the data. Overcomplete dictionaries, which are learned automatically from the data and do not require paired samples, are then used to reconstruct the data from a different scanner, removing variability present in the source scanner in the process. We use the publicly available database from an international challenge to evaluate the method, which was acquired on three different scanners and with two different protocols, and propose a new mapping towards a scanner agnostic space. Results show that the effect size of the four studied diffusion metrics is preserved while removing variability attributable to the scanner. Experiments with alterations using a free water compartment, which is not simulated in the training data, shows that the effect size induced by the alterations is also preserved after harmonization. The algorithm is freely available and could help multicenter studies in pooling their data, while removing scanner specific confounds, and increase statistical power in the process.

**Keywords:** Diffusion MRI, Harmonization, Scanner variability, Dictionary learning, Cross-validation, Akaike information criterion

## 6.1 Introduction

Diffusion weighted magnetic resonance imaging (dMRI) is a noninvasive imaging technique that can indirectly infer the microstructure of tissues based on the displacement of water molecules. As dMRI only offers an indirect way to study, e.g. the brain microstructure, analysis of dMRI datasets includes multiple processing steps to ensure adequate correction of acquisition artifacts due to subject motion or eddy current induced distortions, amongst others (Tournier et al., 2011). Quantitative scalar measures of diffusion can be extracted from the acquired datasets, such as the apparent diffusion coefficient (ADC) or fractional anisotropy (FA) as computed from diffusion tensor imaging (DTI) (Basser, Mattiello, et al., 1994; Basser and Pierpaoli, 1996), with a plethora of other measures and diffusion models nowadays available (Assemlal et al., 2011; Tournier, 2019). These measures are subject to normal variability across subjects and potentially abnormal values or features extracted from dMRI datasets may yield essential information to support analysis of controls and patients cohorts (Johansen-Berg and Behrens, 2009; Jones, 2011).

As small changes in the measured signal are ubiquitous due to differences in scanner hardware (Sakaie et al., 2018), software versions of the scanner or processing tools (Gronenschild et al., 2012; Sakaie et al., 2018), field strength of the magnet (Huisman et al., 2006) or reconstruction methods in parallel MRI and accelerated imaging (Dietrich et al., 2008; St-Jean, De Luca, et al., 2018), non-negligible effects may translate into small differences in the subsequently computed diffusion metrics. Subtle confounds affecting dMRI can even be due to measuring at different time points in the cardiac cycle, leading to changes in the measured values of pseudo-diffusion over the cardiac cycle (De Luca et al., 2019; Federau et al., 2013). In the presence of disease, these small variations in the measured signal are entangled in the genuine biological variability, which is usually the main criterion of interest to discover or analyze subsequently. This can lead to confounding effects and systematic errors that could be mistaken for purely biologically driven variations amongst subjects. To mitigate these issues, large-scale studies try to harmonize their acquisition protocols across centers to further reduce these potential sources of variability (Duchesne et al., 2019) or may only use a single scanner without upgrading it for long term studies (Hofman, Grobbee, et al., 1991; Hofman, Brusselle, et al., 2015). The stability brought by keeping the same scanning hardware is however at the cost of potentially missing on improved, more efficient sequences or faster scanning methods becoming common in MRI (Feinberg et al., 2010; Larkman et al., 2001; Lustig et al., 2007). Even by carefully controlling all these sources of variability as much as possible, there still remain reproducibility issues between scanners of the same model or in scan-rescan studies of dMRI metrics (Kristo et al., 2013; Magnotta et al., 2012; Vollmar et al., 2010). Over the years, many algorithms have been developed to mitigate the variability attributed to non-biological effects in dMRI, e.g. in order to combine datasets from multiple studies and increase statistical power, see e.g. (Tax et al., 2019; Zhu et al., 2019) for reviews. Common approaches consist in harmonizing the raw dMRI datasets themselves (Cetin Karayumak et al., 2019; Mirzaalian et al., 2016)

or the computed scalar metrics (Fortin et al., 2017; Pohl et al., 2016) to reduce variability between scanners. Recently, a dMRI benchmark database containing ten training subjects and four test subjects datasets acquired on three scanners with two acquisition protocols was presented at the computational diffusion MRI (CDMRI) 2017 challenge (Tax et al., 2019). The publicly available CDMRI database was previously used to compare five harmonization algorithms, including a previous version of the algorithm we present here, which we use for evaluation.

In this work, we propose a new algorithm based on adaptive dictionary learning to mitigate the unwanted variability caused by different scanner hardware while preserving the natural biological variability present in the data. Expanding upon the methodology presented in St-Jean, Coupé, et al. (2016) and St-Jean, Viergever, et al. (2017), overcomplete dictionaries are learned automatically from the data for a given target scanner with an automatic tuning of the regularization parameter. These dictionaries are then used to reconstruct the data from a different source scanner, removing variability present in the source scanner in the process. Mapping across different spatial resolutions can be obtained by adequate subsampling of the dictionary. Additional experiments beyond the original challenge show that the harmonization algorithm preserves alterations made on the test subjects while removing scanner variability, but without altering the training datasets, by mapping all the datasets towards a global “scanner space”. The algorithm does *not* require paired datasets for training, making it easy to apply for hard to acquire datasets (e.g. patients with Alzheimer’s, Parkinson’s or Huntington’s disease) or when pooling datasets from unrelated studies that are acquired in separate centers. This makes our proposed method readily applicable for pre-existing and ongoing studies that would like to remove variability caused by non-biological or systematic effects in their data analyzes.

## 6.2 Theory

### 6.2.1 The dictionary learning algorithm

Dictionary learning (Elad and Aharon, 2006; Mairal, Bach, Ponce, and Sapiro, 2010) aims to find a set of basis elements to efficiently approximate a given set of input vectors. This formulation optimizes both the representation  $\mathbf{D}$  (called the dictionary or the set of atoms) and the coefficients  $\alpha$  of that representation (called the sparse codes) as opposed to using a fixed basis (e.g. Fourier, wavelets, spherical harmonics). A dictionary can be chosen to be overcomplete (i.e. more column than rows) as the algorithm is designed to only select a few atoms to approximate the input vector with a penalization on the  $\ell_1$ -norm of  $\alpha$  to promote a sparse solution. Applications in computer vision with the goal to reduce visual artifacts include demosaicking (Mairal, Bach, Ponce, Sapiro, and Zisserman, 2009), inpainting (Mairal, Bach, Ponce, and Sapiro, 2010) and upsampling (Yang, Wang, et al., 2012; Yang, Wright, et al., 2010) amongst others.

In practice, local windows are used to extract spatial and angular neighborhoods of diffusion weighted images (DWIs) to create the set of vectors required for dictionary learning as in St-Jean, Coupé, et al. (2016). This is done by first extracting a small 3D region from a single DWI, which we now refer to as a patch. To include angular information, a set of patches is taken at the same spatial location across DWIs in an angular neighborhood (as defined by the angle between their associated b-vector on the sphere). This considers that patches from different DWIs at the same spatial location, but which are in fact not too far on the sphere, exhibit self-similarity that can be exploited by dictionary learning. Once this process is done, every set of patches is concatenated to a single vector  $\mathbf{X}$ . All of these vectors  $\mathbf{X}_n$  are then put in a 2D matrix  $\Omega = \{\mathbf{X}_1, \dots, \mathbf{X}_n, \dots\}$ , where  $n$  denotes one of the individual set of patches.

Once the set of patches  $\Omega$  has been extracted,  $\mathbf{D}$  can be initialized by randomly selecting  $N$  vectors from  $\Omega$  (Mairal, Bach, Ponce, and Sapiro, 2010). With this initial overcomplete dictionary, a sparse vector  $\alpha_n$  can be computed for each  $\mathbf{X}_n$  such that  $\mathbf{D}$  is a good approximation to reconstruct  $\mathbf{X}_n$ , that is  $\mathbf{X}_n \approx \mathbf{D}\alpha_n$ . This initial approximation can be refined iteratively by sampling randomly  $N$  new vectors  $\mathbf{X}_n \in \Omega$  and updating  $\mathbf{D}$  to better approximate those vectors. At the next iteration, a new set  $\mathbf{X}_n \in \Omega$  is randomly drawn and  $\mathbf{D}$  is updated to better approximate this new set of vectors. This iterative process can be written as

$$\arg \min_{\mathbf{D}, \alpha} \frac{1}{N} \sum_{n=1}^N \left( \frac{1}{2} \|\mathbf{X}_n - \mathbf{D}\alpha_n\|_2^2 + \lambda_i \|\alpha_n\|_1 \right) \text{ s.t. } \|\mathbf{D}_{\cdot p}\|_2^2 = 1 \quad (6.1)$$

with  $\alpha_n \in \mathbb{R}^{p \times 1}$  an array of sparse coefficients and  $\mathbf{D}$  the dictionary where each column is constrained to unit  $\ell_2$ -norm to prevent degenerated solutions.  $\lambda_i$  is a regularization parameter used at iteration  $i$  (which is further detailed in Section 6.2.2) to balance the  $\ell_2$ -norm promoting data similarity and the  $\ell_1$ -norm promoting sparsity of the coefficients  $\alpha_n$ . Iterative updates using Eq. (6.1) alternate between refining  $\mathbf{D}$  (and holding  $\alpha$  fixed) and computing  $\alpha$  (with  $\mathbf{D}$  held fixed) for the current set of  $\mathbf{X}_n$ . As updating  $\alpha$  needs an optimization scheme, this can be done independently for each  $\alpha_n$  using coordinate descent (Friedman et al., 2010). For updating  $\mathbf{D}$ , we use the parameter-free closed form update from Mairal, Bach, Ponce, and Sapiro (2010, Algorithm 2), which only requires storing intermediary matrices of the previous iteration using  $\alpha$  and  $\mathbf{X}_n$  to update  $\mathbf{D}$ . Building dictionaries for the task at hand has been used previously in the context of diffusion MRI for denoising (Gramfort et al., 2014; St-Jean, Coupé, et al., 2016) and compressed sensing (Gramfort et al., 2014; Merlet et al., 2013; Schwab et al., 2018) amongst other tasks. Note that it is also possible to design dictionaries based on products of fixed basis or adding additional constraints such as positivity or spatial consistency to Eq. (6.1), see e.g. (Schwab et al., 2018; Vemuri et al., 2019) and references therein for examples pertaining to diffusion MRI.

## 6.2.2 Automatic regularization selection

Eq. (6.1) relies on a regularization term  $\lambda_i$  which can be different for each set of vectors  $\mathbf{X}_n$  at iteration  $i$ . It is, however, common to fix  $\lambda_i$  for all  $\mathbf{X}_n$  depending on some heuristics such as the size of  $\mathbf{X}_n$  (Mairal, Bach, Ponce, and Sapiro, 2010), the local noise variance (St-Jean, Coupé, et al., 2016) or through a grid search (Gramfort et al., 2014). In the present work, a search through a sequence of candidates  $\{\lambda_0, \dots, \lambda_s, \dots, \lambda_{\text{last}}\}$ , which is automatically determined for each individual  $\mathbf{X}_n$ , is instead employed using either 3-fold cross-validation (CV) and minimizing the mean squared error or by minimizing the Akaike information criterion (AIC) (Akaike, 1974; Zou et al., 2007). For the AIC, the number of non-zero coefficients in  $\alpha_n$  provides an unbiased estimate of degrees of freedom for the model (Tibshirani and Taylor, 2012; Zou et al., 2007). We use the AIC for normally distributed errors in least-squares problems from Burnham and Anderson (2004)

$$AIC_{\lambda_i} = \arg \min_{\lambda_s} m \log \left( \frac{\|\mathbf{X}_n - \mathbf{D}\alpha_{\lambda_s}\|_2^2}{m} \right) + 2\text{df}(\alpha_{\lambda_s}) \quad (6.2)$$

with  $m$  the number of elements of  $\mathbf{X}_n$ . In practice, this sequence of  $\lambda_s$  is chosen automatically on a log scale starting from  $\lambda_0$  (providing the null solution  $\alpha_{\lambda_0} = 0$ ) up to  $\lambda_{\text{last}} = \epsilon > 0$  (providing the regular least squares solution) (Friedman et al., 2010). The solution  $\alpha_n$  at  $\lambda_s$  is then used as a starting estimate for the next value of  $\lambda_{s+1}$ . The process can be terminated early if the cost function Eq. (6.1) does not change much (e.g. the difference between the solution at  $\lambda_s$  and  $\lambda_{s+1}$  is below  $10^{-5}$ ) for decreasing values of  $\lambda_s$ , preventing computation of similar solutions.

## 6.3 Methods

### 6.3.1 Building an optimal representation across scanners

For harmonization based on dictionary learning, all 3D patches of small spatial and angular local neighborhoods inside a brain mask were extracted from the available training datasets for a given scanner as done in (St-Jean, Coupé, et al., 2016; Tax et al., 2019). Since different patch sizes are used depending on the reconstruction task, Sections 6.3.2 and 6.3.5 detail each case that we study in this manuscript. Only patches present inside a brain mask were used for computation and reconstruction. These patches were reorganized as column arrays  $\Omega = \{\mathbf{X}_1, \dots, \mathbf{X}_n, \dots\}$  with each  $\mathbf{X}_n \in \mathbb{R}^{m \times 1}$  represented as vectors of size  $m$ . Each volume was mean subtracted and each patch  $\mathbf{X}_n$  was scaled to have unit variance (Friedman et al., 2010). Subsequently, features were automatically created from the target scanner datasets using dictionary learning as detailed in Section 6.2.1. A dictionary  $\mathbf{D} \in \mathbb{R}^{m \times p}$  was initialized with  $p$  vectors  $\mathbf{X}_{m \times 1} \in \Omega$  randomly chosen, where  $\mathbf{D}$  is set to have twice

as many columns as rows (i.e.  $p = 2m$ ). Updates using Eq. (6.1) were carried for 500 iterations using a batchsize of  $N = 32$ . The coefficients  $\alpha_n$  were unscaled afterwards.

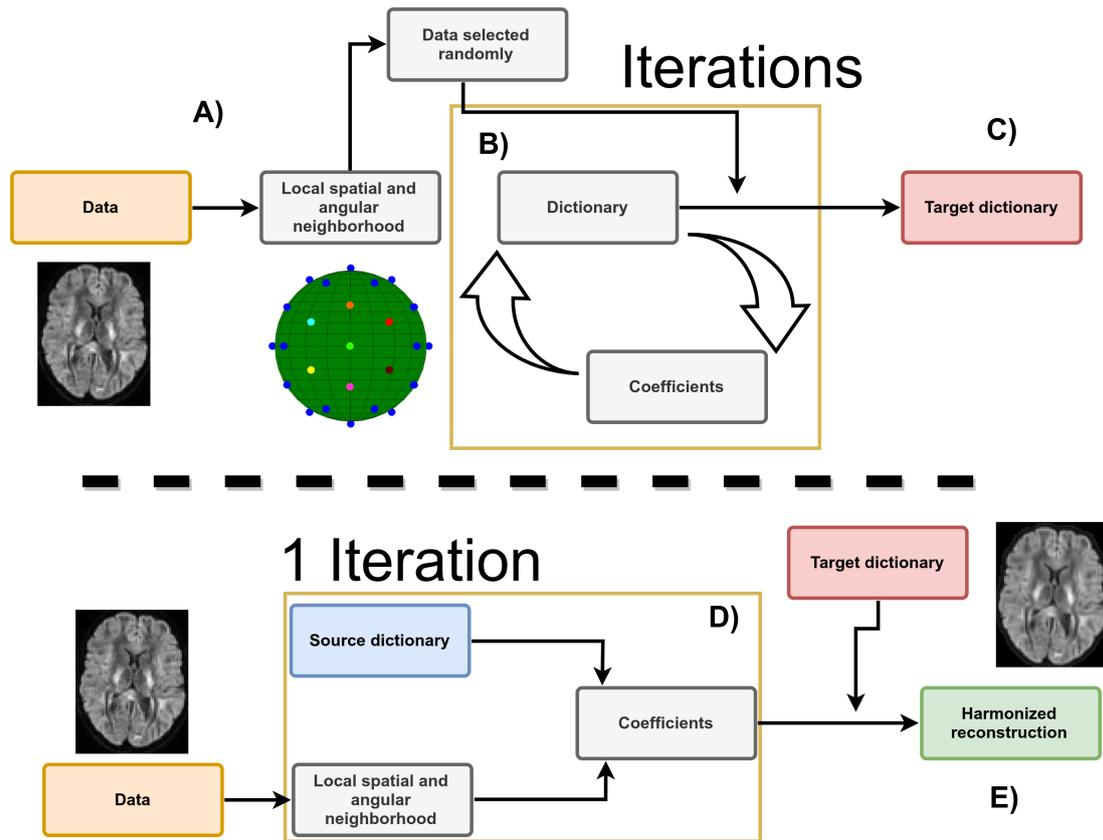
Once a dictionary  $\mathbf{D}$  has been computed, the new, harmonized representation (possibly from a *different* scanner) can be obtained by computing  $\alpha_n$  for every  $\mathbf{X}_n \in \Omega$ . As  $\mathbf{D}$  was created to reconstruct data from a chosen target scanner, it contains generic features tailored to this specific target scanner that are not necessarily present in the set of patches  $\Omega$  extracted from a different scanner. As such, reconstruction using  $\mathbf{D}_{\text{target}}$  created from  $\Omega_{\text{target}}$  can be used to map  $\Omega_{\text{source}}$  towards  $\Omega_{\text{target}}$ , that is  $\mathbf{X}_{n_{\text{harmonized}}} = \mathbf{D}_{\text{target}}\alpha_n$  by using  $\mathbf{X}_{n_{\text{source}}}$  and holding  $\mathbf{D}_{\text{target}}$  fixed while solving Eq. (6.1) for  $\alpha_n$ . These specially designed features from  $\Omega_{\text{target}}$  are not necessarily present in  $\Omega_{\text{source}}$ , therefore eliminating the source scanner specific effects, as they are not contained in  $\mathbf{D}_{\text{target}}$ .

Downsampling  $\mathbf{D}_{\text{target}}$  into  $\mathbf{D}_{\text{small}}$  can also be used to reconstruct data at a different resolution than initially acquired by creating an implicit mapping between two different spatial resolutions. This is done by finding the coefficients  $\alpha$  by holding  $\mathbf{D}_{\text{small}}$  fixed when solving Eq. (6.1), but using  $\mathbf{D}_{\text{target}}$  for the final reconstruction such that  $\mathbf{X}_{n_{\text{harmonized}}} = \mathbf{D}_{\text{target}}\alpha_n$ . This reconstruction with the full sized dictionary provides an upsampled version of  $\mathbf{X}_n$ , the implicit mapping being guaranteed by sharing the same coefficients  $\alpha_n$  for both reconstructions. A similar idea has been exploited previously for the 3D reconstruction of T1w images by Rueda et al. (2013) and in diffusion MRI by St-Jean, Viergever, et al. (2017) in the context of single image upsampling. The general reconstruction process for the harmonization of datasets between scanners is illustrated in Fig. 6.1. Our implementation of the harmonization algorithm is detailed in Section 6.7 and also available in both source form and as a Docker container<sup>1</sup> (St-Jean, Viergever, et al., 2019).

