



**University of
Nottingham**

UK | CHINA | MALAYSIA

On noise, uncertainty and inference for computational diffusion MRI

José Pedro Manzano Patrón

Thesis submitted to the University of Nottingham
for the degree of Doctor of Philosophy

July 2023

*"We were born too late to explore the Earth,
born too early to explore the galaxy,
but born just in time to explore the brain." - Anonymous*

Abstract

Diffusion Magnetic Resonance Imaging (dMRI) has revolutionised the way brain microstructure and connectivity can be studied. Despite its unique potential in mapping the whole brain, biophysical properties are inferred from measurements rather than being directly observed. This indirect mapping from noisy data creates challenges and introduces uncertainty in the estimated properties. Hence, dMRI frameworks capable to deal with noise and uncertainty quantification are of great importance and are the topic of this thesis.

First, we look into approaches for reducing uncertainty, by de-noising the dMRI signal. Thermal noise can have detrimental effects for modalities where the information resides in the signal attenuation, such as dMRI, that has inherently low-SNR data. We highlight the dual effect of noise, both in increasing variance, but also introducing bias. We then design a framework for evaluating denoising approaches in a principled manner. By setting objective criteria based on what a well-behaved denoising algorithm should offer, we provide a bespoke dataset and a set of evaluations. We demonstrate that common magnitude-based denoising approaches usually reduce noise-related variance from the signal, but do not address the bias effects introduced by the noise floor. Our framework also allows to better characterise scenarios where denoising can be beneficial (e.g. when done in complex domain) and can open new opportunities, such as pushing spatio-temporal resolution boundaries.

Subsequently, we look into approaches for mapping uncertainty and design two inference frameworks for dMRI models, one using classical Bayesian methods and another using more recent data-driven algorithms. In the first approach, we build upon the univariate random-walk Metropolis-Hastings MCMC, an extensively used sampling method to sample from the posterior distribution of model parameters given the data. We devise an efficient adaptive multivariate MCMC scheme, relying upon the assumption that groups of model parameters can be jointly estimated if a proper covariance matrix is defined. In doing so, our algorithm increases the sampling efficiency, while preserving accuracy and precision of estimates. We show results using both synthetic and in-vivo dMRI data.

In the second approach, we resort to Simulation-Based Inference (SBI), a data-driven approach that avoids the need for iterative model inversions. This is achieved by using neural density estimators to learn the inverse mapping from the forward generative process (simulations) to the parameters of interest that have generated those simulations. By addressing the problem via learning approaches offers the opportunity to achieve inference amortisation, boosting efficiency by avoiding the necessity of repeating the inference process for each new unseen dataset. It also allows inversion of forward processes (i.e. a series of processing steps) rather than only models. We explore different neural network architectures to perform conditional density estimation of the posterior distribution of parameters. Results and comparisons obtained against MCMC suggest speed-ups of 2-3 orders of magnitude in the inference process while keeping the accuracy in the estimates.

Acknowledgements

During the writing of this thesis, we have witnessed significant global events, including the definitive departure of the UK from the EU, a global pandemic, an economic crisis, and even a war within the borders of this idyllic bubble called Europe. These events are accompanied by the rapid advancements in technology, bringing about significant changes. It is estimated that up to 75% of the jobs that will exist in 2050 are completely unknown at present. Climate change and paradigm shifts in society will reshape our known world. Furthermore, all these transformations are occurring within a single lifespan, highlighting the unprecedented speed at which the world is changing. Now, better than ever, "it's difficult to predict, especially the future" as Niels Bohr aptly said. It seems very timely then to undertake a thesis on uncertainty.

In "Human Knowledge: Its Scope and Limits", Bertrand Russell makes an analysis from the logic of all existing disciplines, from microbiology to astronomy, and concluded the book by saying "All human knowledge is uncertain, inaccurate and partial". While this realization can lead to apathy or a sense of futility, it can also foster us to avoid dogmatism, encourage critical thinking, acknowledge limitations, and gradually build on prior knowledge to illuminate the remaining unknowns. This, to me, is the essence of science. It is not about certainty, not about facts fetched into stone tablets. Rather, science is about painstakingly paying down epistemic uncertainty: sharpening our hypotheses to be "as simple as possible, but no simpler". Werner Heisenberg once said "What we observe is not nature itself but nature exposed to our method of questioning". It is in this direction that I aim to contribute with this thesis, offering tools to navigate the uncertainties that surround us, particularly in understanding a little better the complexities of the brain or, rather, highlighting the limitations we have to know it, and so judge accordingly.

Within this endeavour, first and foremost, I express my deepest gratitude to Stam for his unwavering patience, invaluable guidance, and endless support. Thank you for being an exceptional source of ideas, and constantly pushing me forward at every step. This thesis would not have been possible without you. I am also immensely grateful to my second supervisor, Theodore Kypraios, for his valuable research suggestions, encouragement, and assistance in addressing multiple questions.

I extend warm thanks to all the individuals who have, in some way, contributed to providing me with this opportunity. To my colleagues and professors in Spain who always

encouraged and motivated me to pursue a PhD, to Austin for introducing me to the world of MRI, and to my colleagues and supervisors at CSIC and CNIC for revealing to me the thrilling world of research. I would like to mention as well Dorothee Auer, the Beacon Precision Imaging and SPMIC staff for their daily efforts that allows us to keep doing what we love. I am also very grateful to my collaborators Steen Moeller and Essa Yacoub, and to my assessors Matteo Bastiani, Michael Chappell and Christophe Lenglet, for their feedback and comments. It is both tremendously humbling and motivating to share a conversation with you and to realise how much I can (and I still have to) learn.

I am truly fortunate to have experienced an ideal scientific environment in the birthplace of MRI. I am grateful to all the researchers who have become my friends. If I am getting something better than my PhD after these 4 years is the experience with all my labmates, both those who have left, such as Ellie, Asante, Jiv, Christoph, Cheryl, Sarina, Will,... and those who continue to work alongside me, including Stefan, Shaun, Ali, Andrea, Stephania, Kanella, and many others. It has been a pleasure to travel this path and enjoy lunches, coffee breaks, and after-work beers with such brilliant and diverse individuals. You are the ones that shared the daily burden of the PhD with me, thank you.

A special thanks goes to Angela. I am honoured to have you reading these words. You know firsthand the experiences we have shared and how important you have been in this adventure. I have no doubt that you will achieve great things in your life, and I hope to be there to witness them.

Pero la dedicación más importante debo hacerla en español, ya que es para mi familia. A pesar de no entender muy bien aún a lo que me dedico, siempre me han apoyado incondicionalmente. En ellos no hay lugar para la incertidumbre. Gracias por las incontables oportunidades y la inagotable confianza. Siguen siendo la mayor fuente de motivación. Pablo, mis últimas palabras van para ti: "As we know, there are known knowns; there are things we know that we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns—the ones we don't know we don't know. And if one looks throughout the history, it is the latter category that tend to be the difficult ones." - Estas últimas son las que tienes que perseguir, aunque no sepas donde están. No olvides que lo mejor que podrás ser de mayor es seguir siendo curioso. Observa, investiga, piensa, experimenta, aprende de tus errores, y vuelve a empezar. Algunos te lo venderán como el método científico pero no. Eso, querido Pablo, se llama vivir. Así que sal ahí fuera y vive.

Contents

Abstract	i
Prologue and Acknowledgements	iii
1 Introduction	1
1.1 Organization of the thesis	4
1.2 Computing Infrastructure and Software	6
2 Noise and modeling in diffusion MRI	8
2.1 Thermal noise in MRI	9
2.1.1 Where does thermal noise come from?	9
2.1.2 Statistical properties of MRI thermal noise	11
2.1.3 Modern MRI protocols. Effects on SNR and noise properties	13
2.2 Diffusion MRI and biophysical modeling	16
2.2.1 Diffusion phenomena	16
2.2.2 How can we use MRI to measure diffusion?	19
2.2.3 From dMRI signals to probing tissue microstructure	23
2.2.4 Biophysical Modeling of dMRI data	25
2.2.5 From fibre orientation to Streamline Tractography	31
2.3 How does noise affect dMRI model estimates?	33
2.3.1 Noise-floor signal rectification and bias	33
2.3.2 Variance and Uncertainty	34
3 Uncertainty Mapping and Inference	37
3.1 Uncertainty in dMRI	37
3.1.1 Sources of epistemic uncertainty in dMRI	39
3.2 Why mapping uncertainty can be useful in dMRI?	41
3.2.1 Optimising experimental design	41
3.2.2 Spatial propagation of uncertainty - Probabilistic tractography	42

3.3	Model Estimation and Uncertainty Mapping Techniques	44
3.3.1	Least Squares Estimation (LSE)	44
3.3.2	Bootstrapping	45
3.3.3	Bayesian inference	47
3.3.4	Machine Learning for inference in dMRI	49
3.3.5	Summary	51
4	EDDEN: A framework for Evaluating Diffusion MRI DENoising approaches	54
4.1	Introduction	54
4.2	Theory	57
4.2.1	Powder Averaging	60
4.2.2	Non-Local Means (NLM) Denoising	60
4.2.3	Marchenko-Pastur Principal Component Analysis (MPPCA) Denoising	61
4.2.4	NOise Reduction with DIstribution Correction (NORDIC)	64
4.2.5	Patch2Self	65
4.3	Methods	66
4.3.1	Defining a good denoising algorithm	66
4.3.2	Data	72
4.4	Results	74
4.4.1	Raw signal quality assessments	74
4.4.2	Noise-floor and Signal statistical properties	77
4.4.3	Spatial smoothing	80
4.4.4	Convergence in high-SNR regimes	81
4.4.5	Improvements in modelling performance	83
4.4.6	Capitalising on increased SNR to push spatial resolution	85
4.4.7	Capitalising on increased SNR to reduce scan time	87
4.5	Discussion	90
5	Mapping uncertainty in dMRI using MCMC: A hybrid approach	99
5.1	Introduction	99
5.2	Theory	101
5.2.1	Markov-Chain Monte-Carlo (MCMC) sampling	101

5.2.2	Mixing and sampling efficiency in random-walk MCMC	105
5.2.3	Model selection and automatic relevance determination	107
5.2.4	Improving random-walk MCMC: Block proposals and the Laplace approximation	109
5.3	Methods	112
5.3.1	Inference for the Ball&Sticks model using random-walk MH-MCMC	112
5.3.2	Alternative designs to random-walk MH-MCMC	116
5.3.3	Data	120
5.3.4	Evaluations	123
5.4	Results	126
5.4.1	Simulations	126
5.4.2	In-vivo data	138
5.5	Discussion	148
6	Amortised simulation-based inference in diffusion MRI	153
6.1	Introduction	153
6.2	Theory	155
6.2.1	Simulation-Based Inference	155
6.2.2	Density Estimation using Neural Networks	157
6.2.3	Neural Posterior Estimation (NPE)	160
6.3	Methods	166
6.3.1	Training	168
6.3.2	Evaluations	170
6.4	Results	174
6.4.1	Design, training and evaluation of the SBI network	174
6.4.2	Comparisons of SBI to MCMC – Synthetic data	178
6.4.3	Comparisons of SBI to MCMC – In-vivo brain data	179
6.5	Discussion	182
6.5.1	Uncertainty mapping in SBI	188
6.5.2	Limitations and Future Work	190
7	Conclusions and future directions	192
7.1	Main Contributions	192
7.2	Future Directions	194

7.2.1	Benchmarking for dMRI denoising	195
7.2.2	Direct mapping from raw signal to model estimates	196
7.2.3	Fusing information from multiple sources	197
Bibliography		199
Appendices		225
A List of Abbreviations		225
B Initialization and other MCMC considerations		229
B.1	Spherical to Cartesian coordinate conversion	229
B.2	MCMC Initialization	230
B.2.1	Tensor model fitting	230
B.2.2	Non-linear fitting	231
B.3	Defining priors for the MCMC	232
B.4	Integrating out precision τ from likelihood	233
B.5	MCMC Posterior Energy	233
B.6	Implementation of the multi-fibre model with ARD	234
B.6.1	Beta based ARD prior for f_i	234
B.6.2	Gaussian based ARD prior for f_i	235
C Laplace Approximation		236
C.1	Demonstration 1: Laplace Approximation is located around the posterior mode	236
C.1.1	Computing the posterior mean	237
C.2	Demonstration 2. Hessian matrix for the Ball&Sticks model with 3 com- partments	237
C.2.1	First Derivatives	238
C.2.2	Second Derivatives	241
D Hybrid-MCMC results		244
D.0.1	Dispersion maps in Dataset C - 3 fibres	244
D.1	Probabilistic Tractography and uncertainty propagation	244
D.1.1	Comparisons to the RWM	244
D.1.2	Comparisons to the HCP White-Matter population average atlas . .	247

List of Tables

4.1	Summary table of the datasets acquired for the evaluation of denoising approaches and included in EDDEN.	73
4.2	Summary table of the features of the denoising methods evaluated.	74
4.3	Spatial smoothing	81

List of Figures

1.1	From measurements to biophysical properties	4
2.1	Additive White Gaussian Noise (AWGN)	11
2.2	Main MR imaging acceleration techniques	14
2.3	Scheme of thermal noise introduction in the MRI signal and how it affects the statistical properties	17
2.4	Normalized distribution of the average propagator	19
2.5	Pulse Gradient Sequential Echo (PGSE) sequence	22
2.6	B-values and diffusivity effects in the dMRI measured signal	24
2.7	Examples of common fiber configurations	27
2.8	Microstructural dMRI models	30
2.9	Comparison between DTI and Ball&Sticks fibre orientation estimates	31
2.10	Streamline tractography	32
2.11	Noise-floor bias	35
2.12	Noise-related variance	36
3.1	Uncertainty arisen from the modeling process	40
3.2	Comparison between deterministic and probabilistic tractography	43
4.1	Growth of denoising diffusion MRI publications	55
4.2	Local patch-based denoising	59
4.3	Post-reconstruction denoising methods	59
4.4	The Marchenko-Pastur distribution	63
4.5	NORDIC denoising	64
4.6	EDDEN Datasets	73
4.7	Dataset A (2mm) - Denoised	75
4.8	Dataset B (1.5mm) - Denoised	76

4.9	Dataset C (0.9mm) - Denoised	76
4.10	Gains in Angular Contrast to Noise Ratio (CNR) by denoising	77
4.11	Difference maps between RAW and denoised data	78
4.12	Noise-floor preservation and signal rectification	80
4.13	Estimated voxel resolution after denoising	82
4.14	Convergence at high SNR - Tract correlations with the UKB population- average atlas	83
4.15	A comparison of modelling performance between RAW and denoised data .	84
4.16	Tractography - Maximum Intensity Projection examples	86
4.17	Spatial tract correlation with the HCP atlas (Dataset C - 0.9mm)	86
4.18	Capitalising on increased SNR to reduce scan time	87
4.19	Analysis of the convergence to the multiple-averages at ultra-high resolution	88
4.20	Preservation of the noise-floor in the multiple averages	89
4.21	Convergence to the multiple averages in low attenuation regions	90
5.1	The Metropolis-Hastings MCMC algorithm	103
5.2	Examples of Laplace Approximation (LA)	111
5.3	Scheme of the simulated datasets.	122
5.4	MCMC traceplots	127
5.5	Histograms of the posterior samples	128
5.6	Posterior width obtained by each MCMC algorithm across realisations . . .	128
5.7	Autocorrelation Function (ACF)	129
5.8	Acceptance ratios of each algorithm	130
5.9	Number of fibres correctly detected by each algorithm	131
5.10	Error maps - Dataset A (1 fibre configuration)	132
5.11	Scalar error maps - Dataset B (2 fibres configuration)	134
5.12	Fibre orientation error maps - Dataset B (2 fibre-crossings configuration) .	135
5.13	Dispersion maps - Dataset B (2 fibre-crossings configuration)	136
5.14	Comparison of errors between RWM and Hybrid1 - Dataset C (3 crossing- fibres configuration)	137
5.15	Computational speed-up	138
5.16	MCMC traceplots exemplar obtained in in-vivo brain data	139
5.17	Exemplar posterior distributions obtained in in-vivo brain data	140

5.18 Comparison of mean maps obtained in in-vivo brain data for scalar parameters	141
5.19 Fibre orientations estimates comparison in the Centrum Semiovale	142
5.20 Comparison of fibre orientation maps obtained in in-vivo brain data	143
5.21 Comparison of dispersion maps obtained in in-vivo brain data	144
5.22 Maximum Intensity Projection of some exemplar tracts in subjects from the UKBiobank dataset	146
5.23 Spatial tract correlations between the RWM and the Hybrid1 in 15 subjects from the UKBiobank dataset	147
5.24 Exemplar of non-reconstructed tract	148
5.25 Spatial tract correlations against the HCP population-averaged atlas	149
6.1 Comparison between Classical Bayesian Inference and Simulation-based In- ference	158
6.2 Mixture Density Network (MDN) scheme	162
6.3 Auto-regressive property	162
6.4 Masked Autoencoder Density Estimation (MADE) scheme	163
6.5 Normalising Flows (NFs) scheme	164
6.6 NPE scheme implementation	167
6.7 Uniform sampling in the sphere	170
6.8 Sensitivity analysis of the posterior	172
6.9 Comparison of density estimators in test synthetic data	174
6.10 Mean error in parameter estimates for different training data sizes and diffusivity prior distributions	175
6.11 Variations in the width of the posteriors in function of the SNR level	175
6.12 Variations in the width of the posteriors in the function of the number of posterior samples level	176
6.13 Posterior fit evaluation	176
6.14 Training size vs training time	177
6.15 Posterior Predictive Check - MCMC vs SBI	178
6.16 Error maps in synthetic data - MCMC vs SBI	180
6.17 Dispersion maps in synthetic data - MCMC vs SBI	181
6.18 Qualitative comparison of MCMC and SBI estimates in in-vivo brain data .	183

6.19	Quantitative comparison between MCMC and SBI estimates in in-vivo brain data	184
6.20	Comparison of probabilistic tractography obtained from MCMC and SBI estimates in in-vivo brain data	184
B.1	Angle convention followed	229
D.1	Dispersion maps - Dataset C (3 fibre-crossings configuration)	245
D.2	Spatial tract correlations between UKBiobank subjects with increased num- ber of jumps in the Hybrid1 MCMC parameters	246
D.3	Spatial tract correlations against the HCP population-averaged atlas with increased number of jumps in the MCMC parameters of both algorithms . .	248

Chapter 1

Introduction

Relative to body size, the human brain is the largest across all vertebrate brains and has been described as *one of the most complex systems in the universe* [Koch et al., 2002]. Its complexity emerges from a never-resting network of more than 10^{11} neurons with more than 10^{16} connections between them [Kandel et al., 2000]. Therefore, it is no surprise that **understanding the brain** has been the topic of exploration and research for a very long time, and one of the greatest scientific challenges of the 21st century.

The approaches used to address such a scientific challenge have evolved dramatically over time (e.g., see [Abraham and Feng, 2011; Leeds and Kieffer, 2000; Morecraft et al., 2014] for historical reviews). The invention of the microscope at the end of the 16th century paved the way for cell theory and allowed the study of individual nerve fibres (e.g., [Ramon y Cajal, 1911]). The development of histochemical tracers in the 1950s generated an explosion of very precise connectivity data in animal models at the mesoscale providing the vast majority of our current knowledge of white-matter connections. However, given the manual, time-consuming and delicate procedures involved, these micro/mesoscopic methods are limited to small number of samples, low-throughput and challenges in quantifiability; while their invasive nature limits them to *post-mortem* studies, mostly of the non-human brain.

Some key technological developments during the 1970s and 1980s led to the introduction of imaging techniques, such as **Magnetic Resonance Imaging (MRI)** [Lauterbur, 1973; Mansfield and Maudsley, 1977], which have revolutionised neuroscience. Compared

to other non-invasive alternatives (such as electroencephalography (EEG), magnetoencephalography (MEG) or Computed Tomography (CT)), MRI is unique in being able to provide multiple contrasts that probe different aspects of tissue structure, function and physiology. Morphologic characteristics and gross anatomy [Fischl, 2012], brain connectivity and fine microstructural properties [Basser, 1995; Jbabdi et al., 2015], haemodynamic changes emerged from neural activity [Biswal et al., 1995], chemical decomposition [Clarke et al., 2021], perfusion or metabolism [Chappell et al., 2013] can all be probed using the same machine, non-invasively and in-vivo. Compared to the microscopic imaging techniques, MRI operates at a different scale (the macroscale, i.e. at the systems level [Behrens and Sporns, 2012; Sporns et al., 2005]). It can however be used to study multiple living humans [Miller et al., 2016], relate imaging features with behaviour [Smith et al., 2015], genetics and function [Smith et al., 2013], and longitudinally study disease, development and aging (e.g., [Cabral et al., 2012; Drysdale et al., 2017; Evangelou, 2000]).

Despite its unique potential, **MRI is inherently *indirect*** and this creates a number of **challenges** in dealing with and interpreting the data [Dubois and Adolphs, 2016; Jbabdi et al., 2015; Schilling et al., 2019]. MRI scanners are not optical imaging devices and measure properties directly related to hydrogen protons, which are in abundance in water molecules within the body. For this reason -and excluding allometric and morphological features- the biophysical properties one is interested in are often indirectly *inferred* from the data, rather than measured directly. For instance, neuronal activity is probed through oxyhaemoglobin changes in the blood flow of entire regions [Biswal et al., 1995], or tissue microstructural properties can be probed by the scatter pattern of ensembles of water molecules that randomly diffuse within tissue [Basser, 1995]. This indirect nature of measurements is exacerbated by the macroscopic spatial-temporal scale of MRI measurements (e.g. measure at millimeter scale to probe properties of processes that are at the micron scale, or measure at the second scale to probe neuronal activity that occurs at the millisecond scale). In order to provide a mapping from (coarse-scale) MRI measurements to (potentially finer-scale) brain properties of interest *biophysical models* are needed (Figure 1).

Using models to estimate parameters of interest from MRI observations is known as the

inverse problem. Crucially, this inverse/estimation process can be challenging as it is characterised by **uncertainty**. There are many sources that can contribute to uncertainty during the inference. Firstly, **MRI is inherently noisy**. Due to the way the MRI signal is acquired, the signal-to-noise ratio (SNR) can be low and, additionally, there is a trade-off between the SNR, the spatio/temporal image resolution (the higher the resolution, the lower the SNR) and the scan time affordable. Low SNR can affect the precision and accuracy of estimated model parameters (e.g., [Jones and Basser, 2004]). Secondly, most models are phenomenological (i.e. tend to explain the behaviour of the measured signal rather than its true sources). Failure in having the correct representation of "truth" in models (attempt to link macroscopic measurements to microscopic compartments), can therefore lead to errors and inaccuracies [Jelescu et al., 2020; Murray et al., 2018; Novikov et al., 2018].

For the above reasons, frameworks in MRI data processing that deal with noise and uncertainty mapping are of paramount importance. At its extreme, uncertainty in MRI modeling can affect the interpretability and reproducibility of findings [Griffanti et al., 2016] and the inability to properly characterise it can have detrimental effects. **In this thesis** we explore a number of frameworks for MRI denoising and develop evaluation and uncertainty mapping procedures. We focus on a particularly noisy MRI modality, diffusion MRI, which allows brain microstructure and connectivity to be estimated. Diffusion MRI is SNR-limited as the measured signal of interest is a signal attenuation, i.e. the measurement "destroys" rather than generates signal, therefore leading to inherently noisy measurements.

We first develop **a framework for evaluating denoising algorithms** in an objective manner and consider the inherent properties of the diffusion MRI signal. We then consider a Bayesian approach for uncertainty mapping in model parameter estimation, approached in two ways. Firstly, by **revisiting a random-walk scheme of the traditional Markov-Chain-Monte-Carlo (MCMC)** and demonstrate how computational efficiency can be increased with better algorithmic designs. Secondly, by employing a data-driven approach based on Artificial Neural Networks (ANN), to **design an amortised inference framework** where data are used to learn an approximation of the inverse map-

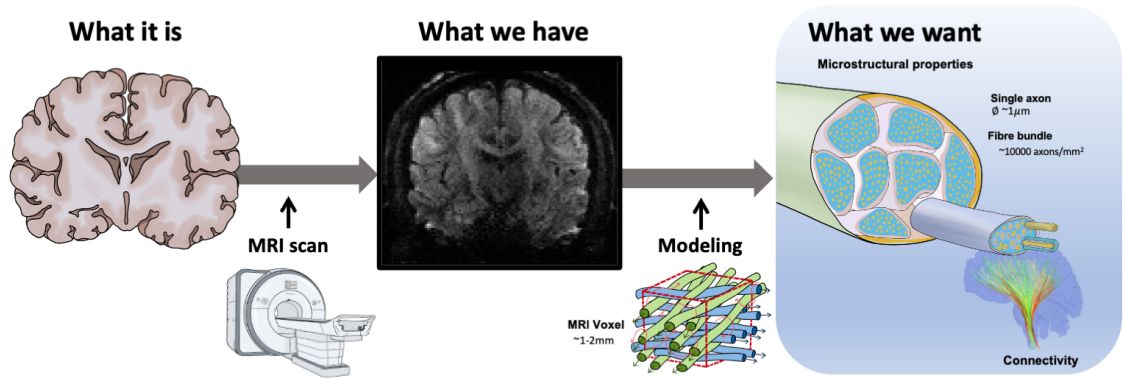


Figure 1.1: From measurements to biophysical properties - To understand the underlying mechanisms of the brain (*What it is*), the properties of interest (*What we want*) are indirectly inferred from biophysical models applied to MRI measurements (*What we measure*). In the figure, an example of diffusion MRI is represented, where microstructural properties of the white matter and structural connectivity between regions are commonly studied.

ping. The approaches developed in this thesis open new opportunities for dealing with noise and uncertainty in modeling dMRI data for estimating tissue microstructure and brain connections.