### 6.3.2 Reconstruction tasks of the challenge

For the reconstruction in task 1 (matched resolution scanner-to-scanner mapping), the dictionary  $\mathbf{D}_{\text{target}}$  was created using patches of size  $3 \times 3 \times 3$  with 5 angular neighbors and one  $b = 0$  s/mm<sup>2</sup> image in each block. Optimization for constructing  $\mathbf{D}_{\text{target}}$  with Eq. (6.1) was performed using 3-fold CV and reconstruction of the final harmonized datasets was done with either CV or minimizing the AIC with Eq. (6.2) in two separate experiments. The datasets from the GE scanner were reconstructed using the dictionary built from the Prisma or Connectom scanner datasets for their respective harmonization task. For the reconstruction in task 2 (spatial and angular resolution enhancement), patches of different spatial sizes were extracted from the images at higher resolution (patches of size  $5 \times 5 \times 5$  for the Prisma scanner and  $6 \times 6 \times 6$  for the Connectom scanner) and used for the dictionary learning algorithm as described in Section 6.2.1. Under the hypothesis that a larger patch is a good representation for its lower resolution counterpart when downsampled, each column of the optimized dictionary  $\mathbf{D}_{\text{target}}$  was resized to a spatial dimension of  $3 \times 3 \times 3$  and the

<sup>1</sup><https://github.com/samuelstjean/harmonization>



**Figure 6.1:** Schematic representation of the harmonization between scanners with adaptive dictionary learning. **A)** Local patches are decomposed into vectors  $X_n$  and a random subset is used to initialize the dictionary  $D$ . **B)** A new set of patches is drawn at every iteration and the dictionary is refined iteratively by alternating updates for the coefficients  $\alpha$  and the dictionary  $D$  using Eq. (6.1). **C)** After a set number of iterations, this target dictionary  $D$  can now be used to reconstruct data from a potentially different dataset. **D)** A set of coefficients is computed for each patch  $X_n$  of the input dataset with a source dictionary. For harmonization tasks, the source and target dictionary from step **C)** are identical. For upsampling tasks, the source dictionary is a downsampled version of the target dictionary. **E)** The harmonized reconstruction for each patch  $X_n$  is obtained by multiplying the target dictionary  $D$  and the coefficients  $\alpha_n$ .

coefficients  $\alpha$  computed for this lower resolution dictionary  $D_{\text{small}}$ . The patches were finally reconstructed by multiplying the original dictionary  $D_{\text{target}}$  with the coefficients  $\alpha$ . This creates a set of upsampled patches from the GE scanner that are both harmonized and at the same spatial resolution as either the Prisma or the Connectom datasets. All reconstruction tasks were computed overnight on our computing server using 100 cores running at 2.1 GHz. On a standard desktop with a 4 cores 3.5 GHz processor, rebuilding one dataset took approximately two hours and 30 minutes with the AIC criterion.

### 6.3.3 Evaluation framework of the challenge

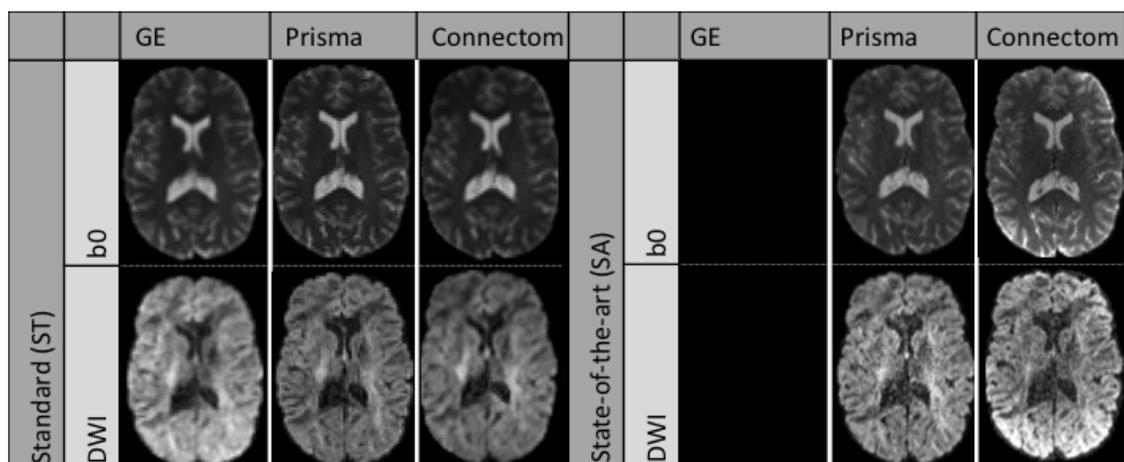
The original challenge requested from the participants to match the original gradient directions from the source to the target datasets and evaluated various scalar metrics on the diffusion weighted images. In our original submission, this matching was done with the truncated spherical harmonics (SH) basis of order 6 (Descoteaux et al., 2007) on the source dataset and sampling the basis at the gradient directions from the target scanner. In the present manuscript, we chose instead to evaluate the metrics directly in the original gradient directions as they are rotationally invariant, saving one interpolation step in the process as it could potentially introduce unwanted blurring of the data. The metrics used in the original evaluation were the apparent diffusion coefficient (ADC) and the fractional anisotropy (FA) from diffusion tensor imaging (DTI) and the rotationally invariant spherical harmonic (RISH) features of order 0 (RISH 0) and order 2 (RISH 2) of the SH basis, see Tax et al. (2019) for additional details. As our evaluation framework is slightly different, we compare our new approach with our initial version of the harmonization algorithm and with a baseline reference prediction created by trilinear interpolation from the source to the target scanner in the spirit of the original challenge.

### 6.3.4 Datasets and experiments

We used the datasets from the MICCAI 2017 harmonization challenge (Tax et al., 2019), consisting of ten training subjects and four test subjects acquired on three different scanners (GE, Siemens Prisma and Siemens Connectom) using different gradient strength (40 mT/m, 80 mT/m and 300 mT/m, respectively) with two acquisition protocols. Experiments are only reported for the four test subjects, which are later on denoted as subjects 'H', 'L', 'M' and 'N'. The standard protocol (ST) consists of 30 DWIs acquired at 2.4 mm isotropic with a b-value of  $b = 1200 \text{ s/mm}^2$ , 3  $b = 0 \text{ s/mm}^2$  images for the GE datasets, 4  $b = 0 \text{ s/mm}^2$  images for the Siemens datasets and  $TE = 98 \text{ ms}$ . Note that the TR is cardiac gated for the GE datasets while the Siemens datasets both use  $TR = 7200 \text{ ms}$ . The state-of-the-art (SA) protocol for the Siemens scanners contains 60 DWIs with a b-value of  $b = 1200 \text{ s/mm}^2$  and 5  $b = 0 \text{ s/mm}^2$  images. The Prisma datasets were acquired with a spatial resolution of 1.5 mm isotropic and  $TE / TR = 80 \text{ ms} / 4500 \text{ ms}$ . The Connectom datasets were acquired with a spatial resolution of 1.2 mm isotropic and  $TE / TR = 68 \text{ ms} / 5400 \text{ ms}$ . Most of the acquisition parameters were shared for the SA protocol which are listed in Table 6.1 with full details of the acquisition available in Tax et al. (2019). Standard preprocessing includes motion correction, EPI distortions corrections and image registration for each subject across scanners. The SA protocols were additionally corrected for gradient nonlinearity distortions. These datasets are available upon request<sup>2</sup> from the organizers. Fig. 6.2 shows an example of the acquired datasets for a single subject.

---

<sup>2</sup><https://www.cardiff.ac.uk/cardiff-university-brain-research-imaging-centre/research/projects/cross-scanner-and-cross-protocol-diffusion-MRI-data-harmonisation>



**Figure 6.2:** Example  $b = 0$  s/mm<sup>2</sup> images (top row) and  $b = 1200$  s/mm<sup>2</sup> images (bottom row) for a single subject acquired on the three scanners after preprocessing. The standard protocol (ST) is shown on the left and the state-of-the-art protocol (SA) is shown on the right. Note that the challenge asked participants to harmonize the GE ST protocol towards the two other scanners, but no SA protocol is available for the GE scanner. The figure is adapted from Tax et al. (2019), available under the CC-BY 4.0 license.

Scanner	GE 40 mT/m	Siemens Prisma 80 mT/m		Siemens Connectom 300 mT/m	
Protocol	Standard (ST)	Standard (ST)	State-of-the-art (SA)	Standard (ST)	State-of-the-art (SA)
Sequence	TRSE	PGSE	PGSE	PGSE	PGSE
# directions per b-value	30	30	60	30	60
TE [ms]	89	89	80	89	68
TR [ms]	Cardiac gated	7200	4500	7200	5400
$\Delta/\delta$ [ms]		41.4/26.0	38.3/19.5	41.8/28.5	31.1/8.5
$\delta_1 = \delta_4/\delta_2 = \delta_3$ [ms]	11.23/17.84				
Acquired voxel size [mm <sup>3</sup> ]	2.4 x 2.4 x 2.4	2.4 x 2.4 x 2.4	1.5 x 1.5 x 1.5	2.4 x 2.4 x 2.4	1.2 x 1.2 x 1.2
Reconstructed voxel size	1.8 x 1.8 x 2.4	1.8 x 1.8 x 2.4	1.5 x 1.5 x 1.5	1.8 x 1.8 x 2.4	1.2 x 1.2 x 1.2
SMS factor	1	1	3	1	2
Parallel imaging	ASSET 2	GRAPPA 2	GRAPPA 2	GRAPPA 2	GRAPPA 2
Bandwidth [Hz/Px]	3906	2004	1476	2004	1544
Partial Fourier	5/6	–	6/8	6/8	6/8
Coil combine		Adaptive combine	Sum of Squares	Adaptive combine	Adaptive combine
Head coil	8 channel	32 channel	32 channel	32 channel	32 channel

**Table 6.1:** Acquisition parameters of the datasets for the three different scanners. TRSE: twice-refocused spin-echo, PGSE: pulsed-gradient spin-echo. The table is adapted from Tax et al. (2019).

### 6.3.5 Simulations beyond the challenge

To further make our proposed harmonization algorithm widely applicable, we designed additional experiments beyond the challenge to harmonize data towards a common scanner space. As the MICCAI challenge focused on harmonization of datasets from a source scanner to a target scanner, the organizers essentially provided matching datasets of all subjects across all scanners. This data collection would be appropriate, for example, in a

longitudinal study design with scanner hardware upgrades during the study and subsequent data analysis. However, such an experimental setup might not be available in practice when harmonizing datasets from multiple centers or studies where data collection is done only once per subject e.g. to reduce costs associated with scan time or reduce traveling of the participants.

The additional experiments consist of harmonizing all the datasets from the ST protocol at once and predicting their harmonized version using this common basis instead of creating one dictionary per scanner and per protocol. To ensure that the scanner effects are properly removed, the test datasets were also altered in a small region with a simulated free water compartment as described in [Section 6.3.6](#). As these altered datasets were never “seen” by the harmonization algorithm, we can now quantify if the induced effects are properly reconstructed, as they were not present in the training set in the first place. This experiment is similar to creating a common space on a larger set of healthy subjects and finally harmonizing data from the remaining healthy subjects and “patients” towards this common space. In our current setup, the harmonization algorithm is not aware that the datasets are in fact from matched subjects and, by design, could also be used on unpaired training datasets.

### 6.3.6 Alterations of the original datasets

To create the altered version of the test datasets, a region of 3000 voxels ( $15 \times 20 \times 10$  voxels) in the right hemisphere was selected at the same spatial location in image space. Every voxel in the selected region was separately affected by a free water compartment to mimic infiltration of edema according to

$$S_{b_{\text{altered}}} = S_b + fS_0 \exp(-bD_{\text{csf}}) \quad (6.3)$$

with  $S_{b_{\text{altered}}}$  the new signal in the voxel,  $S_b$  the original signal in the voxel at b-value  $b$  and  $S_0$  the signal in the  $b = 0$  s/mm<sup>2</sup> image,  $f$  is the fraction of the free water compartment, which is drawn randomly for every voxel from a uniform distribution  $U(0.7, 0.9)$  and  $D_{\text{csf}} = 3 \times 10^{-3}$  mm<sup>2</sup>/s is the nominal value of diffusivity for free water (e.g. cerebrospinal fluid (CSF)) at 37°celsius (Pasternak et al., 2009; Pierpaoli and Jones, 2004). Since the individual subjects are not aligned, but all scans from a given subject are registered, this introduces normal variability in terms of the number of white matter and gray matter voxels that would be affected by edema and their location in a patient subject.

### 6.3.7 Evaluation metrics

**Error and accuracy of predicted metrics** We reproduced parts of the analyses conducted in the original CDMRI challenge from Tax et al. (2019), namely the per voxel error for each metric as computed by the mean normalized error (MNE) and the voxelwise error.

Denoting the target data to be reproduced as *acquired* (Prisma or Connectom scanners) and the source data to be harmonized as *predicted* (GE scanner), the MNE is defined as  $MNE = |(\text{predicted} - \text{acquired})| / \text{acquired}$  and the error is defined as  $\text{error} = \text{predicted} - \text{acquired}$ . The original challenge reports values taken either globally in a brain mask, in FreeSurfer regions of interest (ROI) and excluding poorly performing regions or the median value computed in sliding windows. Since the masks of these ROIs were not released for the challenge, we instead report boxplots of the two metrics using the brain masks from the challenge as this reports the global median error in addition to the global mean error and additional quantiles of their distribution. To prevent outliers from affecting the boxplots (particularly located at the edges of the brain masks), we clip the MNE and error values at the lowest 0.1% and largest 99.9% for each dataset separately.

**Kullback-Leibler divergence as a measure of similarity** As the voxelwise difference may not be fully indicative of the global trend of the harmonization procedure between datasets (e.g. due to registration errors), we also computed the Kullback-Leibler (KL) divergence (Kullback and Leibler, 1951) between the distributions of each harmonized dataset from the GE scanner and its counterpart from the target scanner for each of the four metrics. The KL divergence is a measure of similarity between two probability distributions  $P(x)$  and  $Q(x)$  where lower values indicate a higher similarity and  $KL(P, Q) = 0$  when  $P(x) = Q(x)$ . In its discrete form, the Kullback-Leibler divergence is given by

$$KL(P, Q) = \sum_k P_k \log \left( \frac{P_k}{Q_k} \right), \quad (6.4)$$

where  $P_k$  is the “candidate” probability distribution,  $Q_k$  the true probability distribution and  $k$  represents the number of discrete histogram bins. The measure is not symmetric, that is  $KL(P, Q) \neq KL(Q, P)$  in general. We instead use the symmetric version of the KL divergence as originally defined by Kullback and Leibler (1951)

$$KL_{sym} = KL(P, Q) + KL(Q, P). \quad (6.5)$$

In practice, a discrete distribution can be constructed from a set of samples by binning and counting the data. By normalizing each bin so that their sum is 1, we obtain a (discrete) probability mass function. For each metric, the discrete distribution was created with  $k = 100$  equally spaced bins. We also remove all elements with probability 0 from either  $P_k$  or  $Q_k$  (if any) to prevent division by 0 in Eq. (6.4).

**Statistical testing and effect size in the presence of alterations** We conducted Student’s t-test for paired samples for each subject separately between each scanner in the predefined region of 3000 voxels with simulated changes (Student, 1908). This was done on both the normal datasets (testing between scanners) and the altered datasets (testing between

scanners and additionally between the normal and altered datasets). The p-values from the tests were subsequently corrected for the false discovery rate (FDR) at a level of  $\alpha = 0.05$  (Benjamini and Hochberg, 1995). In addition, we also report the effect size of those paired t-tests as computed by Hedges'  $g$  (Hedges, 1981; Lakens, 2013), which we redefine as

$$g = \frac{|\mu_1 - \mu_2|}{(\sigma_1 + \sigma_2)/2} \times \left(1 - \frac{3}{4(n_1 + n_2) - 9}\right), \quad (6.6)$$

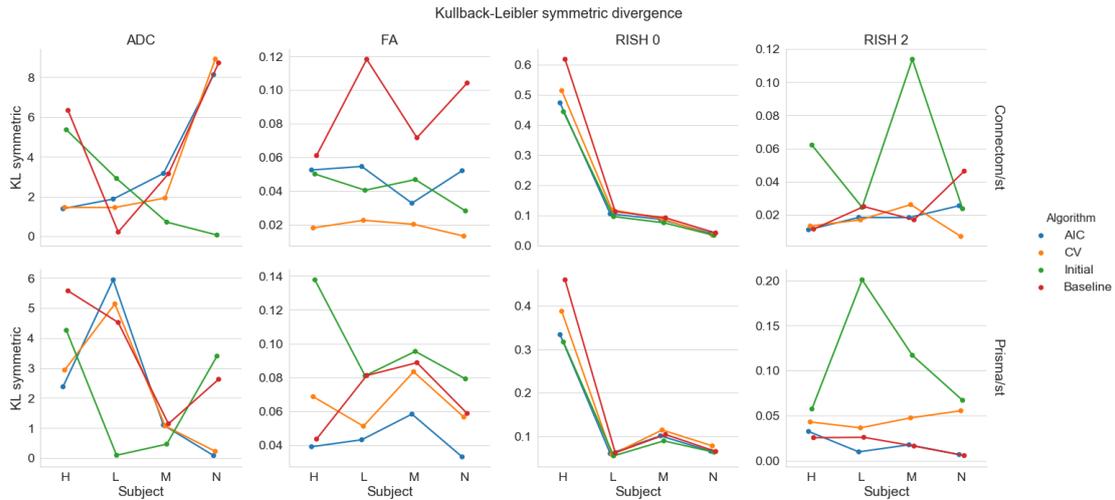
where  $\mu_i$ ,  $\sigma_i$  and  $n_i$  are the mean, the standard deviation, and the size of sample  $i$ , respectively. A value of  $g = 1$  indicates that the difference between the means is of one standard deviation, with larger values indicating larger effect sizes as reported by the difference in the group means. In the original definition of Hedges (1981),  $g$  is not enforced to be positive. We instead report the absolute value of  $g$  as we do not know *a priori* which mean is larger than the other, but are only interested in the magnitude of the effect rather than its sign. With this definition, values of  $g$  reported for the test between a given subject for two different scanners which are lower than the reference method indicate an improvement by removing scanner specific effects. On the other hand, similar values of  $g$  between the reference and the harmonized dataset for a given subject and its altered counterpart on the same scanner indicates preservation of the simulated effects as it is the *only* difference between these two datasets by construction.

## 6.4 Results

### 6.4.1 Results from the challenge

**Mapping between scanners for matched acquisition protocols** Fig. 6.3 shows the KL symmetric divergence as presented in Section 6.3.7 for the standard protocol. In general, the baseline has a higher KL value than the other methods on the Connectom scanner. The CV based method is generally tied or outperforms the AIC based method. For the Prisma scanner, results show that the AIC performs best with the CV based method following the baseline reference. In the case of the ADC metric, our initial algorithm outperforms the three other methods for some subjects.

Fig. 6.4 shows the distribution (as boxplots) in the absolute mean normalized error and mean error of the four metrics for the standard protocol. The MNE is almost tied or slightly higher for the baseline method than the alternatives for both scanners. For the FA and RISH 2 metrics, the baseline error is tied or larger than the other methods. For the voxelwise error, all methods underestimate the ADC and overestimate the RISH 0 on average while the FA and RISH 2 metrics show a different pattern depending on the scanner. For the Connectom scanner, the CV based method generally has an average error around 0 for the FA while the AIC and our initial algorithm generally overestimate the metric. The baseline is on the other spectrum and generally underestimates the FA. On



**Figure 6.3:** KL symmetric divergence (where lower is better) for the harmonization task at the same resolution between the GE ST datasets and the Connectom ST (top row) or the Prisma ST (bottom row) datasets on the four test subjects ('H', 'L', 'M' and 'N'). Each metric is organized by column (ADC, FA, RISH 0 and RISH 2) for the four compared algorithms (AIC in blue, CV in orange, our initial version of the harmonization algorithm in green and the baseline comparison in red).

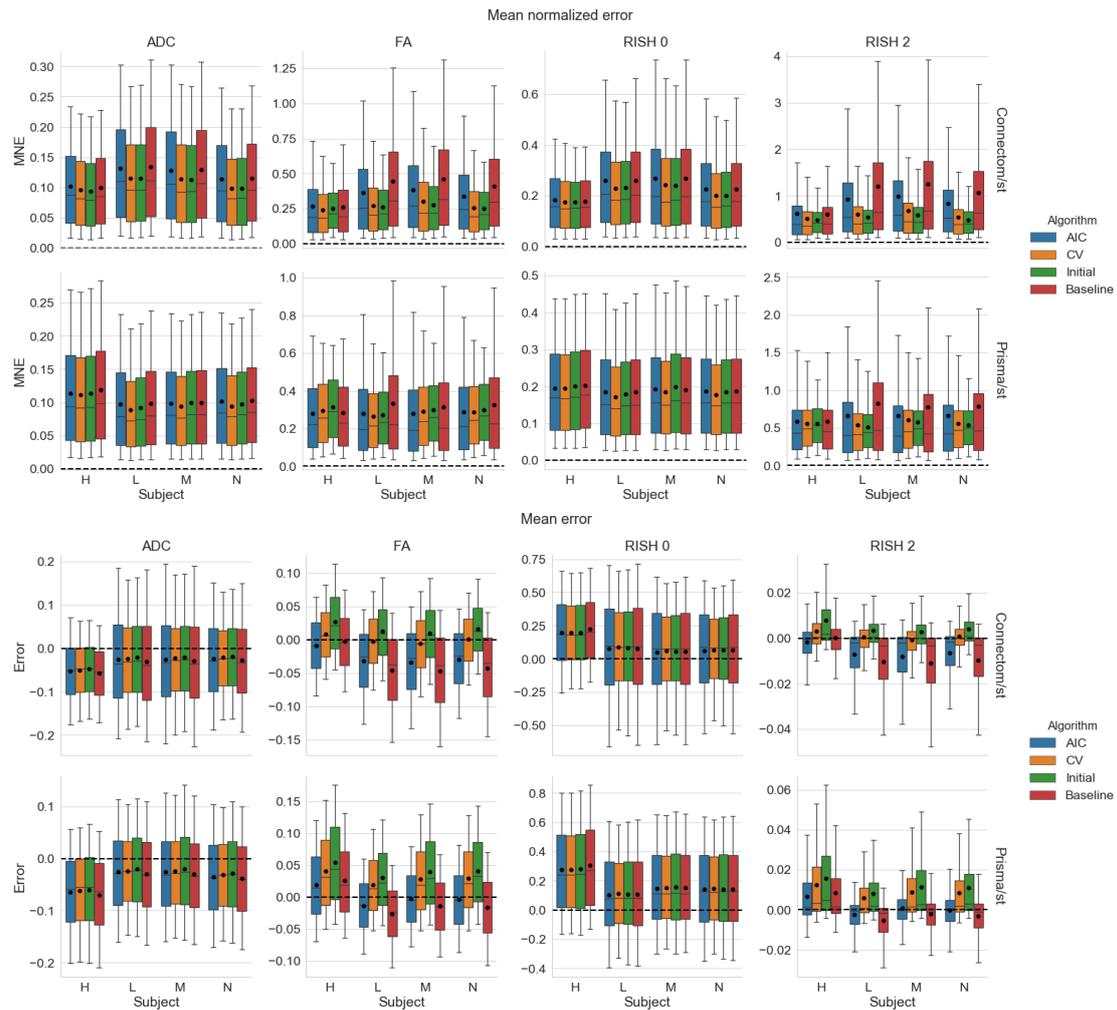
the Prisma scanner, the effect is reversed; there is a general overestimation of the FA while the error committed by the AIC based method is in general close to 0. The RISH 2 error follows the same pattern as the FA error on both scanners for the four compared methods.

**Mapping between scanners across spatial resolutions** Fig. 6.5 shows the KL symmetric divergence for the second task of the challenge, mapping the GE ST protocol datasets to the SA protocols of the Prisma or Connectom scanners. For the Connectom scanner, the AIC based algorithm and our initial algorithm, which is also AIC based, performs best in most cases. The CV based algorithm also outperforms the baseline method for the ADC and RISH 0 metrics. For the Prisma scanner, the AIC outperforms most of the compared methods or is tied with the CV. Notably, the baseline ranks second for the FA and RISH 2 metrics, but is the worst performer for the ADC and the RISH 0 metrics.

Fig. 6.6 shows results for the absolute mean normalized error and mean error for all algorithms on harmonizing the SA protocol. For the Connectom scanner, the baseline ranks last for most subjects on the isotropy metrics (ADC and RISH 0) while it only performs slightly better than the CV based algorithm for the anisotropy metrics (FA and RISH 2). On the Prisma scanner, results are similar for the ADC and RISH 0 metrics. For the FA metrics, the best performance is obtained with the AIC based method while the baseline is better for harmonizing the RISH 2 metric for three of the subjects.

Now looking at the mean error, results show that the ADC metric is underestimated for all methods and on both scanners with the three methods usually outperforming the

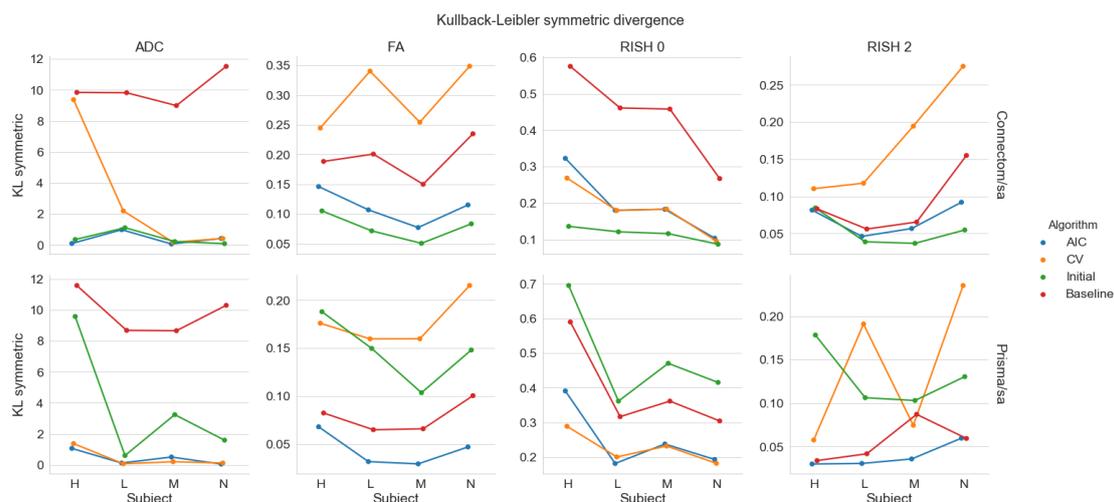
## 6.4. Results



**Figure 6.4:** Boxplots of the voxelwise mean normalized error (top) and error (bottom) for each metric, following the same conventions detailed in Fig. 6.3. The black dot shows the mean error and the dashed line indicates an error of 0, representing a perfect match between the harmonized GE dataset and the dataset for the target scanner.

baseline comparison. The FA, RISH 0 and RISH 2 metrics are instead overestimated. For the FA metric, the AIC and our initial algorithm commit less error on average than the baseline on the Connectom scanner. On the Prisma scanner, only the AIC has an average error lower than the baseline. All methods perform better or almost equal on average to the baseline comparison for the RISH 0 metric. The RISH 2 metric shows a scanner dependent pattern; on the Connectom scanner, the best performing method is our initial algorithm followed by the AIC based algorithm while on the Prisma scanner, the lowest error is achieved by the AIC based method.

In general, results show that the isotropy metrics (ADC and RISH 0) are subject to



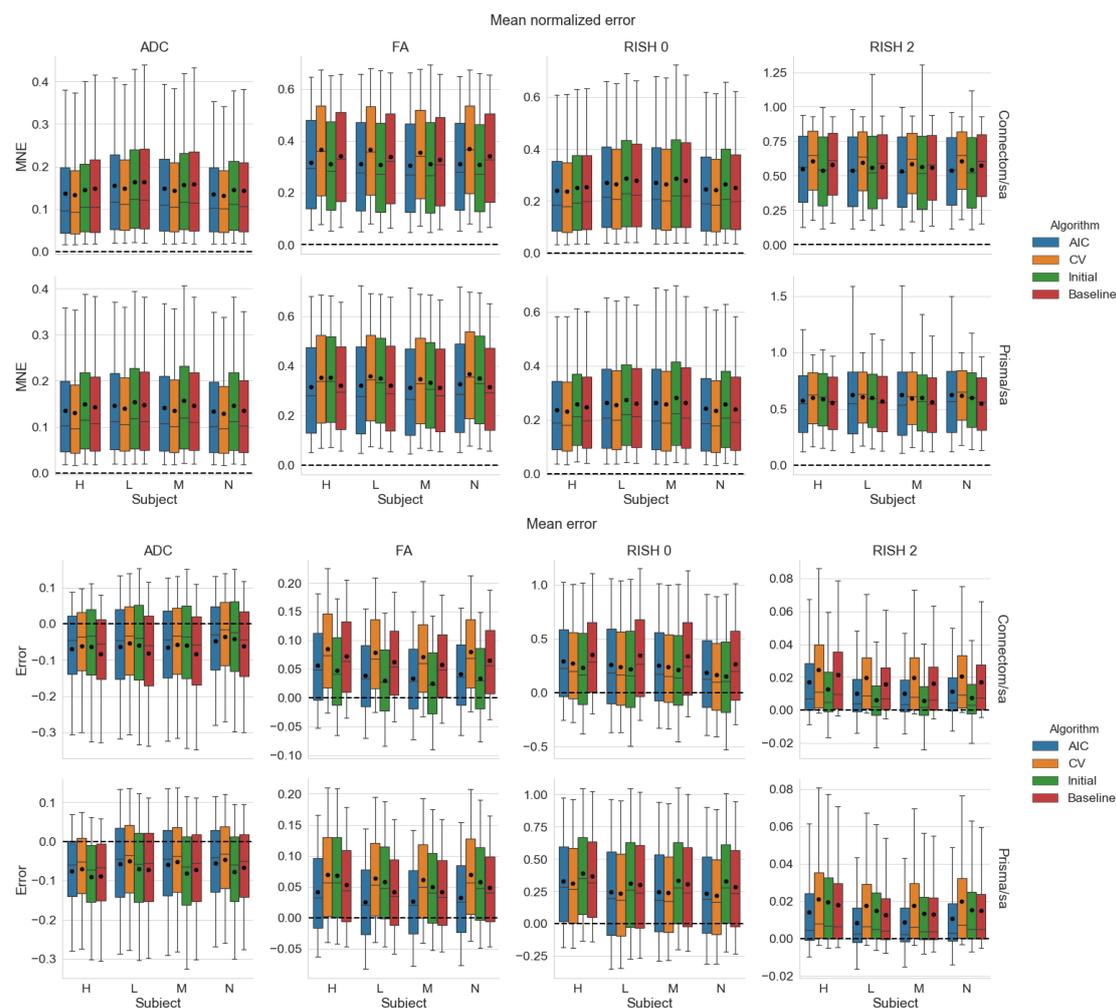
**Figure 6.5:** Symmetric KL divergence (where lower is better) for the harmonization task across resolution between the GE ST datasets and the Connectom SA (top row) or the Prisma SA (bottom row) datasets. The organization is the same as previously used in Fig. 6.3.

global scanner effects while the anisotropy metrics (FA and RISH 2) may be subject to orientation dependent effects. These effects are also likely different for each scanner since the gradient strength and timings are different, even if the b-values are matched. In these experiments, the target scanner is untouched and therefore still contains its own scanner effect when computing the voxelwise error of each harmonization algorithm.

#### 6.4.2 Mapping original and altered datasets towards a common space

In these experiments, alterations were made to the test set as previously described in Section 6.3.6. As these altered datasets were never used for training, we can quantify the removal of scanner effects and the preservation of the alterations by comparing solely the altered regions with their original counterpart in each subject, free of processing effects. In these experiments, the baseline comparison is to not process the datasets at all since the datasets are altered versions of themselves, therefore not requiring any interpolation or resampling. As these experiments are outside of the challenge’s scope, they are not covered by our initial algorithm and therefore the “previous” category is not presented in this section. Fig. 6.7 shows the original and altered metrics for one subject on the raw data and after harmonization with the AIC and CV based algorithms and Fig. 6.8 shows the relative percentage difference between the raw datasets and their harmonized counterpart. We define the relative percentage difference as  $\text{difference} = 100 \times (\text{harmonized} - \text{raw}) / \text{raw}$ . The alterations is mostly visible on the  $b = 0 \text{ s/mm}^2$  image while the  $b = 1200 \text{ s/mm}^2$  image is only slightly affected due to the high diffusivity of the CSF compartment. However, the differences are visible on the diffusion derived maps, seen as an increase in ADC and a

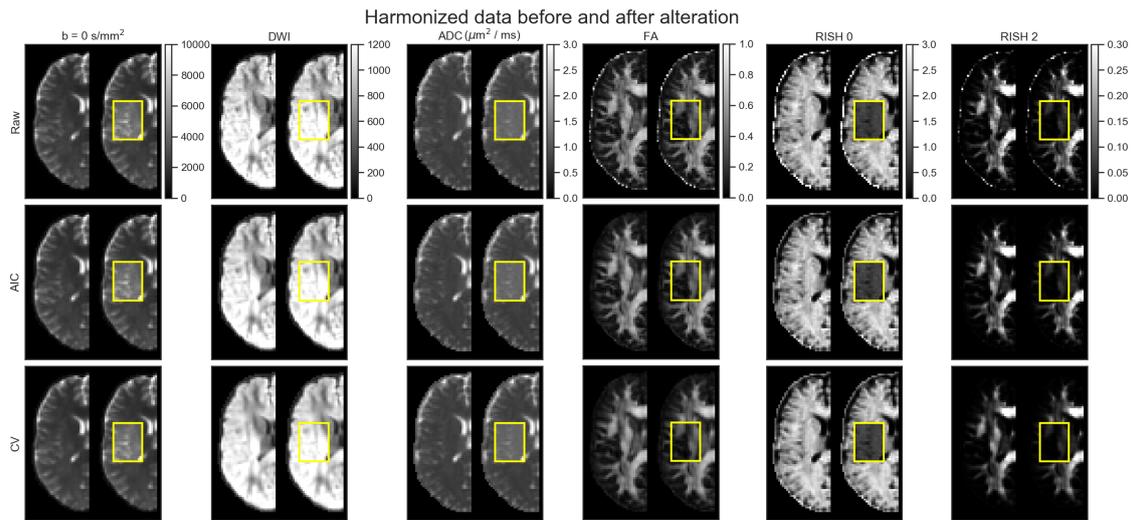
## 6.4. Results



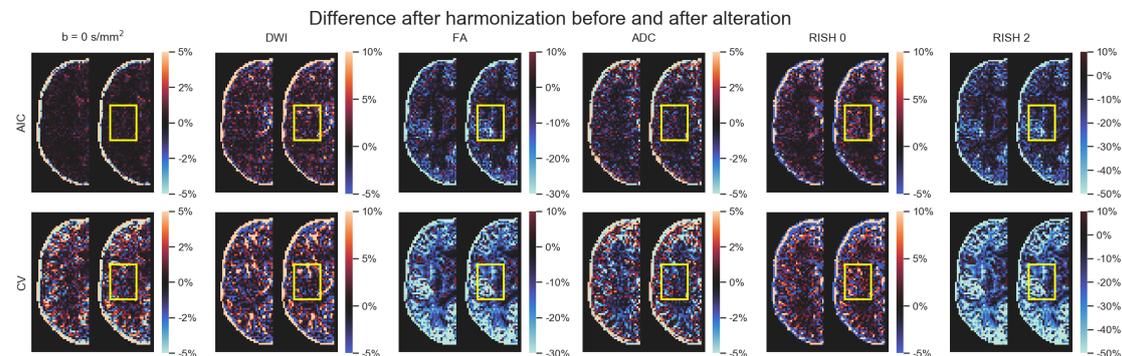
**Figure 6.6:** Boxplots of the voxelwise mean normalized error (top) and error (bottom) of each metric for the four algorithms. The black dot shows the mean error and the dashed line indicates an error of 0. The organization follows the conventions of Fig. 6.4.

reduction for the FA, RISH 0 and RISH 2 metrics. Visually, harmonized datasets do not seem different from their original counterpart, but the difference maps show that small differences are present with the CV method generally showing larger differences than the AIC method. Notably, the anisotropy metrics (FA and RISH 2) are lower after harmonization while the difference for the isotropy metric (ADC and RISH 0) is distributed around 0.

Fig. 6.9 shows boxplots of the effect size as computed by a paired t-test after harmonization towards a common space for all scanners. Tests were conducted for every subject between each scanner in addition to the altered versions of the datasets as previously described in Section 6.3.7. For the ADC metric, both methods yield a lower effect size on average than the raw, unprocessed data and preserve the effect size due to the alterations



**Figure 6.7:** Exemplar slice of subject 'H' on the GE scanner as original (left half) and altered (right half) metrics. Only the affected portion of the data (yellow box) is analyzed in paired statistical testing against the same location in the original dataset. Each column shows (from left to right) a  $b = 0 \text{ s/mm}^2$  image, a DWI at  $b = 1200 \text{ s/mm}^2$ , the FA, ADC, RISH 0 and RISH 2 metrics with a common colorbar per column. The top row shows the raw data, the middle row shows the data harmonized using the AIC and the bottom row shows the harmonized data using the CV. The  $b = 0 \text{ s/mm}^2$  image, the DWI and the ADC map are increased after adding the free water compartment while the FA, RISH 0 and RISH 2 metrics are instead lower in their altered counterpart.