1.1 Organization of the thesis

This thesis is organized into seven chapters. The first two introduce background knowledge about noise in MRI, the basics of diffusion MRI (dMRI), and existing estimation approaches to address the inverse problem in neuroimaging with a focus on diffusion MRI biophysical models. The next three chapters present the original research contributions of this thesis. Finally, the last chapter summarizes the results presented and discusses future perspectives. More specifically:

Chapter 2 introduces an overview of the physical sources and properties of thermal noise that affects MRI measurements. It further summarises the basics of diffusion MRI, where water diffusion patterns are probed by measurements of signal attenuation. The application of biophysical dMRI models to estimate brain properties such as microstructural features of the white matter (WM) or structural connectivity are reviewed as well. Finally, the chapter links the dMRI signal and noise properties reviewed, and provides some examples of how uncertainty quantification can be crucial during the inference process, particularly when the signal level is very low.

Chapter 3 reviews the main methods to solve the inverse problem, estimate parameters from data and map uncertainty in microstructural modeling in dMRI. Noise and other sources of uncertainty can be accumulated throughout the modeling process. Bootstrapping and Bayesian inference are reviewed as particular methods broadly used for parameter estimation, both in parametric and non-parametric contexts. Applications of uncertainty mapping are shown for experimental design, but also for probabilistic tractography algorithms used to map brain connections. Finally, the chapter overviews new directions for inference based on Machine Learning (ML) that have been recently introduced as potential alternatives to overcome limitations in current approaches.

Chapter 4 is the first original chapter and is concerned with denoising approaches in diffusion MRI. It starts with a review of denoising methods specific to diffusion MRI, an existing gap in the current literature. It then presents a set of considerations for evaluating denoising approaches and a new framework for evaluating them objectively. This includes evaluations on the impact in the raw signal to effects in advanced modeling estimates. Different brain dMRI datasets at different resolution and SNR regimes have been acquired to support this framework. The framework is then used to compare a number of existing denoising approaches that operate either in the magnitude MRI signal or in the complex MRI signal domain.

Chapter 5 is concerned with mapping uncertainty in dMRI biophysical models through classical Bayesian inference¹. It builds upon the Metropolis-Hasting Markov Chain Monte Carlo (MH-MCMC), a sampling method that requires thousands of iterations per model fit and, therefore, it is highly time-consuming and inefficient. A hybrid MCMC is developed by introducing a novel block-design, allowing groups of multiple parameters to be inferred simultaneously and reducing the number of iterations needed. Using real human brain data, comparisons between the results obtained with the classical MH-MCMC and the method proposed shows that the hybrid MCMC accelerates the inference process while achieving similar accuracy and precision in the estimates.

¹We use the term "classical Bayesian inference" throughout the thesis to refer traditional methods to perform Bayesian inference, e.g. MCMC or VI; it does not refer to "classical" (i.e. frequentist) vs. Bayesian approaches.

Chapter 6 extends the Bayesian inference framework to non-classical designs, using modern artificial intelligence algorithms. The idea of simulation-based inference (SBI) in dMRI is explored and adapted. This novel approach allows the integration of Bayesian principles of inference and uncertainty quantification into Neural Networks capable of learning the inverse mapping using synthesised datasets. This learning process avoids the necessity of repeating the inference process for each new data point or subject, also known as inference amortisation. As never evaluated before in neuroimaging, a formal introduction of SBI is provided. This is followed by tests in both simulated and experimental data, comparing the results obtained by the SBI with validated classical approaches such as the MH-MCMC. Results obtained suggest ground-breaking speed-ups in the inference process while keeping the accuracy in the estimates.

Finally, **Chapter 7** provides a summary of the results in chapters 4-6, conclusions and take-home messages from this research and potential lines of future work.

1.2 Computing Infrastructure and Software

The primary computational resource used for this thesis was the University of Nottingham High-Performance Computing (HPC) cluster. The cluster partitions used consisted of 600 CPU cores in total with memory capabilities ranging from 192 GB to 1536 GB depending on the node and 10 GP-GPUs (6 Nvidia Tesla V100 and 4 Nvidia Tesla P100). A secondary processing system was also used in this work, a Linux interactive server (Intel Xeon CPU E5-2698 v4, 2.20 GHz) with CentOS Linux v7 (630 GB of memory and dual GP-GPUs).

General image processing was performed using the FMRIB software library, FSL (version 6.0.2 and beyond) [Jenkinson et al., 2012]. Statistical analysis and visualisations were performed using the Python programming language (v3.9.12) (Python Software Foundation, <https://www.python.org>) along with several community-developed python libraries for general purpose (mainly NumPy, SciPy, NiBabel, Matplotlib, Seaborn, PyTorch). Python libraries and other software tools dedicated for particular purposes (e.g., denoising or simulation-based inference) are specified in the corresponding *Methods* section of each chapter. Code to reproduce the results is available under the GitHub repository

<https://github.com/josepman>.

Chapter 2

Noise and modeling in diffusion

MRI

The signal of interest in MRI comes from spins that are intentionally excited. However, there are multiple sources that can cause variations in the image intensity. In this chapter, we are concerned with unstructured random variations due to **thermal noise**, which is an inherent property of electronic systems. This can be particularly harmful for MRI modalities when the signal intensity can be relatively low with respect to noise, such as in diffusion MRI (dMRI). Unstructured noise differs from artifacts in the images (e.g., geometric distortions) and physiological confounds (e.g., head motion); those are sources of structured noise, i.e. they affect particular regions in a particular manner and, although also hamper the quantitative analysis of the MRI, they require a different treatment [Le Bihan et al., 2006; Tax et al., 2021]. We overview here sources of thermal noise and its statistical properties in MRI images.

Subsequently, a review of the basic principles of dMRI is provided, the main imaging modality this thesis is concerned with. An introduction to dMRI physics is followed by a section on biophysical modeling of dMRI data. The last section links the above by showing how thermal noise can have a dual effect on dMRI signals and model estimates that a) can lead to numerical biases in model estimates, and b) increases the signal variance producing higher uncertainty in the estimates.

2.1 Thermal noise in MRI

In an MRI experiment, thermal noise is due to both hardware, but also to the presence of a subject with the scanner (subject loading). We will briefly overview the sources of these two noise components, provide a description of the expected statistical properties of thermal noise and comment on how these can be altered in modern MRI protocols.

2.1.1 Where does thermal noise come from?

Thermal noise is unavoidable in any body populated by electrons, such as measurement systems or a human body. Similar to gas laws of diffusion and basic thermodynamics [Einstein, 1956], there is a movement of charged particles within electronic sources. These agitations and collisions between electrons give rise to a **thermal process** known as Johnson-Nyquist noise or **thermal noise** [Johnson, 1928; Nyquist, 1928]. Based on the Equipartition Theorem of Boltzmann and Maxwell, using the idea that electrons are independent harmonic oscillators, and relating it to the Planck Distribution in thermodynamics, Nyquist proved that the power spectral density of the noise in a conductor is given by

$$S_n(w) = \frac{\hbar\omega}{\pi(e^{\frac{\hbar\omega}{2k_B T}} - 1)} \approx 2K_B T \quad [per\ Hz] \quad (2.1)$$

where $\hbar = 6.62607015 \cdot 10^{-34} [m^2 \cdot kg/s]$ is the Planck constant, $K_B = 1.38 \cdot 10^{-23} [J/K]$ is the Boltzmann constant, T is the temperature (in K). The relationship shown in eq.2.1 indicates that, for central frequencies $\omega < k_B T / \hbar \approx 6000 [GHz]$ (i.e., MRI systems are in the order of 3 T \sim 128 MHz, so unless we go to 3000T magnets eq.2.1 holds), the power spectral density of the noise can be represented by an uniform distribution, independent of the frequency. I.e., the total amount of noise in a conductor increases with temperature and it is constant in the spectrum affecting all the frequencies equally, a property known as **white** noise (as all spectral components resulted from splitting off the white light by Newton). The 2 factor is introduced to take into account the negative frequencies (respecting the power conservation laws). If $S_n(w)$ is integrated over the bandwidth B (in Hz) of interest (i.e. the pulse bandwidth in MRI), it is possible to get the mean square voltage $\langle v^2 \rangle$ of the noise in the transmission line (i.e. the background noise of the image):

$$\langle v^2 \rangle = 4 \cdot K_B \cdot T \cdot B \cdot R \quad (2.2)$$

E.g., for a given resistance of $R = 50 \Omega$, there is a root mean square (RMS) noise voltage of $v_{RMS} = 1 \text{ } [nV/\sqrt{Hz}]$. As specified in eq.2.2, part of a current flowing through a conductor will be partially dissipated into heat because the conductor has some resistance R , so this type of noise will be always present in the hardware electrical resistance R_{HW} from the receiver coils, data cables and other electronic sources within the MR scanner.

On the other hand, human bodies are mainly made by water, which is a conductor. Only a tiny fraction of all the protons in the body are excited; the rest remain in thermal equilibrium, moving between high and low energy states, and contributing to the background noise. Furthermore, the current flowing in the coil and electronic components will produce a magnetic field in the object being scanned that, by the Faraday-Lenz's law, will induce a current of the opposite sign in the coil, increasing its effective resistance R_{HW} and, therefore, the hardware noise. This is known also as **subject loading**. Hence, subjects can be seen also as a conductor with an associated resistance R_S and, in fact, the subject's loading is more important than the intrinsic hardware noise for systems above 1T: $R = R_{HW} + R_S \approx R_S$ [Redpath, 1998]. The sample noise R_S is also dependent on elements such as the coils (geometry, placement, number channels, types and size of the coils, etc.) or the bandwidth (which differs in each pulse sequence). E.g. the less copper wound in the coil, the lower skin depth, the less it couples with the body, and the lesser the sample noise [Navest et al., 2019].

Finally, there is another non-thermal process due to the random fluctuations in the amount of quantised electrons flowing through conductive pieces that produce what is known as *shot-noise* (apart from Eddy currents). It has no dependence on temperature or resistance, only on the amount of current flowing. In pure conductors at equilibrium, as electrons move equally freely in every direction, movements in opposite directions are cancelled out and the net average is 0; the shot-noise is not observed. However, in a semiconductor, where an electron is not free to move equally in every direction, shot-noise is manifested. Thus, the effects of transistors (semiconductor) within typical pre-amplifier circuits of the scanners present shot-noise. Nevertheless, it follows a Poisson process, so for large enough

currents (as in MRI), it can be approximated by a Gaussian and added to the thermal noise from the circuit and the samples being measured, contributing to the overall noise level of the system.

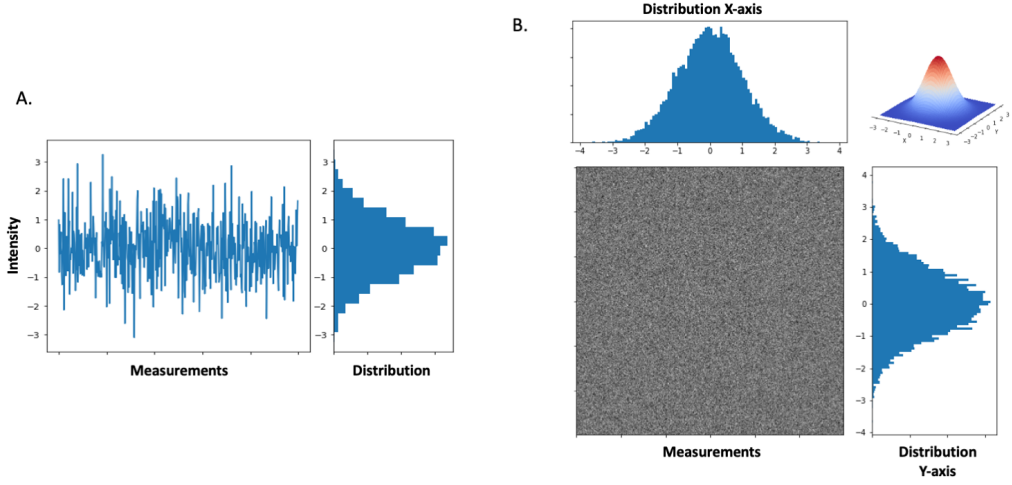


Figure 2.1: Representation of the Additive White Gaussian Noise (AWGN) properties of the thermal noise in A) 1 dimension, and B) 2 dimensions.

2.1.2 Statistical properties of MRI thermal noise

All in all, background noise is the result of summing a large number of different sources, most of them originated by thermal noise (e.g., see Fig.2.3). It cannot be represented by an analytical function but by using statistical terms. By means of the Central Limit Theorem (CLT), it can be demonstrated that **thermal noise**, which is assumed to be the sum of independent random fluctuations, is well approximated by a **Gaussian** random variable with **mean** $\mu_N = 0$ (the random displacement of electrons cancels out the net average) and **standard deviation** σ_N . It can be described also as **Additive White Gaussian Noise (AWGN)** (see Fig.2.1), where the *additive* term means that the magnitude of the noise does not depend on the magnetization magnitude or RF excitations. Furthermore, it can be assumed to be temporally **stationary** (statistical properties of the noise distribution do not depend on time) and **ergodic** (statistical properties can be reliably estimated from *long enough* samples).

The MRI signal is quite small in general, it is therefore typically measured using two-channel quadrature detectors (real and imaginary) [Callaghan et al., 1988], which allow an increase in SNR by $\sqrt{2}$. Measurements are performed at distinct spatial frequencies (k) giving rise to arrays of continuous complex numbers $S(k)$ [Callaghan et al., 1988]:

$$S(k) = S_R(k) + N_R(k) + i * (S_I(k) + N_I(k)) \quad (2.3)$$

where R and I sub-indices denote the real and imaginary part of the signal (S) and the noise (N), respectively, $i = \sqrt{-1}$ and k is the location in k-space in which the signal is acquired at. Apart from the statistical properties commented above for the thermal noise, it is generally assumed that the noise components $N_R(k)$ and $N_I(k)$ are **independent, identically distributed (i.i.d.)** and, as they arise from the same scanner, they are zero-mean Gaussian distributions with equal variance σ_N for the two components.

Given the phase aberrations in the measurements (e.g., because of Eddy currents induced by the gradients) and/or phase-shifts (produced by involuntary physiological motion or inhomogeneities in the magnetic susceptibility) that can cause local signal cancellations across acquisitions, forcing all the image information into the real component or working with **magnitude-only images** has become the standard. At the expense of lower SNR, by simply discarding the phase information they prevent these phase artifacts and allow easier clinical interpretations. In these cases, the magnitude m of the complex signal is given by:

$$m = \sqrt{((S_R(k) + N_R(k))^2 + (S_I(k) + N_I(k))^2)} \quad (2.4)$$

Although the Inverse DFT (IDFT) used to reconstruct the images conserves the properties of the noise distribution, the **magnitude calculation** introduces a non-linear transformation. As a consequence, the distribution of the voxel intensities is not Gaussian anymore on magnitude images. The statistical description of the magnitude of two independent Gaussian random variables that are in phase quadrature was initially developed by Rice [Rice, 1945]. After some calculus, it can be demonstrated that any magnitude values derived from complex arrays, as in 2-channels MRI data, follow a **Rician distribution**:

$$P(s; A, \sigma_N) = \frac{S}{\sigma_N^2} \cdot \exp\left(\frac{-(S^2 + A^2)}{2\sigma_N^2}\right) \cdot I_0\left(\frac{S \cdot A}{\sigma_N^2}\right) \quad (2.5)$$

where S is the signal measured, A is the signal in the absence of noise, σ_N is the standard deviation of the Gaussian noise in the real and imaginary images (assumed to be equal) and I_0 is the modified zero-th order Bessel function of the first kind. In the case of no

signal applied (e.g., **in the background** of the image where there is no response signal produced by the gradients), the intensity values of a voxel represented will correspond only to the noise and will follow a **Rayleigh distribution**, which can be seen as a particular case of the Rician distribution with $A = 0$, as stated by [Edelstein et al., 1984] and later [Gudbjartsson and Patz, 1995], among others:

$$P(s; A, \sigma_N)|_{A=0} = \frac{S}{\sigma_N^2} \exp\left(\frac{-S^2}{2\sigma_N^2}\right) \quad (2.6)$$

2.1.3 Modern MRI protocols. Effects on SNR and noise properties

Current acquisition and reconstruction protocols of MR images do not rely on single-channel coils anymore. The pressure for reducing scanning time led to the development of accelerated imaging strategies. One of the most straightforward hardware acceleration approaches is the use of phased arrays coils or **multi-channel coils** [Roemer et al., 1990]. Here, different coils are placed around the object being scanned to acquire images simultaneously, so this can be seen as a technique that increases the SNR by averaging (and, therefore, allowing to reduce scan time needed to achieve the same SNR). Reconstruction techniques that are able to properly combine the images from all coils are required, such as the **Root-Sum of Squares (RSoS)** approach [Larsson et al., 2003; Roemer et al., 1990].

Given that imaging speed acquisition is limited by physical (gradient amplitude and slew-rate) and physiological (nerve stimulation) constraints, software acceleration approaches were proposed at the end of the 90's to take advantage of the redundancy and symmetric properties of the k-space. These methods aim to reconstruct the same image with less k-space measurements, as in **partial-Fourier** sampling. One of the main techniques is **Parallel Imaging (PI)** [Deshmane et al., 2012; Pruessmann et al., 1999]. In PI, phase-encode direction is sampled every R lines (acceleration factor), allowing to reduce the scanning time by R . However, when k-space is undersampled, the Nyquist criterion is violated, and Fourier reconstructions exhibit aliasing artifacts. To address this issue, different implementations are available, such as the *Generalized Autocalibrating Partial Parallel Acquisition (GRAPPA)* [Griswold et al., 2002] which operates in k-space, or *Sensitivity Encoding (SENSE)* [Pruessmann et al., 1999] that operates in image space.

Although the approaches and their performance in different scenarios can differ, they are based on similar principles: a weighted non-linear reconstruction of the image using the coil spatial sensitivity profiles. Furthermore, motivated by the widespread success of data compression in natural images, approaches such as **Compressed Sensing** [Lustig et al., 2007] have been applied to MRI. Here, by taking incoherent measurements of the k-space, the subsampling is designed to produce noise-like artifacts in both x- and y-directions. Due to the image redundancy, and given some mathematical properties, it is possible to recover the image from a compressed version represented by a reduced number of (sparse) components in the k-space. These components can be obtained by projections, such as the wavelet transform, Discrete Cosine Transform (DCT) or any other basis function set. Recently, **Machine Learning (ML)**, and especially Deep Learning (DL) (e.g., see [Chen et al., 2022; Pal and Rathi, 2022; Zeng et al., 2021] for reviews), have been successfully applied to learn the optimal k-space undersampling and improve aspects of the techniques above, such as the generalization of reconstruction methods when using different coils [Beauferris et al., 2021].

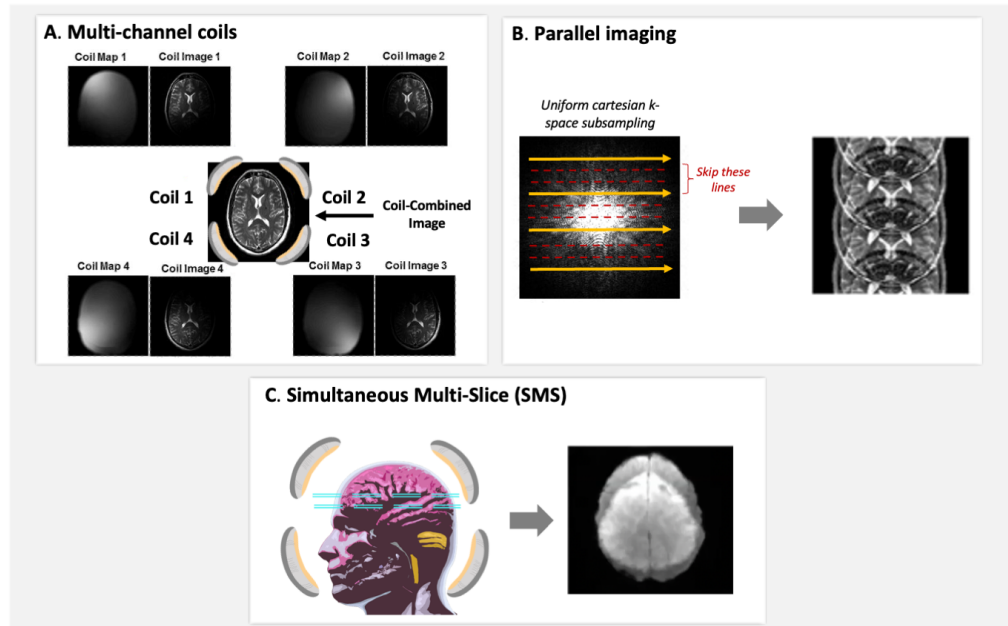


Figure 2.2: Main MR imaging acceleration techniques. Figures adapted from [Hamilton et al., 2017]

In general, as all these methods provide a shorter scan time, fewer data points are acquired and averaged, so there is also an inherent **reduction in the SNR**. Furthermore, given the nature of reconstruction, modern acquisitions lead also to **spatial inhomogeneities** of noise, i.e. noise variance being different at different locations. This spatial variability

can be captured by the geometry factor (**g-factor**), which provides a map of noise amplification at different parts of the image. For instance, the resultant SNR with and without acceleration in parallel imaging is:

$$SNR_{PI} = \frac{SNR_{non-PI}}{g\sqrt{R}} \quad (2.7)$$

In addition to these complexities, modern MRI protocols can severely modify the statistical properties of the noise, making their definition much more complex [Aja-Fernández et al., 2011; Dietrich et al., 2007; Sotiropoulos et al., 2013c]. Similar to the calculation of the magnitude images, non-linear and iterative reconstruction methods no longer preserve Gaussian properties of the signal; magnitude MR images obtained from accelerated protocols generally exhibit properties that can be described by **approximations to the non-central Chi distributions family**, which includes the Rician distribution as a special case (with $\sigma = 1$ in eq.2.5 and $k = 2$ in eq.2.8).

$$P(s; k, \lambda) = e^{-(S^2 + \lambda^2)/2} \frac{S^k \lambda}{(\lambda S)^{k/2}} I_{k/2-1}(\lambda S) \quad (2.8)$$

where k specifies the degrees of freedom (i.e., the number of S_i datapoints) and λ relates to the mean of S as:

$$\lambda = \sqrt{\sum_{i=1}^k \left(\frac{\mu_i}{\sigma_{N_i}} \right)^2}$$

Some properties of this distribution, such as the degrees of freedom (DOF) k , will depend on factors such as the number of independent multi-channel coil receivers and their sensitivity profiles, the reconstruction technique used to combine the signals from the different coils (e.g., RSoS), the parallel imaging method applied (e.g., GRAPPA, SENSE), or some of the possible filters applied in k-space (e.g., Hanning, elliptical, etc.) [Aja-Fernández et al., 2013; Dietrich et al., 2007, 2008; Sotiropoulos et al., 2013c]. For instance, using the RSoS to reconstruct multichannel data results in a magnitude signal that follows a non-central Chi distribution in case of non-accelerated imaging, and to an approximate non-central Chi distribution when accelerated [Constantinides et al., 1997; Kellman and McVeigh, 2005]. Its degrees of freedom will depend on the number n of channels. The greater number of channels, the more deviation from the Rician distribution. This holds if the signals obtained from the channels are independent. In real experiments with high-density phased-array coils, the non-central Chi distribution has effectively fewer DOF k

and larger variance than the ones predicted for independent channels [Aja-Fernández et al., 2011].

All the above challenges in noise characterisation apply to and are very relevant to diffusion MRI, as the measured information resides in the signal attenuation, an inherently small and noisy signal. In some regions, the measured dMRI signal can be as low as the background noise level, also known as the **noise-floor**. In such cases, the signal can be rectified by noise [Jones and Basser, 2004], leading subsequently to biased quantification of model parameters and erroneous estimations. In the next sections, we review the main principles behind dMRI acquisition and modeling that will allow us to further explore the effects of thermal noise in dMRI quantitative analysis.