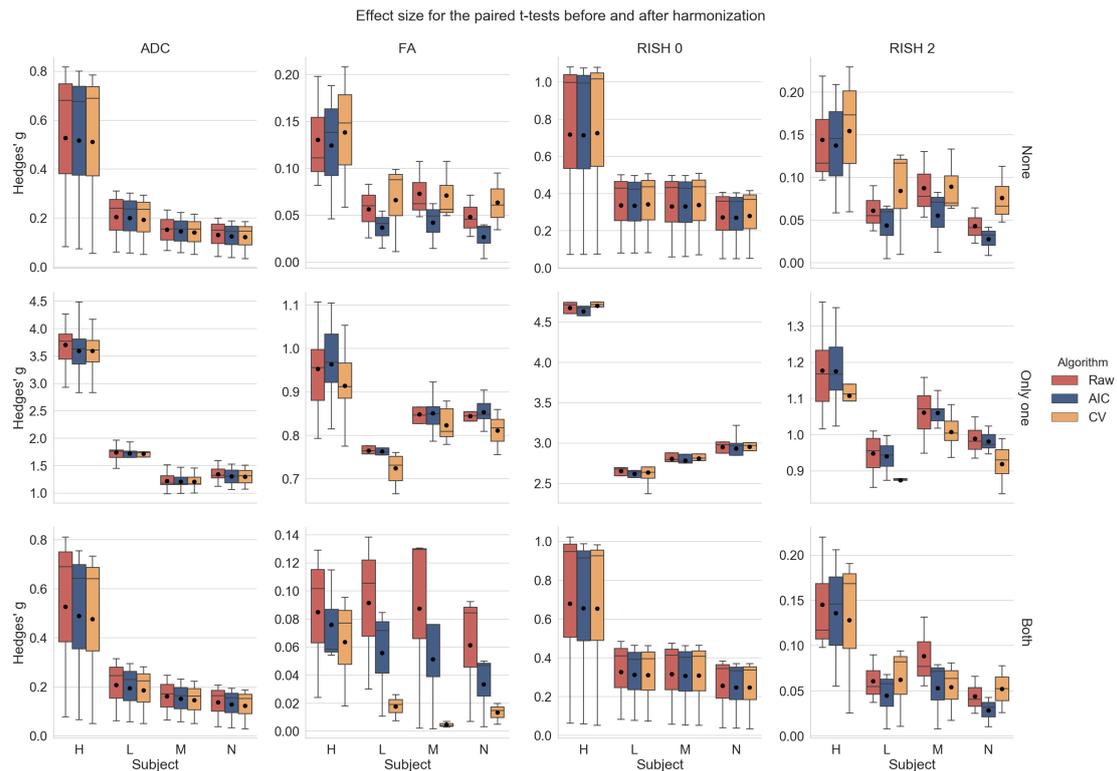


**Figure 6.8:** Exemplar slice of subject 'H' on the GE scanner as original (left half) and altered (right half) metrics. Each column shows (from left to right) a  $b = 0 \text{ s/mm}^2$  image, a DWI at  $b = 1200 \text{ s/mm}^2$ , the FA, ADC, RISH 0 and RISH 2 metrics with a common colorbar per column as in Fig. 6.7. The top row (resp. the bottom row) shows the relative percentage difference between the harmonized data using the AIC (resp. the CV) and the raw data.

as shown in the middle row. The RISH 0 metric shows similar behavior with the CV based method producing an average effect size slightly higher than the raw datasets. Now looking at the anisotropy metrics (FA and RISH 2), the effect size is reduced or equal on

## 6.4. Results

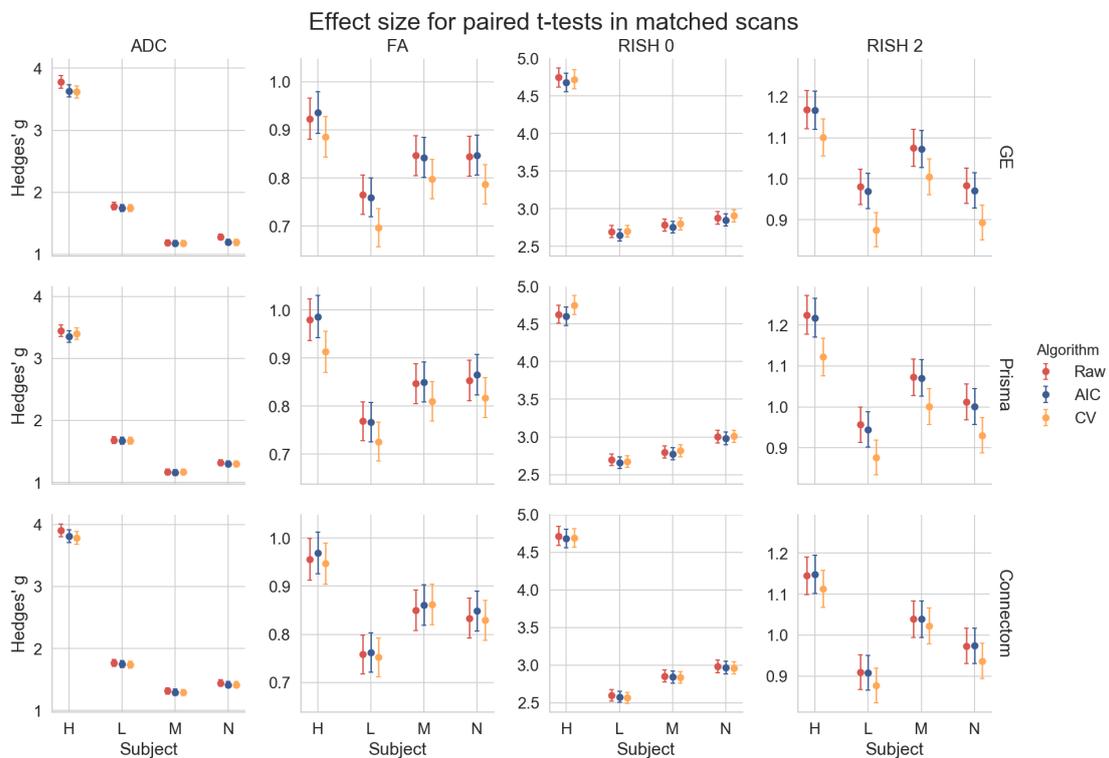
average in most cases (except for subject 'H' when only one scan is altered) when scans are harmonized with the AIC algorithm. The CV based algorithm shows a higher effect size for harmonization between scans and a lower effect size when both scans are altered. As we only report the absolute value of the effect size, this is due to both a lower group mean and group standard deviation than the raw datasets. The harmonization process is likely only removing scanner effects present in each dataset as the middle row (where only one of the compared dataset is affected) shows similar reductions in effect size, but is still on the same magnitude as the raw datasets since the alteration is preserved.



**Figure 6.9:** Boxplots of Hedges'  $g$  effect size for each metric with the mean value as the black dot. The raw data is shown in red (no harmonization), the data harmonized with the AIC in blue and finally the data harmonized with the CV in orange, similarly to the previous figures. The top row shows the effect size when both datasets are in their original version (**None** of the datasets are altered), the middle row when **only one** of the dataset is altered and the bottom row when **both** datasets are altered as indicated on the right of each row. The top and bottom row are only affected by scanner effects. The middle row shows larger effects size due to one of the compared dataset being altered in addition to the scanner effects.

Fig. 6.10 shows the effect size, with a 95% confidence interval (CI), for the paired t-test between the original and altered datasets on each scanner. While Fig. 6.9 showed the general trend for all results, we instead now focus on the effect size attributable solely to the alterations we previously induced. Results show that the ADC and RISH 0 metrics have

the smallest CI, showing the lowest variability in the 3000 voxels in the altered region. All CI are overlapping and therefore have a 95% chance of containing the true mean effect size for every case. The FA and RISH 2 metrics have both larger CI, showing larger variability in their sample values, but are overlapping with the raw datasets CI in most cases. Only the CV based harmonization method CI is outside the raw datasets CI for two cases. This shows that the effect size is likely preserved after applying the harmonization algorithm in most cases since the only source of variability is the effects we induced in that region to create the altered datasets. Supplementary materials 1 contains the individual effect sizes, p-values and other intermediary statistics for every tested combination that generated the boxplots shown in Fig. 6.9.



**Figure 6.10:** Hedges' g effect size for each metric between the original and altered datasets on the same scanner with a 95% CI. The top row shows the effect size between the original and altered dataset on the GE scanner, the middle row for the Prisma scanner and the bottom row for the Connectom scanner. Most of the CI are overlapping except for the CV in the cases of subject 'L' on the GE scanner and subject 'H' on the Prisma scanner. This effect size is only due to the alterations performed in the experiments and is free of any other source of variability, such as registration error or scanner effects.

## 6.5 Discussion

### 6.5.1 Reducing variability across scanners

We have presented a new algorithm based on dictionary learning to harmonize datasets acquired on different scanners using the benchmark database from the CDMRI 2017 harmonization challenge (Tax et al., 2019). The flexibility of the method lies in its ability to adapt the regularization parameter  $\lambda_i$  automatically to each subset of training examples in Eq. (6.1), ensuring that the relevant information to reconstruct the data is encoded in the dictionary  $\mathbf{D}$ . Only features deemed important to the reconstruction are stored as the  $\ell_1$  norm on the coefficients  $\alpha$  encourages a sparse reconstruction and forces most of the coefficients to zero (Candès et al., 2008; Daubechies et al., 2010; St-Jean, Coupé, et al., 2016). In the reconstruction step, a new value of  $\lambda_i$  is automatically selected for each reconstructed patch, ensuring that the regularization is tuned uniquely so that the reconstruction matches the original patch, but using only features found in the target scanner. This is of course at the cost of additional computations since a least-square problem needs to be solved for each candidate value  $\lambda_i$ , but convex and efficient numerical routines reusing the previous solution as a starting point can be used to alleviate computational issues (Friedman et al., 2010). To the best of our knowledge, this is the first case where an automatic search of the regularization parameter has been used in both stages of the optimization.

For the reconstruction step, we introduced two alternatives to compute  $\lambda_i$  through the AIC or CV using held out parts of the signal. While other choices are possible, such as the Bayesian information criterion (Schwarz, 1978), we chose here the AIC for simplicity and because it is in fact equivalent to leave one out CV in the asymptotic limit (Stone, 1977). Cross-validation was done with a classical approach as done in statistics, predicting the signal on parts of the current reconstructed patch as opposed to e.g. reconstructing a completely separate patch with the same value of  $\lambda_i$  as may be done in machine learning. This could explain why the AIC based method performed better than the CV criterion for the anisotropy metrics in the SA protocol since the held out data, which is selected at random for every case, may sometimes unbalance the angular part of the signal because of the subsampling. The AIC would not be affected as it can access the whole data for prediction but instead penalizes reconstructions that do not substantially reduce the mean  $\ell_2$  error and are using too many coefficients—a likely situation of overfitting. This also makes the AIC faster to compute since there is no need to refit the whole model from the beginning unlike the CV.

One major advantage of the harmonization approach we presented is that raw datasets are directly harmonized without the requirement of paired samples during training. In fact, the data was given at random for the training phase and we mixed patches from all subjects and all scanners altogether in the additional experiments we described in Section 6.3.5, preventing overfitting to a particular scanner in the process. While other harmonization approaches have been developed, most of them require paired samples (Cetin Karayumak et

al., 2019; Mirzaalian et al., 2016) or harmonize only the extracted scalar maps from diffusion MRI instead (Alexander et al., 2017; Fortin et al., 2017), limiting their applicability in studies that do not account for these requirements at first in their design. Moreover, it is not clear if the mapping developed for a particular scalar map is in fact similar between metrics as scanner effects may behave differently, e.g. isotropy metrics may be subject to global effects while anisotropy metrics may exhibit orientational bias due to low SNR in some given gradient directions. We also observed in our experiments that the error for the ADC and RISH 0 metrics were similar for most methods while the error was larger for the FA and RISH 2 metrics for the baseline method, which are orientation dependent. This shows that the “optimal” mapping function could likely be task dependent if one wants to harmonize directly the scalar maps between scanners, which could complicate interpretation between studies that are not using a matched number of b-values or gradient orientation.

In the additional experiments, we introduced the idea of creating a neutral “scanner space” instead of mapping the datasets towards a single target scanner. We also harmonized datasets that had been altered towards that common space and shown that the induced effect sizes are preserved while at the same time preserving normal anatomical variability. This approach has the benefit of removing variability attributable to both scanners, instead of trying to force the source scanner to mimic variability that is solely attributable to the target scanner. It is also important to mention here that a good harmonization method should remove unwanted variability due to instrumentation, all the while preserving genuine anatomical effects as also pointed out previously by Fortin et al. (2017). While this statement may seem obvious, success of harmonization towards a common space is much more difficult to quantify than between scanners since we can not look at difference maps between harmonized datasets anymore. since we can not look at difference maps between harmonized datasets anymore. As a thought experiment, a harmonization method that would map all datasets towards a constant value would show no difference between the harmonized datasets themselves, therefore entirely removing all variability. It would however commit very large errors when compared against the original version of those same datasets. From Fig. 6.7, we see that the harmonized datasets are similar to their original version, but Fig. 6.8 shows that the CV based algorithm has larger relative differences with the data before harmonization. It is however not obvious if the CV based algorithm is removing too much variability by underfitting the data or if the AIC based method is not removing enough, overfitting the data. Fig. 6.10 suggests that the CV criterion might underfit the data due to the lower effect size, but this could be due to using only 3 fold CV in our experiments to limit computation time. Results might be improved by using more folds as the AIC approximates the CV as we have previously mentioned.

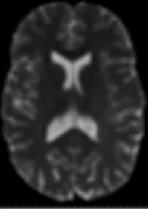
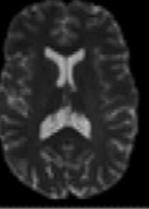
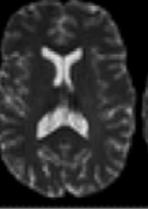
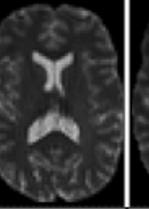
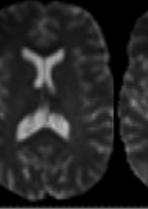
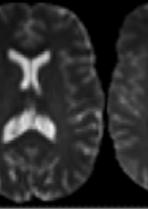
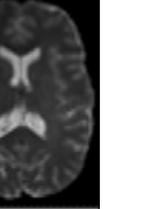
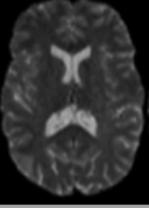
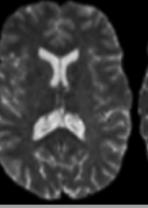
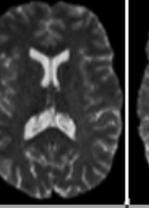
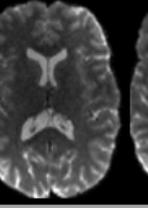
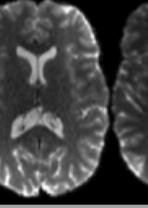
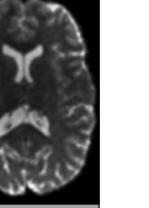
### 6.5.2 Dependence of isotropy and anisotropy metrics on scanning parameters

While it is usually advocated that protocols should use similar scanning parameters as much as possible to ensure reproducibility, this is not always easily feasible depending on the se-

quences readily available from a given vendor and differences in their implementations. Subtle changes in TE and TR influence the measured signal as shown in Fig. 6.11 by changing the relative T2 and T1 weighting of the measured diffusion signal, respectively. While dMRI local models are usually applied on a per voxel basis, changes in these weightings will yield different values of the diffusion metrics, which makes comparisons between scans more difficult as the weighting depends on the different (and unknown) values of T1 and T2 of each voxel (Brown et al., 2014, Chap. 8). Even if these changes are global for every voxels, matched b-values are not sufficient to ensure that the diffusion time is identical between scans as changes in TE influence diffusion metrics such as increased FA (Qin et al., 2009), but this effect may only manifest itself at either long or very short diffusion times in the human brain (Clark et al., 2001; Kim et al., 2005). Proper care should be taken to match the diffusion time beyond the well-known b-value, which may not always be the case if different sequences are used e.g. PGSE on the Siemens scanners and TRSE on the GE scanner as used in this manuscript. Additional effects due to gradients and b-values spatial distortions (Bammer et al., 2003) could also adversely affect the diffusion metrics, especially on the Connectom scanner as it uses strong gradients of 300 mT/m (Tax et al., 2019). Isotropy metrics are not necessarily free of these confounds as gradients nonlinearity create a spatially dependent error on the b-values (Paquette et al., 2019). This could explain the larger mean error for the CV and baseline methods on the Connectom scanner harmonization task, especially for the anisotropy metrics. While correcting for these effects is not straightforward, gradient timings should be reported in addition to the usual parameters (e.g. TE, TR, b-values and number of gradient directions) in studies to ease subsequent harmonization. Accounting for these differences during analysis could be done e.g. by using a (possibly mono-exponential) model including diffusion time and predicting the diffusion metrics of interest at common scanning parameters values between the acquisitions to harmonize.

### 6.5.3 Limitations

**Limitations of harmonization** As Burnham and Anderson (2004) stated, “in a very important sense, we are not trying to model the data; instead, we are trying to model the information in the data”. This is indeed the approach taken in the challenge by the participants, the four other entries relying on deep learning and neural networks for the most part with all methods (including ours) optimizing a loss function which considered the difference between the original and the harmonized dataset. With the rapid rise of the next generation of deep learning methods such as generative adversarial networks (GAN) and extensions (Goodfellow et al., 2014), it is now possible to instead model directly the *distribution* of the data. This allows generation of datasets from a completely different imaging modality such as synthesizing target CT datasets from source MRI datasets (Wolterink et al., 2017). However, if proper care is not taken to represent truthfully the distribution of the data (e.g. not including enough tumor samples in a harmonization task between

	GE	Prisma			Connectom		
Standard (ST)							
TE/TR (ms)	89/CG	89/7200	80/7200	89/13000	89/7200	68/7200	89/13000
State-of-the-art (SA)							
TE/TR (ms)		80/4500	80/7200	89/7200	68/5400	68/7200	89/7200

**Figure 6.11:** Example  $b = 0$  s/mm<sup>2</sup> images for the standard protocol (top row) and the state-of-the-art protocol (bottom row) for a single subject acquired on the three scanners at different combinations of TE and TR. Note that the  $b = 0$  s/mm<sup>2</sup> image for the GE scanner was only acquired at a single TE with a cardiac gated (CG) TR. The figure is adapted from Tax et al. (2019), available under the CC-BY 4.0 license.

datasets with pathological data), this can lead to severe issues. Cohen et al. (2018) recently showed that in such a case, GAN based methods could try to remove the pathology in the data to match the distribution of healthy subjects that the method previously learned, precluding potential applications to new datasets or pathological cases not represented “well enough” in the training set. The same concept would likely apply to systematic artifacts; if every dataset from a target scanner is corrupted by e.g. a table vibration artifact, it may very well be possible that a harmonization algorithm will try to imprint this artifact to the source datasets to match the target datasets. The same remark would apply to our harmonization algorithm; if systematic artifacts are in the data, the learned dictionary may very well try to reconstruct these systematic artifacts. However, when rebuilding the source dataset using this corrupted target dictionary, we expect that the artifact would be mitigated since it would not appear in the source dataset and hence should not be reconstructed by Eq. (6.1) as it would penalize the  $\ell_2$  norm part of the cost function. While offering a promising avenue, care must be taken when analyzing harmonization methods to ensure that they still faithfully represent the data as optimal values of the cost functions themselves or “good” reconstruction of the diffusion metrics only may not ensure this fact (Rohlfing, 2012).

**Limitations of our algorithm and possible improvements** Our additional experiments with simulated free water have shown how harmonization can, to a certain extent, account

for data abnormalities not part of the training set. However, the presence of CSF and the boundary between gray matter and CSF (or a linear combination of those elements) may yield enough information for the reconstruction to encode these features in the dictionary. This can provide new elements that are not used for the reconstruction of normal white matter but may be useful for the altered data in the experiments. It is not necessarily true that this property would also be valid for other neurological disorders such as tumors, especially if their features are not well represented in the training data as we have mentioned previously in [Section 6.5.3](#). Another aspect that we did not explicitly cover is multishell data i.e. datasets acquired with multiple b-values, which was in fact part of the following CDMRI challenge (Ning et al., 2019). Nevertheless, our method can still be used on such datasets, but would not be aware of the relationship between DWIs beyond the angular domain. Other approaches to build the dictionary could be used to inform the algorithm and potentially increase performance on such datasets by explicitly modeling the spatial and angular relationship (Schwab et al., 2018) or using an adaptive weighting considering the b-values in the angular domain (Duits et al., 2019) amongst other possible strategies. Modeling explicitly the angular part of the signal could also be used to sample new gradients directions directly, an aspect we covered in the original CDMRI challenge by using the spherical harmonics basis (Descoteaux et al., 2007). Correction for the nature of the noise distribution could also be subsequently included as a processing step before harmonization since reconstruction algorithms vary by scanner vendor (Dietrich et al., 2008; St-Jean, De Luca, et al., 2018), leading to differences between scans due to changes in the noise floor level (Sakaie et al., 2018). Improvements could also potentially be achieved by considering the group structure shared by overlapping patches when optimizing [Eq. \(6.1\)](#) (Simon et al., 2013). While this structure would need to be explicitly specified, optimizing jointly groups of variables has recently led to massive improvements in other applications of diffusion MRI such as reduction of false positive connections in tractography (Schiavi et al., 2019).