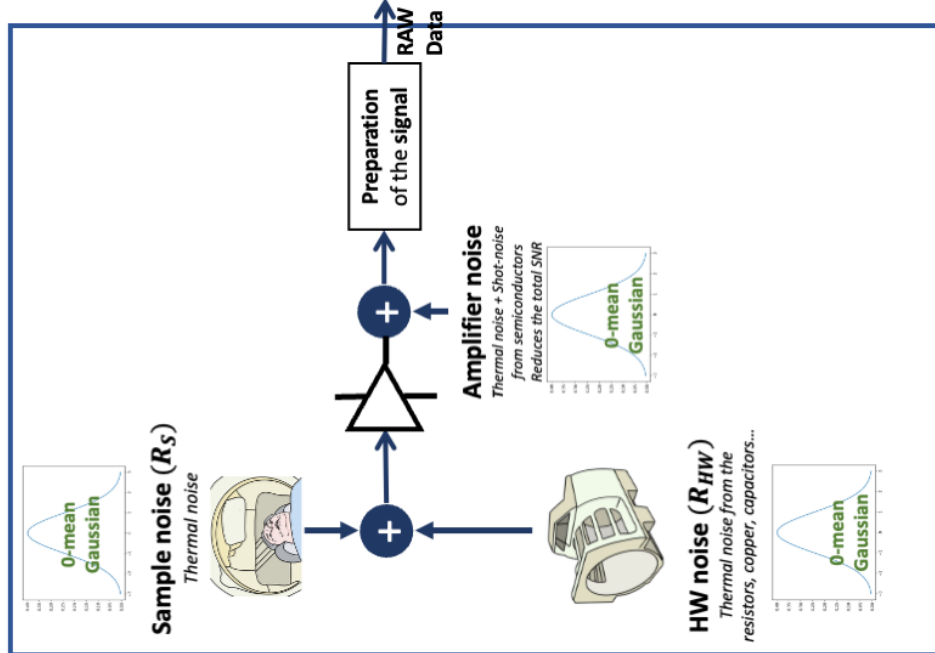
2.2 Diffusion MRI and biophysical modeling

MRI emerged from the technique of nuclear magnetic resonance (NMR), widely used in chemistry and biochemistry for compound and structure identification. Thus, MR images are formed by contrast mechanisms. Although contrast can be achieved by multiple elements (Carbon, Sodium, ...), Hydrogen concentrations are much higher in human bodies and, as it only has spin $\pm 1/2$, it is easier to operate with. Thus, diffusion MRI (dMRI) uses the self-diffusion of the excited spins of water molecules (H_2O) within the different tissues as an imaging contrast mechanism [Bihan et al., 1986; Dietrich et al., 2010]. This diffusion of particles is used as a proxy to describe the microstructural properties of the imaged tissues. In the following sections, we will describe what drives the phenomena of diffusion and how MRI can use these principles to depict structural and connectional features of the brain.

2.2.1 Diffusion phenomena

At a macroscopic level, the diffusion process can be described by the Fick's laws [Fick, 1855]. The first law states that diffusion can be seen as a net flux J from high to low-concentration regions. The second law describes the spatio-temporal dynamics of this diffusion process, also known as the Heat equation or the Diffusion equation. In a homogeneous region, the flux J is given by:

1. Pre-acquisition



2. During and post-acquisition

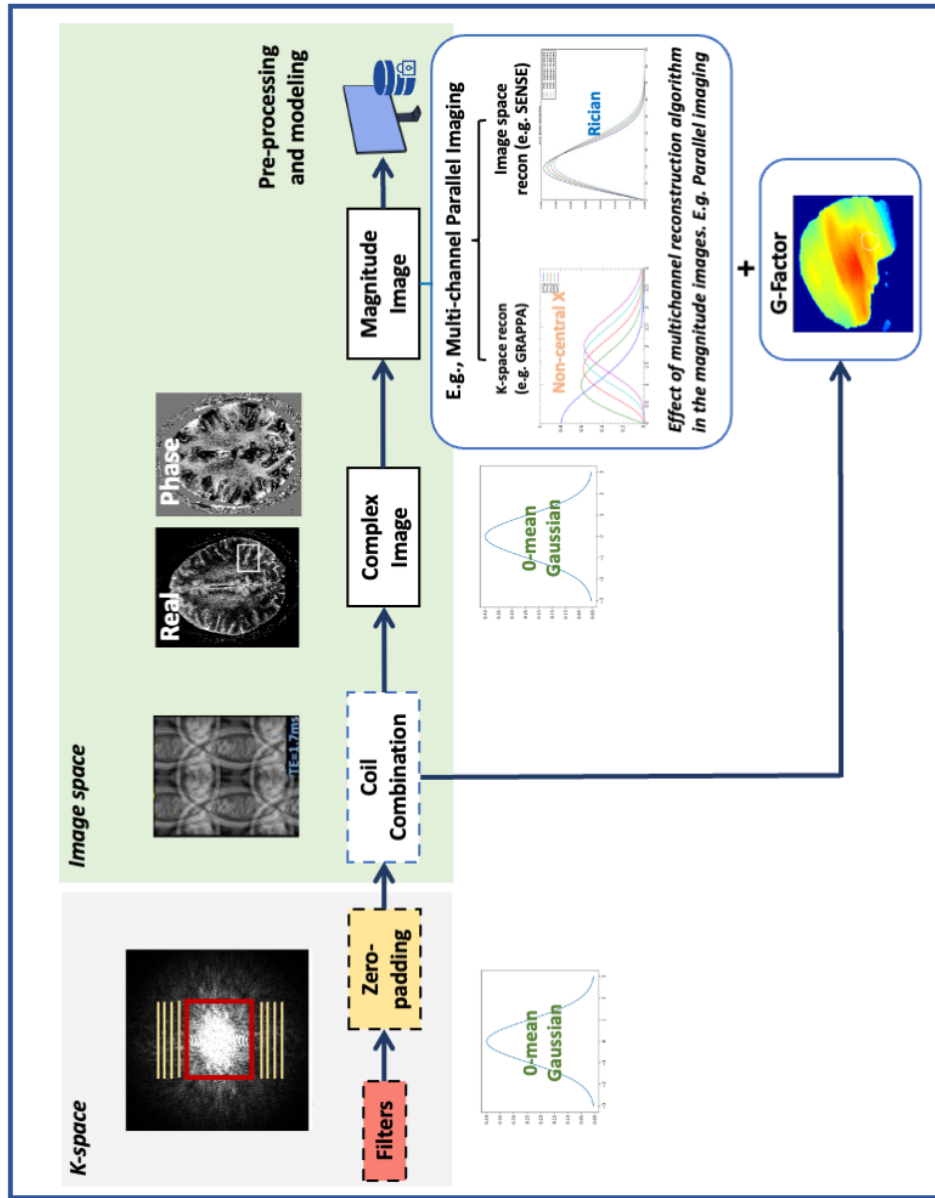


Figure 2.3: Scheme of noise properties and factors that have a relevant influence on it, ranging from the scanner and acquisition parameters to reconstruction and pre-processing steps. Here, we give an example of multi-channel parallel imaging. While in the complex space, the thermal noise generally follows a 0-mean Gaussian distribution. However, when performing the multiple coils reconstruction and getting the magnitude of the image, noise distribution properties are modified. This modification will depend on factors like the type of reconstruction employed (e.g. GRAPPA or SENSE) or the number of coils. These can interfere also with In-plane, out-of-plane acceleration techniques. In fact, at this stage not only the distribution of noise changes but also the variance in this distribution is spatially-variable, represented by the G-factor map.

$$J = \frac{d\phi}{dt} = -D \cdot \nabla^2 \phi \quad \left[\frac{mol}{m^2 s} \right] \quad (2.9)$$

where ϕ is the concentration in the high particle concentration region, and D is the diffusion coefficient (in m^2/s), a scalar constant that depends on the medium viscosity, the particle size and the temperature [Hobbie and Roth, 2007]. Upon Fick's law, Einstein extended in 1956 the definition of diffusion into the microscopic level [Einstein, 1956]. Here, the diffusion process can be approximated as thermally-driven random motion, or Brownian motion. Having a population of molecules sufficiently large in a homogeneous and barrier-free medium, the mean squared diffusion displacement $\langle r^2 \rangle$, averaged over the total number of molecules in the ensemble, is directly proportional to the observation time τ (in *sec.*) and the number of dimensions n :

$$\langle r^2 \rangle = 2 \cdot n \cdot D \cdot \tau \quad [mm^2] \quad (2.10)$$

In this way, the averaged displacement of a flux of particles is defined in statistical terms. When both the diffusing particles and the medium in which they diffuse are of the same type, the process is called self-diffusion (and D the self-diffusion coefficient). At human body temperature (37 °C), the self-diffusion coefficient of water is $3 \cdot 10^{-3} [mm^2/s]$.

The diffusion profile of a hydrogen proton (spin), at location x_0 , can be expressed in terms of conditional probability density function as $P_s(x|x_0, \tau)$, i.e., the probability of diffusion from x_0 to x in a time τ for a single spin s . For an ensemble of particles, the average probability for any particle in the ensemble of having a relative diffusion displacement r after time τ is given by [Callaghan, 1991]:

$$P(r, \tau) = P(-r, \tau) = \int P_s(x_0 + r|x_0, \tau) \cdot \rho(x_0) \cdot dx_0 \quad (2.11)$$

with ρ defining the spin density at a specific location. This probability is also known as the diffusion propagator, **ensemble average propagator (EAP)** or diffusion scatter pattern, and plays an essential role in dMRI. The EAP is the object of interest when modeling dMRI signals in every voxel. It has been widely studied and it can be proved that, in a barrier-free medium and having all the particles concentrated in the same point at time $\tau = 0$, the integral in eq.2.11 reduces to a multivariate Gaussian distribution

[Basser et al., 1994]:

$$P(r, \tau) = \frac{1}{\sqrt{(4\pi D\tau)^3}} \cdot \exp\left(\frac{-r^2}{4D\tau}\right) \quad (2.12)$$

I.e., the EAP is a 3D function, (and r is a vector, either $[x, y, z]$ in cartesian or $[r, \theta, \phi]$ in spherical coordinates), where mean $\mu = 0$ indicates that in a barrier-free medium the total displacement is 0 because random motion in every direction cancels out each other, and the variance $\sigma^2 \sim 2D\tau$ for each direction depends on the observation time τ and the coefficient D .

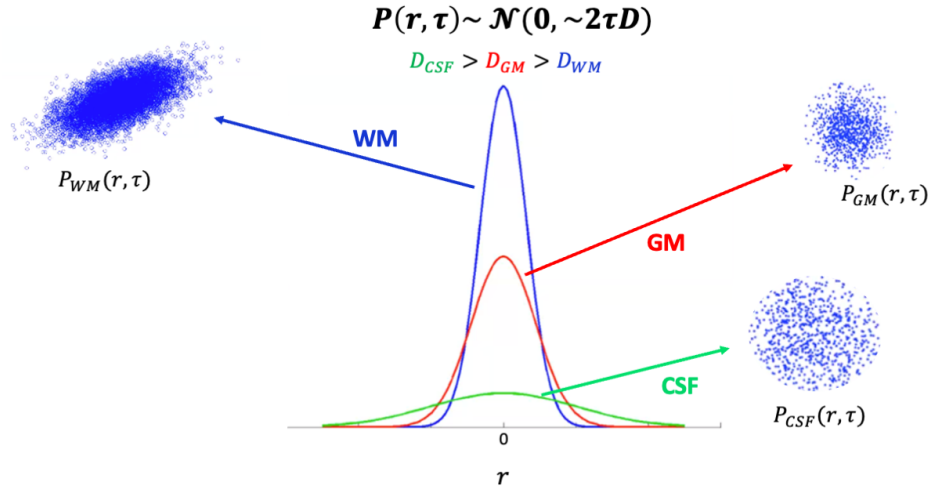


Figure 2.4: *The normalized distribution of the average propagator indicates the probability of (an ensemble of) molecules of having a relative displacement r (in m.) in a time τ (in sec) for 1-D. Depending on the tissue, the diffusion is completely free (isotropy), as in the Cerebro-Spinal Fluid (CSF, in green), or restricted in some specific direction (anisotropy), as in the White-Matter (WM, in blue). From this profile of diffusion or scattering, the EAP $P(r, \tau)$ can be constructed. Image adapted from the FSL Course (<https://fsl.fmrib.ox.ac.uk/fslcourse/>).*

2.2.2 How can we use MRI to measure diffusion?

The details of MRI physics for image acquisition and image formation are complex, so we only give a brief summary here.

A permanent magnetic field B_0 (in T) aligns all chemically equivalent protons that have the same resonant frequency (e.g., all the protons from water) into the same orientation and makes them precess at the Larmor frequency, $f_0 = \gamma_H \cdot B_0$ (in Hz), with $\gamma_H = 42.576 [MHz/T]$ the gyromagnetic ratio of the hydrogen protons. If the system is irradiated with energy from a transverse magnetic field B_1 at the Larmor frequency, nuclear magnetic resonance will occur. The B_1 field is typically turned *on* for only a few

milliseconds, during which the net magnetisation M tips away from the main magnetic field orientation and start precessing on a different plane (transverse plane). When B_1 is turned off, the net magnetization continues to precess around B_0 and starts to relax to return to the equilibrium state again. Subsequently, a receiver coil capable of sensing changes in magnetic flux within the transverse plane is employed to detect the transverse component of M , denoted as M_{xy} . A small current is generated in this coil via the Faraday-Lenz induction principle. To detect protons from the different anatomic parts of the body, as they resonate at the same frequency, a spatially-selective magnetic field gradient is introduced in the magnetic field to spatially encode the image. This process effectively uses the resonant frequency to encode (i.e., represent) the spatial position in the gradient direction. The pattern of B_1 excitations and M_{xy} signal readouts is known as a sequence. One of the most common sequences is the Spin-Echo sequence, where excitation that brings spins to the transverse plane (90° excitation) is followed by another excitation that flips the spins (180° excitation). That way any spin dephasing during relaxation is flipped and an "echo" is formed.

In dMRI, in order to measure the amount and direction of the diffusion of water molecules within the brain, new sequences were built on top of Spin-Echo, originating what is known as **Diffusion Weighted Imaging (DWI)**. More specifically, Stejskal and Tanner introduced additional diffusion-sensitizing gradients to the classical SE, known as the **pulsed gradient spin echo (PGSE)** sequence [Stejskal and Tanner, 1965]. The PGSE applies two identical diffusion-sensitizing magnetic field gradients, one on either side of the 180° excitation (see Figure 2.5). These lead to phase differences that depend on the starting and end position of a spin during the application of the first and second gradient respectively.

Specifically, during the application of a gradient, the spins located at a position x will experience a Larmor frequency offset $\omega_G(x, t) = \gamma_H \cdot x \cdot G(t)$ (in Hz) along the direction of the gradient G , being $G(t) = [G_x(t), G_y(t), G_z(t)]$. The accumulated phase during the application of a diffusion-sensitising gradient will be:

$$\Phi_G(x) = - \int_0^\infty \omega_G(x, t) dt = -\gamma_H \cdot x \int_0^\infty G(t) dt \quad (2.13)$$

Let

$$q = \gamma_H \int_0^\infty G(t) \cdot dt, \quad (2.14)$$

what is known as the q-vector. In the case of a constant gradient of magnitude \mathbf{G} and duration δ , q reduces to

$$q = \gamma_H \cdot \mathbf{G} \cdot \delta \quad (2.15)$$

and the phase φ_i acquired by a spin i at location x during the application of this gradient:

$$\varphi_i(x, t) = -q \cdot x \quad (2.16)$$

i.e., the phase accumulated will depend on the q-vector. Therefore, the dMRI measurements are acquired in a frequency domain also known as the **q-space** [Callaghan, 1991], where diffusivity values are based on gradient magnitude, directions and duration.

Based on these concepts, PGSE applies two identical gradients on either side of a 180° RF pulse (see Fig.2.5). The 180° RF pulse will reverse the phase change that occurred prior to its application. So, being x_0 the position of a spin at $t = 0$ and x_1 the position at $t = \Delta$, there are two possible situations after the two gradients are applied (see fig.2.5):

- (a) In the case of **stationary spins**, $x_1 = x_0$. The two phases cancel each other and there will be a complete rephasing: $\varphi_i(x, t) = -q \cdot (x_1 - x_0) = 0$
- (b) If **spins move** between the first and second gradient, $x_1 \neq x_0$, a phase will be acquired. In an ensemble of spins, different random displacements will lead to different dephasing, what is known as phase dispersion. As a consequence, the measured signal $S(q, \tau)$ will be an attenuated version of the baseline signal S_0 (when no diffusion-sensitizing gradients are applied).

In both cases, the signal attenuation can be measured $E(q, \tau) = S(q, \tau)/S_0$ by integrating over the accumulated induced phases obtained by all possible diffusion displacements in an ensemble of spins. The **q**-vector provides a Fourier-conjugate for the diffusion displacement **r**, i.e. $E(q, \tau)$ and the diffusion propagator $P(r, \tau)$ can be related by the Fourier Transform \mathcal{F} :

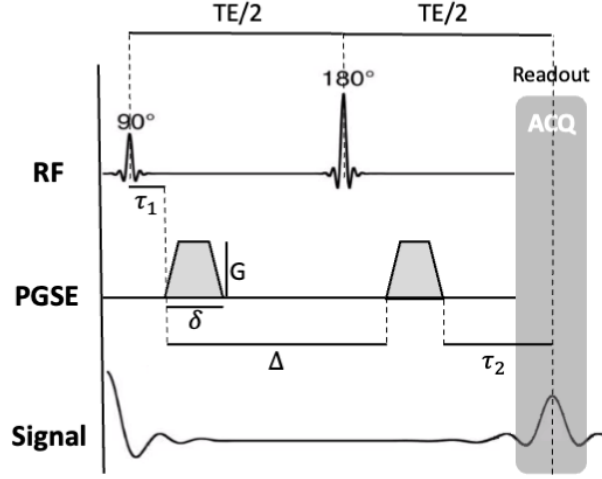


Figure 2.5: Pulse Gradient Sequential Echo (PGSE) sequence - If there is no displacement of spins, complete rephasing occurs and there is no signal loss during the readout. On the other hand, if spins displace from location x_0 to x_1 between gradients, dephasing occurs, such as $\varphi_i(x, t) = -q \cdot (x_1 - x_0)$, and produces a signal loss. In the image, δ is the pulse width, Δ is the pulse separation, G is the gradient vector, TE is the echo time, τ_1 is the time between the RF pulse (90°) and the beginning of the first gradient, and τ_2 is the time between the end of the 180° RF pulse and the readout.

$$E(q, \tau) = \mathcal{F}[P(q)] = \int P(r, \tau) \cdot e^{-iqr} dr \quad (2.17)$$

Measurements at high "frequencies" q will be more sensitive to subtle diffusion displacements, while low q values will capture larger and coarser displacements. By using this Fourier relationship, the diffusion propagator P can be estimated from diffusion measurements in the q -space. Assuming homogeneous diffusivity and barrier-free diffusion, eq.2.17 can be reformulated to obtain the Stejskal and Tanner equation [Stejskal and Tanner, 1965], where the measured signal $S(q, \tau)$ is modelled as:

$$S(q, \tau) = S_0 \cdot e^{-|q|^2 D \tau} \quad (2.18)$$

where S_0 is the signal when no gradients are applied. Let

$$b = \tau \cdot |q|^2 = \tau \cdot \gamma_H^2 \cdot G^2 \cdot \delta^2 \quad (2.19)$$

and, therefore

$$S(q, \tau) = S_0 \cdot e^{-bD} \quad (2.20)$$

This b-value has units of s/mm^2 and determines the amount of diffusion weighting applied

during acquisition. It can be controlled by the amplitude G , the duration of the diffusion-sensitizing gradients δ or the effective diffusion time τ , governed by the time between the two gradients (typically denoted by Δ).

2.2.3 From dMRI signals to probing tissue microstructure

At this point, one may wonder *how will dMRI reveal brain structural features?* In the brain, cell bodies (soma) and dendrites tend to cluster together and form the gray matter (GM); while axons cluster together to form axonal bundles (tracts or fasciculi) in the white matter (WM). When measuring diffusion of water molecules within neuronal tissue, this is hindered by such tissue microstructure (cell membranes, axons and myelin sheaths), that act as barriers to motion. In regions like the **ventricles**, this motion can happen in every direction with the same probability (**isotropic**), as there are no barriers. In the case of **GM**, there are barriers, but there is no systematic hindrance towards any direction, so diffusion is also isotropic. For **WM**, as there is a systematic organisation, water molecules will mainly diffuse along the tracts directions rather than perpendicular to the tracts (**anisotropic**). All the above were first confirmed by Le Bihan et al. [Bihan et al., 1986], who used the PGSE sequence to obtain for the first time a dMRI of the human brain. As they demonstrated, the diffusion in the brain was hindered by the various tissue compartments and when enough diffusion time was allowed, particles hit the boundaries and the diffusion D observed was lower than the expected self-diffusion coefficient in water, i.e. isotropic diffusion like in a free-barrier medium. During the experiments, they also realised that they were not only measuring spin dephasing due to sensitizing-gradients, but also other incoherent artifacts like vasculature pulsations, bulk motion, etc. will contribute to signal attenuation. Hence, the estimated diffusivity was re-termed as *apparent diffusion coefficient* (ADC).

From eq.2.20, it is possible to deduce some important relationships. The diffusion-weighted signal intensity mainly depends on the diffusion gradient parameters, and the mean diffusivity of D , which reflects the underlying tissue configuration. For instance, given the same gradient excitation, each tissue diffusivity will return a different intensity signal, contributing to the image contrast of dMRI. In the CSF and the GM, as there is isotropic random diffusion, the signal measured barely varies with the different gradient orientations applied.

In the WM, on the other hand, there are considerable changes in the intensity signal depending on the fibre orientations and the gradient applied; if gradient direction and fibre orientation are parallel, the attenuation of the signal is maximum (see Fig.2.6). Typical average diffusivity values observed in the human brain are $D_{WM} \sim D_{GM} \sim 7-8 \cdot 10^{-4} \text{ mm}^2/\text{s}$ and $D_{CSF} = 3 \cdot 10^{-3} \text{ mm}^2/\text{s}$, reflecting the lack of microstructural barriers in the ventricles compared to within the brain tissue).

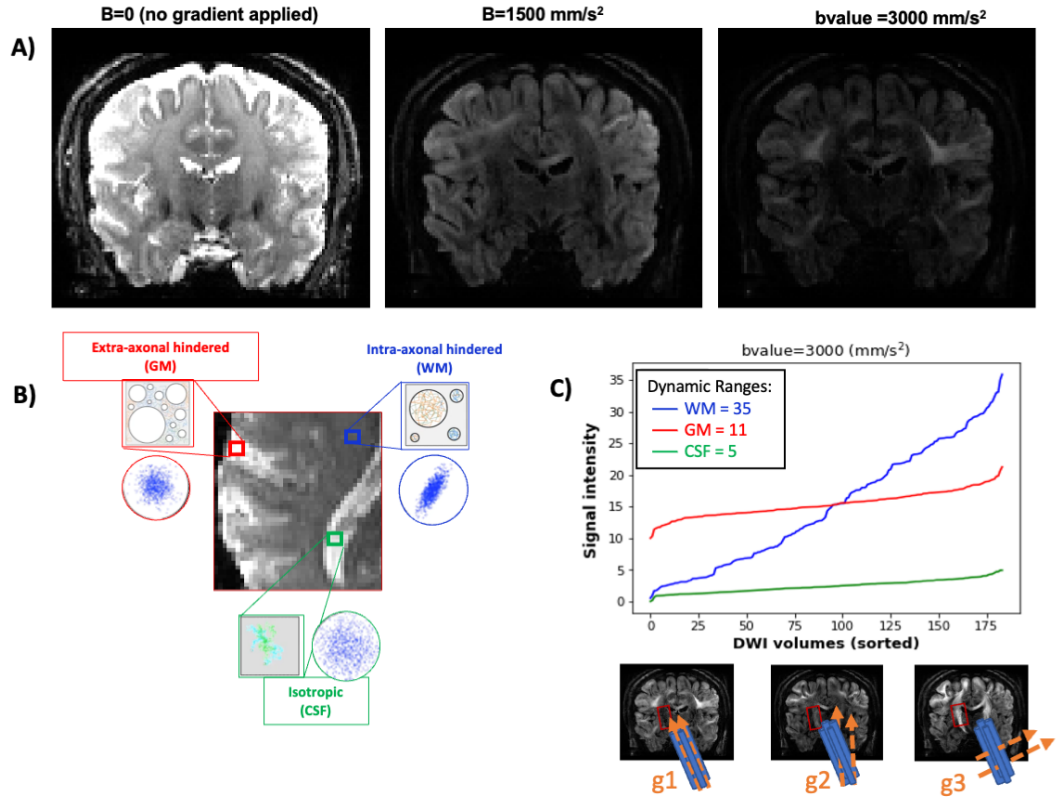


Figure 2.6: *B-values and diffusivity effects in the dMRI measured signal* - A) dMRI at different *b*-values. The higher the *b*-value, the higher the angular contrast but also the attenuation of the signal (and, therefore, the lower the SNR). B) Diffusion profiles in the different brain tissues. C) Example of a dMRI signal measured in each tissue for a *b*-value=3000(mm/s²). The signal is reordered from parallel (highest attenuation) to perpendicular (lowest attenuation) gradient orientations g_i respect to an exemplar high-anisotropic WM voxel (e.g., located in the Cortico Spinal Tract (CST)). As it can be observed, the variations in the WM signal are much higher than in the GM or the CSF because of its anisotropic characteristics.

Without any gradients G applied, the S_0 signal reflects a T2-weighted contrast. As D reflects the microstructural environment, it is by adjusting the combination of *b*-values and the diffusion-sensitizing gradient orientations that this T2-weighted contrast can be manipulated. For instance, when diffusion weighting is applied ($b > 0$), the signal intensity $S \propto e^{-bD}$ is attenuated and, therefore, the SNR is also lower than the T2-weighted image. For a typical *b*-value of $b = 1000 \text{ (s/mm}^2\text{)}$ (common *b*-values range from 0 to 3000

(s/mm^2) in modern acquisition protocols), the signal intensity in the WM will decay by a factor of 0.5 (i.e., $e^{-bD} \sim 0.4965$). The higher the b-value, the better angular contrast (higher \mathbf{q}), but also the higher the signal attenuation (i.e. the lower the signal-to-noise-ratio (SNR)).