## 6.6 Conclusions

In this paper, we have developed and evaluated a new harmonization algorithm to reduce intra and inter scanner differences. Using the public database from the CDMRI 2017 harmonization challenge, we have shown how a mapping from one scanner to another can be constructed automatically through dictionary learning using unpaired training datasets. This can also be done for different spatial resolutions through careful matching of the spatial patch size used to build the dictionary from the target scanner. We also introduced the concept of mapping datasets towards an arbitrary “scanner space” and used the proposed algorithm to reconstruct altered versions of the test datasets corrupted by a free water compartment, even if such data was not part of the training datasets. Results have shown that the effect size due to alterations is preserved after harmonization, while removing variability attributable to scanner effects. We also provided recommendation when harmonizing

protocols, such as reporting the gradient timings to inform subsequent harmonization algorithms that could exploit these values across studies. As perfect matching of scanner parameters is difficult to do in practice due to differences in vendor implementations, an alternative approach could be to account for these differences through models of diffusion using these additional parameters. Nevertheless, as the algorithm is freely available, this could help multicenter studies in pooling their data while removing scanner specific confounds and increase statistical power in the process.

## 6.7 Appendix: The harmonization algorithm

This appendix outlines the harmonization algorithm in two separate parts. [Algorithm 6.1](#) first shows how to build a target dictionary as depicted in the top part of [Fig. 6.1](#). The bottom part of the diagram shows how to rebuild a dataset given the dictionary and is detailed in [Algorithm 6.2](#). Our implementation is also freely available at <https://github.com/samuelstjean/harmonization> (St-Jean, Viergever, et al., 2019).

---

**Algorithm 6.1:** The proposed harmonization algorithm - building a target dictionary.

---

**Data:** Datasets, patch size, angular neighbor

**Result:** Dictionary  $\mathbf{D}$

**Step 1 :** *Extracting patches from all datasets;*

**foreach** *Datasets* **do**

    Find the closest angular neighbors;

    Create a 4D block with a  $b = 0$  s/mm<sup>2</sup> and the angular neighbors;

    Extract all 3D patches and store the result in an array  $\Omega$ ;

**end**

**Step 2 :** *Build the target dictionary;*

**while** *Number of max iterations not reached* **do**

    Randomly pick patches from  $\Omega$ ;

    Solve [Eq. \(6.1\)](#) for  $\alpha$  with  $\mathbf{D}$  fixed;

    Solve [Eq. \(6.1\)](#) for  $\mathbf{D}$  with  $\alpha$  fixed;

**end**

---

---

**Algorithm 6.2:** The proposed harmonization algorithm - reconstruction of the harmonized data.

---

**Data:** Dataset, dictionary

**Result:** Harmonized dataset

**Step 1 :** *Extracting patches from the dataset to harmonize;*

**foreach** *Dataset* **do**

    | Find the closest angular neighbors;  
    | Create a 4D block with a  $b = 0$  s/mm<sup>2</sup> and the angular neighbors;  
    | Extract all overlapping 3D patches and store the result as  $\Omega$ ;

**end**

**if** *Matching across spatial resolution* **then**

    | Downsample  $\mathbf{D}$  into  $\mathbf{D}_{\text{small}}$  spatially before reconstruction;

**else**

    |  $\mathbf{D}_{\text{small}} = \mathbf{D}$ ;

**end**

**Step 2 :** *Find the harmonized patch;*

**foreach** *patches*  $\in \Omega$  **do**

    | Find the coefficients  $\alpha$  by solving Eq. (6.1) for  $\mathbf{D}_{\text{small}}$  fixed;  
    | Find the harmonized representation  $\mathbf{X} = \mathbf{D}\alpha$ ;

**end**

**foreach** *patches*  $\in \Omega$  **do**

    | Put back each patch at its spatial location and average overlapping parts;

**end**

---

# Bibliography

- [1] H. Akaike. “A new look at the statistical model identification”. In: *IEEE Transactions on Automatic Control* 19.6 (Dec. 1974), pp. 716–723 (cit. on p. 151).
- [2] D. C. Alexander et al. “Image quality transfer and applications in diffusion MRI”. In: *NeuroImage* 152 (May 2017), pp. 283–298 (cit. on p. 167).
- [3] H.-E. Assemlal, D. Tschumperlé, L. Brun, and K. Siddiqi. “Recent advances in diffusion MRI modeling: Angular and radial reconstruction.” In: *Medical image analysis* 15.4 (Aug. 2011), pp. 369–96 (cit. on p. 148).
- [4] R. Bammer et al. “Analysis and generalized correction of the effect of spatial gradient field distortions in diffusion-weighted imaging”. In: *Magnetic Resonance in Medicine* 50.3 (Sept. 2003), pp. 560–569 (cit. on p. 168).
- [5] P. J. Basser, J. Mattiello, and D. LeBihan. “MR diffusion tensor spectroscopy and imaging.” In: *Biophysical journal* 66.1 (Jan. 1994), pp. 259–67 (cit. on p. 148).
- [6] P. J. Basser and C. Pierpaoli. “Microstructural and Physiological Features of Tissues Elucidated by Quantitative-Diffusion-Tensor MRI”. In: *Journal of Magnetic Resonance, Series B* 111.3 (June 1996), pp. 209–219 (cit. on p. 148).
- [7] Y. Benjamini and Y. Hochberg. “Controlling the false discovery rate: a practical and powerful approach to multiple testing”. In: *Journal of the Royal Statistical Society* 57.1 (Nov. 1995), pp. 289–300 (cit. on p. 158).
- [8] R. W. Brown, E. M. Haacke, Y.-C. N. Cheng, M. R. Thompson, and R. Venkatesan. *Magnetic resonance imaging: physical principles and sequence design*. John Wiley & Sons, 2014 (cit. on p. 168).
- [9] K. P. Burnham and D. R. Anderson. “Multimodel Inference”. In: *Sociological Methods & Research* 33.2 (Nov. 2004), pp. 261–304 (cit. on pp. 151, 168).
- [10] E. J. Candès, M. B. Wakin, and S. P. Boyd. “Enhancing Sparsity by Reweighted l1 Minimization”. In: *Journal of Fourier Analysis and Applications* 14.5-6 (Oct. 2008), pp. 877–905 (cit. on p. 166).
- [11] S. Cetin Karayumak et al. “Retrospective harmonization of multi-site diffusion MRI data acquired with different acquisition parameters”. In: *NeuroImage* 184.May 2018 (Jan. 2019), pp. 180–200 (cit. on pp. 148, 166).
- [12] C. A. Clark, M. Hedehus, and M. E. Moseley. “Diffusion time dependence of the apparent diffusion tensor in healthy human brain and white matter disease”. In: *Magnetic Resonance in Medicine* 45.6 (June 2001), pp. 1126–1129 (cit. on p. 168).
- [13] J. P. Cohen, M. Luck, and S. Honari. “Distribution Matching Losses Can Hallucinate Features in Medical Image Translation”. In: *Medical Image Computing and Computer Assisted Intervention – MIC-CAI 2018*. Vol. 57. 2. Springer International Publishing, 2018, pp. 529–536 (cit. on p. 169).
- [14] I. Daubechies, R. Devore, M. Fornasier, and C. S. Güntürk. “Iteratively reweighted least squares minimization for sparse recovery”. In: *Communications on Pure and Applied Mathematics* 63 (2010), pp. 1–38 (cit. on p. 166).

## Bibliography

---

- [15] A. De Luca, S. Franklin, C. Lucci, J. Hendrikse, M. Froeling, and A. Leemans. “Investigation of the dependence of free water and pseudo-diffusion MRI estimates on the cardiac cycle”. In: *International Society for Magnetic Resonance in Medicine (ISMRM)*. 2019, p. 344 (cit. on p. 148).
- [16] M. Descoteaux, E. Angelino, S. Fitzgibbons, and R. Deriche. “Regularized, fast, and robust analytical Q-ball imaging”. In: *Magnetic Resonance in Medicine* 58.3 (Sept. 2007), pp. 497–510 (cit. on pp. 154, 170).
- [17] O. Dietrich, J. G. Raya, S. B. Reeder, M. Ingrisch, M. F. Reiser, and S. O. Schoenberg. “Influence of multichannel combination, parallel imaging and other reconstruction techniques on MRI noise characteristics.” In: *Magnetic resonance imaging* 26.6 (July 2008), pp. 754–62 (cit. on pp. 148, 170).
- [18] S. Duchesne et al. “The Canadian Dementia Imaging Protocol: Harmonizing National Cohorts”. In: *Journal of Magnetic Resonance Imaging* 49.2 (Feb. 2019), pp. 456–465 (cit. on p. 148).
- [19] R. Duits, E. St-Onge, J. Portegies, and B. Smets. “Total Variation and Mean Curvature PDEs on the Space of Positions and Orientations”. In: *Lecture Notes in Computer Science*. Ed. by A. M. Bruckstein, B. M. ter Haar Romeny, A. M. Bronstein, and M. M. Bronstein. Vol. 11603. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2019, pp. 211–223 (cit. on p. 170).
- [20] M. Elad and M. Aharon. “Image Denoising Via Sparse and Redundant Representations Over Learned Dictionaries”. In: *IEEE Transactions on Image Processing* 15.12 (Dec. 2006), pp. 3736–3745 (cit. on p. 149).
- [21] C. Federau et al. “Dependence of Brain Intravoxel Incoherent Motion Perfusion Parameters on the Cardiac Cycle”. In: *PLoS ONE* 8.8 (Aug. 2013). Ed. by J. Hendrikse, e72856 (cit. on p. 148).
- [22] D. a. Feinberg et al. “Multiplexed echo planar imaging for sub-second whole brain fmri and fast diffusion imaging”. In: *PLoS ONE* 5.12 (2010) (cit. on p. 148).
- [23] J.-P. Fortin et al. “Harmonization of multi-site diffusion tensor imaging data”. In: *NeuroImage* 161 (Nov. 2017), pp. 149–170 (cit. on pp. 149, 167).
- [24] J. Friedman, T. Hastie, and R. Tibshirani. “Regularization Paths for Generalized Linear Models via Coordinate Descent.” In: *Journal of statistical software* 33.1 (2010), pp. 1–22 (cit. on pp. 150, 151, 166).
- [25] I. J. Goodfellow et al. “Generative Adversarial Networks”. In: *Proceedings of the International Conference on Neural Information Processing Systems (NIPS 2014)*. 2014, pp. 2672–2680 (cit. on p. 168).
- [26] A. Gramfort, C. Poupon, and M. Descoteaux. “Denoising and fast diffusion imaging with physically constrained sparse dictionary learning.” In: *Medical image analysis* 18.1 (Jan. 2014), pp. 36–49 (cit. on pp. 150, 151).
- [27] E. H. B. M. Gronenschild et al. “The Effects of FreeSurfer Version, Workstation Type, and Macintosh Operating System Version on Anatomical Volume and Cortical Thickness Measurements”. In: *PLoS ONE* 7.6 (June 2012). Ed. by S. Hayasaka, e38234 (cit. on p. 148).
- [28] L. V. Hedges. “Distribution Theory for Glass’s Estimator of Effect size and Related Estimators”. In: *Journal of Educational Statistics* 6.2 (June 1981), pp. 107–128 (cit. on p. 158).
- [29] A. Hofman, D. E. Grobbee, P. T. V. M. De Jong, and F. A. Van den Ouweland. “Determinants of disease and disability in the elderly: The Rotterdam elderly study”. In: *European Journal of Epidemiology* 7.4 (July 1991), pp. 403–422 (cit. on p. 148).
- [30] A. Hofman, G. G. O. Brusselle, et al. “The Rotterdam Study: 2016 objectives and design update”. In: *European Journal of Epidemiology* 30.8 (Aug. 2015), pp. 661–708 (cit. on p. 148).

- 
- [31] T. A. G. M. Huisman et al. “Quantitative diffusion tensor MR imaging of the brain: field strength related variance of apparent diffusion coefficient (ADC) and fractional anisotropy (FA) scalars”. In: *European Radiology* 16.8 (Aug. 2006), pp. 1651–1658 (cit. on p. 148).
- [32] H. Johansen-Berg and T. E. J. Behrens. *Diffusion MRI*. Elsevier, 2009, p. 502 (cit. on p. 148).
- [33] D. K. Jones. *Diffusion MRI: Theory, Methods, and Applications*. Cambridge: Cambridge University Press, 2011, p. 784 (cit. on p. 148).
- [34] S. Kim, G. Chi-Fishman, A. S. Barnett, and C. Pierpaoli. “Dependence on diffusion time of apparent diffusion tensor of ex vivo calf tongue and heart”. In: *Magnetic Resonance in Medicine* 54.6 (Dec. 2005), pp. 1387–1396 (cit. on p. 168).
- [35] G. Kristo, A. Leemans, M. Raemaekers, G.-J. Rutten, B. de Gelder, and N. F. Ramsey. “Reliability of two clinically relevant fiber pathways reconstructed with constrained spherical deconvolution”. In: *Magnetic Resonance in Medicine* 70.6 (Dec. 2013), pp. 1544–1556 (cit. on p. 148).
- [36] S. Kullback and R. A. Leibler. “On Information and Sufficiency”. In: *The Annals of Mathematical Statistics* 22.1 (Mar. 1951), pp. 79–86 (cit. on p. 157).
- [37] D. Lakens. “Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs”. In: *Frontiers in Psychology* 4.NOV (2013), pp. 1–12 (cit. on p. 158).
- [38] D. J. Larkman, J. V. Hajnal, A. H. Herlihy, G. A. Coutts, I. R. Young, and G. Ehnholm. “Use of multicoil arrays for separation of signal from multiple slices simultaneously excited”. In: *Journal of Magnetic Resonance Imaging* 13.2 (Feb. 2001), pp. 313–317 (cit. on p. 148).
- [39] M. Lustig, D. Donoho, and J. M. Pauly. “Sparse MRI: The application of compressed sensing for rapid MR imaging”. In: *Magnetic Resonance in Medicine* 58.6 (Dec. 2007), pp. 1182–1195 (cit. on p. 148).
- [40] V. A. Magnotta et al. “MultiCenter Reliability of Diffusion Tensor Imaging”. In: *Brain Connectivity* 2.6 (Dec. 2012), pp. 345–355 (cit. on p. 148).
- [41] J. Mairal, F. Bach, J. Ponce, and G. Sapiro. “Online Learning for Matrix Factorization and Sparse Coding”. In: *Journal of Machine Learning Research* 11 (Aug. 2010), pp. 19–60 (cit. on pp. 149–151).
- [42] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman. “Non-local sparse models for image restoration”. In: *2009 IEEE 12th International Conference on Computer Vision*. Vol. 2. ICCV. IEEE, Sept. 2009, pp. 2272–2279 (cit. on p. 149).
- [43] S. Merlet, E. Caruyer, A. Ghosh, and R. Deriche. “A computational diffusion MRI and parametric dictionary learning framework for modeling the diffusion signal and its features”. In: *Medical Image Analysis* 17.7 (May 2013), pp. 843–830 (cit. on p. 150).
- [44] H. Mirzaalian et al. “Inter-site and inter-scanner diffusion MRI data harmonization”. In: *NeuroImage* 135 (July 2016), pp. 311–323 (cit. on pp. 148, 167).
- [45] L. Ning et al. “Multi-shell Diffusion MRI Harmonisation and Enhancement Challenge (MUSHAC): Progress and Results”. In: ed. by E. Kaden, F. Grussu, L. Ning, C. M. W. Tax, and J. Veraart. *Mathematics and Visualization*. Cham: Springer International Publishing, 2019, pp. 217–224 (cit. on p. 170).
- [46] M. Paquette, C. Eichner, and A. Anwender. “Gradient Non-Linearity Correction for Spherical Mean Diffusion Imaging”. In: *International Symposium on Magnetic Resonance in Medicine (ISMRM’19)*. 2019, p. 550 (cit. on p. 168).
- [47] O. Pasternak, N. Sochen, Y. Gur, N. Intrator, and Y. Assaf. “Free water elimination and mapping from diffusion MRI”. In: *Magnetic Resonance in Medicine* 62.3 (Sept. 2009), pp. 717–730 (cit. on p. 156).

## Bibliography

---

- [48] C. Pierpaoli and D. K. Jones. “Removing CSF Contamination in Brain DT-MRIs by Using a Two-Compartment Tensor Model”. In: *Proceedings of the 12th Annual Meeting of ISMRM, Kyoto*. 2004, p. 1215 (cit. on p. 156).
- [49] K. M. Pohl et al. “Harmonizing DTI measurements across scanners to examine the development of white matter microstructure in 803 adolescents of the NCANDA study”. In: *NeuroImage* 130 (Apr. 2016), pp. 194–213 (cit. on p. 149).
- [50] W. Qin et al. “Effects of echo time on diffusion quantification of brain white matter at 1.5T and 3.0T”. In: *Magnetic Resonance in Medicine* 61.4 (Apr. 2009), pp. 755–760 (cit. on p. 168).
- [51] T. Rohlfing. “Image similarity and tissue overlaps as surrogates for image registration accuracy: Widely used but unreliable”. In: *IEEE Transactions on Medical Imaging* 31.2 (2012), pp. 153–163 (cit. on p. 169).
- [52] A. Rueda, N. Malpica, and E. Romero. “Single-image super-resolution of brain MR images using overcomplete dictionaries”. In: *Medical Image Analysis* 17.1 (Jan. 2013), pp. 113–132 (cit. on p. 152).
- [53] K. Sakaie, X. Zhou, J. Lin, J. Debbins, M. Lowe, and R. J. Fox. “Technical Note: Retrospective reduction in systematic differences across scanner changes by accounting for noise floor effects in diffusion tensor imaging”. In: *Medical Physics* 45.9 (Sept. 2018), pp. 4171–4178 (cit. on pp. 148, 170).
- [54] S. Schiavi, M. Barakovic, M. O. Pineda, M. Descoteaux, J.-P. Thiran, and A. Daducci. “Reducing false positives in tractography with microstructural and anatomical priors”. In: *bioRxiv* (2019) (cit. on p. 170).
- [55] E. Schwab, R. Vidal, and N. Charon. “Joint spatial-angular sparse coding for dMRI with separable dictionaries”. In: *Medical Image Analysis* 48 (Aug. 2018), pp. 25–42 (cit. on pp. 150, 170).
- [56] G. Schwarz. “Estimating the dimension of a model”. In: *The Annals of Statistics* 6.2 (1978), pp. 461–464 (cit. on p. 166).
- [57] N. Simon, J. Friedman, T. Hastie, and R. Tibshirani. “A Sparse-Group Lasso”. In: *Journal of Computational and Graphical Statistics* 22.2 (Apr. 2013), pp. 231–245 (cit. on p. 170).
- [58] S. St-Jean, P. Coupé, and M. Descoteaux. “Non Local Spatial and Angular Matching: Enabling higher spatial resolution diffusion MRI datasets through adaptive denoising”. In: *Medical Image Analysis* 32.2016 (Aug. 2016), pp. 115–130 (cit. on pp. 149–151, 166).
- [59] S. St-Jean, A. De Luca, M. A. Viergever, and A. Leemans. “Automatic, Fast and Robust Characterization of Noise Distributions for Diffusion MRI”. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*. Ed. by A. F. Frangi, J. A. Schnabel, C. Davatzikos, C. Alberola-López, and G. Fichtinger. Springer International Publishing, May 2018, pp. 304–312 (cit. on pp. 148, 170).
- [60] S. St-Jean, M. A. Viergever, and A. Leemans. “A unified framework for upsampling and denoising of diffusion MRI data”. In: *25th Annual Meeting of ISMRM* (2017), p. 3533 (cit. on pp. 149, 152).
- [61] S. St-Jean, M. A. Viergever, and A. Leemans. *samuelstjean/harmonization: Harmonization of diffusion MRI datasets with adaptive dictionary learning*. Sept. 2019.  
URL: <https://doi.org/10.5281/zenodo.3385925> (cit. on pp. 152, 171).
- [62] M. Stone. “An Asymptotic Equivalence of Choice of Model by Cross-Validation and Akaike’s Criterion”. In: *Journal of the Royal Statistical Society: Series B (Methodological)* 39.1 (Sept. 1977), pp. 44–47 (cit. on p. 166).
- [63] Student. “The Probable Error of a Mean”. In: *Biometrika* 6.1 (1908), pp. 1–25 (cit. on p. 157).

- 
- [64] C. M. W. Tax et al. “Cross-scanner and cross-protocol diffusion MRI data harmonisation: A benchmark database and evaluation of algorithms”. In: *NeuroImage* 195 (July 2019), pp. 285–299 (cit. on pp. 148, 149, 151, 154–156, 166, 168, 169).
- [65] R. J. Tibshirani and J. Taylor. “Degrees of freedom in lasso problems”. In: *The Annals of Statistics* 40.2 (Apr. 2012), pp. 1198–1232 (cit. on p. 151).
- [66] J.-D. Tournier. “Diffusion MRI in the brain – Theory and concepts”. In: *Progress in Nuclear Magnetic Resonance Spectroscopy* 112–113 (June 2019), pp. 1–16 (cit. on p. 148).
- [67] J.-D. Tournier, S. Mori, and A. Leemans. “Diffusion tensor imaging and beyond.” In: *Magnetic Resonance in Medicine* 65.6 (2011), pp. 1532–1556 (cit. on p. 148).
- [68] B. C. Vemuri et al. “A Geometric Framework for Ensemble Average Propagator Reconstruction from Diffusion MRI”. In: *Medical Image Analysis* (July 2019) (cit. on p. 150).
- [69] C. Vollmar et al. “Identical, but not the same: Intra-site and inter-site reproducibility of fractional anisotropy measures on two 3.0 T scanners”. In: *NeuroImage* 51.4 (July 2010), pp. 1384–1394 (cit. on p. 148).
- [70] J. M. Wolterink, A. M. Dinkla, M. H. F. Savenije, P. R. Seevinck, C. A. T. van den Berg, and I. Išgum. “Deep MR to CT Synthesis Using Unpaired Data”. In: *Lecture Notes in Computer Science*. Vol. 10557 LNCS. 2017, pp. 14–23 (cit. on p. 168).
- [71] J. Yang, Z. Wang, Z. Lin, S. Cohen, and T. Huang. “Coupled Dictionary Training for Image Super-Resolution”. In: *IEEE Transactions on Image Processing* 21.8 (Aug. 2012), pp. 3467–3478 (cit. on p. 149).
- [72] J. Yang, J. Wright, T. Huang, and Y. Ma. “Image Super-Resolution Via Sparse Representation”. In: *IEEE Transactions on Image Processing* 19.11 (Nov. 2010), pp. 2861–2873 (cit. on p. 149).
- [73] A. H. Zhu, D. C. Moyer, T. M. Nir, P. M. Thompson, and N. Jahanshad. “Challenges and Opportunities in dMRI Data Harmonization”. In: *Computational Diffusion MRI MICCAI 2019*. Ed. by E. Kaden, F. Grussu, L. Ning, C. M. W. Tax, and J. Veraart. Mathematics and Visualization. Cham: Springer International Publishing, 2019, pp. 157–172 (cit. on p. 148).
- [74] H. Zou, T. Hastie, and R. Tibshirani. “On the “degrees of freedom” of the lasso”. In: *The Annals of Statistics* 35.5 (Oct. 2007), pp. 2173–2192 (cit. on p. 151).

- *Il nous reste un modèle de l'an passé, c'est sûr yé 1500\$ moins cher, mais vous avez pas le choix des couleurs.*
- *M'en sacre des couleurs si yé moins cher. Yé quelle couleur?*
- *Ok, hum, brun avec des barres brunes, le siège est turquoise avec des télétubbies en relief. Pis en avant s't'écrit 'vroum vroum pipi boum-boum'.*
- *Ben sais-tu j'pense que j'va...*
- *En lettrage sport.*
- *On va en prendre un de l'année, hein Mona?*

Le gars qui magasine



## Summary and discussion

### 7.1 Summary

Through this thesis, we have seen how diffusion MRI can complement anatomical MRI by providing information about the diffusivity of in vivo tissues in a noninvasive way. By devising adequate biophysical models explaining the measured signal, these diffusivity values can then be used to produce scalar maps and infer the architecture and abnormalities of the underlying tissues. Analysis of the statistical properties of the signal (e.g., the moments and the kurtosis) can also be used to produce scalar maps that are not intrinsically tied to a particular model but can still reveal abnormalities in the underlying tissues.

Chapter 2 showed how high resolution diffusion MRI datasets acquired on a standard scanner can be used to improve anatomical accuracy, even though the low signal level associated with smaller voxels would normally preclude a direct analysis. Through a new denoising algorithm exploiting spatial and angular redundancy of the multiple volumes obtained in routine diffusion MRI acquisitions, we have employed a dictionary learning algorithm coupled with signal bias correction and an iterative  $\ell_1$  reweighting scheme. The reconstruction naturally discards the unwanted noise associated with low SNR through using an upper bound on the  $\ell_2$  norm, which depends on the local noise variance. Experiments on synthetic datasets have shown that the error in the recovered diffusion metrics (such as the FA and the ADC) by the method is lower than the error produced by the compared denoising algorithms. This property of the algorithm qualitatively transferred to in vivo data, where additional anatomical details and enhanced tractography are easily identifiable on a 1.2 mm isotropic dataset when compared to the original, noisy 1.2 mm version or on a comparable (in terms of acquisition time) 1.8 mm dataset of the same subject. Note that the lower resolution dataset had an increased SNR and 64 diffusion volumes, whereas the 1.2 mm dataset only contained 40 diffusion volumes.

Chapters 3 and 4 presented an enhancement to along-tract types of analysis. Chapter 3 first showed that direct geometrical averaging based on the Cartesian coordinates of metrics extracted along-tract can lead to a mismatch notably in the presence of splitting (e.g. arcuate fasciculus) or fanning (e.g. corticospinal tract) fiber bundle configurations. Assigning values using orthogonal cross-sections instead can help to resolve this issue since points are now assigned locally towards a representative streamline instead of considering their absolute coordinates in space. After extraction of a representative streamline for each subject, Chapter 4 presented a new algorithm for the realignment of those representative streamlines. As tractography and subsequent extraction of bundles of interest are realized on every subject separately, there is no guarantee that coordinates in the 1D space of analysis are matching for every subject. By finding a template candidate amongst all subjects, only the overlapping segments (up to a user-defined threshold) are kept after realignment for further statistical analysis. We have shown that such a strategy reduces the coefficient of variation of the metrics of interest (in this case the mean diffusivity, fractional anisotropy and apparent fiber density) in synthetic experiments and on a database of 100 in vivo datasets. On those same in vivo datasets, experiments inducing alterations on half of the subjects conducted on realigned profiles helped to uncover the region affected by those alterations, which was not always possible in the non-realigned case. This conclusion also holds when the alterations in the extracted scalar values are large and only cover a small region of the whole tract length.

Chapter 5 brought us back to the acquisition by presenting a new automated method to estimate the signal distribution from repeated magnitude MRI measurements, such as the multiple diffusion weighted images required for local modeling in diffusion MRI. The main advantage of the method lies in the fact that it does not rely on external measurements, such as coil sensitivities or reconstruction matrices, which are usually not recorded at acquisition time. Multiple experiments on simulated phantoms with different coil simulations (including two different accelerated parallel imaging algorithms) have shown that the noise distribution can be recovered successfully without information about the acquisition process. Experiments on an acquired phantom of a water bottle showed that the method is robust to signal leakage due to multiband imaging as long as the acceleration factor is not too high i.e. about 3 in our experiments. We also analyzed two in vivo datasets acquired in two separate centers, including four repetitions of the same subject, which are publicly available online. Results showed that the proposed method is stable on those four datasets and identified the expected signal distribution, while at the same time discarding voxels contaminated by artifacts. The second in vivo dataset also gave results in line with the expected theoretical signal distribution according to the parallel MRI reconstruction that was used during the acquisition.

Finally, Chapter 6 presented a new algorithm to harmonize diffusion MRI datasets acquired with different scanners. Using dictionary learning, we have shown how features can be automatically extracted from datasets acquired on a target scanner and used to re-

construct datasets from a different source scanner, removing variability attributable to both scanners in the process while preserving anatomical variability. Experiments with a publicly available database showed how the algorithm reduces variability on common diffusion MRI metrics and can even be used if the spatial resolution of the scanners is not identical through adequate spatial subsampling of the dictionary. To verify if the removed variability was not due to genuine anatomical differences, we also generated altered datasets contaminated artificially by free water and mimicking edema. The results supported the hypothesis of the variability originating solely from the scanner as it decreased between scanners while preserving the effect size when comparing the altered dataset with their original, unaltered counterpart using paired t-tests on the four studied diffusion metrics. The effect size was, on average, also in the same 95% confidence interval as the untouched datasets, confirming that the harmonization process did not remove anatomical variability in its reconstruction when comparing harmonized versus original datasets.

## 7.2 Discussion

### 7.2.1 Theory

Each chapter of this thesis introduced a new method or algorithm to enhance the analysis and comprehension of diffusion MRI datasets. While the results contained in this thesis are a good first step, it is obviously not the last stopping point for diffusion MRI in general. One of the central goals of diffusion MRI is to predict the diffusivity of the tissues, namely through local models applied in each voxel. While there is no shortage of said models, there is unfortunately no consensus on one model to use in particular. A few models are the usual “go to” candidates such as diffusion tensor imaging (DTI) (Basser et al., 1994), diffusion kurtosis imaging (DKI) (Jensen et al., 2005) and constrained spherical deconvolution (CSD) (Tournier et al., 2007) as they provide easy to understand metrics or orientational information, respectively. More advanced choices such as the neurite orientation dispersion and density imaging (NODDI) (Zhang et al., 2012) or the mean apparent propagator (MAP) MRI (Özarslan, Koay, Shepherd, et al., 2013) can further provide specific measures not covered by the earlier models. Interestingly, those four models besides NODDI are in fact an approximation of the diffusion signal as measured in q-space, with NODDI instead explicitly modeling the diffusivities using a compartmental model. Nevertheless, the denoising and harmonization methods presented in Chapters 2 and 6 are applied directly on the DWIs and are therefore compatible with any model of interest. While we primarily used DTI and CSD throughout this thesis for their stability and ability to work with single shell datasets, the release of high-quality multishell datasets such as the human connectome project (HCP) database (Van Essen, Ugurbil, et al., 2012; Van Essen, Smith, et al., 2013) can use (and benefit from) advanced models (e.g., DKI, MAP MRI) and the additional information, such as additional scalar maps, which can be extracted using those models. However, usage of such advanced datasets is not necessarily common outside of

the circles developing new dMRI methods due to the inherently longer scan times needed to collect more data and the longer TE required to achieve higher diffusion weighting, leading to a further decrease in SNR for the higher b-value images (Froeling et al., 2017).

This chicken-and-egg problem leads to conservative usage of DTI in clinical studies, or even only measuring the ADC by geometrically averaging measurements, as these measures are robust due to their lower number of parameters. Advanced models are harder to analyze for the uninitiated scientist and can be subject to numerical issues due to the complexity and interdependence of the parameter space (Jelescu et al., 2016; Novikov, Veraart, et al., 2018). To widen the usage of advanced models and their promise of increased specificity, more work is needed by the diffusion community to validate their usage, applicability and showcase how they can be used in e.g. diseased population by providing specific biomarkers. We are not there yet as no consensus in the community exists to suggest a default model to use, much less on the acquisition strategy required to support such a hypothetical model. Providing “textbook” acquisition schemes that are agreed upon for a few models could be a great start to widen their usage, just as it is usually recommended to acquire at least 30 DWIs for DTI acquisitions (Jones and Basser, 2004). Acquiring a lower number of DWIs is also possible as long as it can support the desired diffusion model (Lebel et al., 2012). Nevertheless, collecting additional DWIs helps in reducing variability due to measurements, but the tipping point where acquiring more data only lengthens the acquisition is not always clear, especially e.g. in the presence of motion, signal dropouts or artifacts as they may render useless a portion of the collected datasets (Jones, Knösche, et al., 2013; Tax et al., 2015). The denoising algorithm presented in Chapter 2 of this thesis can however be used on these single-shell datasets with nothing precluding its application, from a theory point of view, to more advanced acquisitions. Since the algorithm presented in Chapter 6 also uses the same foundation, applicability to multishell datasets should also be straightforward, although some extensions to exploit this new information are discussed in the chapter itself. This would obviously require further validation to ensure that the improvements showcased here for single-shell datasets apply as well to multishell datasets. As for the noise estimation algorithm from Chapter 5, it has already been shown to work on both single- and multishell datasets. This is because it will automatically identify voxels from the background distribution, which are similar since the signal is independent of the diffusion weighting in these regions. While it would indeed be useful to have a version that is additionally valid over signal regions, the distribution of each and every voxel would be different, especially across diffusion weighting. Such an extension would be easier to include in the context of local modeling, as its goal is to explain the measured diffusion signal itself, by replacing the value of the signal by its desired parametric formulation. It could be possible, for example, to initialize such an algorithm with the values computed from our proposed algorithm as a first step and refine the estimation of the signal distribution locally afterward. Finally, the optimal assignment strategy and realignment algorithm from Chapters 3 and 4 can also be used with any diffusion model as they come into play after tractography, making them

both independent of prior choices regarding data processing steps.

Unfortunately, the perfect “press button” acquisition scheme and assorted local model that tells us everything we would like to know from diffusion MRI with a two minutes acquisition protocol does not exist yet. While new developments using deep learning may enable faster imaging (Golkov et al., 2016), there is still prior work to be done in validation (e.g. in tractography to reduce false positive connections (Maier-Hein et al., 2017)), and in fundamental theoretical modeling of diffusion MRI (Novikov, Kiselev, et al., 2018), which may come through new sequences beyond the original 1965 Stejskal-Tanner experiment (Stejskal and Tanner, 1965).

### 7.2.2 Recent developments in diffusion MRI

Recently, new developments in diffusion MRI acquisition have highlighted limitations in the classical Stejskal-Tanner sequence (Westin et al., 2016). Designed under the umbrella term of multidimensional diffusion MRI, these sequences allow, e.g., measuring the covariance of a distribution of diffusion tensors, instead of the classical measures which result in the average of the distribution over each voxel. Additional information can be obtained through higher moments of the diffusion propagator, which is not available to the classical single diffusion encoding sequence (Novikov, Fieremans, et al., 2019). While it may help to provide more specific information about the macroscopic content of a given voxel through new metrics such as the micro FA, preliminary results on clinical scanners have shown encouraging results in, e.g., differentiating meningioma from glioblastoma. Szczepankiewicz et al. (2015) have shown that the FA maps were lower for both type of tumors, but the micro FA map showed that the meningioma exhibited coherent structure, whereas the glioblastoma did not, resulting in high and low micro FA values, respectively.

The sequence from Szczepankiewicz et al. (2015) was using a gradient strength of 80 mT/m and a voxel size of 3 mm isotropic to preserve acceptable SNR due to the much larger TE of 160 ms and b-value up to  $b = 2800 \text{ s/mm}^2$  required for the encoding. For comparison, the experiments from Chapter 2 used a gradient strength of 45 mT/m and a voxel size of 1.8 mm isotropic with a TE of 63 ms for the baseline scan. Furthermore, the 1.2 mm isotropic sequence still only required a TE of 104 ms while keeping the b-value at  $b = 1000 \text{ s/mm}^2$ . These multidimensional encoding sequences still require stronger diffusion gradients hardware than available on most current clinical scanners (about 40 mT/m) to achieve an acceptable SNR, but are still fairly new and could be further optimized for efficiency (Westin et al., 2016). As they also enable access to additional scanning parameters, these sequences are prime candidates for the harmonization algorithm presented in Chapter 6 if they are to be used for large-scale studies with various scanners and reconstruction methods. The denoising algorithm from Chapter 2 could also be employed to circumvent the rather large voxel size and increase the spatial sensitivity of the method, similarly to the processing used on the high spatial resolution dataset of 1.2 mm isotropic to recover additional anatomical details which were hidden by the low SNR. The noise estimation

method from [Chapter 5](#) could also be used to inform the denoising algorithm or even the model fitting procedure subsequently used in multidimensional diffusion MRI.

A 3 minutes protocol has also recently been presented on a database of 42 patients with intracranial tumors to measure anisotropic and isotropic kurtosis, bringing these new developments closer to clinical applications (Nilsson et al., 2019). As orientational information required for tractography is, so far, not obtained from these sequences, relying on other algorithms using multiple b-values such as multi-tissue CSD (De Luca et al., 2019; Jeurissen et al., 2014) can be used as a substitute since the multidimensional encoding also typically uses multiple b-values. Combining tractography with these additional covariance measures of diffusion could be analyzed using the along-tract framework extensions presented in [Chapters 3](#) and [4](#) instead of the classical ROI analysis (Nilsson et al., 2019). This would provide additional information (e.g. for neurosurgical applications) such as the location of the affected region in a white matter bundle or even if a tumor is infiltrative or is instead “pushing back” the white matter structure (Chamberland et al., 2014).

### 7.2.3 Future directions

While there is recently renewed interest in revisiting the standard pulse sequences used in diffusion MRI, another aspect worth exploring is the field of numerical methods and optimization. Indeed, newer models oftentimes require nonlinear optimization due to the complexity and relationship between their parameters, requiring additional constraints to increase numerical stability (Novikov, Kiselev, et al., 2018; Novikov, Veraart, et al., 2018). Recent analytical developments have shown how a single fiber population can be written as the product of the spherical harmonics (SH) basis and a Legendre polynomial representation of the kernel for this single fiber population (Novikov, Veraart, et al., 2018). This leads to a set of nonlinear equations based on the moments with fewer parameters, but still describing the full single fiber population and the extracellular space around it for each voxel. Unfortunately, this representation also highlighted a fundamental degeneracy in the parameter estimation space where parameters are in fact correlated and interdependent, yielding spurious solutions that are invalid, but still biophysically plausible. Determining which set of parameters is correct, as it likely varies for each voxel, is currently an active topic of research.