2.2.4 Biophysical Modeling of dMRI data

After measuring dMRI signals at several points in q-space (e.g. with a fixed b-value and varying gradient orientations), multiple models are available to link the measured signal $S(q, \tau)$ in each voxel with the underlying tissue microstructure (see [Alexander et al., 2019] for a review). Here we focus on a subset of models that are used throughout this thesis: the Diffusion Tensor model, crossing-fibre approaches and, in particular, the Ball&Sticks model.

Diffusion Tensor Imaging (DTI)

While the simplest dMRI model assumes that a single diffusion coefficient can describe the diffusion scatter pattern in a voxel (see eq.2.20), this limits to descriptions of only isotropic patterns. The first model and most heavily used for capturing anisotropy in the brain is **Diffusion Tensor Imaging (DTI)**, proposed by [Basser et al., 1994]. It extends the solution of the Einstein's equation from eq.2.11 to 3D models, where a positive-definite symmetric diffusion matrix \mathbf{D} is used instead of a scalar coefficient:

$$\mathbf{D} = \begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{yx} & D_{yy} & D_{yz} \\ D_{zx} & D_{zy} & D_{zz} \end{bmatrix} \quad (2.21)$$

This tensor is estimated using all measured signals S_i (acquired for diffusion-sensitizing gradient direction g_i and with a constant b value) in each voxel:

$$S_i(g_i, b) = S_0 \cdot \exp(-b \cdot \mathbf{g}_i^T \cdot \mathbf{D} \cdot g_i) \quad (2.22)$$

where S_0 is the signal with no diffusion gradients applied, and \mathbf{D} the unknown diffusion tensor. If the logarithm is taken on both sides, eq.2.22 becomes a linear system that can be

solved using ordinary least squares (OLS). Therefore, to estimate the unknown tensor, at least 6 signals in different directions and one non-diffusion-weighted image ($b = 0$) need to be acquired. Typical protocols acquire 32 or 64 weighted images, providing more accurate estimations of \mathbf{D} [Jones, 2004].

DTI models the diffusion propagator \mathbf{P} as a **Gaussian distribution** for each axis (x,y,z), where the covariance matrix is proportional to the diffusion tensor \mathbf{D} . Making the assumption that each voxel has a principal direction of diffusion in WM (i.e., an ellipsoidal shape of \mathbf{P} , see Fig.2.7), the orientation can be estimated by calculating the eigenvectors \mathbf{e} of \mathbf{D} . The eigenvalues λ of \mathbf{D} provide the diffusivity of the 3 orthogonal directions of the diffusion distribution. It is assumed that the eigenvector associated with the largest eigenvalue (i.e. the direction of fastest diffusion) corresponds to the principal fibre orientation in that voxel. From the diffusion tensor, a number of scalar rotationally invariant metrics can be constructed [Basser, 1995], such as:

- **Mean Diffusivity (MD)**: given by the mean of the tensor's eigenvalues, that is equal to the average of the ADC measured in the three orthogonal directions. It reflects the average mobility of water molecules. Such, it creates images with relatively continuous values within a tissue, with high MD in regions with less boundaries (e.g. CSF) and lower MD in regions with more microstructure (e.g. GM and WM).

$$MD = \frac{D_{xx} + D_{yy} + D_{zz}}{3} = \frac{\lambda_1 + \lambda_2 + \lambda_3}{3} \quad (2.23)$$

- **Fractional Anisotropy (FA)**: reflects the normalized variance of the tensor eigenvalues. It ranges between FA=0, in cases of perfect isotropy ($\lambda_1 = \lambda_2 = \lambda_3$); and FA=1, in cases of perfect anisotropy ($\lambda_1 \neq 0, \lambda_2 = \lambda_3 = 0$). Therefore, WM regions will have high FA, while CSF and GM will have lower FA values.

$$FA = \sqrt{\frac{3 \cdot \sum_{i=1}^3 (\lambda_i - \lambda_{avg})^2}{2 \cdot \sum_{i=1}^3 \lambda_i^2}} \quad (2.24)$$

DTI maps provide good contrast in **distinguishing between WM (highly anisotropic,**

$\lambda_1 \gg \lambda_2, \lambda_3$) and GM and CSF (highly isotropic, $\lambda_1 \approx \lambda_2 \approx \lambda_3$). However, a lack of DTI anisotropy can also be erroneously observed in WM. In regions with complex fibre configurations where there are more than one main fibre orientation, such as **fibres crossing, fanning, bending**, etc., the DTI model fails to provide good estimates of the fibre orientations and the anisotropy (see fig.2.7). As it is expected that, at least, 80% of the voxels in WM contain fibre-crossings [Jeurissen et al., 2013; Schilling et al., 2017], this leads to numerous challenges in using DTI faithfully in many WM regions [Behrens et al., 2007]. To solve these challenges, more complex models exist for estimating multiple fibre orientations.

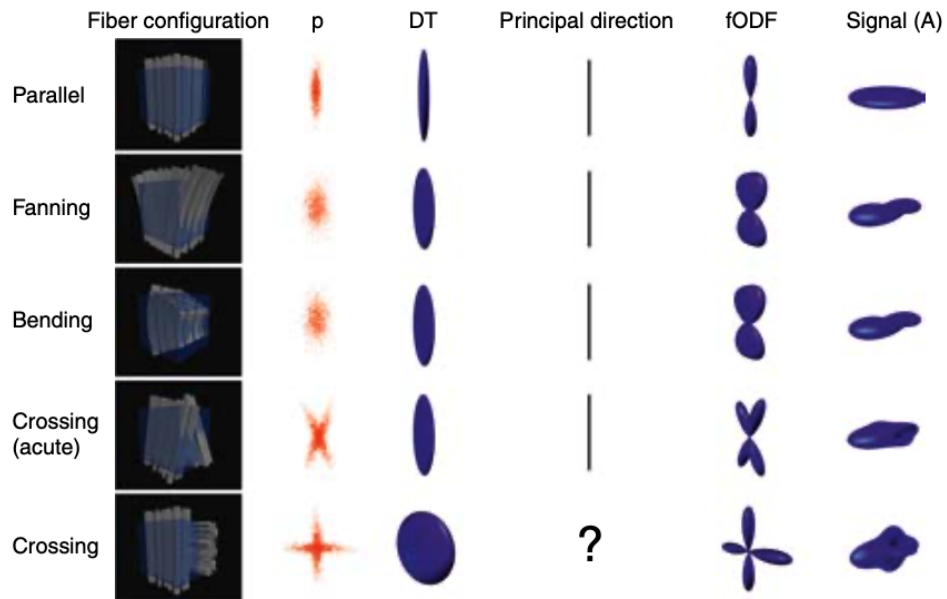


Figure 2.7: Examples of common fiber configurations - First column: Examples of axon fibers configurations. Second column: scatter pattern expected from each of the fibre configurations in the first column. Third column: typical DTI fit. Fourth column: principal direction of the DTI. Fifth column: fibre orientation distribution function (fODF) for each configuration. Sixth column: directional variation of the diffusion-weighted signal for fixed diffusion weighting. For perpendicular crossings (last row), the DTI cannot distinguish a principal direction of diffusion (oblate profile). Image extracted from [Seunarine and Alexander, 2009]

Beyond the Diffusion Tensor

Apart from the inability of DTI to estimate complex fibre configurations, microstructural properties obtained from the tensor model are known to lack specificity, i.e., multiple structural sources can equally modify their parameter values but there is no mechanism to disentangle the true source of it [Farquharson et al., 2013]. For instance, both demyelination, axonal loss and oedema can result in an increase of the radial diffusivity estimated by DTI, but the model is not able to capture the source of it. To address these issues, more

advanced dMRI models have been proposed during the last 20 years (see Fig.2.8). One common classification distinguishes between **model-free approaches** (or phenomenological models), and **biophysical models** or (mechanistic models).

Model-free methods are generally based on the decomposition of the diffusion signal as a linear combination of different functional bases (e.g., spherical harmonics (SH)). Hence, no particular model is assumed for the different types of tissues. One of the first quantities used for characterising multiple fibre orientations was the diffusion ODF (dODF). Diffusion Spectrum Imaging (DSI) [Wedeen et al., 2005] proposed to estimate the dODF using the Fourier relationship between the applied diffusion weighting gradient and the q-space measured signal. Q-ball imaging (QBI) [Tuch, 2004] reformulated DSI for a spherical acquisition scheme, i.e. sampling different encoding directions while keeping a constant b-value. It operated with the Funk-Radon transform (FRT) instead of the full Fourier transform and, although it provided a lower angular resolution than DSI, the reduced amount of acquired data needed and its faster reconstruction made it more popular than DSI. Many other techniques have been proposed since then for an improved characterization of the diffusion propagator or the dODF, such as (3D-SHORE) [Özarslan et al., 2013a], MAP-MRI [Özarslan et al., 2013b], spherical polar Fourier imaging [Cheng et al., 2010] or diffusion propagator imaging [Descoteaux et al., 2009b].

On the other hand, there are **mechanistic models** that incorporate a higher degree of biological description of the tissue microstructural properties. Most of these models rely on the signal compartmentalization assumption, i.e. that the measured dMRI signal can be described as a weighted sum of signals representing water diffusing in different free/hindered/restricted micro-compartments. Thus, these models are also known as a mixture or **multi-compartmental models**. The many methods that have been proposed to date differ with respect to the number and type of compartments they consider (e.g. intra-axonal, extra-axonal, partial volume compartments) and the detail with which they represent each of these compartments; there is no thing such as a generalized standard model but many proposals that focus on different tissue properties or regimes of validity. An example is the Neurite Density Model (NDI) [Jespersen et al., 2007, 2010], which was later simplified to make it feasible in the clinical practice, resulting in the neurite ori-

entation dispersion and density imaging (NODDI) [Zhang et al., 2012]) model. Another example is the **Ball&Sticks** model [Behrens et al., 2003], which we review in the following section. Biophysical models can provide more specific information than DTI about tissue structure, but they should be interpreted carefully (e.g. assumptions and constraints can limit estimates' specificity as well).

If the quantity of interest to estimate is the fibre orientations, the diffusion ODF provides only blurred versions [Descoteaux et al., 2009a] of the true quantity of interest, the fibre ODF (fODF). An approach to estimate the fODF is by representing the measured dMRI signal as the convolution over the sphere of two functions: the fODF, $f(\theta, \phi)$, and a single fibre response function, $R(\theta, \phi)$. The first represents what is the likelihood to see a fibre in a certain orientation $[\theta, \phi]$ in space; the latter describes the signal that would be measured in response to a single fibre.

$$\frac{S_i}{S_0} = FOD \otimes R = \int_0^{2\pi} \int_0^\pi f(\theta, \phi) \cdot R(\theta, \phi) \cdot \sin(\theta) d\theta d\phi \quad (2.25)$$

A number of deconvolution approaches, therefore, exist to estimate the fODF directly. Some of them are closer to model-free methods, as the Constrained Spherical Deconvolution (CSD) [Tournier et al., 2004, 2007], where $R(\theta, \phi)$ is directly estimated from the data (avoiding any model) and uses a spherical harmonics representation for the fODF (the *model assumptions*, if any, rely on deciding the specific fibre orientation from the fODF estimated). Similarly, model-based deconvolution approaches exist based on the multi-compartment model [Dell'Acqua et al., 2007]. In fact, this is a convenient formalization for most of the mechanistic models, as it provides an analytical expression; then, it is the definition and estimation of the parameters of interest $f(\theta, \phi)$ and $R(\theta, \phi)$ what actually differs between them. For instance, in the NDI model the interest lies in the parameters of $R(\theta, \phi)$ (e.g. diffusivities of individual compartments and/or their volume fractions) while in the CSD the aim is to estimate the fODF for tractography purposes, generally.

The Ball&Sticks model

The **Ball&Sticks model** [Behrens et al., 2003] can be seen as a special case of a Multi-tensor model, a generalization of the DTI model, where the Gaussian model for P is replaced with a mixture of $N + 1$ Gaussian densities. This allows the signal in each

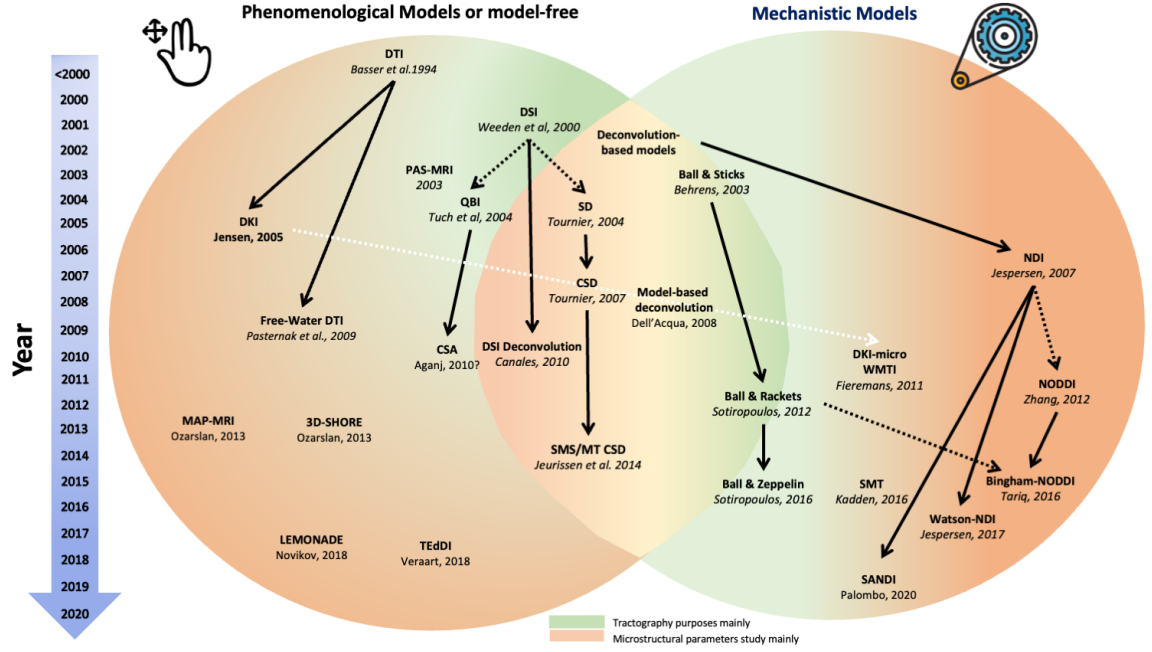


Figure 2.8: Sample example of the microstructural dMRI models proposed, ordered chronologically (from top to bottom). Note that this scheme is orientative as categorization boundaries are not strictly defined in the literature. E.g. the Ball&Sticks model can be seen both as a mechanistic model and as a deconvolution method that inform about microstructural properties and can be used for tractography purposes.

voxel to be represented as originated from N distinct groups of fibre populations (Partial Volume Model), each one modelled by a separate diffusion tensor (D) driven by a Gaussian diffusion. The tensors can either be perfectly anisotropic (the N “sticks”) or a spherical tensor that describes a perfectly isotropic diffusion (the “ball”). The signal S_j measured at each voxel j is modeled by:

$$S_j = S_0 \cdot \left[\left(1 - \sum_{i=1}^N f_i\right) \cdot e^{-b \cdot d} + \sum_{i=1}^N f_i \cdot e^{-b \cdot d \cdot (g \cdot v_i^T)^2} \right], \quad (2.26)$$

where S_0 is the signal with no diffusion gradients applied, b is the b-value, g_i represents the applied diffusion-sensitizing magnetic field gradient (b-vector, a unit direction vector), d is the diffusivity (assumed to be constant in the voxel), v_i is vector describing the i^{th} fibre-stick orientation, f_i is the proportional volume fraction of the i^{th} fibre (where $\sum f_i = 1$).

The model can be also seen as a version of parametric spherical deconvolution. If in eq.2.25 the fODF is assumed to be a sum of delta functions and the response kernel to be the signal obtained from a stick compartment, then by replacing

$$f(\theta, \phi) = \sum_{i=1}^N \delta(\theta_i, \phi_i) \quad (2.27)$$

and

$$R_i(\theta, \phi) = e^{-b \cdot d \cdot (g \cdot v_i^T)^2}, \quad (2.28)$$

as $\int \delta(t) dt = 1$, it reduces to the fibre compartment (stick) sum of eq.2.26. More refined models can be derived from this, like the Ball&Rockets [Sotiropoulos et al., 2012], which models each fibre population as a continuous Bingham distribution of orientations, accounting for within-voxel fibre dispersion.

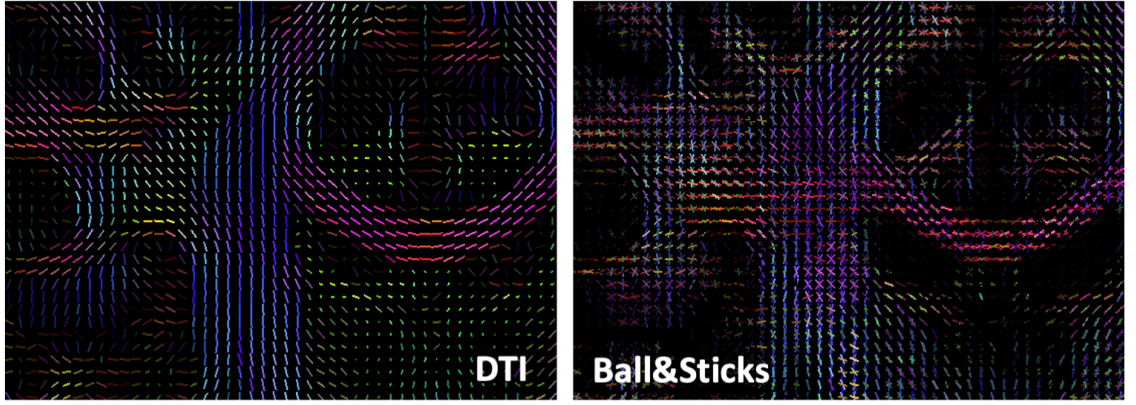


Figure 2.9: Comparison of the fibre orientation estimates obtained by DTI (only 1 fibre direction) and Ball&Sticks (detects crossing-fibres) in an HCP dataset image acquired at 1.5mm.

2.2.5 From fibre orientation to Streamline Tractography

The fibre orientations estimated locally within a voxel can be used to extract long-range white matter bundles. **Tractography** algorithms allow to perform in-vivo *virtual dissection* of white matter bundles [Catani et al., 2002]. By running tractography over the whole brain it is possible to build a **connectome**, a comprehensive map of macroscopic connections as estimated by diffusion MRI [Sotiropoulos and Zalesky, 2019; Sporns et al., 2005; Yeh et al., 2019].

In the simplest tractography approach, it can be assumed that there is a single predominant orientation per voxel. The set of voxel fibre orientations defines a vector field v . Thus, starting from a seed-point p_0 , tractography returns the streamlines r by finding the curves maximally tangent to this vector field. The trajectory of these streamlines can be then

estimated numerically by solving the differential equation:

$$\frac{\partial r(p)}{\partial p} = v(r(p)) \quad (2.29)$$

where $\mathbf{r}(p)$ is the position along the streamline and v represents the fibre orientation. Eq.2.29 can be solved numerically using finite differences, obtaining what is known as the Euler approximation [Conturo et al., 1999]: $r(p_{n+1}) = r(p_n) + h \cdot v(r(p_n))$, being h a small step size (relative to the trajectory). As this is continuous, it assumes there are orientation estimates for every location $r(p)$; however, fibre orientations are voxel-wise. Then, interpolation methodology (e.g., trilinear interpolation, splines, nearest neighbours, etc.) is applied to estimate values at any intermediate arbitrary location. Voxel by voxel, this procedure is followed until some stopping criterion is reached, such as low anisotropy to avoid propagation within CSF or GM regions (e.g., FA drops below 0.2), or high streamline curvature to avoid unrealistic sharp turns (e.g., larger than 45°) [Jeurissen et al., 2019] or the edges of the brain tissue. As a result, the fibre tracking algorithm returns a streamline using only the voxel fibre orientations; the ensemble of pathways reconstructed is known as *tractogram*.

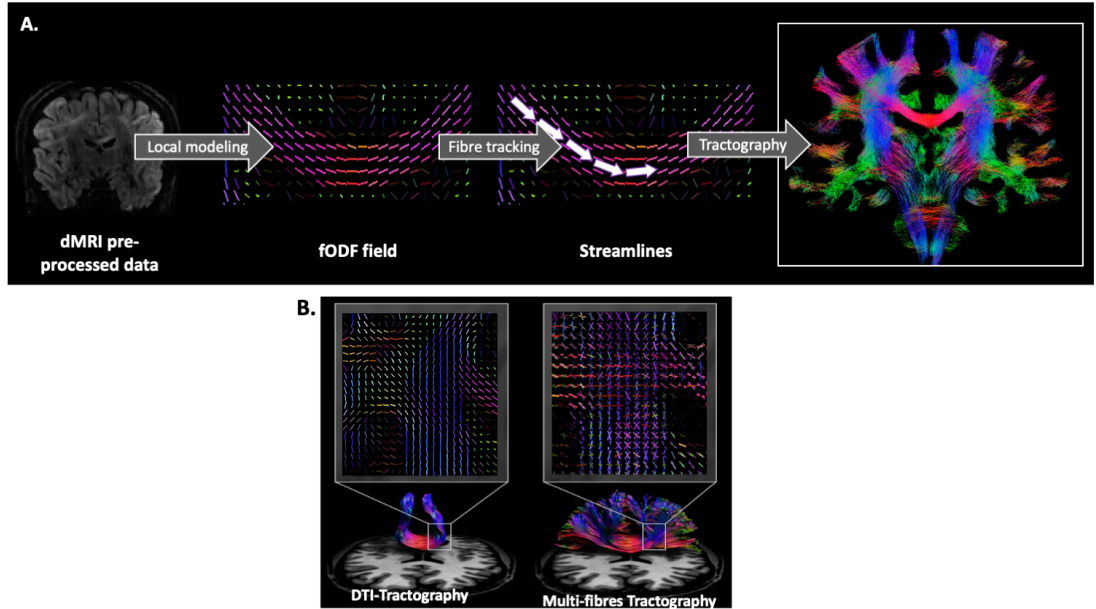


Figure 2.10: A) Streamline tractography propagation. B) Comparison of tractograms obtained in the Corpus Callosum by DTI-tractography and multi-fibre tractography. Adapted from [Jeurissen et al., 2019]

One of the most important aspects of fibre tracking is the model used to infer the local fibre orientation from the dMRI signal. In contrast to the DTI-tractography, by using as

a propagator the estimated fODF from any of the deconvolution approaches seen in the previous section, improvements have been observed in white matter regions with complex fibre architectures (see Fig.2.10.B).

Since the introduction of this technique in the late 90's [Mori et al., 1998], many tracking algorithms have been proposed and nowadays different tractography approaches can be identified. For instance, the general approach explained above is also known as **deterministic streamline tractography**. Streamlines are propagated voxel-by-voxel, independently from each other, and considering discrete orientations. However, these approaches are subject to errors in the local fibre orientation estimates due to imaging noise, artifacts or model inaccuracies. An alternative is to follow a **global** approach, where all tracks are reconstructed simultaneously by finding the configuration that best describes the measured dMRI data in their entirety, according to some optimisation criteria, avoiding local error propagation (e.g., [Christiaens et al., 2015; Jbabdi et al., 2007; Kreher et al., 2008]). Another alternative that will be more relevant for this thesis is **probabilistic streamline tractography**, which allows for incorporating uncertainty in fibre orientation estimation (e.g., [Behrens et al., 2007]), as we will see later.

2.3 How does noise affect dMRI model estimates?

As seen previously, thermal noise modifies the MRI signal in convoluted ways and dMRI is no exception. In fact, as dMRI is inherently a low SNR imaging modality, noise can have a **dual effect** in the model parameter estimates: it can introduce **biases** and increase variance and **uncertainty**.

2.3.1 Noise-floor signal rectification and bias

Whereas the noise in the complex domain is zero-mean Gaussian, the non-central-Chi family of distributions that can result into the magnitude domain by modern protocols are asymmetric, non-negative distributions and, hence, the average noise intensity is increased from 0 to a positive value; i.e., the minimum value or **noise-floor is not 0** anymore [Edelstein et al., 1984]. Although the Rician distribution can be fairly well approximated by a Gaussian distribution for $\text{SNR} > 3$ [Gudbjartsson and Patz, 1995; Salvador et al., 2005], the more the deviation to a non-central-Chi, the higher the noise floor. As a consequence,

this elevated noise floor can interact with the signal when the SNR is low, subsequently affecting the accuracy of model estimates.

As dMRI information is encoded in signal attenuations, acquisition protocols can return signals as low as the noise floor in a) regions where the signal attenuation is sufficiently high (e.g. CSF-filled areas or highly anisotropic regions like in the Corpus Callosum), and/or b) when b-value/spatial resolution is pushed to high limits. In such cases, the signal can hit the noise-floor and experience a non-linear rectification by the noise-floor. In that case, as the signal attenuation observed is lower than it would have been in the absence of noise, it leads to an underestimation (**bias**) of estimated diffusivity in regions with such low-SNR signal.