Representation of the signal using the SH basis has applications in diffusion MRI such as modeling the orientation distribution function (ODF) (Descoteaux, Angelino, et al., 2007; Tournier et al., 2007) or the diffusion propagator (Descoteaux, Deriche, et al., 2011; Özarslan, Koay, and Basser, 2013). These approaches rely on linear algebra, which can be solved using classical least-squares methods, and scalar maps of interest are subsequently derived from the coefficients. Explicit products of the SH part with Legendre polynomials can also be used to model directly the diffusion signal instead of separating the kernel and ODF (Jespersen et al., 2007). The spherical mean technique (SMT) instead factors out the ODF by averaging the measurements, enabling quantification of diffusivity of the fiber

population without confounds due to dispersion (Kaden et al., 2016). The SMT and the approach of Jespersen et al. (2007), however, rely on nonlinear optimization, which is prone to degeneracies and numerical issues since changing one parameter will affect differently the values of the remaining parameters, in addition to the increased computational burden when compared to linear methods. However, it may very well be possible to combine these ideas in a framework that relies on linear optimization, thereby reducing the complexity and simplifying analysis of the error propagation. The answer to these questions may in fact come from other domains of research where fitting of exponential data is ubiquitous. Namely, approaches using a polynomial representation (such as Legendre polynomials) to fit exponential decays through differential equations (Knisley and Glenn, 1996; Martin et al., 1994) may offer new insights about the difficulties faced currently in diffusion MRI. It is possible, using solely the coefficients of the polynomial, to solve a linear system of equations and obtain the exponential decay rate associated with the original signal. In the case of diffusion MRI, the 1D signal can be represented as a Legendre polynomial. Coefficients of the polynomial can be used to derive the diffusivities with a combination of linear algebra and root finding. Preliminary results (St-Jean, 2019) have shown how this approach can be used to obtain ADC maps for a two compartments model in the whole brain in about 2 minutes. This is done on a 1D version of the diffusion signal obtained by geometrically averaging the DWIs at each b-value. Solving explicitly the same equation with traditional nonlinear optimization requires about 10 hours and yields visually identical results. Using the SH basis, this 1D preliminary approach could likely be extended to also incorporate the ODF since the product of the SH basis and Legendre polynomials results in another set of Legendre polynomials (Jespersen et al., 2007). This would lead, in theory, to a per-tissue diffusivity measure (and assorted scalar maps) and ODF, without the need to fix a per-tissue kernel as is done in spherical deconvolution (De Luca et al., 2019; Jeurissen et al., 2014). As the process is fully determined once the polynomial coefficients have been computed, designing time-optimal acquisition schemes regarding the number of required DWIs and the set of optimal b-values to ensure the stability of the fitting procedure becomes feasible. This could help widen usage of the concepts discussed previously since it is both faster than traditional nonlinear fitting and the numerical stability of the process is easier to understand as only linear sets of equations are involved. Coupled with multidimensional MRI, these ideas could offer new and promising insights in studying the brain while reducing the acquisition burden at the scanner and subsequent processing time to obtain relevant information from diffusion MRI.

# Bibliography

- [1] P. J. Basser, J. Mattiello, and D. LeBihan. “MR diffusion tensor spectroscopy and imaging.” In: *Biophysical journal* 66.1 (Jan. 1994), pp. 259–67 (cit. on p. 180).
- [2] M. Chamberland, K. Whittingstall, D. Fortin, D. Mathieu, and M. Descoteaux. “Real-time multi-peak tractography for instantaneous connectivity display.” English. In: *Frontiers in neuroinformatics* 8.May (Jan. 2014), p. 59 (cit. on p. 183).
- [3] A. De Luca, F. Guo, and A. Leemans. “Estimation of multiple fiber orientation distributions (mFODs) from diffusion MRI data using spherical deconvolution”. In: *International Society for Magnetic Resonance in Medicine (ISMRM)*. 2019, p. 163 (cit. on pp. 183, 184).
- [4] M. Descoteaux, E. Angelino, S. Fitzgibbons, and R. Deriche. “Regularized, fast, and robust analytical Q-ball imaging”. In: *Magnetic Resonance in Medicine* 58.3 (Sept. 2007), pp. 497–510 (cit. on p. 183).
- [5] M. Descoteaux, R. Deriche, D. Le Bihan, J.-F. Mangin, and C. Poupon. “Multiple q-shell diffusion propagator imaging”. In: *Medical Image Analysis* 15.4 (Aug. 2011), pp. 603–621 (cit. on p. 183).
- [6] M. Froeling, C. M. Tax, S. B. Vos, P. R. Luijten, and A. Leemans. ““MASSIVE” brain dataset: Multiple acquisitions for standardization of structural imaging validation and evaluation”. In: *Magnetic Resonance in Medicine* 77.5 (May 2017), pp. 1797–1809 (cit. on p. 181).
- [7] V. Golkov et al. “q-Space Deep Learning: Twelve-Fold Shorter and Model-Free Diffusion MRI Scans”. In: *IEEE Transactions on Medical Imaging* 35.5 (May 2016), pp. 1344–1351 (cit. on p. 182).
- [8] I. O. Jelescu, J. Veraart, E. Fieremans, and D. S. Novikov. “Degeneracy in model parameter estimation for multi-compartmental diffusion in neuronal tissue”. In: *NMR in Biomedicine* 29.1 (2016), pp. 33–47 (cit. on p. 181).
- [9] J. H. Jensen, J. A. Helpert, A. Ramani, H. Lu, and K. Kaczynski. “Diffusional kurtosis imaging: The quantification of non-gaussian water diffusion by means of magnetic resonance imaging”. In: *Magnetic Resonance in Medicine* 53.6 (June 2005), pp. 1432–1440 (cit. on p. 180).
- [10] S. N. Jespersen, C. D. Kroenke, L. Østergaard, J. J. Ackerman, and D. A. Yablonskiy. “Modeling dendrite density from magnetic resonance diffusion measurements”. In: *NeuroImage* 34.4 (Feb. 2007), pp. 1473–1486 (cit. on pp. 183, 184).
- [11] B. Jeurissen, J.-D. Tournier, T. Dhollander, A. Connelly, and J. Sijbers. “Multi-tissue constrained spherical deconvolution for improved analysis of multi-shell diffusion MRI data”. In: *NeuroImage* 103 (Dec. 2014), pp. 411–426 (cit. on pp. 183, 184).
- [12] D. K. Jones, T. R. Knösche, and R. Turner. “White matter integrity, fiber count, and other fallacies: The do’s and don’ts of diffusion MRI”. In: *NeuroImage* 73 (June 2013), pp. 239–254 (cit. on p. 181).
- [13] D. K. Jones and P. J. Basser. ““Squashing peanuts and smashing pumpkins”: how noise distorts diffusion-weighted MR data.” In: *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine* 52.5 (Nov. 2004), pp. 979–93 (cit. on p. 181).

## Bibliography

---

- [14] E. Kaden, F. Kruggel, and D. C. Alexander. “Quantitative mapping of the per-axon diffusion coefficients in brain white matter”. In: *Magnetic Resonance in Medicine* 75.4 (Apr. 2016), pp. 1752–1763 (cit. on p. 184).
- [15] J. R. Knisley and L. L. Glenn. “A linear method for the curve fitting of multiexponentials”. In: *Journal of Neuroscience Methods* 67.2 (Aug. 1996), pp. 177–183 (cit. on p. 184).
- [16] C. Lebel, T. Benner, and C. Beaulieu. “Six is enough? Comparison of diffusion parameters measured using six or more diffusion-encoding gradient directions with deterministic tractography”. In: *Magnetic Resonance in Medicine* 68.2 (Aug. 2012), pp. 474–483 (cit. on p. 181).
- [17] K. H. Maier-Hein et al. “The challenge of mapping the human connectome based on diffusion tractography”. In: *Nature Communications* 8.1 (Dec. 2017), p. 1349 (cit. on p. 182).
- [18] J. L. Martin, D. J. Maconochie, and D. E. Knight. “A novel use of differential equations to fit exponential functions to experimental data”. In: *Journal of Neuroscience Methods* 51.2 (Mar. 1994), pp. 135–146 (cit. on p. 184).
- [19] M. Nilsson et al. “Tensor-valued diffusion MRI in under 3 minutes: an initial survey of microscopic anisotropy and tissue heterogeneity in intracranial tumors”. In: *Magnetic Resonance in Medicine* (Sept. 2019), mrm.27959 (cit. on p. 183).
- [20] D. S. Novikov, E. Fieremans, S. N. Jespersen, and V. G. Kiselev. “Quantifying brain microstructure with diffusion MRI: Theory and parameter estimation”. In: *NMR in Biomedicine* 32.4 (Apr. 2019), e3998 (cit. on p. 182).
- [21] D. S. Novikov, V. G. Kiselev, and S. N. Jespersen. “On modeling”. In: *Magnetic Resonance in Medicine* 79.6 (June 2018), pp. 3172–3193 (cit. on pp. 182, 183).
- [22] D. S. Novikov, J. Veraart, I. O. Jelescu, and E. Fieremans. “Rotationally-invariant mapping of scalar and orientational metrics of neuronal microstructure with diffusion MRI”. In: *NeuroImage* 174 (July 2018), pp. 518–538 (cit. on pp. 181, 183).
- [23] E. Özarslan, C. G. Koay, and P. J. Basser. “Simple Harmonic Oscillator Based Reconstruction and Estimation for One-Dimensional q-Space Magnetic Resonance (1D-SHORE)”. In: *Excursions in Harmonic Analysis, Volume 2*. Ed. by T. D. Andrews, R. Balan, J. J. Benedetto, W. Czaja, and K. A. Okoudjou. Vol. 1. Boston: Birkhäuser Boston, 2013, pp. 373–399 (cit. on p. 183).
- [24] E. Özarslan, C. G. Koay, T. M. Shepherd, et al. “Mean apparent propagator (MAP) MRI: A novel diffusion imaging method for mapping tissue microstructure”. In: *NeuroImage* 78 (2013), pp. 16–32 (cit. on p. 180).
- [25] S. St-Jean. “Towards analytic quantitative MRI”. In: *ISI Group meeting Unpublished results* (2019) (cit. on p. 184).
- [26] E. O. Stejskal and J. E. Tanner. “Spin Diffusion Measurements: Spin Echoes in the Presence of a Time-Dependent Field Gradient”. In: *The Journal of Chemical Physics* 42.1 (Jan. 1965), pp. 288–292 (cit. on p. 182).
- [27] F. Szczepankiewicz et al. “Quantification of microscopic diffusion anisotropy disentangles effects of orientation dispersion from microstructure: Applications in healthy volunteers and in brain tumors”. In: *NeuroImage* 104 (Jan. 2015), pp. 241–252 (cit. on p. 182).
- [28] C. M. W. Tax, W. M. Otte, M. A. Viergever, R. M. Dijkhuizen, and A. Leemans. “REKINDLE: Robust Extraction of Kurtosis INDices with Linear Estimation”. In: *Magnetic Resonance in Medicine* 73.2 (2015), pp. 794–808 (cit. on p. 181).

- 
- [29] J.-D. Tournier, F. Calamante, and A. Connelly. “Robust determination of the fibre orientation distribution in diffusion MRI: non-negativity constrained super-resolved spherical deconvolution.” In: *NeuroImage* 35.4 (May 2007), pp. 1459–72 (cit. on pp. 180, 183).
- [30] D. C. Van Essen, K. Ugurbil, et al. “The Human Connectome Project: A data acquisition perspective”. In: *NeuroImage* 62.4 (2012), pp. 2222–2231 (cit. on p. 180).
- [31] D. C. Van Essen, S. M. Smith, D. M. Barch, T. E. Behrens, E. Yacoub, and K. Ugurbil. “The WU-Minn Human Connectome Project: An overview”. In: *NeuroImage* 80 (Oct. 2013), pp. 62–79 (cit. on p. 180).
- [32] C.-F. Westin et al. “Q-space trajectory imaging for multidimensional diffusion MRI of the human brain”. In: *NeuroImage* 135 (July 2016), pp. 345–362 (cit. on p. 182).
- [33] H. Zhang, T. Schneider, C. A. Wheeler-Kingshott, and D. C. Alexander. “NODDI: Practical in vivo neurite orientation dispersion and density imaging of the human brain”. In: *NeuroImage* 61.4 (July 2012), pp. 1000–1016 (cit. on p. 180).

# Samenvatting

Dit proefschrift laat zien hoe diffusie MRI een aanvulling kan zijn op anatomische MRI door informatie te verschaffen over de diffusie van in vivo weefsels op een niet-invasieve manier. Door geschikte biofysische modellen te ontwikkelen die het gemeten signaal verklaren, kunnen deze diffusie eigenschappen vervolgens worden gebruikt om de architectuur en afwijkingen van de onderliggende weefsels af te leiden. Analyse van de statistische eigenschappen van het signaal (bijvoorbeeld de momenten en de kurtosis van dit signaal) kan ook worden gebruikt om informatie te verkrijgen die niet intrinsiek gekoppeld is aan een bepaald model, maar nog steeds afwijkingen in de onderliggende weefsels kan onthullen.

Na het inleidend [Hoofdstuk 1](#), laat [Hoofdstuk 2](#) zien hoe hoge-resolutie diffusie MRI datasets, verkregen met een klinische scanner, kunnen worden gebruikt om de anatomische nauwkeurigheid van afgebeelde structuren te verbeteren, ondanks het lage signaalniveau gerelateerd aan kleinere voxels, wat een directe analyse uitsluit. Door een nieuwe methode van ruisonderdrukking, die gebruik maakt van ruimtelijke en angulaire redundantie van informatie van de meerdere diffusie MRI volumes verkregen bij diffusie MRI acquisities, hebben we een “dictionary learning” algoritme kunnen combineren met signaalbiascorrectie en een iteratief  $\ell_1$ -weging schema. De reconstructie negeert de ongewenste ruis geassocieerd met lage SNR door gebruik te maken van een limiet op de  $\ell_2$ -norm die afhankelijk is van de lokale ruisvariantie. Experimenten met synthetische datasets tonen aan dat de fout in de door de methode gevonden diffusiematen (zoals FA en ADC) lager is dan die verkregen uit de andere algoritmen voor ruisonderdrukking. Deze eigenschap van het algoritme blijkt ook op te gaan voor isotrope 1.2 mm in vivo data, waarbij meer anatomische details en verbeterde tractografie herkenbaar zijn dan in de originele, ruizige versie van de 1.2 mm data of de vergelijkbare (wat betreft acquisitietijd) 1.8 mm data van dezelfde proefpersoon. De dataset met lagere resolutie had een verhoogde SNR en 64 diffusie MRI-volumes, terwijl de 1.2 mm dataset slechts 40 diffusie MRI-volumes bevatte.

[Hoofdstukken 3](#) en [4](#) presenteren een verbetering van de analyse waarbij langs het traject van de vezelbundels wordt gekeken. [Hoofdstuk 3](#) toont eerst aan dat directe geometrische middeling op basis van de Cartesiaanse coördinaten van diffusie eigenschappen, geëxtraheerd langs het pad, kan leiden tot een mismatch, met name in de aanwezigheid van splitsende (bijvoorbeeld bij de fasciculus arcuatus) of uitwaaiierende (bijvoorbeeld bij de corticospinale vezelbanen) vezelbundelconfiguraties. Door in plaats daarvan orthogonale doorsneden te gebruiken, kan dit probleem worden opgelost: datapunten kunnen nu lokaal worden toegewezen aan een representatief pad in plaats van dat hun absolute coördinaten in de ruimte beschouwd worden.

[Hoofdstuk 4](#) beschrijft een nieuw algoritme voor de herschikking van representatieve hersenvezelpaden die niet gealigneerd zijn. Aangezien tractografie en daaropvolgende extractie van specifieke bundels voor elke persoon afzonderlijk worden gerealiseerd, is er geen garantie dat coördinaten in de 1D analyseruimte tussen personen overeenkomen. Door het vinden van een voor alle personen representatief sjabloon, wordt na het aligneren alleen het overlappend segment (met een drempel door de gebruiker gedefinieerd) bewaard voor verdere statistische analyse. We tonen met simulaties en een database van 100 personen aan dat een dergelijke strategie de variatie van de relevante diffusie eigenschappen (e.g., gemiddelde diffusie en fractionele anisotropie) vermindert. Met diezelfde in vivo datasets kunnen we aantonen dat onze methode verschillen tussen groepen kan identificeren, wat zonder deze correctie niet mogelijk was.

In [Hoofdstuk 5](#) wordt een nieuwe geautomatiseerde methode gepresenteerd die de signaalverdeling van herhaalde magnitude MRI metingen, zoals de meerdere diffusie-gewogen beelden die nodig zijn voor lokale modellering in diffusie MRI, kan schatten. Het belangrijkste voordeel van de methode ligt in het feit dat deze niet afhankelijk is van externe metingen, zoals spoelgevoeligheden of reconstructiematrices, die meestal niet worden geregistreerd op het moment van acquisitie. Meerdere experimenten met gesimuleerde fantomen met verschillende spoelsimulaties (inclusief twee verschillende versnelde parallelle beeldvormingsalgoritmen) tonen aan dat de ruisverdeling met succes kan worden bepaald zonder informatie over het acquisitieproces. Experimenten met een fantoom van een waterfles tonen aan dat de methode robuust imperfecties als gevolg van “multiband” beeldvorming kan signaleren, zolang de versnellingsfactor niet te hoog is, d.w.z. ongeveer 3 in onze experimenten. We hebben ook twee publiek beschikbare in vivo datasets geanalyseerd die zijn verkregen in verschillende instituten, waaronder een met vier acquisities van dezelfde persoon. De resultaten tonen aan dat de voorgestelde methode stabiel is op deze vier datasets en de verwachte signaalverdeling identificeert, terwijl voxels die door artefacten zijn aangetast, worden weggelaten. De tweede in vivo dataset geeft ook resultaten die overeenkomen met de verwachte theoretische signaalverdeling volgens de parallelle MRI reconstructie die werd gebruikt tijdens de acquisitie.

Ten slotte laat [Hoofdstuk 6](#) een nieuw algoritme zien voor het harmoniseren van diffusie MRI datasets die op verschillende scanners zijn verkregen. Met behulp van “dictionary learning” tonen we aan hoe eigenschappen automatisch kunnen worden geëxtraheerd uit datasets die zijn verkregen op een specifieke MRI scanner en kunnen worden gebruikt om datasets van een andere scanner te reconstrueren. Hierbij worden de variaties die aan de scanners kunnen worden toegeschreven verwijderd, met behoud van de anatomische variaties. Experimenten met een in vivo database laten zien hoe het algoritme de variaties in diffusie MRI eigenschappen vermindert en zelfs kan worden gebruikt als de ruimtelijke resolutie van de scanners niet identiek is. Om te controleren of de verwijderde variaties niet te wijten zijn aan echte anatomische verschillen, hebben we datasets gegenereerd die kunstmatig zijn aangepast door vrije water diffusie en oedeem na te bootsen. De resulta-

ten ondersteunen de hypothese dat de verwijderde variaties alleen afkomstig waren van de reconstructie.

# Résumé

Dans cette thèse, nous avons vu comment l'IRM de diffusion peut compléter l'IRM anatomique en présentant de l'information additionnelle concernant la diffusivité des tissus *in vivo* de façon non invasive. En développant des modèles biophysiques expliquant le signal, ces diffusivités peuvent ensuite être utilisées pour inférer l'architecture ou l'anormalité des tissus sous-jacents. L'analyse des propriétés statistiques du signal (par exemple les moments ou le kurtosis) peut aussi être utilisée pour obtenir des cartes scalaires, cartes qui ne sont pas intrinsèquement liées à un modèle particulier, mais peuvent tout de même révéler les tissus anormaux.