Previous studies have shown the **effects of noise-floor** in the downstream quantitative analysis and modeling of dMRI data, including tractography. For instance, [Jones and Basser, 2004] reported the following biases in DTI model estimates, especially in anisotropic regions of the WM (peanut-shaped ADC profiles): 1) MD underestimation, 2) overestimation of the FA, 3) a correlation between MD and FA (which are independent in theory), 4) a non-linear relationship of these issues with the b-values and/or lower SNR, and 5) non-gaussian diffusion profiles that wrongly produces multi-compartment estimations. Some of these observations were also previously reported by [Pierpaoli and Basser, 1996]. Dietrich et al. [Dietrich et al., 2001, 2007] made similar observations; in directions with lower SNR, the ADC is underestimated and there exists eigenvalue repulsion that leads to the overestimation of the diffusion anisotropy. [Sotiropoulos et al., 2013c] shows also how the rectification affects the **fibre orientation estimation** in both model-based and model-free approaches, especially in very anisotropic voxels (e.g., mid of Corpus Callosum).

2.3.2 Variance and Uncertainty

An inherent impact of any type of noise is the increase of the signal variability compared to a noise-free scenario. This leads to increased **uncertainty** in any parameter estimate extracted from the signal; the higher the noise, the higher the variance in the signal and the uncertainty around an estimated model parameter.

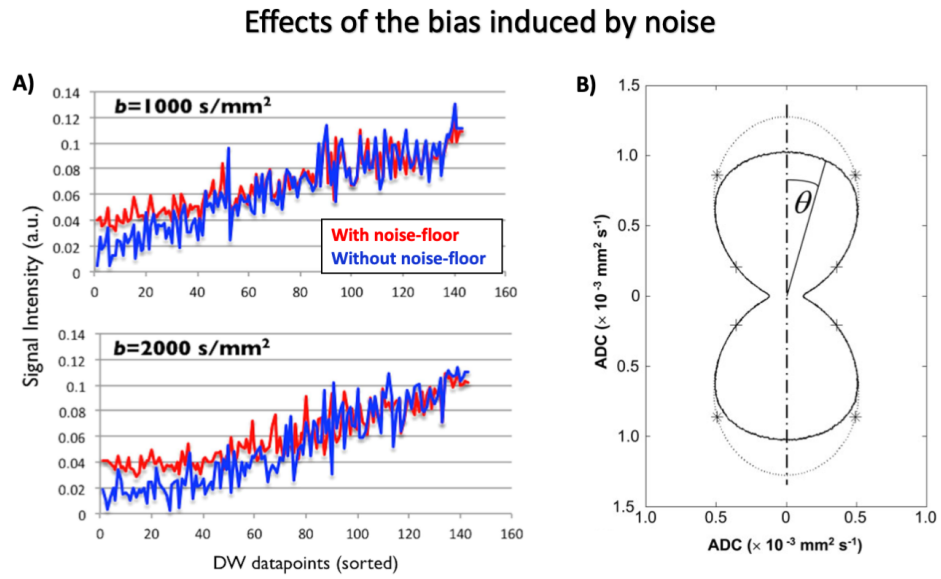


Figure 2.11: Noise-floor bias - A) Noise-floor rectified signal (red) and noise-floor-free signal (blue) for different b -values at a voxel in the midbody of the Corpus Callosum. Data points are sorted according to the angular distance of the respective gradient direction and the principal fiber orientation of the voxel (i.e., from parallel to perpendicular to the major fiber orientation). Adapted from [Sotiropoulos et al., 2013c]. Different types of image reconstruction can mitigate the effects of the noise floor, as shown in the blue line plot. B) Polar representation of the ADC profile, adapted from [Jones and Basser, 2004]. The dotted line corresponds to the noise-free ADC profile, whereas the solid line corresponds to the mean of the noisy estimates of ADC. The asterisks mark the predicted points at which the noisy profile should deviate from the noise-free profile.

A nice example of this lack of precision for DTI model parameters was presented by [Jones, 2003]. They used a statistical approach to map the voxel-wise uncertainty around fiber orientations and visualised them as "cones of uncertainty", which represent the 95% confidence interval for the principal direction of diffusion in the DTI model. Lower SNR is expected to make these cones wider, i.e. to increase the uncertainty. However, it was observed that not only regions with high level of noise presented high uncertainty, but also regions with complex fibre patterns (see Fig.2.12 B). As it will be explained in the next chapter, the uncertainty captured by model estimates does not come only from noise, but there are many other factors that can contribute to it, such as modeling errors or data sampling efficiency.

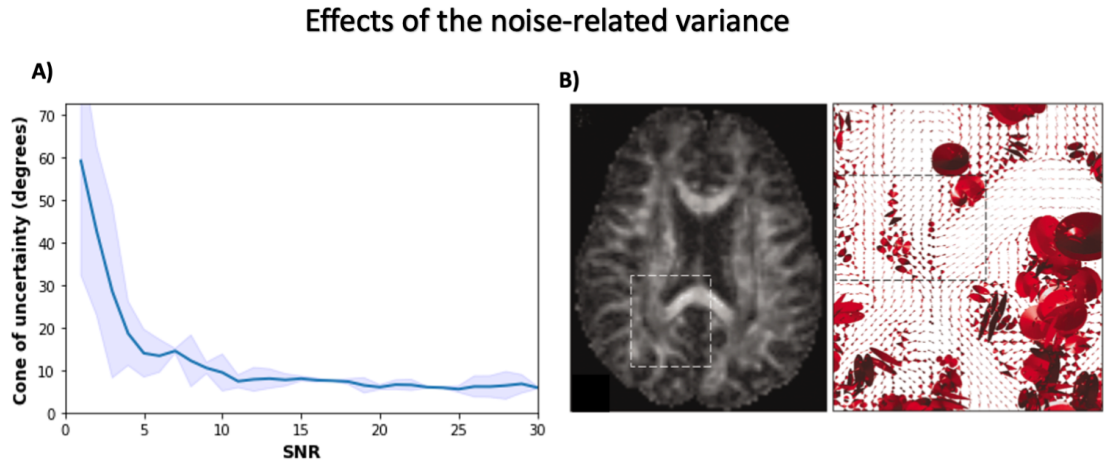


Figure 2.12: Noise-related variance - Examples of how the variance in the signal induced by noise can affect the uncertainty in model parameter estimates. A) The variation and the value of the cones of uncertainty (95% CI) in the estimates of the first fibre orientation reduce as the SNR increases. Results were obtained from 100 synthetic dataset simulated for each SNR value from 1 to 30 (in steps of 1). B) Cones of uncertainty estimated in real brain data. Both noise and model complexity can increase the uncertainty in the estimates. Adapted from [Jones, 2003].

To reduce the noise-related variance of the signal and move to a higher-SNR regime, it is not uncommon to improve SNR by averaging multiple identical acquisitions, also known as *powder average*; the signal measured is assumed to be the same while the random variations of the noise will cancel out (i.e., $\text{sum}(S)/\text{cancel}(\text{std}) = \text{higher SNR}$). However, this implies the acquisition of N repeats at the expense of N -fold increase of scan time. Over the last decade, dMRI denoising approaches have drawn considerable attention aiming to mimic the effect of powder averaging without the cost of increased scan time. A detailed review and a framework for evaluating their impact are provided in Chapter 4.

Chapter 3

Uncertainty Mapping and Inference

In Chapter 2, an overview of diffusion MRI has been presented, along with the main principles behind commonly used biophysical models that map dMRI measurements to microstructural parameters. Furthermore, the effects of thermal noise as an important and inherent source of uncertainty in MRI and, particularly, in dMRI where it can lead to biases, have been summarised. Nevertheless, there are other sources of uncertainty in MRI that can contribute to the lack of precision in dMRI model estimates, such as modeling imperfections and oversimplifications arising from the indirect nature of the techniques used to measure microstructural brain properties.

The first section of this chapter reviews some of these **uncertainty sources**. The second section highlights why **mapping this uncertainty** of estimates can be important and shows some exemplar cases of how this information can be used. Following that, an overview of **approaches used to map uncertainty** in order to infer knowledge about microstructural brain properties while providing confidence estimates is given. Limitations of existing methods and potential for new techniques are discussed.

3.1 Uncertainty in dMRI

Every time there is an estimation process (i.e. fitting a model to some given noisy data), there is some degree of uncertainty associated with the results. Uncertainty is generally decomposed into two sources: *aleatory uncertainty*, which refers to the inherent ran-

domness of natural phenomena, and *epistemic uncertainty*, which refers to the lack of knowledge about the system being modelled and its complexity.

Epistemic uncertainty comprises all the imprecision introduced by approximations and parameter/method selections at every step of the pre-processing and modeling stages, inadequate data to evaluate the model and/or to represent the full space of parameters [Brun et al., 2011; Hüllermeier and Waegeman, 2021]. Hence, aleatory uncertainty usually presents stationary statistical characteristics while epistemic uncertainty, which arises from ignorance (“*unknown unknowns*”), may result in non-stationary residual characteristics that, if it cannot be represented explicitly, might lead to biases or overconfidence in inference. As opposed to uncertainty caused by randomness, uncertainty caused by ignorance can in principle be reduced on the basis of additional information. In other words, epistemic uncertainty generally refers to the reducible part of the (total) uncertainty, whereas aleatory uncertainty refers to the irreducible part. For example, a straightforward approach to reduce epistemic uncertainty is by obtaining more observations (so it reduces the approximation uncertainty; the model and parameters uncertainty will remain). Nevertheless, aleatory and epistemic uncertainty should not be seen as absolute notions. For instance, random thermal noise, which is supposed to be the main source of irreducible aleatory uncertainty, can be actually removed by denoising approaches (see Chapter 4).

This distinction between uncertainty sources is not easy to make in the MRI practice; some specific artefacts are known and can be partially modelled but a characterization and quantification of every source of uncertainty would require expensive experiment designs that make it infeasible in practice. Hence, uncertainty is generally regarded and quantified as a whole in dMRI. While most of the aleatory uncertainty can be attributed to the random thermal noise (reviewed in the previous chapter), in the next section we will review some of the most important sources of epistemic uncertainty in dMRI.

3.1.1 Sources of epistemic uncertainty in dMRI

One of the main sources of epistemic uncertainty in dMRI is the **modeling process**. The same observations of the same underlying microstructure can be represented in very different ways by different models, reflecting different assumptions or different non-parametric strategies (in case of model-free approaches). For instance, there is no clear consensus about which modeling strategy is best: multi-compartment model or model-free approach? An example is shown in Fig.3.1. A number of simulated dispersing fibre patterns are depicted in considerably different ways by different modeling approaches in the absence of thermal noise. Each of these estimates will reflect different levels of “confidence” in the represented pattern compared to the observed ground-truth, even if there is no noise introduced in the system. All differences arise from how the model itself and the modeling assumptions.

This is a known problem in the literature. Modeling assumptions can be also driven by **computational limitations**; models have to be complex enough to link the signal measured to the microstructural properties, but not too complex in order to be fitted in a feasible time. Even different implementations of the same approach can be more adequate for different scenarios (e.g., see [Canales-Rodríguez et al., 2019] for a comparison of spherical deconvolution approaches). Microscopic studies have revealed that microstructure of the brain is really complex and irregular, and representing it by compartments with regular and fixed morphology relies on assumptions that may be partially true or that do not hold for every tissue or region of the brain. For instance, the response function in the CSD model for each tissue and for each b-shell is usually assumed to be the same across the whole brain but pathological situations can break the assumptions, [Christiaens et al., 2020; Veraart et al., 2019]. Furthermore, the modeling process aims to map features in the order of microns (e.g., axon diameter) while observations are in the order of millimetres, requiring certain simplifications in the models [Novikov, 2021].

More refined models have been developed over years, e.g., incorporating partial volume effects in model-free approaches [Jeurissen et al., 2014], considering asymmetries in the fODF [Bastiani et al., 2017], etc. There have been surveys and reviews comparing multiple methods that co-exist (e.g., [Canales-Rodríguez et al., 2019; Davis et al., 2022; Ferizi

et al., 2014; Jelescu et al., 2020; Novikov et al., 2019]). However, results converge to the conclusion that there is not one model that can fit all problems and that uncertainty in the modeling parameters is an inherent property of these systems that needs to be quantified.

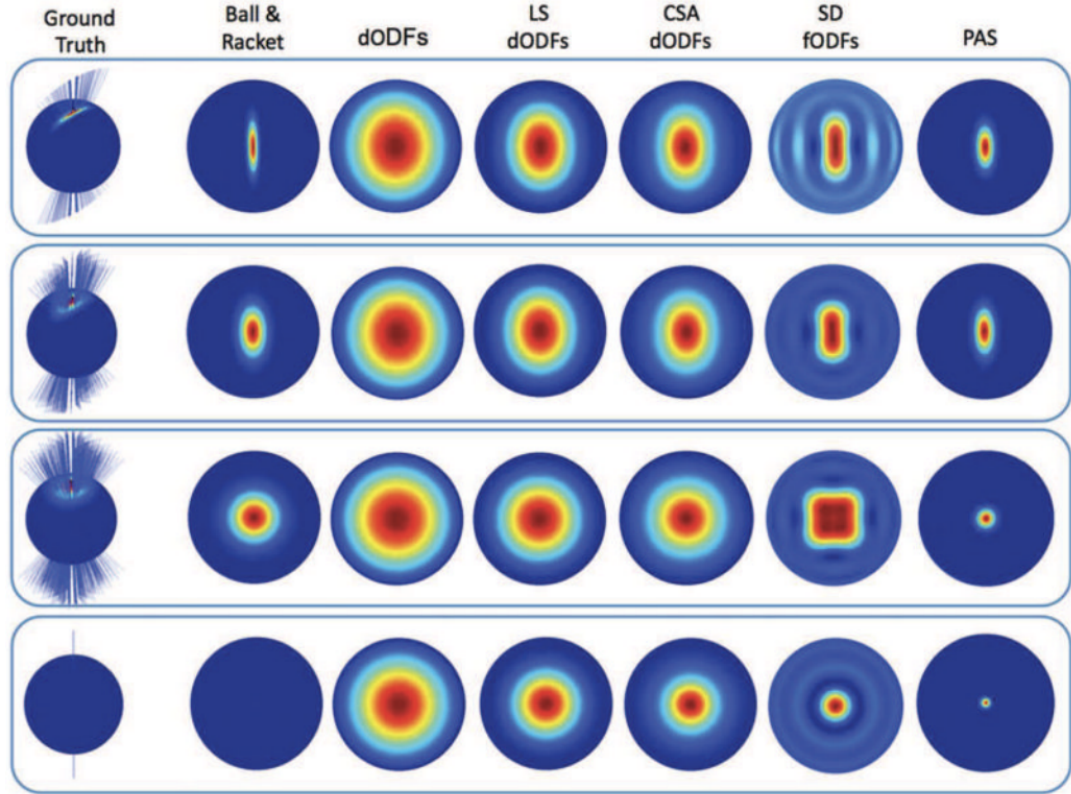


Figure 3.1: Uncertainty arised from the modeling process - Noise-free reconstructions of four fanning patterns using (from left to right): The ball and rackets model, dODFs, Laplacian-sharpened (LS) dODFs, constant solid angle (CSA) ODFs, spherical deconvolution (SD) fODFs and the persistent angular structure (PAS). The ground truth images show collectively all the fiber orientations present in the grid. Image extracted from [Sotiropoulos et al., 2012]

Another potential source of epistemic uncertainty is **data pre-processing**. A number of processing steps are needed to remove systematic artefacts and distortions from dMRI (e.g. MRI geometrical distortions, involuntary subject motion, physiological noise). However, these directly modify the data and it is not clear how they affect statistical noise properties and subsequently modeling assumptions to alter precision in estimates. For instance, spatial normalization can reduce random variance in neighbour voxels (reduces uncertainty) but can also lead to loss of information that will be transmitted to downstream modeling.

3.2 Why mapping uncertainty can be useful in dMRI?

As explained in the previous sections, there are multiple factors that can lead to reduced confidence in dMRI model parameter estimates, from thermal noise effects to modelling errors. Being able to detect and quantify the **reliability** of the results is essential to assess and **interpret** them and, finally, to make informed decisions. What is more, uncertainty quantification can be used as a feedback in order to **improve the modeling** and inference process. For example, uncertainty can be used to compare models with similar accuracy, or to indicate rare classes in classification problems and improve the generalization ability for imbalanced classes. To show relevant applications in dMRI, the following subsections provide two exemplar cases of how mapping the uncertainty can be employed to improve protocol acquisitions and to obtain reliability measurements in the spatial modeling of tracts.

3.2.1 Optimising experimental design

Surrogate errors could be developed from the uncertainty in the parameters to indicate in what regions tissue microstructure estimates are reliable for downstream analysis. Or how data acquisition can be optimised to minimise uncertainty for the parameters of a certain model. [Alexander, 2008] proposed a generic framework towards this direction, finding the acquisition settings that minimize variance in model parameters. They used the Fisher information matrix and the Cramer-Rao Lower Boundary (CRLB) as error metrics for experiment design. The CRLB provides a lower bound on the variance for the unbiased estimators of a fitted model parameter that often correlates closely with the true variance. Applying a Markov-Chain Monte Carlo (MCMC) inference approach to simulated data (see Chapter 5 for a more detailed description of the MCMC), they observed under which acquisition and sampling protocols the model behaved with greatest sensitivity, and when the posterior distribution of the parameters were broader (i.e., higher uncertainty).

Similar in spirit, [Jones et al., 1999] showed in previous work that a uniform coverage q-space sampling scheme minimises variance and also avoids orientational dependence in DTI parameters (i.e., single-shell data). Here, the standard deviation in the estimate of the trace of the diffusion tensor \mathbf{D} served as an indicator of the precision of its estimation. Combining concepts from these works, [Caruyer et al., 2013] extended the method

to design multi-shell acquisitions (i.e. multiple spheres) with uniform angular coverage.

All in all, by using uncertainty quantification, these approaches not only achieved a lower variance in the model estimates (or bounded in what conditions the model is reliable) but they also found an optimal combination of acquisition parameters that provided the same results with a lower number of measurements (shorter acquisitions).

3.2.2 Spatial propagation of uncertainty - Probabilistic tractography

Another very common application that inherently uses uncertainty is probabilistic tractography. Uncertainty of voxel-wise model parameters is used in the reconstruction of global pathways in white matter and propagated spatially, in order to construct a spatial distribution.

Streamline tractography presented in section 2.2.5 does not provide any intrinsic way to assign a confidence measure to a reconstructed path and provides a deterministic or "point" estimate per seed location. However, streamline propagation is very prone to errors caused by experimental noise [Lazar and Alexander, 2003] and potentially by the method used for orientation estimation [De Luca et al., 2021; Jelescu et al., 2015; Schilling et al., 2018]. Therefore some **uncertainty measure** is needed to assess how **reproducible** a particular estimate is given the observations. Ideally, such confidence measure could be obtained by repeating an experiment N times, reconstruct the streamline for each of those experiments and subsequently check the overlap across repeats, as shown in [Jones, 2003]. As this is not feasible in the practice, **probabilistic tractography** was introduced simultaneously by [Behrens et al., 2003] and [Parker et al., 2003] to tackle some of these issues in an indirect way. Probabilistic tractography aims to estimate a spatial distribution of pathways arising from a seed, by utilising uncertainty in voxel-wise fibre orientation estimates; thus, providing a probability distribution on the most probable location of the underlying fibre bundle given the data. To achieve that, in each propagation step of the streamline, a random perturbation is introduced (given the uncertainty in the current fibre orientation). Once N_R streamlines have been generated between a starting seed S and an arbitrary point B , a probability index of connectivity can be obtained. The different probabilistic approaches mainly differ in the way the orientation uncertainty is assessed.

As seen in the previous chapter, (probabilistic) tractography is built upon fiber orientation estimates, which can be expressed in form of continuous density functions, the fODF. This is a biophysical property, i.e. the proportion of fibers leaving region A along a given orientation. However, the **fODF cannot be measured directly**. Instead, our measurements are a function of the dODF. To derive these fiber orientation profiles from it, dMRI models rely on the assumption that diffusion is given along the fibre orientations, without water exchange. If this were true, then the dODF should be identical to the fODF. In reality, there is still significant diffusion along other orientations, not just along the main direction of diffusion, even perpendicular to the axons orientations. As a result, the dODF profile is generally broader than the fODF and estimating the exact fibre orientation from it implies some degree of uncertainty.

This modeling uncertainty, as well as uncertainty due to noise on the fODF shape, are reflected in the voxel-wise **uncertainty ODF (uODF)**. The uODF does not represent a biophysical property of the imaged system, as the dODF or the fODF, but a statement about our confidence in our measurements. Probabilistic tractography uses these local uODFs to estimate streamlines uncertainty at a more global scale. In the next sections, we will see how such local uncertainty can be mapped during the model estimation/inference process.

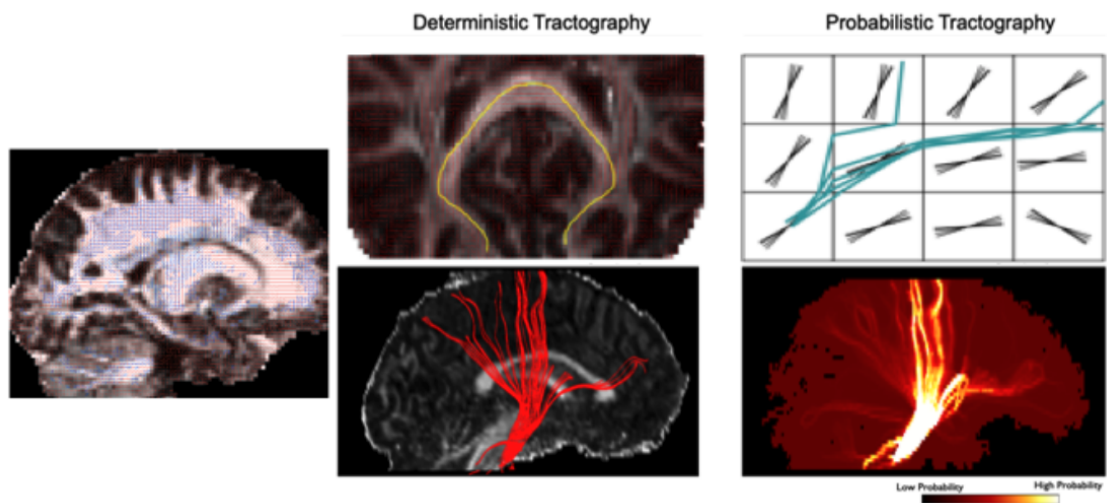


Figure 3.2: Comparison between deterministic and probabilistic tractography. Image extracted from the FSL Course.

3.3 Model Estimation and Uncertainty Mapping Techniques

A forward model comprises of parameters and assumptions that link these parameters to some quantities of interest (typically unobserved directly). A model also represents a measurement process as a function of these parameters, such that for some fixed parameter values, a model prediction/simulation of what the measurements would be can be obtained. Given a set of measurements and a model, the inverse process can be used. A search can be performed of what values of the parameters in the model are most likely to have produced such observations. This process is also known as **inference**, the **inverse problem** or model inversion. Fitting experimental data to an analytical model can be seen as an optimization problem where, e.g. in the simplest case the ω parameter values that minimize a given objective function are found e.g., $\min\{S_{measured} - S_{model}(\omega)\}$, $S_{measured}$ being the observed data and $S_{model}(\omega)$ the model signal prediction.

As part of the inference, uncertainty around the exact values of ω can be mapped. In the following subsections we will review the main approaches that have been applied in dMRI to solve the inverse problem while providing some uncertainty assessment of the parameter estimates. More specifically, we will review the application of **least-squares** fitting, **Bootstrapping** and **Bayesian Inference** into dMRI modeling. We will also briefly overview the idea of **Machine Learning (ML)** to learn the inverse mapping and provide a faster and more flexible alternative approach to traditional optimisation techniques.

3.3.1 Least Squares Estimation (LSE)

A common approach to estimate model parameters is by *least squares fitting*. This attempts to find the set of model parameter values that minimise the sum of squared residuals, between the observations $S_{measured}^i$ and the model predictions $S_{model}^i(\omega)$, i.e. $\sum_i (S_{measured}^i - S_{model}^i(\omega))^2$, across i observations. Depending on the model, this function can be a linear or non-linear function of the parameters ω . Some of the most popular approaches and their performance in dMRI models are described and compared elsewhere (e.g., see [Harms et al., 2017; Veraart et al., 2013]).

For linear models and assuming zero-mean Gaussian noise, LSE fitting reduces to a linear

matrix inversion. For non-linear models, there are generally two main strategies to solve least squares fitting problems: gradient-free and gradient-based algorithms. Depending on the algorithm chosen, the accuracy and quality of the solution will vary [Harms et al., 2017]. Also, due to the existence of multiple local minima, noise can make different set of parameter values equally probable to explain the measured dMRI signal, leading to model degeneracy problems. For instance, at very low SNR, e.g., due to the use of high b-values, all least squares estimators are inherently biased [Veraart et al., 2013]. Given certain statistical assumptions (i.e. of Gaussian noise), LSE approaches can provide error measures and confidence interval of the estimates **analytically**.

3.3.2 Bootstrapping

LSE approaches provide a point-estimate for each model parameter, i.e. a single value and some error boundaries. An alternative is to use approaches that provide distributions of values for each model parameter. High uncertainty will then be reflected in a wide distribution and low uncertainty will result to a narrow distribution for a given parameter. Bootstrapping is amongst the simplest techniques in this group of methods.

Bootstrapping is a statistical approach to provide confidence assessment about the parameter estimates calculated on a sample of observations S . It mimics the scenario of running a measurement multiple times and looking into overlap of results. To do that, bootstrapping methods iteratively create subsets B of the data S , by randomly drawing samples from S and allowing duplicates (*sampling with replacement*).

Its application to dMRI was introduced by [Pajevic and Bassar, 2003] to estimate the uncertainty of DTI model parameters. In their approach, the superset S was formed by a number of n repeated scans of G gradient orientations each. Then, for $i = 1, 2, ..G$, the volume i was randomly picked from one of the n repeats of the superset. That formed a single bootstrap sample. The process is repeated R times (with $R < n^G$) to get R new bootstrap datasets. The DTI model is then fit for each of the R bootstrap samples, providing R estimates for each model parameter, i.e., a distribution of values for each of them. In [Jones and Pierpaoli, 2005], $n = 9$ repeated acquisition of $G = 64$ dMRI volumes were acquired. Then, using that *superset*, $R = 1000$ bootstrapped new datasets were

constructed. This information can be later used in tractography. For instance, [Jones and Pierpaoli, 2005; Lazar and Alexander, 2004] used the distributions of parameters to propagate streamlines from a given seed, giving to the deterministic tractography algorithm applied a new probabilistic fashion.

Classical bootstrapping approaches are non-parametric (i.e. have no explicit assumptions on noise structure), but require multiple repetitions. [O’Gorman and Jones, 2006] used Monte Carlo simulations to determine that, at least, $n = 4$ dMRI datasets were needed in order to obtain accurate and unbiased uncertainty estimates. This expensive and timing limitation led to investigate alternative approaches of bootstrapping, like the **model-residual bootstrap** [Berman et al., 2008; Haroon et al., 2009]. Here, to construct a new bootstrap sample we just need to 1) fit the model to the acquired data, 2) calculate the normalised residuals e_{r_i} for all the predicted values, 3) permute the residuals e_{r_i} randomly with the residuals $e_{r'_i}$ from other subsets R , and 4) add them back to the model predicted values: $\hat{y}_i = y_{r_i} - e_{r_i} + e_{r'_i}$. Then, repeat this process R times to get R bootstrapped samples. The key assumption here is that, even in heteroskedastic models with unknown distribution of noise, the normalised residuals would return identical distributions (same noise variance), so permutations can be done freely among them [Davison and Hinkley, 1997]. A special case of residual bootstrap is the **wild bootstrapping**, which also can operate in non-i.i.d. models with heteroskedasticity [Liu, 1988]. Here, instead of just permuting the residuals, it also multiplies the permuted residuals by a random variable w_i , such as $\hat{y}_i = y_{r_i} - e_{r_i} + w_i \cdot e_{r'_i}$. For instance, $w_i \sim -1, 1$, which simply randomly flips the sign half of the time, forcing the residual distribution to be symmetric, has been applied to DTI model estimates [Whitcher et al., 2008] and in tractography [Jones, 2008]. Results in both residual-based approaches have shown comparable results to the non-parametric bootstrapping, while using a fraction of the data (e.g., in [Jones, 2008] only 1/9 of the data of [Jones and Pierpaoli, 2005] is used).

However, further assessments of bootstrapping approaches have revealed some biases when used in dMRI data [Chung et al., 2006]. Bootstrapping can be dependent on factors such as the sample size of the dataset as well as the complexity of the underlying fibre patterns modeled. E.g., in non-anisotropic regions we can find larger dispersion and a downward

bias for the estimated orientations. Another limitation in modern acquisitions is the non-interchangeability of residuals from different shells, as this can produce biases hard to correct [Sjölund et al., 2018]. Nevertheless, residual bootstrapping is still commonly used to estimate confidence intervals and error quantification for dMRI model estimates (e.g., [Bernstein et al., 2019; Gu et al., 2019; Ning et al., 2021]).

3.3.3 Bayesian inference

A powerful tool to solve inverse problems is **Bayesian inference**. Bayesian inference can be seen as the main representative of probabilistic methods. As with bootstrapping, rather than a point-wise estimate for each model parameter, a distribution of parameter values is obtained. Following Bayes theorem, this distribution reflects the posterior probability density of the model parameters given the data. These posterior densities intrinsically provide a quantification of the estimate uncertainty.

Given a model with parameters ω and an assumed (additive) noise distribution, one can define the likelihood function $\pi(Y|\omega)$, the conditional probability of observations Y given the model parameters ω . Using **Bayes theorem**, the posterior probability $\pi(\omega|Y)$ of particular values of the model parameters given the data is:

$$\pi(\omega|Y) = \frac{\pi(\omega) \cdot \pi(Y|\omega)}{\int_{\omega} \pi(\omega) \cdot \pi(Y|\omega) \cdot d\omega} \quad (3.1)$$

The **posterior** is a conditional distribution from where we want to estimate the values of the parameters of interest ω , given some data Y . This posterior can be seen as an equivalent to uODFs when the parameters of interest are defined in spherical space (such as a fiber orientations). Bayes theorem provides a simple rule to relate this conditional distribution from the **likelihood** function of the data, $\pi(Y|\omega)$ (i.e., the model). Furthermore, Bayesian inference is considered a subjective method because it allows us to introduce some prior knowledge about the parameters, $\pi(\omega)$, to help us calculate the posterior distribution.

Once the signal, noise models and prior assumptions have been chosen, there are a number of strategies to solve Bayes equation [MacKay et al., 2003], such as analytical solutions, mathematical optimizations, heuristic approximations, and simulation approaches, among

others. For instance, [Sotiropoulos et al., 2010] proposed an analytical solution to estimate the orientational uncertainty of the dODF and use it to get samples from it, as probabilistic tractography does not require a complete representation of the uODF but only samples from it. However, analytical solutions only work for certain noise models and are in general uncommon in real world problems. In the general case, the calculation of the denominator in Equation 3.1, which is known as the **normalising constant** or evidence (the actual probability of the data given the model chosen), is most of the times not possible to solve analytically (high-dimensional integral with dimensions as the number of model parameters), and **approximation methods** are needed for these cases ¹.

Most of the Bayesian frameworks applied for dMRI applications have been based on approximations, simulations and sampling approaches. The most common approach is **Markov Chain Monte Carlo (MCMC)** [Gilks et al., 1995], used e.g. in [Behrens et al., 2003, 2007; Kaden and Kruggel, 2012; Sotiropoulos et al., 2013a]). MCMC is an iterative sampling approach that targets the posterior up to a proportionality constant, as it avoids the computation of the normalising constant. It provides in theory guaranteed convergence to the true sample distribution, although in practice it is challenging to obtain metrics that inform about how close we are to that true distribution (more details about the MCMC method in Chapter 5).

Popular alternatives to MCMC and sampling methods are **Variational Bayes (VB)** [Blei et al., 2017] and **Approximate Bayesian Computation (ABC)** [Sisson et al., 2018]. In VB, rather than iterative sampling as in MCMC, a global approximation to the posterior is done under the constraints of an analytical functional form (typically using Normal distributions) [Chappell et al., 2009; Kaden et al., 2008]. Although it is considerably faster than MCMC, the posterior distributions provided are less specific. In ABC, the likelihood function is approximated by simulations (*synthetic likelihood*). In this way, it bypasses the explicit evaluation of the likelihood, which is often expensive or not available. ABC suffers from the *curse of dimensionality* (with more than a few parameters), but the ideas

¹Often *approximation methods* refer to methods which are *not* accessing the posterior distribution but instead an approximation of it, e.g. Variational Bayes. In this thesis, the term *approximation methods* will be used elsewhere with a more general perspective to refer also to methods that, rather than providing an analytic expression of the posterior density, produce samples from it and do so with the need of knowing/calculating the evidence (e.g., MCMC), or they learn a surrogate model that aims to produce samples from the (approximated) exact posterior (e.g. SBI).

behind it will pave the basis for the Simulation-based inference (SBI) [[Cranmer et al., 2020](#)] approach introduced in this thesis, a promising strategy of combining simulations, density estimators and neural networks to solve the inverse problem by Bayesian principles (more details about the ABC and SBI in 6).

In summary, Bayesian approaches offer a natural framework where prior knowledge can be introduced to constrain the problem into plausible solutions. Furthermore, Bayesian modeling has shown to be especially useful when data is limited, as it helps avoiding overfitting and can model uncertainty on parameters estimates. However, and although great speed-ups have been achieved (e.g., adapting the methods to GPUs [[Hernandez-Fernandez et al., 2019](#)], Bayesian techniques can be very computationally demanding. In that regards, Machine Learning approaches open a promising route to learn the inverse problem and avoid the repetition of iterative likelihood calculations, like in MCMC.

3.3.4 Machine Learning for inference in dMRI

Machine Learning has provided many effective tools to capture hidden patterns in a data-driven manner and to solve highly non-linear optimisation problems in complex scenarios where no model is available, relying on few assumptions. Indeed, the number of papers published in dMRI using ML is already comparable to the number of papers based on mathematical modelling, and since 2010 their numbers increased much faster [[Ravi et al., 2019](#)]. The range of applications in dMRI is diverse: from image reconstruction (see [[Knoll et al., 2020](#)] for a short review) and pre-processing (e.g., [[Muckley et al., 2021](#)]), to classification between healthy subjects and different pathologies [[Schnyer et al., 2017](#)] or tractography (see [[Poulin et al., 2019](#)] for a review).

Of particular interest for this thesis is to learn how to directly map the (pre-processed) diffusion signal measured to the diffusion parameters of interest. Because ML approaches are mainly model-free, ML methods have the potential to reduce sources of error by avoiding to rely on sub-optimal mathematical models of the diffusion signal (e.g., see sec 3.1.1) and increase efficiency by learning how to incorporate both local and non-local information during the learning process. Yet, its application to solve the inverse problem in biophysical MRI models has been explored in very few studies so far. One of the first of these stud-

ies, done by [Nedjati-Gilani et al., 2017], trained a **Random Forest (RF)** regressor to predict tissue microstructure from the Karger model using rotationally invariant features computed from diffusion signals. Similar approaches are presented in [Fick et al., 2017] to map axon diameter with an RF trained using histological data, in [Alexander et al., 2017] to map diffusion tensor to microstructure parameters from NODDI and SMT models, and more recently in [Hill et al., 2021] for an axonal permeability model applied to mouse and validated with histology. However, despite being powerful for fast biophysical model fitting, RF regressors have shown a limitation in their poor generalizability outside the training set.

Artificial Neural Networks (ANN) have been also explored to estimate **scalar diffusion parameters**. This line of work was initiated by [Golkov et al., 2016], where authors proposed a method called q-space deep learning (q-DL) that allows mapping parameters, such as diffusion kurtosis or orientation dispersion, directly from the q-space signals (i.e., without explicit diffusion models). Q-DL have been refined by Ye et al. in recent works by introducing the sparsity of diffusion signals into the deep network design and by taking patches (instead of single voxels) to further incorporate the information in the spatial domain (MEDN+, [Ye, 2017]), and by allowing adaptive incorporation of historical information with modified long short-term memory (LSTM) units (MESC-Net, [Ye et al., 2019]). In addition, unlike in previous works where the estimation is designed specifically for a model, the network in [Ye et al., 2019] is not limited to a particular signal model. A parallel line of research has focused on the **estimation of the fODF** and the number of compartment models directly from the signal, without imposing any mathematical model or physical assumption [Lin et al., 2019; Nath et al., 2019]. For instance, in [Koppers and Merhof, 2016; Koppers et al., 2017] CNNs are exploited to reconstruct fibre orientations. Here, instead of solving the inverse problem, the network is trained to solve a classification problem in each voxel in order to obtain the correct orientation. [Karimi et al., 2021] is one of the most recently proposed methods that outperformed both traditional and previous ML approaches.

Despite their potential, the flexibility shown by these ML models may lead as well to the need of tuning a large number of **hyper-parameters** (i.e. parameters not of biophysical

interest, but necessary to make the framework work) and to the risks that sub-optimal choices may affect the results. For instance, the authors in [Gyori et al., 2022] tested the bias and variance in the estimates from a traditional non-linear optimization (via non-linear least squares) and common implementations of a RF and an ANN. Particularly, they assessed the effect of the training distributions in these supervised ML approaches and demonstrated that: 1) parameter estimates obtained from traditional model fitting are overall more accurate and less-biased than the estimates obtained from the ML models, 2) the more restricted the distribution of training samples, the higher the bias, i.e., better to train with uniform distributions of plausible parameter value, and 3) at low SNR, traditional fitting suffers from high uncertainty, while ML models can suffer from "overconfidence".

Overall, these works have shown the possibility to obtain similar estimates using a fraction of the measurements typically needed in traditional fitting approaches, what makes it interesting for clinical applications [Aliotta et al., 2019; Gibbons et al., 2019; Golkov et al., 2016; Karimi et al., 2021; Lin et al., 2019]. Although promising, more work is needed to address challenges such as biases and uncertainty mapping. In this thesis we present contributions towards this direction, by proposing and testing an ANN-based framework for Bayesian inference in Chapter 6.

3.3.5 Summary

In this chapter, different approaches to map uncertainty have been reviewed but *how do they compare with each other?*

The **analytical approach** is a deterministic process that provides a point-estimate for each parameter and relies on assumptions about the model and the residuals that are rarely met (e.g., normal distribution of the residuals, non-co-linearity between parameters, etc.). Violation of these assumptions can lead to biases or inaccuracies. For instance, it has been observed that the analytical method generally underestimates the uncertainties of statistics of interest compared to bootstrapping [Wu et al., 2018].

Bootstrapping is a resampling method that provides a distribution of the statistics of interest or model parameters. This sampling distribution is then used to map uncertainty and confidence intervals. Classical bootstrapping (but not model-based bootstrapping) does not involve any assumptions about the data (e.g., prior distributions on model parameters or noise, etc.), so it is less restricted than analytical approaches. Furthermore, it has showed to be sensitive to sources of variability that could not be included parametrically in the models, like physiological noise or motion [Jones and Pierpaoli, 2005]. On the other hand, residual-based bootstrapping are an efficient alternative that have shown similar results in neuroimaging model-fitting problems.

Bayesian approaches have been widely adopted in the neuroimaging community because their flexibility in modeling and noise assumptions (any noise model will work without any change in the framework), and incorporation of prior knowledge. They not only provide uncertainty quantification, but a principled framework to solve the inverse problem where the probabilities can be directly interpreted as degree of confidence (e.g., the more “peaked” this distribution, the less uncertainty), while frequentist approaches define a relative frequency (that can have multiple influence factors). This can be actually reformulated as the debate between two types of uncertainty: Epistemic uncertainty is the subjective Bayesian interpretation, the kind of uncertainty that can be reduced by learning. Aleatory uncertainty is the kind of uncertainty you accept and work around in frequentist approaches. This perspective highlights the interaction between prior knowledge and data. The possible outcomes, and both aleatory and epistemic uncertainty, not only depend on the data but also on the way in which prior knowledge (the model and hypothesis space) and data interact with each other. In that regards, Bayesian frameworks are advantageous as they allow for the introduction of inductive biases (prior knowledge) that can help to reduce the uncertainty. E.g., the more restrictive the model assumptions are, the smaller the uncertainty will be, and the less data is needed. Nevertheless, most of the Bayesian algorithms that can be applied to real-world problems are highly iterative and demand elevated computational resources. Therefore, bootstrapping still provides a much faster and simpler alternative for uncertainty quantification. To address this challenge, in Chapter 5 we propose an optimal design to improve the computational performance of the classical Metropolis-Hastings MCMC.

In **Machine Learning**, since these techniques are associational and applied to observational data, they are prone to picking up spurious influences from data, making their predictions potentially unreliable. Furthermore, except [Ye et al., 2020], all the ML methods presented in Sec.3.3.4 return point-estimate predictions. [Tanno et al., 2021] proposed a method to characterise different sources of uncertainty in DL methods in the context of medical image enhancement applications. There, the predictive uncertainty in modeling can be described as the combination of the complexity of the task, referred as *intrinsic uncertainty*, and the suitability of the model used to describe the data, or *model uncertainty* (which at the same time can arise from the parameter uncertainty and the model bias). Although the framework proposed in [Tanno et al., 2021] is not directly applied for mapping biophysical dMRI models, it shows how the uncertainty estimates can improve predictions (e.g., by regularising decision based on uncertainty info) and the experimental design (e.g. if the parameter uncertainty is high but intrinsic uncertainty is low, this indicates that collecting more training data would be beneficial).

Of particular interest for this thesis is the re-emergence of Simulation-based Inference (SBI) in recent years. The introduction of ML methods into these frameworks has helped to overcome computational limitations of Bayesian Inference [Cranmer et al., 2020]. Here, simulations can be used to train an unsupervised approach (e.g. NNs) that learns an approximation to the exact posterior distribution without the necessity of iterative model inversion (inference amortisation). This offers a promising method where high-dimensional problems can be addressed while obtaining uncertainty quantification of the estimates. In Chapter 6, we build upon these ideas and present a novel framework to learn the mapping between model forward predictions and the posterior probability of model parameters given the data, and test this approach against classical MCMC in the context of dMRI biophysical modeling. To the best of our knowledge, only [Jallais et al., 2022] have recently explored this approach to solve microstructural dMRI models and results obtained in this thesis are complementary to theirs.

Chapter 4

EDDEN: A framework for Evaluating Diffusion MRI DENoising approaches

4.1 Introduction

In the previous chapters, we reviewed different sources of uncertainty and their potential implications in the quantitative modelling of dMRI signals. One of the main sources of uncertainty in dMRI is originated from thermal noise, which is unavoidable in MRI and has a dual effect: 1) increases the variance of the signal producing a reduced precision, and 2) can modify signal properties and lead to biases and lack of accuracy in imaging-derived measures. To overcome this challenge, a number of approaches have been developed to **reduce the effects of thermal noise** (even being considered as a source of aleatory uncertainty). This is further motivated by current research directions of the field, like the need for pushing the boundaries in in-vivo (e.g. the HCP [Moeller et al., 2021b]) and post-mortem acquisitions [Roebroek et al., 2019b]; the convenience for reducing scanning times, of particular interest in the clinical daily routine; or the interest in higher b-shells for higher angular contrasts.

A field of particular interest in this context is dMRI **denoising**. Denoising can be defined as any signal processing method that extracts signal from a mixture of signal and noise, thus preserving the useful information and, consequently, **increasing the SNR**. As shown

before (Fig.2.3), there are multiple steps where **thermal noise** is introduced or propagated and, therefore, where improvements can be introduced to avoid or reduce it. On the other hand, the growing landscape of DL-based algorithms used to denoise natural images is starting to be transferred into medical images as well [Lundervold and Lundervold, 2019; Tamada, 2020; Tian et al., 2020]. All the above combined has led to a considerable rise in the number of works published about denoising methods in dMRI in recent years (see Fig.4.1).

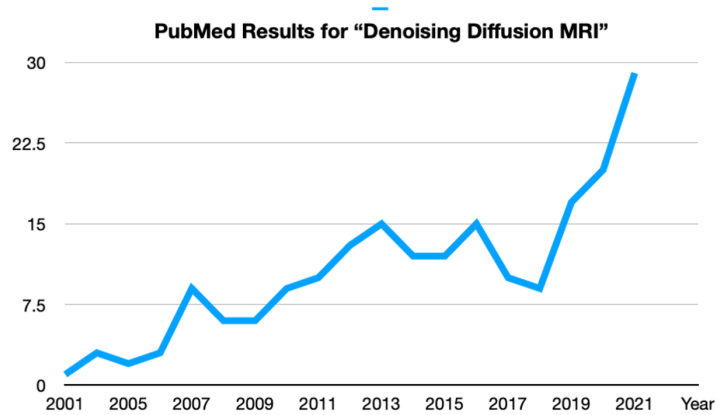


Figure 4.1: Growth in the number of publications per year in Pubmed containing "Denoising Diffusion MRI" in the title.

However, **despite progress and interest in this area, gaps in our knowledge still exist**. The modifications introduced in the dMRI measurements by the denoising process are still not fully characterised in an objective manner. As a consequence, open questions remain such as *how should one denoise dMRI data? Are there any undesired effects caused by denoising? Are the theoretical properties of the dMRI signal preserved? What dMRI denoising method to use and why?*. Even the fundamental aspect of whether **"to denoise or not"** is still debatable, as arguably signal modelling already performs some type of *denoising* as noise is an implicit part of any model.

There have been efforts in developing frameworks to evaluate denoising and answer similar questions on multiple fields, such as for natural images [Alkinani and El-Sakka, 2017; Fan et al., 2019; Goyal et al., 2020; Ilesanmi and Ilesanmi, 2021], medical images [Mohd Sagheer and George, 2020], and even anatomical MRI [Annavaarapu and Borra, 2020; Heo et al., 2020; Mishro et al., 2021] and functional MRI [Huang et al., 2021; Kay, 2022]. However,

despite the explosive growth of denoising approaches for dMRI, there are **no frameworks specific to dMRI denoising**. Two factors seem to be limiting progress in that direction: **the lack of datasets and the lack of objective and comprehensive evaluation criteria**.

Simulating brain data, especially by using the Brainweb framework [Collins et al., 1998], has been a standard approach for evaluating denoising approaches in MRI. However, even having as many parameters to fine-tune as this framework have, the simulation approaches available nowadays cannot reproduce real brain complexity, inhomogeneities and acquisition interactions, especially for very high-resolution regimes. Thus, results obtained in simulated datasets cannot be fully extrapolated to real cases; they need further validation and assessment of their generalization in real acquired data. Furthermore, the models available in the BrainWeb framework provide only structural MRI but not diffusion MRI, so the impact of denoising in the dMRI signal cannot be directly assessed. Therefore, both a dMRI simulation framework and a bespoke real brain data/clinical set are still needed to evaluate denoising.

For the **evaluation** of denoising, many of the studies reviewed validated results in MRI denoising using some type of qualitative assessment (e.g., visual inspection of the image and the residuals). While this may be sufficient for natural images, where the visualization of more pleasing or visually cleaner images can be the ultimate goal, it can be problematic for medical imaging, especially for modalities where modelling and post-processing are needed to extract information of interest, rather than visual inspection. Quantifiable Image Quality Assessment (IQA) can be employed to provide metrics about both denoising effectiveness (e.g., SNR, PSNR, CNR, MSE, Coefficient of Variation, etc.) and structural details preservation (e.g., SSIM [Chow and Paramesran, 2016; Wang et al., 2004]). Most of these IQA's are generally designed from classical image processing and most of them require the availability of the true noise-free image, which is not available in real brain dMRI. Still, they are inherited and extensively used when validating and comparing MRI denoising methods, especially in anatomical images and tasks such as segmentation or cortical surface extraction. However, the interest in dMRI does not only rely on image features such as edges and other sharp structures, typical aspects evaluated by most of

these IQA's, but more importantly in the information contained in the different points of the q-space that will indirectly inform about the microstructural properties of the brain. Finally, as we have overviewed in the previous chapters, noise can cause challenges with precision/uncertainty, but also introduce challenges with bias/accuracy, due to the noise-floor and the rectification effects on the dMRI signal. However, existing evaluation criteria have mostly focused on the reduction of variance after denoising, leaving unassessed the dual effect of noise on dMRI.

For all these reasons, we have developed and presented a novel framework for evaluating **Diffusion MRI DENoising approaches (EDDEN)**. EDDEN comprises a set of considerations and criteria for performance evaluation, as well as bespoke datasets aimed to sample different SNR regimes and complementary dMRI signal properties. The new framework is aimed to address some of the limitations commented on above by providing more insights into the nature of denoising and how its effects can be formally characterised for the particular case of diffusion MRI. The framework is applied in a representative set of existing denoising approaches, highlighting aspects of their behaviour, formally characterising their performance, and comparing alternatives.

The following section overviews theory on current post-reconstruction denoising methods for (d)MRI thermal noise ¹, focusing on the ones that will be used in this chapter. Section 4.3 provides the definition of what a good denoising algorithm should provide and, according to this, a detailed description of the criteria proposed in EDDEN, with methodology on how they are tested. Results follow in section 4.4, where we compare some of the most common denoising strategies using this framework. The chapter ends with a discussion about the results obtained, contributions and limitations, and the potential future steps of this line of research in section 4.5.

4.2 Theory

Denoising approaches can be broadly grouped based on whether they are applied **during acquisition or early reconstruction (k-space)** versus those applied at a **post-**

¹Denoising methods addressed in this thesis refer to thermal-noise-related components of the dMRI measured signal, i.e. we do not address other sources of artefacts or variability such as motion, susceptibility distortions and Eddy currents, physiological noise, etc.

reconstruction stage (imaging space). The former include approaches such as modified acquisition sequence and reconstruction protocols and are particularly useful to avoid or reduce (rather than removing) the magnitude of the noise or other types of artefacts, such as eddy currents or motion. Deep Learning based image reconstruction methods have gained mode interest in recent years within this category (e.g., see [Pal and Rathi, 2022; Zeng et al., 2021] for recent reviews). These are designed for learning an optimal subsampling of the k-space that allows for the accelerated acquisition and considerable reduction of noise introduced in the images.

On the other hand, **post-reconstruction denoising** methods aim to remove undesired noise contained in the reconstructed image. Early work focused on applying filters on model estimates, but results showed that these produce more distorted images than operating on the dMRI signals directly [Basu et al., 2006; Parker et al., 2000; Zhang et al., 2017]. Initial proposals in dMRI, such as fundamental filters, were directly inherited and adapted from traditional digital signal processing. Nowadays, more domain-specific alternatives try to exploit different known properties of MR images, such as information sparseness and randomness of noise. Another important property is the spatial self-similarity, i.e. images display similarity in neighbour voxels (local similarity), but also in non-adjacent regions (non-local similarity). This redundancy of information is even more evident in dMRI as each acquisition is composed by dozens or even hundreds of volumes representing the same brain (with a different diffusion-sensitising gradient). Therefore, the majority of algorithms try to exploit the spatial self-similarity by performing some type of neighbourhood filtering or **performing denoising in spatial patches** (see Fig.4.2) rather than voxel-wise [Alkinani and El-Sakka, 2017].