Le [Chapitre 2](#) montra comment une image d'IRM de diffusion à haute résolution spatiale acquise sur un scanner standard peut être utilisée pour accroître la précision anatomique, même si le ratio signal sur bruit généralement associé avec des voxels plus petits prévient normalement une analyse directe. Nous avons présenté un nouvel algorithme de débruitage exploitant la redondance spatiale et angulaire des multiples volumes régulièrement acquis en IRM de diffusion, couplant un algorithme d'apprentissage automatique de dictionnaire combiné à une correction du biais du signal et un processus itératif pondéré de norme  $\ell_1$ . Le processus de reconstruction écarte naturellement le bruit non voulu associé à un faible ratio signal sur bruit en utilisant une borne sur la norme  $\ell_2$  du signal reconstruit dépendant de la variance locale du bruit. Les expériences sur des données synthétiques ont démontré que l'erreur des métriques d'IRM de diffusion (par exemple l'anisotropie fractionnaire ou la diffusivité moyenne) calculée avec l'aide de la méthode proposée est plus basse que l'erreur commise en utilisant les algorithmes comparés. Cette propriété de l'algorithme est aussi validée qualitativement sur des données *in vivo*, où l'on retrouve des détails anatomiques additionnels. On remarque aussi que la tractographie est plus fidèle à la réalité sur des données à 1.2 mm isotrope lorsque comparée aux données bruitées originales à 1.2 mm ou encore sur des données comparables (en termes de temps d'acquisition) de 1.8 mm isotrope du même sujet. Nous notons également que les données à plus faible résolution spatiale avaient un signal ratio sur bruit plus élevé et comprenant 64 volumes d'IRM de diffusion contre 40 volumes pour les données à 1.2 mm.

Les [Chapitres 3](#) et [4](#) présentèrent une amélioration au concept d'analyse le long d'un segment de fibres. Le [Chapitre 3](#) montra en premier lieu que la moyenne géométrique des coordonnées cartésiennes des métriques extraites peut mener à une disparité, notamment en présence d'embranchement (par exemple dans le faisceau arqué) ou encore de configuration en éventail (par exemple le faisceau corticospinal). Une affectation des valeurs utilisant des plans orthogonaux permet au contraire de résoudre ce problème puisque les points sont

désormais assignés localement envers une fibre représentative au lieu de considérer leurs coordonnées absolues. Après avoir extrait une fibre représentative par sujet, le [Chapitre 4](#) présente un nouvel algorithme pour réaligner ces fibres représentatives. Puisque la tractographie et l'extraction des faisceaux de la matière blanche sont réalisées pour chaque sujet séparément, il n'existe pas de moyen garantissant que les coordonnées de l'espace 1D correspondent pour chaque sujet. En trouvant automatiquement un candidat modèle parmi tous les sujets disponibles, seulement les portions de segments qui correspondent (selon un seuil défini par l'utilisateur) sont conservées après le réalignement pour être subséquentement analysées. Nous avons montré que cette stratégie permet de réduire le coefficient de variation des métriques de diffusion étudiées (diffusivité moyenne, anisotropie fractionnaire et densité de fibres apparente) avec des expériences synthétiques et avec une base de données composée de 100 sujets. Sur cette dernière, les expériences où la moitié des sujets avaient été altérés ont démontré que l'algorithme de réalignement permet de retrouver les régions affectées par ces mêmes altérations, alors que ce n'était pas toujours possible dans le cas où les sujets ne sont pas réalignés. Cette conclusion est aussi valide lorsque les altérations réalisées sur les valeurs scalaires sont majeures et couvrent seulement une petite région du faisceau complet.

Le [Chapitre 5](#) nous ramena à l'acquisition en présentant une nouvelle méthode pour estimer automatiquement la distribution du signal provenant d'une série d'acquisitions d'IRM de magnitude, tel que les multiples volumes acquis en IRM de diffusion qui sont requis pour supporter les modèles locaux. L'avantage principal de l'algorithme repose sur le fait qu'il ne nécessite pas d'information externe, tel qu'une carte de sensibilité de l'antenne ou une matrice de reconstruction, qui n'est habituellement pas enregistrée lors de l'acquisition. Des expériences multiples sur des données synthétiques avec un nombre varié de canaux (incluant deux algorithmes d'accélération parallèle différents) ont démontré que la distribution du bruit peut être retrouvée efficacement sans nécessiter d'information *a priori* sur le protocole d'acquisition. Les expériences réalisées sur les images d'une bouteille d'eau montrèrent que la méthode est robuste aux contaminations du signal causées par l'accélération multibande tant et aussi longtemps que le facteur d'accélération reste modéré, c'est-à-dire environ un facteur 3 pour nos expériences. Nous avons aussi analysé deux jeux de données *in vivo* acquis dans deux centres différents, incluant quatre répétitions d'un même sujet publiquement disponible en ligne. Les résultats ont démontré que la méthode proposée est stable sur ces quatre jeux de données et permet d'identifier la distribution du signal théorique, tout en évitant les voxels contaminés par des artefacts. Les résultats du second jeu de données *in vivo* indiquèrent aussi que la distribution du signal calculée était similaire à celle dictée par la théorie selon l'algorithme de reconstruction utilisé lors de l'acquisition.

Finalement, le [Chapitre 6](#) présente un nouvel algorithme pour harmoniser les données d'IRM de diffusion acquises sur différents scanners. En utilisant un algorithme d'apprentissage de dictionnaire, nous avons montré comment des caractéristiques peuvent être automatiquement extraites des données acquises sur un scanner cible. Ces dernières sont ensuite

utilisées pour reconstruire les données d'un scanner source différent, tout en éliminant la variabilité attribuée aux deux scanners et en préservant la variabilité anatomique. Les expériences sur un jeu de données publiquement disponible montrèrent comment l'algorithme réduit la variabilité sur certaines métriques couramment utilisées en IRM de diffusion et peut même être utilisé si la résolution spatiale ne concorde pas entre les scanners grâce à un sous-échantillonnage spatial du dictionnaire. Pour vérifier si la variabilité retranchée n'était pas due à de véritables différences anatomiques, nous avons aussi généré une version altérée des données en les contaminant artificiellement avec de l'œdème libre. Les résultats supportèrent l'hypothèse de la variabilité comme étant issue des scanners, puisque cette dernière décrut tout en préservant la taille d'effet lorsque les données altérées furent comparées avec leur version originale à l'aide d'un test t de Student pour échantillons appariés sur les quatre métriques de diffusion étudiées. La taille d'effet était, en moyenne, comprise dans le même intervalle de confiance à 95% que les données non altérées, confirmant ainsi que le procédé d'harmonisation n'a pas enlevé de variabilité due à l'anatomie lorsque l'on compare les données harmonisées avec les données originales.

# Publications

## Journal publications

- [1] **S. St-Jean**, M. Chamberland, M. A. Viergever, and A. Leemans. “Reducing variability in along-tract analysis with diffusion profile realignment”. In: *NeuroImage* 199 (2019), pp. 663–679.
- [2] C. M. W. Tax, F. Grussu, E. Kaden, L. Ning, U. Rudrapatna, C. John Evans, **S. St-Jean**, A. Leemans, S. Koppers, D. Merhof, A. Ghosh, R. Tanno, D. C. Alexander, S. Zappalà, C. Charron, S. Kusmia, D. E. Linden, D. K. Jones, and J. Veraart. “Cross-scanner and cross-protocol diffusion MRI data harmonisation: A benchmark database and evaluation of algorithms”. In: *NeuroImage* 195 (2019), pp. 285–299.
- [3] R. N. Henriques, A. Rokem, E. Garyfallidis, **S. St-Jean**, E. T. Peterson, and M. M. Correia. “[Re] Optimization of a free water elimination two-compartment model for diffusion tensor imaging”. In: *ReScience* 3.1 (Apr. 2017).
- [4] K. H. Maier-Hein, P. F. Neher, J.-C. Houde, M.-A. Côté, E. Garyfallidis, J. Zhong, M. Chamberland, F.-C. Yeh, Y.-C. Lin, Q. Ji, W. E. Reddick, J. O. Glass, D. Q. Chen, Y. Feng, C. Gao, Y. Wu, J. Ma, H. Renjie, Q. Li, C.-F. Westin, S. Deslauriers-Gauthier, J. O. O. González, M. Paquette, **S. St-Jean**, G. Girard, F. Rheault, J. Sidhu, C. M. W. Tax, F. Guo, H. Y. Mesri, S. Dávid, M. Froeling, A. M. Heemskerk, A. Leemans, A. Boré, B. Pinsard, C. Bedetti, M. Desrosiers, S. Brambati, J. Doyon, A. Sarica, R. Vasta, A. Cerasa, A. Quattrone, J. Yeatman, A. R. Khan, W. Hodges, S. Alexander, D. Romascano, M. Barakovic, A. Auria, O. Esteban, A. Lemkaddem, J.-P. Thiran, H. E. Cetingul, B. L. Odry, B. Mailhe, M. S. Nadar, F. Pizzagalli, G. Prasad, J. E. Villalon-Reina, J. Galvis, P. M. Thompson, F. D. S. Requejo, P. L. Laguna, L. M. Lacerda, R. Barrett, F. Dell’Acqua, M. Catani, L. Petit, E. Caruyer, A. Daducci, T. B. Dyrby, T. Holland-Letz, C. C. Hilgetag, B. Stieltjes, and M. Descoteaux. “The challenge of mapping the human connectome based on diffusion tractography”. In: *Nature Communications* 8.1 (2017), p. 1349.
- [5] **S. St-Jean**, P. Coupé, and M. Descoteaux. “Non Local Spatial and Angular Matching: Enabling higher spatial resolution diffusion MRI datasets through adaptive denoising”. In: *Medical Image Analysis* 32.2016 (2016), pp. 115–130.
- [6] L. Ning, F. Laun, Y. Gur, E. V. R. DiBella, S. Deslauriers-Gauthier, T. Megherbi, A. Ghosh, M. Zucchelli, G. Menegaz, R. Fick, **S. St-Jean**, M. Paquette, R. Aranda, M. Descoteaux, R. Deriche, L. O’Donnell, and Y. Rathi. “Sparse Reconstruction Challenge for diffusion MRI: Validation on a physical phantom to determine which acquisition scheme and analysis method to use?” In: *Medical Image Analysis* 26.1 (2015), pp. 316–331.

## Preprints

- [1] **S. St-Jean**, A. De Luca, C. M. W. Tax, M. A. Viergever, and A. Leemans. *Automated characterization of noise distributions in diffusion MRI data*. In: *Bioarxiv* (2019). doi: [10.1101/686436](https://doi.org/10.1101/686436).

- [2] **S. St-Jean**, M. A. Viergever, and A. Leemans. *Harmonization of diffusion MRI datasets with adaptive dictionary learning*. In: *Arxiv* (Oct. 2019). <https://arxiv.org/abs/1910.00272>.

## Conference proceedings

- [1] M. Chamberland, **S. St-Jean**, C. M. W. Tax, and D. K. Jones. “Obtaining representative core streamlines for white matter tractometry of the human brain”. In: *Computational Diffusion MRI*. Ed. by E. Bonet-Carne, F. Grussu, L. Ning, F. Sepehrband, and C. M. W. Tax. Granada: Springer International Publishing, 2018, pp. 3–19.
- [2] **S. St-Jean**, A. De Luca, M. A. Viergever, and A. Leemans. “Automatic, Fast and Robust Characterization of Noise Distributions for Diffusion MRI”. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*. Ed. by A. F. Frangi, J. A. Schnabel, C. Davatzikos, C. Alberola-López, and G. Fichtinger. Springer International Publishing, 2018, pp. 304–312.

## Conference abstracts

- [1] **S. St-Jean**, A. De Luca, M. A. Viergever, and A. Leemans. “A separable least squares approach for intravoxel incoherent motion (IVIM) MRI”. In: *ISMRM Benelux*. 2019.
- [2] **S. St-Jean**, A. De Luca, M. Viergever, and A. Leemans. “Investigating noise distribution changes after motion correction and its effects on subsequent diffusion MRI processing”. In: *International Symposium on Magnetic Resonance in Medicine (ISMRM’18)*. 2018.
- [3] C. M. W. Tax, F. Grussu, E. Kaden, L. Ning, U. Rudrapatna, J. Evans, **S. St-Jean**, A. Leemans, S. Puch, M. Rowe, P. Rodrigues, V. Prčkovska, S. Koppers, D. Merhof, A. Ghosh, R. Tanno, D. C. Alexander, C. Charron, S. Kusmia, D. E. J. Linden, D. K. Jones, and J. Veraart. “Cross-vendor and Cross-protocol harmonisation of diffusion MRI data: a comparative study”. In: *ISMRM 2018: 26th Annual Meeting of the International Society for Magnetic Resonance in Medicine*. 2018, p. 471.
- [4] **S. St-Jean**, M. A. Viergever, and A. Leemans. “A unified framework for upsampling and denoising of diffusion MRI data”. In: *25th Annual Meeting of ISMRM*. 2017, p. 3533.
- [5] C. Granziera, A. Daducci, **S. St-Jean**, G. Krueger, and M. Descoteaux. “Denoised diffusion spectrum imaging of white matter tracts in the brain stem”. In: *International Symposium on Magnetic Resonance in Medicine (ISMRM’16)*. 2016.
- [6] **S. St-Jean**, M. A. Viergever, and A. Leemans. “Denoising diffusion MRI data using a path algorithm strategy”. In: *ISMRM workshop : Breaking the barriers of diffusion MRI*. September. 2016.
- [7] **S. St-Jean**, M. A. Viergever, G. J. Biessels, and A. Leemans. “Correcting spatial misalignment between fiber bundles segments for along-tract group analysis”. In: *International Symposium on Magnetic Resonance in Medicine (ISMRM’16)*. 2016.
- [8] **S. St-Jean**, G. Gilbert, and M. Descoteaux. “Connectome-like quality diffusion MRI in 13 minutes – Improving diffusion MRI spatial resolution with denoising”. In: *International Symposium on Magnetic Resonance in Medicine (ISMRM’15)*. Vol. 13. 4. 2015, p. 2249.
- [9] M. Paquette, G. Sanguinetti, **S. St-Jean**, E. Garyfallidis, and M. Descoteaux. “Cartesian 3D-SHORE with Laplacian Regularization”. In: *MICCAI Computational Diffusion MRI workshop challenge*. 2014.

- [10] **S. St-Jean**, P. Coupé, and M. Descoteaux. “Non Local Spatial and Angular Matching : a new denoising technique for diffusion MRI”. In: *International Symposium on Magnetic Resonance in Medicine (ISMRM’14)*. 2014.
- [11] **S. St-Jean**, M. Paquette, E. Garyfallidis, and M. Descoteaux. “Spherical Deconvolution on single shell and multishell combined with denoising”. In: *MICCAI Computational Diffusion MRI workshop challenge*. 2014.
- [12] E. Garyfallidis, M. Paquette, **S. St-Jean**, P. Coupé, and M. Descoteaux. “Deconvolution enhanced Generalized Q-Sampling 2 and DSI deconvolution”. In: *ISBI HARDI reconstruction challenge*. San Francisco, USA, 2013.
- [13] E. Garyfallidis, **S. St-Jean**, M. Paquette, P. Coupé, and M. Descoteaux. “Constrained spherical deconvolution on signal and ODF values”. In: *HARDI Reconstruction Challenge, International Symposium on Biomedical Imaging*. San Francisco, USA, 2013, p. 22.
- [14] M. Paquette, E. Garyfallidis, **S. St-Jean**, P. Coupé, and M. Descoteaux. “Particle Swarm Optimization in Multi-Tensor Imaging”. In: *ISBI HARDI reconstruction challenge*. San Francisco, USA, 2013.

## Software implementations

- [1] **S. St-Jean**. *samuelstjean/dpr: [v0.1.1b] - 2019-03-13*. 2019.  
URL: <https://doi.org/10.5281/zenodo.2592014>.
- [2] **S. St-Jean**, A. De Luca, C. M. W. Tax, M. A. Viergever, and A. Leemans. *samuelstjean/autodmri: First release - 2019-07-17*. 2019.  
URL: <https://doi.org/10.5281/zenodo.3339158>.
- [3] **S. St-Jean**, M. A. Viergever, and A. Leemans. *samuelstjean/harmonization: Harmonization of diffusion MRI datasets with adaptive dictionary learning*. 2019.  
URL: <https://doi.org/10.5281/zenodo.3385925>.
- [4] **S. St-Jean**. *samuelstjean/nlsam: The reference implementation for the Non Local Spatial and Angular Matching (NLSAM) denoising algorithm for diffusion MRI*. 2016.  
URL: <https://github.com/samuelstjean/nlsam/>.

## Available datasets

- [1] **S. St-Jean**, M. Chamberland, M. A. Viergever, and A. Leemans. *Datasets for 'Reducing variability in along-tract analysis with diffusion profile realignment'*. 2018.  
URL: <http://dx.doi.org/10.5281/zenodo.2483169>.
- [2] **S. St-Jean**, A. De Luca, C. M. W. Tax, M. A. Viergever, and A. Leemans. *Datasets for 'Automated characterization of noise distributions in diffusion MRI data'*. 2018.  
URL: <https://zenodo.org/record/2483105>.
- [3] **S. St-Jean**, P. Coupé, and M. Descoteaux. *Datasets for 'Non Local Spatial and Angular Matching: Enabling higher spatial resolution diffusion MRI datasets through adaptive denoising'*. 2016.  
URL: [https://github.com/samuelstjean/nlsam\\_data](https://github.com/samuelstjean/nlsam_data).

## MSc Thesis

- [1] **S. St-Jean.** “Acquisitions d’IRM de diffusion à haute résolution spatiale : nouvelles perspectives grâce au débruitage spatialement adaptatif et angulaire”. Université de Sherbrooke, 2015, p. 149.

# Acknowledgments

And now we are, quite unfortunately, almost at the end of this fine thesis. Of course, there were numerous people along the way to help me out a bit in getting there. I guess the first part started back in Sherbrooke at the end of my bachelor's degree and during my MSc, where I learned about imaging, MRI and the principles underlying diffusion MRI itself in the first place. This was, of course, thanks to everyone else for helping me out with their respective expertise, namely in mathematics, computer science, imaging, and the optimization guys. As this part can be found in my old master's thesis, it would be shorter to mention the people involved in the second part during my PhD in the Netherlands (although some people are indeed present in both locations). This is also the part where we mention a lot of people and forget others at the same time, so no need to feel bad if you were personally left out in a sense.

I guess the first people to thank are the diffusion guys, the master students (since they have the most potential) and everyone who came over for a (short or not) visit along the way. It always helps out to talk over with other people and get a different view on a problem you are working on after all as you may be missing something which is obvious to someone else without knowing it. Following on that thought, I guess the next in line for this page are the physics guys (but it's mostly girls) because I always throw at them very advanced concepts apparently which I always think they should easily understand (especially Mike and Kim). I guess it comes from never understanding physics in the first place (especially the special tricks and simplifications, and when they are allowed or not), even if I can play around with simple equations, as evidenced by the advanced MR physics classes.

The last part of this thank you section goes to everyone at the ISI and related departments since they are always there to cheer people up and help out when your stuff is not working properly (a daily issue for everyone in fact according to the funny posters in our offices). For some reason I became known as the  $\text{\LaTeX}$  guy for the deep learning guys, go figure. I guess the staff should also be mentioned since they always help me out with the tons of administrative issues I regularly have.

## About the author

Samuel St-Jean was born in 1988 in the city of Sherbrooke, Québec, Canada. He did a BSc in mathematics at the Université de Sherbrooke from 2008 to 2012, including research internships namely at the Centre de recherche, de développement et de transfert technologique acéricole (Centre ACER) in Saint-Hyacinthe and at Statistics Canada with the Division des comptes des revenus et dépenses in the federal capital of Ottawa. During his third year, he followed a signal processing class and learned about MRI, followed by a research internship with prof. Maxime Descoteaux. This internship was followed by an MSc in imaging and computer science on diffusion MRI from 2012 to 2015 at the Sherbrooke connectivity imaging lab (SCIL). After receiving a three-year scholarship from the Fonds québécois de la recherche sur la nature et les technologies (FQRNT) for doctoral studies, he enrolled in the Image sciences institute of the University Medical Center Utrecht as a PhD candidate to pursue improvements in diffusion MRI processing and methods. The research conducted during these years is reported in the thesis you are currently reading.