Fig.4.3 presents the range of post-reconstruction denoising methods in diffusion MRI, based on the image statistical properties exploited by each method. Most of the techniques can be further expanded to adaptive, local/non-local, capability/variations to address different types of noise, or even to be combined with methods that exploit another signal property, as some of the assumptions are not mutually exclusive. We offer here a short overview of a representative subset of methods, focusing more on approaches that will be used in this chapter.

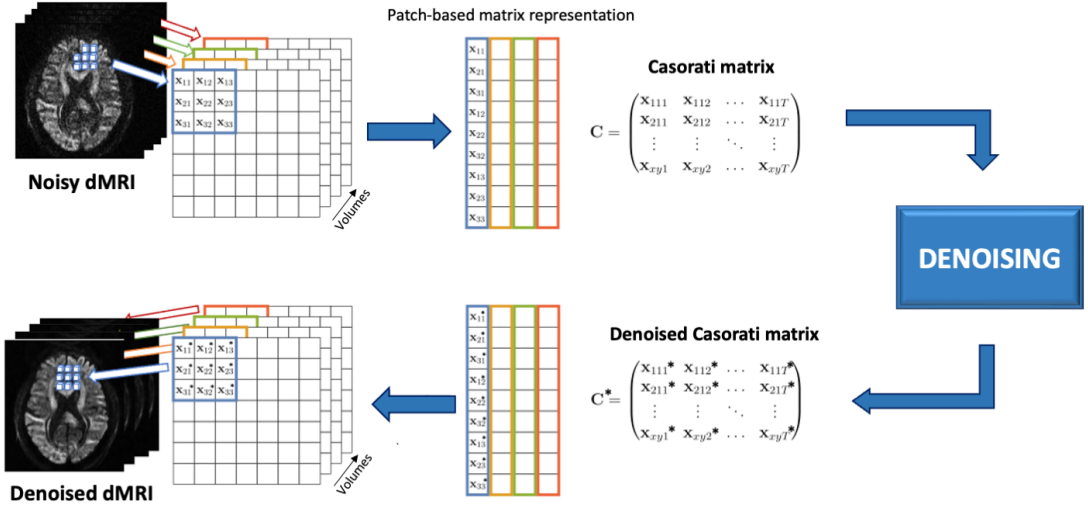


Figure 4.2: Spatial patch-based denoising - In order to exploit the self-similarity property of MRI images, most of denoising methods in dMRI apply the denoising algorithm to local spatial patches. The dMRI image can be iteratively filtered by unfolding 3D patches from the noisy data into a matrix C , denominated as the Casorati matrix by some methods. After the denoising algorithm has operated and filtered this matrix, it can be reshaped again into 3D to reconstruct the denoised image.

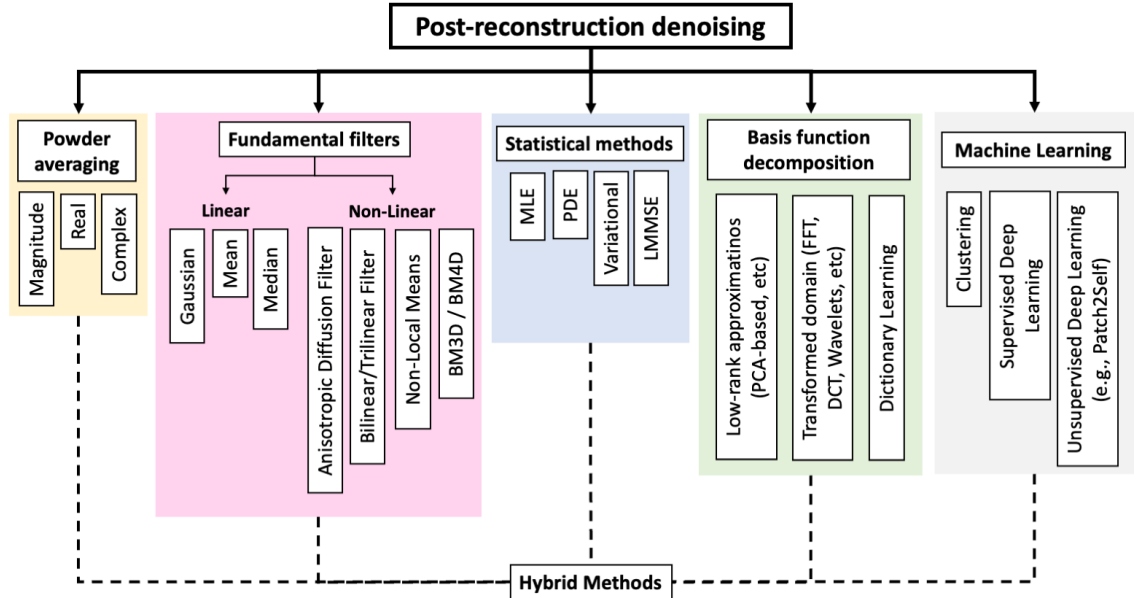


Figure 4.3: Post-reconstruction denoising methods - Classification of denoising approaches adapted and extended from [Mishro et al., 2021] to the particular case of dMRI.

4.2.1 Powder Averaging

The simplest way of denoising is to increase the SNR of the image by some type of **averaging** multiple images acquired with identical image parameters, also known as *powder average*. In principle, this approach allows us to average out the variance associated with random fluctuations originated by thermal noise without manipulating the data. This makes the multiple-averages to be generally considered as the **gold standard**. It can be demonstrated that averaging N signals increases the SNR by \sqrt{N} times [Eichner et al., 2015]:

$$SNR = \frac{S}{\sigma_{noise}} \xrightarrow{N \text{ scans}} \frac{N \cdot S}{\sqrt{N} \cdot \sigma_{noise}} = \sqrt{N} \cdot SNR \quad (4.1)$$

However, averaging has a number of drawbacks. The most evident and limitant is the requirement of multiple acquisitions, which has motivated the research in alternative post-acquisition denoising methods for single images. There is also an inherent resolution blurring when averaging, increased by misalignments and subject motion produced during the registration between sessions.

4.2.2 Non-Local Means (NLM) Denoising

The Non-Local Means (NLM) [Buades et al., 2005] is probably the most used method of the Non-linear Fundamental Filters. These have constituted the traditional approach in pre-processing natural images to either remove noise or select specific features. Such filtering can be applied to any domain (temporal, spatial, frequency, etc.) and is generally obtained as the convolution of a weighted kernel with the patches of the image. Unlike linear models (e.g. mean or median filters), spatially varying noise is assumed in NLM so non-linear weighted kernels are applied here.

More specifically, NLM is built upon two basic principles: 1) The self-similarity usually present in MR images can be used to perform a clever way of averaging out the thermal noise, and 2) this can be done by looking for similar regions all over the image (i.e., *non-local*) instead of averaging all (similar and non-similar) local information. As noise goes down proportionally to the number of averages, localising similar patches provides multiple regions, so more averaging and more denoising, although it may induce also more spatial blurring.

In NLM, the resulting denoised value of the central voxel v_i of a given patch is the weighted average of all the voxels v_j in the image, where the weights depend on the similarity between different patches i and j . The more similar, the larger the weight; these weights can be assigned based on different kernel functions $w(i, j)$, e.g. a Gaussian kernel.

$$v_i = \frac{1}{C_i} \sum_j v_j \cdot w_{i,j} \quad (4.2)$$

While **reducing the signal variance, spatial smoothing is induced** in the image by these filters. Consequently, blurring appears, and edges and fine details of the image can be degraded. Although non-linear filters usually overcome most of the smoothing and no-edge-preservation issues of linear filters, the level of blurring and denoising will still depend on the parametrisation of the filters and the patch size.

[[Manjón et al., 2008](#)] popularized the use of NLM in MRI applications and [[Wiest-Daesslé et al., 2007](#)] adapted it to dMRI. To date, NLM is one of the denoising approaches with more flavours and extensions, e.g. to Rician models [[Wiest-Daesslé et al., 2008](#)], adaptive approaches [[Manjón et al., 2010](#)], and combined with advanced methods such as Singular Value Decomposition (SVD) [[Wu et al., 2019](#)], in the joint k-q space [[Chen et al., 2019](#)], or with Deep Learning [[Manjón and Coupe, 2021](#)].

4.2.3 Marchenko-Pastur Principal Component Analysis (MPPCA) Denoising

It is commonly accepted that thermal noise follows random patterns and cannot be compressed, while noise-free MR images carry redundant information both in the spatial domain (local and non-local self-similarities), and the angular domain in the case of dMRI (correlation between gradient directions) [[Veraart et al., 2016a](#)]. Therefore, there might be a latent space where the information of interest could be summarised by sparse components into a lower dimensionality, while the rest of the components can be assumed as noise and discarded.

The sparsity across each dimension can be modelled as a **low-rank approximation (LRA)** problem, i.e. the approximation of a given multidimensional array or tensor of

dimensionality N , by a sparsely represented tensor \mathcal{A}_k of order k , such that $k < N$. Formally, let $\mathcal{A} \in \mathbb{R}^{L_1 \times \dots \times L_n \times \dots \times L_N}$ be the noisy tensor of order N . If \mathcal{A} contains redundant information, its covariance matrix Σ is rank-deficient, i.e. the rank k is smaller than the dimensionality of \mathcal{A} , while the covariance matrix of the noise Σ_{noise} is full-rank (it is random, not redundant). To obtain the best rank- k approximation \mathcal{A}_k , **Principal Component Analysis (PCA)** is one of the most common techniques applied to decompose this redundancy for matrices ($N = 2$) [Hotelling, 1933; Muresan and Parks, 2003], or the multiway-SVD and higher-order tensor decompositions for $N > 2$ (e.g., [Zare et al., 2018]).

In dMRI, 4D data have been traditionally reformatted as matrices C (also known as Casorati matrix) of $N^3 \times M$ (columns x rows) by taking overlapping local patches of $N = n \times n \times n$ voxels with M representing the number of diffusion-sensitising volumes (b-values x directions)(see Fig.4.2). By standardising the data and assuming they are corrupted by white additive Gaussian noise, PCA can be applied to the variance-covariance matrix to perform the LRA, i.e. to separate correlated components (redundant) into linearly M uncorrelated orthogonal eigenvectors (principal components, PCs), and their associated singular values λ_k (weights or scores). Sorting by descending order of λ_k , the first PCs would describe the largest variance. The informative signal decay in the dMRI signal should be captured by PCs with larger eigenvalues, because the variation in these signals tends to be correlated across voxels and have large variation (corresponding to genuine tissue/image contrast), whereas the PCs with small eigenvalues typically correspond to noise since they are uncorrelated across voxels and have lower variation [Gurney-Champion et al., 2019]. Hence, by nullifying the smaller-eigenvalue PCs with index $k > K_{thr}$, a denoised signal can be reconstructed under the assumption that the filtered PCs correspond to noise [Manjón et al., 2013]. However, **defining an optimal value for K_{thr} is a challenge** as it can vary across the brain, b-values, SNR, etc. Generally, it has been heuristically selected based on experimental results [Manjón et al., 2013], until the criterion was formalized in the **Marchenko-Pastur PCA (MPPCA)** approach [Veraart et al., 2016b].

MPPCA is an elegant approach that exploits a universal law for eigenvalues that Marchenko and Pastur described in 1967 [Marčenko and Pastur, 1967] where, following the Random Matrix Theory (RMT), the asymptotic distribution of non-zero singular values of a large

rectangular random covariance matrix can be estimated. Hence, under the assumption that the noise level is constant and uncorrelated within the local neighbourhood and across the dMRI measurements, it is possible to find the cut-off threshold λ_+ that delimits the \hat{M} pure-noise components from the significant signal-carrying P components (see Fig.4.4). The MPPCA algorithm can derive this threshold from the noise level σ , which needs to be estimated. This noise level can be also used for post-denoising rician-bias correction of the noise-floor (e.g., [Koay and Basser, 2006]). Then, nullifying the \hat{M} components (i.e., imposing a hard thresholding on $\lambda \leq \lambda_+$), the denoised Casorati matrix C^* can be reconstructed undoing the Singular Value Decomposition with only the P components (see Fig.4.2).

$$C^* = \sqrt{N}\mathbf{U}\hat{\Lambda}\mathbf{V}^T \quad (4.3)$$

where \mathbf{U} and \mathbf{V} are unitary matrices whose columns are the left- and right-singular vectors of C , and $\hat{\Lambda}$ are the singular values such that $\lambda > \lambda_+$.

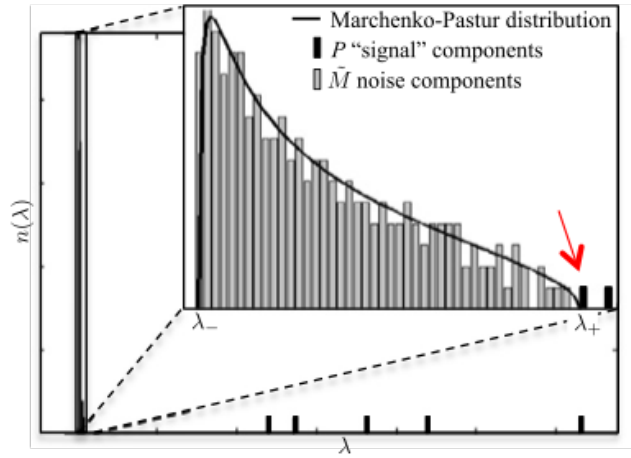


Figure 4.4: *The Marchenko-Pastur distribution fitted to the histograms of eigenvalues λ of C (sorted from lowest to highest). λ_+ sets the threshold between pure-noise M components ($\lambda \leq \lambda_+$) and the P signal components ($\lambda > \lambda_+$). Image obtained from [Veraart et al., 2016b].*

The redundancy $M-P$ (and, therefore, the number of components removed) will increase with factors such as the windows size (typically, N is set to be $N \geq M$) or the number of shells. All in all, MPPCA has been shown to be effective in removing variance solely rooted in thermal noise while preserving anatomical details. It is commonly used as the state-of-the-art denoising approach when multiple acquisitions are not available.

4.2.4 NOise Reduction with DIstribution Correction (NORDIC)

Among others, **MPPCA** relies on two fundamental assumptions about the random covariance matrix: 1) the **properties of noise** are identical across neighbour voxels, and 2) that **signals are zero-mean Gaussian distributed**. However, we have already seen that 1) noise properties in modern protocols are spatially non-stationary, because of the use of undersampled k-space acquisitions [Aja-Fernández et al., 2014], and 2) that the magnitude data commonly used deviate from Gaussian properties in modern protocols [Gudbjartsson and Patz, 1995; Salvador et al., 2005] (see sec.2.1.2).

To address this, **NOise Reduction with DIstribution Correction (NORDIC)** [Moeller et al., 2021a] introduces a few modifications to the original MPPCA algorithm. First, NORDIC operates on complex-valued images; since the assumption that noise is additive, zero-mean Gaussian is full-filled more closely in the complex domain. The Casorati matrix is constructed in the same manner but it is assumed to be $C^* = C + N$, where $N \sim \mathcal{N}(0, \sigma^2)$. Second, data is divided by the *g-factor* map to achieve variance normalization and get a homogeneous spatial distribution, hence fulfilling the assumption about spatially homogeneous noise properties (see Fig.4.5). Unlike MPPCA, NORDIC uses known information from the acquisition to "pre-process" the data to fit the MP law instead of either estimating the necessary information (noise level) or adapting the algorithm to fit the data.

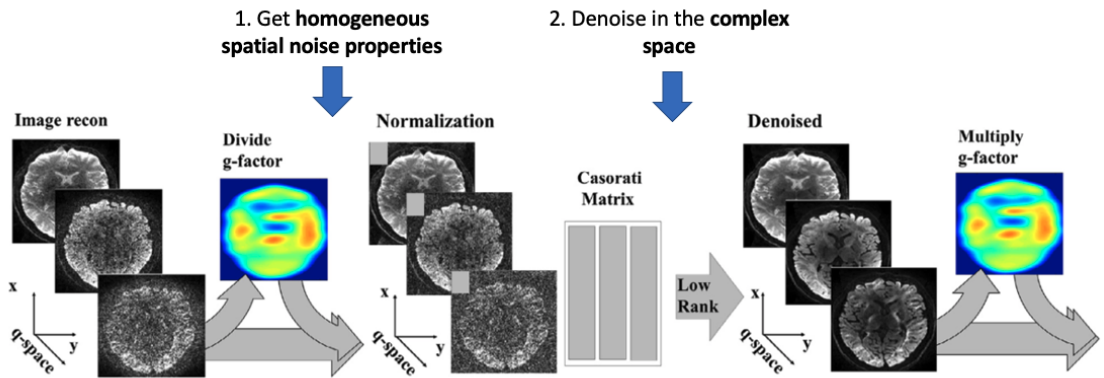


Figure 4.5: Scheme of NORDIC denoising. It introduces 2 steps to adapt the data to meet the Marchenko-Pastur statistical assumptions: 1) Dividing by the *g-factor* to get a homogeneous spatial distribution of noise and 2) denoise in the complex space where data is corrupted by Gaussian noise. Image adapted from [Moeller et al., 2021a]

Using this formalism, results from Random Matrix Theory can be effectively used to devise

a parameter-free objective threshold based on the eigen-spectrum of the Casorati matrix, as MPPCA. However, while MPPCA needs to simultaneously estimate the amount of noise σ and signal components P in magnitude-data in order to remove components that have little contribution to the variance, NORDIC sets this threshold numerically to achieve the removal of all components that cannot be distinguished from Gaussian noise thanks to meeting the MP assumptions. More specifically, as the analytical expression given in the MP law is an asymptotic expression for infinite matrices, Monte-Carlo simulations with the same size of the images are used to find the largest singular value (i.e., λ_+) of a random such with i.i.d. zero-mean gaussian with identical variance σ^2 as in noise images acquired without RF applied. Then, similar to PCA, the denoised data can be obtained by nullifying the $\lambda \leq \lambda_+$ components and applying eq.4.3.

4.2.5 Patch2Self

Recently, Machine Learning principles used for denoising images in other fields have been proposed for denoising MRI. Among the different types of learning, self-supervised methods have arisen as the most suitable approach for MRI given the lack of ground-truth data. Although it has not been extensively applied yet to dMRI (e.g., see [Moreno López et al., 2021] for a review), the Patch2Self (P2S) [Fadnavis et al., 2020, 2022a] method has gained attention in the field.

Patch2Self (P2S) is built upon the Noise2Noise [Lehtinen et al., 2018] and Noise2Self [Batson and Royer, 2019] principles. Shortly, Noise2Noise is based on the assumption of **statistical independence of the noise** of image pixels, so it does not make any assumption about the signal distribution. As noise is random and independent, Noise2Noise needs 2 noisy versions of the same image, using one to learn about the other. This necessity of having at least two images is removed in Noise2Self by relying on a mathematical property called **J-invariance**. A denoising function is J-invariant if *the prediction it makes for each pixel does not depend on the value of that pixel in the original image*, i.e. it can use the information contained in the rest of the image to make the prediction of the denoised voxel. In this way, an image can be denoised only using the image itself.

P2S follows these principles and assumes that thermal noise is random and uncorrelated

across dMRI volumes. So, as the signal has some correlation between volumes and the noise is assumed to be independent between them, P2S try to learn each of the 3D volumes as a combination of the other $N - 1$ volumes. Relying on the J-invariance assumption, the problem can be simplified effectively to some sort of **out-of-sample prediction**: 1) a 3D volume J is taken from the 4D dMRI data and is held out, 2) patches from the rest $N - 1$ dMRI volumes are used to train a model M that predict the central voxel of the patches, 3) use the model M trained on the $N - 1$ volumes to predict the central voxels of patches in volume J . Those predictions replace the original data and constitute the denoised data. This process is repeated for the rest of the volumes (i.e. if you have N volumes, N models need to be trained). The P2S approach can be seen also as **predicting a q-space point as a combination of all the others, but not of itself**. Any method can be used to learn the model M , such as neural networks. Nevertheless, authors suggest that *a linear regression is good enough given the oversampling and redundant information of q-space* [Fadnavis et al., 2020, 2022a].

4.3 Methods

Given the plethora of denoising methods and lack of objective ways for evaluating and comparing them, we propose a set of considerations aimed towards this direction. We call them collectively **EDDEN (Evaluation of Diffusion MRI DENoising)**, comprising of a set of criteria for characterising performance of denoising methods, from signal raw quality to model estimates; and a set of purpose-fit datasets representing different SNR/resolution trade-offs. This work aims to provide more insights into the nature of denoising and how it can be formally characterised. We use EDDEN to compare and characterise a number of denoising algorithms overviewed in the previous section, namely NLM, P2S, MPPCA and NORDIC.

4.3.1 Defining a good denoising algorithm

EDDEN is based on features and performance that a principled and well-behaved denoising algorithm is expected to have. Intuitively, an ideal denoising method should be able to identify and remove as much noise as possible from the raw signal and, as a consequence, also their undesired effects (e.g., the noise-floor). At the same time, this must be done without perturbing the information-carrying components of the signal regardless of the

level of noise contained in the measurements. Translating this into the specific case of dMRI, there are a number of features that can be expected, characterising from raw signal quality to high-level analyses:

- Denoising should improve raw signal quality by removing noise-related variance.
- Denoising should preserve expected statistical properties of the dMRI signal and remove biases introduced by noise-floor.
- Denoising should preserve the spatial resolution of the original data.
- Denoising should converge in the high-SNR regime (e.g. not introduce undesired effects in low-noise scenarios).
- Denoising should be beneficial for modelling performance.
- Denoising should allow us to utilise the gained SNR for a range of SNR-limited applications (e.g., allow for very high spatial resolution, or allow for short scans with a few volumes).

Based on these principles we evaluate the performance of denoising methods, based on corresponding considerations.

Consideration 1: Provide gains on raw signal quality

The signal-to-noise ratio (SNR) and contrast to noise ratio (CNR) are common metrics to characterise dMRI quality. In our work, the SNR is evaluated using repeats of images with the same contrast and, specifically, with no diffusion gradient applied ($b = 0s/mm^2$). The CNR is evaluated in the angular domain (i.e. "diffusion" contrast), across volumes within the same b-shell. By reducing the σ_{noise} , both SNR and CNR increase and, generally, the higher these are the better. For our experiments, we use the voxel-wise estimations provided by *Eddyqc* for each voxel v [Bastiani et al., 2019].

$$SNR_{b=0}^v = \frac{mean(S_j^v)}{std(S_j^v)} \quad (4.4)$$

$$CNR_{b_i}^v = \frac{mean(S_{jpred}^v)}{std(S_j^v - S_{jpred}^v)} \quad (4.5)$$

where S_j^v indicates the signal of voxel v from volumes obtained with b-shell i . For the CNR, the mean of noise-less signal predictions $S_{j_{pred}}^v$ are used as signal, while residuals of predictions to true measurements are used as noise. Predictions are obtained from a Gaussian Process (GP) that represents non-parametrically the signal in each shell per voxel, considering signals from all diffusion-sensitising volumes [Andersson and Sotiropoulos, 2016]. An average SNR and CNR value can then be obtained from the above maps, by considering the voxels of interest (e.g. tissue-based).

Consideration 2: Reduce noise-floor and preserve signal statistical properties

It is essential that a denoising approach preserves the statistical properties of the signal, as these determine the validity of assumptions for downstream applications and analysis. A number of post-processing steps assume that the signal follows a Gaussian/Rician distribution, so it is important to examine whether these prevail after denoising. Of crucial interest here is to determine whether denoising algorithms are able to remove or, at least, to **reduce the noise-floor**, the minimum detectable signal level in the absence of true information. A noise-floor can cause significant biases to dMRI estimates [Jones and Basser, 2004; Sotiropoulos et al., 2013c], so a denoising algorithm is expected to deal with it.

To characterise the above, noise properties such as the variance or the mode should be sampled from the data. Ideally, these would be taken from the background but it has been already commented that noise is also spatially non-stationary. Noise properties can be extracted from the ventricles instead. The diffusion signal there is completely attenuated because of the isotropic propagation and, compared to the background, it can offer closer proximity to WM voxels (and also measurements in the iso-centre of the brain are more robust compared to the periphery) [Dietrich et al., 2008].

Further to characterising noise properties before and after denoising, we also evaluate denoising effects on the noise-floor. Good candidates in WM to explore noise-floor effects are voxels with very large attenuation, such as in the mid-body of the Corpus Callosum. Due to the large anisotropy in these regions, the signal in these voxels has a large dynamic range (from parallel to perpendicular to the fibres) and is also maximally attenuated compared to any other WM region. Such voxels typically suffer from noise-floor effects,

where noise rectifies the true signal level and reduces the true dynamic range [Sotiropoulos et al., 2013c].

Consideration 3: Preserve spatial resolution and do not induce spatial blurring

Given the use of patches or averages in denoising approaches, a question remains whether some loss of spatial resolution and spatial blurring can be expected after denoising. There are few proxies that can be used to evaluate the blurring induced by denoising, such as the frequencies suppressed in the k-space energy densities used in [Veraart et al., 2016b], or introducing Gaussian blurring and estimating what auto-correlation functions match with the denoised image [Moeller et al., 2021a]. Similar to the latter, we use Resolution Elements or *Resels* in this work. This is a standard approach from the Gaussian Random Field theory used, for instance, in fMRI for clustering inference or to correct p-values in multiple comparisons [Hayasaka and Nichols, 2003]. By using derivatives of the spatial covariance, it is possible to explain neighbourhoods of voxels that have similar covariance (i.e., effective resolution). More specifically, to estimate the voxel resolution, a 2D zero-mean Gaussian kernel is fitted to the data; the Full-Width Half-Maximum (FWHM) of the best fitting will provide an approximation to the actual voxel resolution (i.e., if there is spatial smoothing, covariance will be introduced into a neighbourhood of voxels). This kernel can be interpreted as the kernel that would produce the same smoothing as the one we observed in the data.

To estimate the spatial blurring in the filtered data we used FSL’s *smoothest* that estimates the smoothing extent in linear model residuals. It is a proxy, but it can be fairly extrapolated to the smoothing in filtered data before pre-processing, as it is just a linear transformation. We fit the tensor model (DTI) to the data and then obtain the residuals. It is then possible to estimate these Resels by studying the variance (i.e., the FWHM) on the residuals along frequency (x-axis) and phase encoding (y-axis) image directions. We repeat this process before and after denoising to explore whether denoising causes any difference.

Consideration 4: Converge at a high SNR

The convergence at a high SNR regime can be used to verify any unwanted signal modification when the noise levels are too low. Modern protocols operating at relatively low resolutions (e.g. $\sim 2\text{mm}$) have medium-to-high SNR. Denoising algorithms should be able to operate in these scenarios with minimum invasiveness. For instance, denoising should not remove useful information nor introduce any bias or undesired effect, i.e., statistics and performance in raw data and denoising methods should not differ substantially. We performed a number of checks to verify the above before and after denoising on a relatively high-SNR dataset. We evaluated raw signal quality metrics (SNR, CNR), as well as second-level analysis (convergence of tractography results).

Consideration 5: Improvements on modelling performance

The preservation of statistical assumptions and an increased SNR/CNR are expected to allow better performance in estimating dMRI microstructural models, although quantifying improvements in accuracy and precision in real brain data is limited by the lack of ground-truth. An attractive scenario is resolving crossing fibres in White Matter, as there are regions where multiple 3-way crossings are a-priori expected almost everywhere (e.g., Centrum Semiovale [Jeurissen et al., 2013]) and regions where no crossings are expected (e.g., midbody of the Corpus Callosum). For modelling crossing-fibres, we used the Ball&Sticks model (BedpostX tool in FSL [Behrens et al., 2007; Jbabdi et al., 2012]²) with up to 3 fibre compartments and a Gaussian noise model (so not including any term to model the noise-floor). By looking into the number of estimated crossings in the Centrum Semiovale, we can compare modelling performance before and after denoising; by looking into uncertainty estimates in the Corpus Callosum, we can obtain a measure on precision for regions where no fibre crossings are expected (and hence any estimation of complex fibre structure and thus increased uncertainty will likely be artefactual).

Consideration 6: Benefits for pushing spatial resolution and/or reducing scan time

Improvements in signal quality can be capitalised further in a number of ways, by: a) attempting ultra-high spatial resolution images or b) reducing scan time by utilising the

²*burnin*=3000 samples to ensure convergence of the MCMC chain to the parameter region of interest

higher effective SNR per unit time. We have acquired data and designed tests to explicitly assess these two routes for denoising methods.

Firstly, we acquired a sub-millimeter 3T dataset (very noisy data, see sec.4.3.2) and explored whether algorithms that rely on whole-image information, like tractography, can operate in this highly noisy regime, before and after denoising. We relied on a set of standardised tractography protocols, available in XTRACT [Warrington et al., 2020] to see whether we can reconstruct a range of WM bundles, from association to projection and commissural fibres, that all demonstrate a different level of complexity along their route. We have also acquired multiple repeats of this dataset so that the multiple-averages can be used here as subject-level reference, while the high-resolution HCP atlas (1mm isotropic) is used as population-average reference [Warrington et al., 2020].

Secondly, the increased SNR/time ratio expected from denoising opens the possibility to indirectly reduce scan the time needed to achieve certain SNR to achieve similar performance as with the full acquisition by a) either reducing number of repeats, or b) the number of dMRI directions. For that, we evaluated the convergence of different sizes (i.e. scan times) of denoised data to a) the multiple-averages and b) to the complete sequence of a single-repeat in the RAW data. Convergence was quantitatively assessed using agreement in simple DTI estimates (e.g., FA or MD).

Specifically, RAW Datasets B and C (1.5mm and 0.9mm respectively, see sec.4.3.2) were subsampled multiple times with equally increasing sizes. Discarding b0 volumes, Dataset B was subsampled into the first 45, 90, 135, 180 and 225 directions (effectively reducing to 15, 30, 45, 60, 75 and 90 $b=1000\text{mm}/s^2$ volumes for DTI model fits); similarly, Dataset C was subsampled into the first 30, 60, 90, 120, and 150 directions (effectively reduce to 15, 30, 45, 60, 75 and 90 $b=1000\text{mm}/s^2$ volumes for DTI model fits). So, in total, 6 datasets (5 subsampled sets and the full acquisition) were obtained from the first repeat of Dataset B and other 6 from the first repeat of Dataset C. Each of these datasets were denoised and pre-processed independently following the pipeline explained in 4.3.2, so every subsampled dataset also reflected the effect of the respective data size.

4.3.2 Data

All MRI datasets were acquired on the same healthy subject at the Centre for Magnetic Resonance Research (CMRR) at the University of Minnesota in a Siemens 3T Prisma MRI System using a 32-channel head coil. A T1-weighted MPRAGE sequence was run at 0.8x0.8x0.8 mm voxel size (TR=2.4 s, TE=2.22 ms). Three dMRI multi-band (MB) datasets [Moeller et al., 2010] were acquired and reconstructed using a SENSE1 algorithm [Sotiropoulos et al., 2013c] to represent three different SNR regimes:

- **Dataset A:** A UKBiobank-like dMRI dataset [Miller et al., 2016] with 2mm isotropic resolution, TR=3s, TE=92ms, MB=3 (no GRAPPA), and 116 volumes in total: 9 b0 volumes and 107 diffusion encoding orientations, as 53 with b-value of 1000 s/mm^2 and 54 with b-value of 2000 s/mm^2 , phase encoding direction AP and one b0 image with PA for susceptibility correction. This dataset represents a relatively medium-to-high SNR regime.
- **Dataset B:** Five repeats of an HCP-like dMRI dataset [Harms et al., 2018] with 1.5mm isotropic resolution, TR=3.23 s, TE=89.2 ms, MB=4 (no GRAPPA), and 300 volumes in total: 30 b0 volumes and 270 diffusion encoding orientations as 90 with b-value of 1000 s/mm^2 , 90 with b-value of 2000 s/mm^2 and 90 with b-value of 3000 s/mm^2 , phase encoding direction AP and one b0 image with PA for susceptibility correction. This is a low-to-medium SNR dataset, with relatively high resolution and multiple repeats to average and use as a reference.
- **Dataset C:** Three repeats of an ultra-high-res dataset with 0.9mm isotropic resolution, TR=6.569 s, TE=91 ms, MB=3, GRAPPA=2, phase encoding direction AP (with one b0 image with PA for susceptibility correction), and 200 volumes in total: 20 b0 volumes and 180 diffusion encoding gradients, as 90 with b-value of 1000 s/mm^2 and 90 with b-value of 2000 s/mm^2 , phase encoding direction AP and one b0 image with PA for susceptibility correction. The purpose of this data set was to study whether denoising can recover information and connectivity estimates in extremely noisy data, acquired using a clinical scanner.

Each data set was acquired on a different day, minimizing the tiredness and involuntary motion of the subject. All acquisitions were obtained parallel to the anterior and posterior commissure line, covering the entire cerebrum. Each data set was visually inspected before

further processing. No severe artefacts or alterations were observed.

	Protocol	Resolution	B-shells	Directions	Repeats
Dataset A	UKB-Like	2mm	1k/2k	107	1
Dataset B	HCP-Like	1.5mm	1k/2k/3K	270	5
Dataset C	HCP-Like	0.9mm	1k/2k	180	3

Table 4.1: Summary table of the datasets acquired for the evaluation of denoising approaches and included in EDDEN.

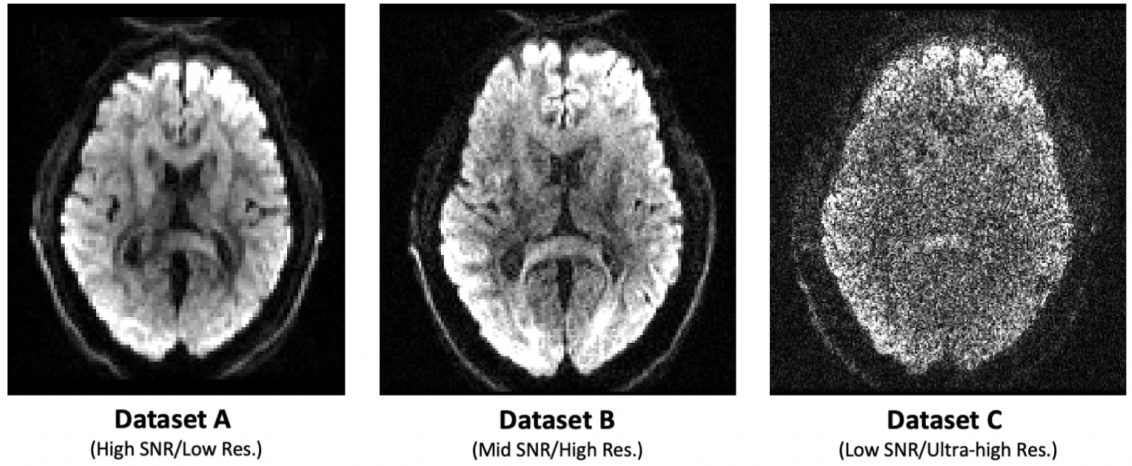


Figure 4.6: Examples axial screenshots of the three datasets acquired for the evaluation of denoising approaches. Each dataset offers a different SNR/resolution scenario.

Data processing

Every dataset has been pre-processed for distortion and motion correction using an in-house implementation [Mohammadi-Nejad et al., 2019] of the dMRI HCP pipeline [Sotiropoulos et al., 2013b]. This includes skull stripping, Eddy current correction, EPI distortion correction, motion correction, non-linear registration to MNI standard space and quality control evaluation. Averages from multiple repeats in datasets B and C were used in some experiments as gold-standard references. In these cases, each repeated acquisition was linearly aligned to the first b0 image of the first dataset acquired in that protocol, pre-processed and averaged afterwards. For some experiments (e.g., criterion 6), this pipeline was slightly different; for such cases, the modifications on the pipeline are specified in the correspondent section.

Denoising was applied to all data sets prior to any distortion correction, described above. As denoising algorithms, we used the Non-Local Means (NLM, using the default implementation in DIPY [Garyfallidis et al., 2014]) as an exemplar of non-linear fundamental filter; the Marchenko-Pastur PCA (MPPCA, using the default MrTrix3 implementation (<https://www.mrtrix.org>), and the NORDIC algorithm (author’s original implementation) as exemplars of PCA-based approaches, which are considered the current state of the art, applied to magnitude and complex domain, respectively; and Patch2Self (P2S) (using the default implementation in DIPY [Garyfallidis et al., 2014]) as exemplar of novel self-supervised approaches that can be combined with Deep Learning methods. NLM, MPPCA and Patch2Self were applied to the reconstructed magnitude data; NORDIC was applied to the magnitude and phase reconstructed data (i.e. complex domain).

	Data	Patch-Size	Algorithm	Multi-shell
NLM	3D-Magnitude	3x3x3	Non-local averaging	No
MPPCA	4D-Magnitude	$\text{int}(n \leq \sqrt{M})^* 3$	Local with overlapping	Yes
NORDIC	4D-Complex	11x11x11	Local with overlapping	Yes
P2S	4D-Magnitude	3x3x3	Local with overlapping	Yes

Table 4.2: Summary table of the features of the denoising methods evaluated.

4.4 Results

In this section, we will show results obtained from evaluating the denoising approaches indicated above. For the representation of the results, the following colour code will be used: RAW data (non-denoised) in orange, NLM in pink, MPPCA in blue, NORDIC in green, P2S in purple, multiple-averages in red.

4.4.1 Raw signal quality assessments

Qualitative demonstrations of diffusion data pre- and post-denoising for each method and for different b-values can be seen in Fig.4.7 (Dataset A - 2mm), Fig.4.8 (Dataset B - 1.5mm), and Fig.4.9 (Dataset C - 0.9mm). Across all SNRs a visual improvement in data quality can be perceived after denoising, particularly noticeable for low SNR raw data. What can be noticed as well is that the signal in the ventricles (noise-floor) is preserved or reduced differently by different methods (see arrows in Fig.4.9, for instance). The ap-

proach that operates in the complex domain (NORDIC), seems to reduce this signal more compared to the others, which is particularly evident for the $b=2000 \text{ s/mm}^2$ in 0.9mm data.

As described in Methods, gains in raw signal quality can be quantified by SNR and angular-CNR (Fig.4.10). In agreement with the qualitative images, all the denoising approaches improved the quantitative metrics as well. These gains are especially noticeable in NORDIC at higher b-shells and higher resolutions, outperforming the rest of the methods.

Gains in signal quality are not shown for P2S as they were highly inconsistent (and very sensitive to seemingly unrelated parameters, like the number of dMRI volumes). EddyQC, the tool used to calculate the SNR and angular CNR after distortion correction, returned very low CNR values, which did not agree with the visual aspect of the denoised images obtained prior to distortion correction (i.e. Figs.4.7, 4.8, and 4.9). EddyQC uses a Gaussian Process predictor and this is tuned by hyperparameters optimised on runtime from the data. It seems that the deterministic nature of P2S predictions (P2S replaces all data with a deterministic prediction) interferes with EddyQC’s Gaussian process tuning and further estimation.

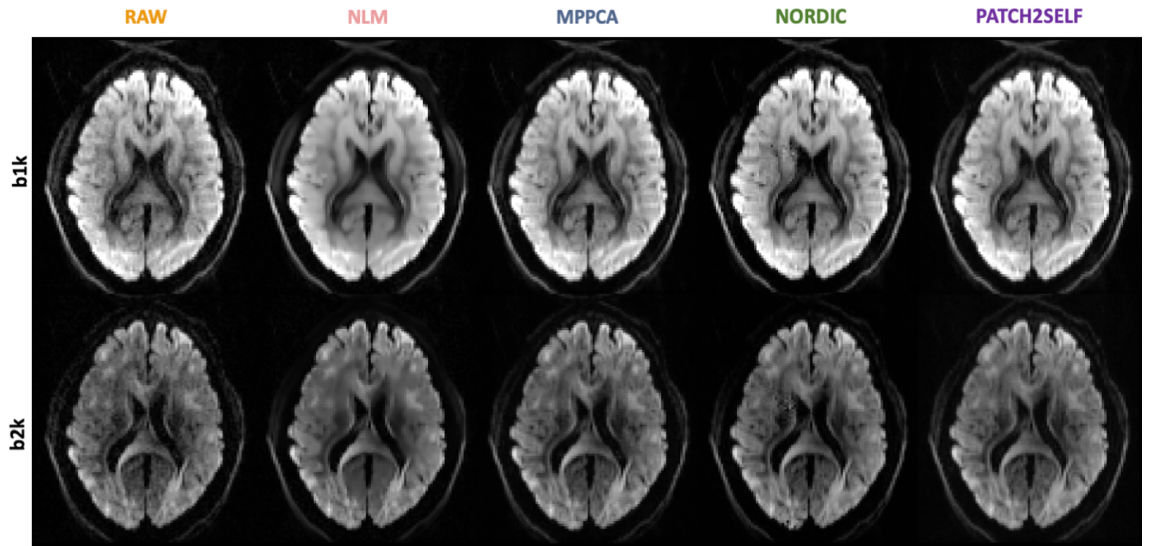


Figure 4.7: Dataset A (2mm) - Denoised - Qualitative maps of the dMRI signals pre- and post-denoising by each method (columns) in a $b=1000 \text{ (s/mm}^2\text{)}$ (top row) and $b=2000 \text{ (s/mm}^2\text{)}$ (bottom row). At this resolution, all denoising methods present similar results except NLM, which returns a smoothed image losing considerable structural information. NORDIC provides greater gains than the rest of the methods, especially at higher resolutions at higher b-values.

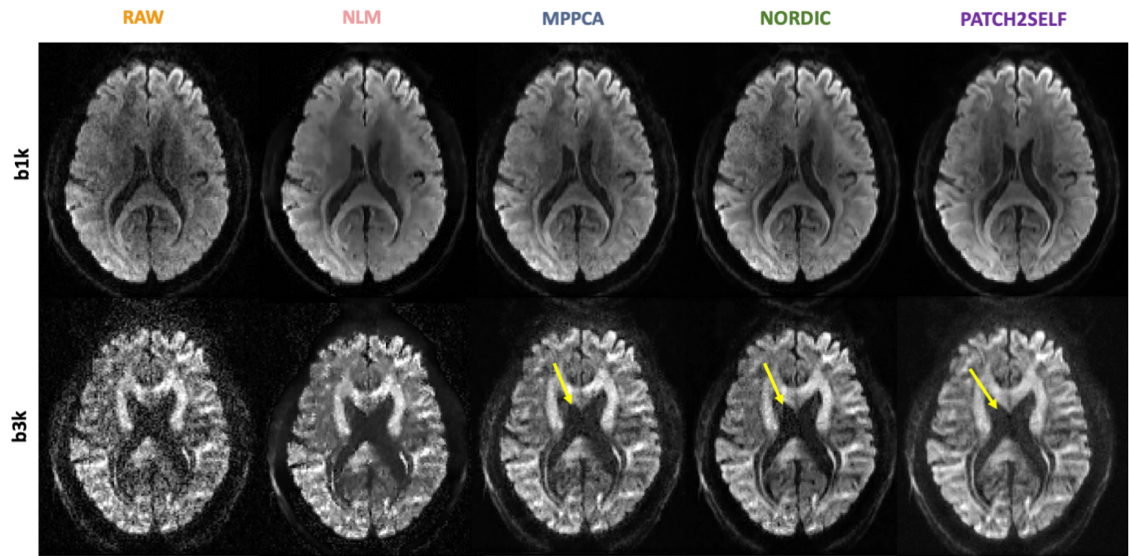


Figure 4.8: Dataset B (1.5mm) - Denoised - Qualitative maps of the dMRI signals pre- and post-denoising by each method (columns) in a $b=1000$ (s/mm^2) (top row) and $b=3000$ (s/mm^2) (bottom row). At high b -values, the noise effect is higher and elevated noise-floor preservation in the ventricles can be observed (yellow arrows).

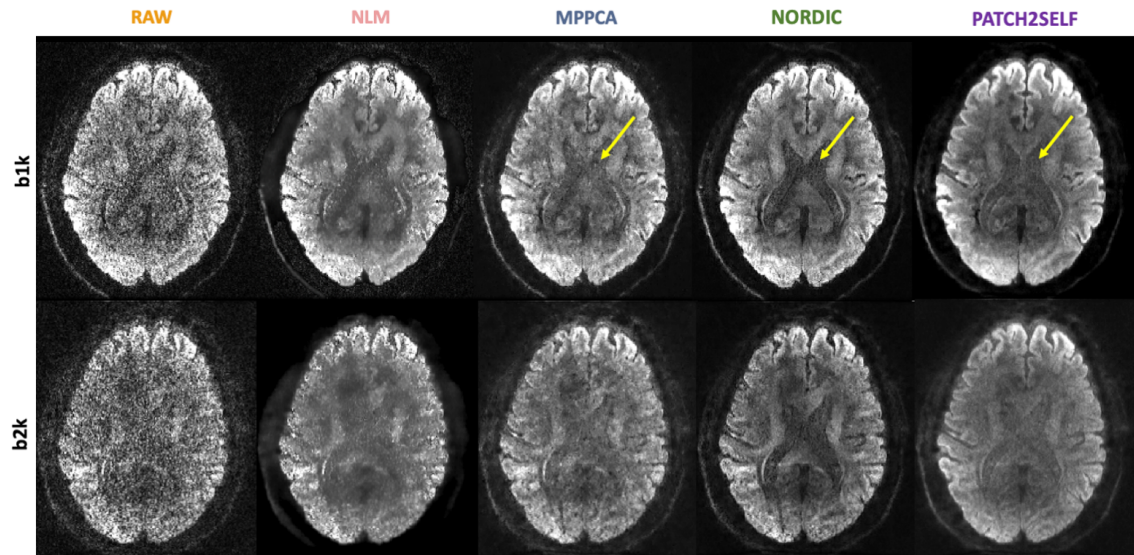


Figure 4.9: Dataset C (0.9mm) - Denoised Qualitative maps of the dMRI signals pre- and post-denoising by each method (columns) in a $b=1000$ (s/mm^2) (top row) and $b=2000$ (s/mm^2) (bottom row). In this case, the elevated noise-floor in the ventricles can be observed at every b -value.

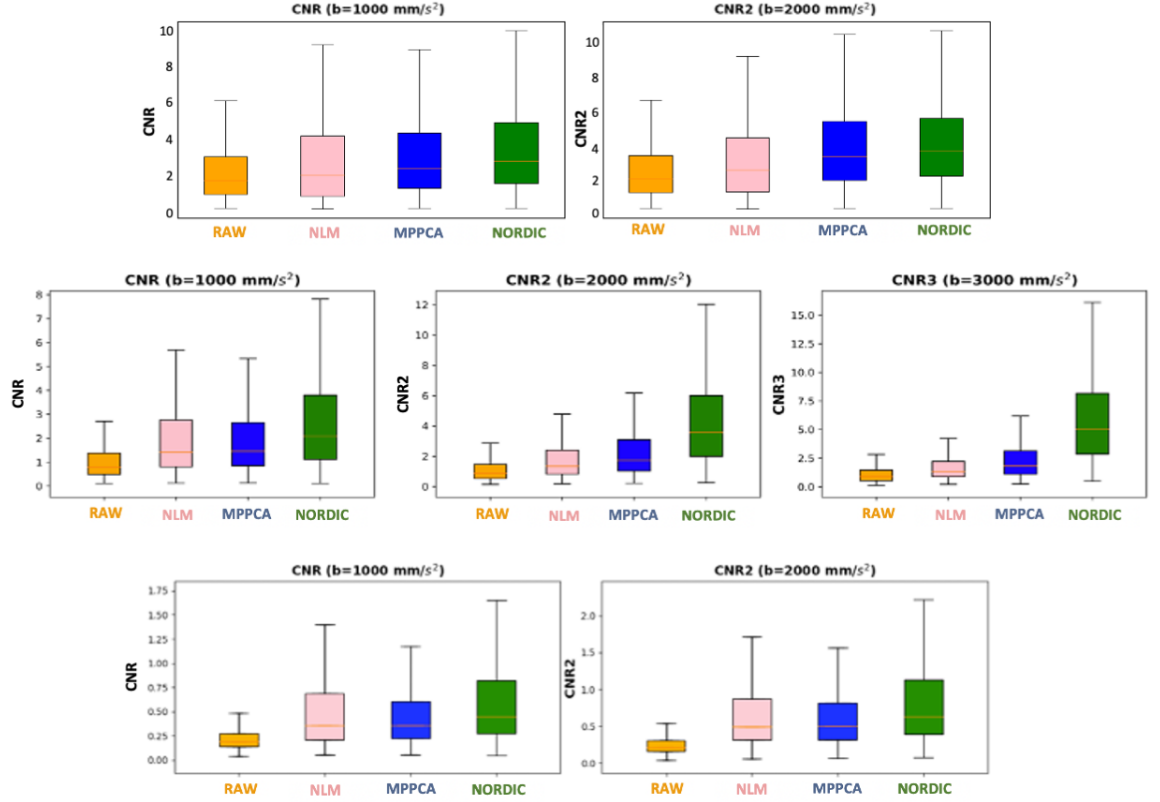


Figure 4.10: Gains in Angular Contrast to Noise Ratio (CNR) by denoising - Top row: Dataset A - 2mm; Middle row: Dataset B - 1.5mm; Bottom row: Dataset C 0.9mm. All methods show gains in angular CNR regardless of the resolution

Fig.4.11 shows difference maps and their corresponding histogram estimated between the RAW image and the denoised image for each case. As these should arise from the thermal noise, they are expected to be spatially random and zero-mean Gaussian distributed. This is the behaviour mostly observed in MPPCA and NORDIC denoised data. However, in P2S and especially in NLM, some structural information remains in the difference maps, which is translated into deviations from zero-mean and skewed distributions. This agrees with the smoothing effect observed in Figs.4.7, 4.8, and 4.9, and with results from previous studies [Mishro et al., 2021] (see also sec.4.4.3). Hence, only PCA-based approaches seem to perform a principled separation between signal and noise, regardless of the scenario.

4.4.2 Noise-floor and Signal statistical properties

Fig.4.12 shows the signal intensity distribution from voxels in the ventricles. The signal is maximally attenuated in these CSF-filled regions, so what is depicted is effectively the distribution of noise, before and after denoising. The histograms of all denoised datasets show an evident reduction of the signal variance (more restricted range of values), as expected.

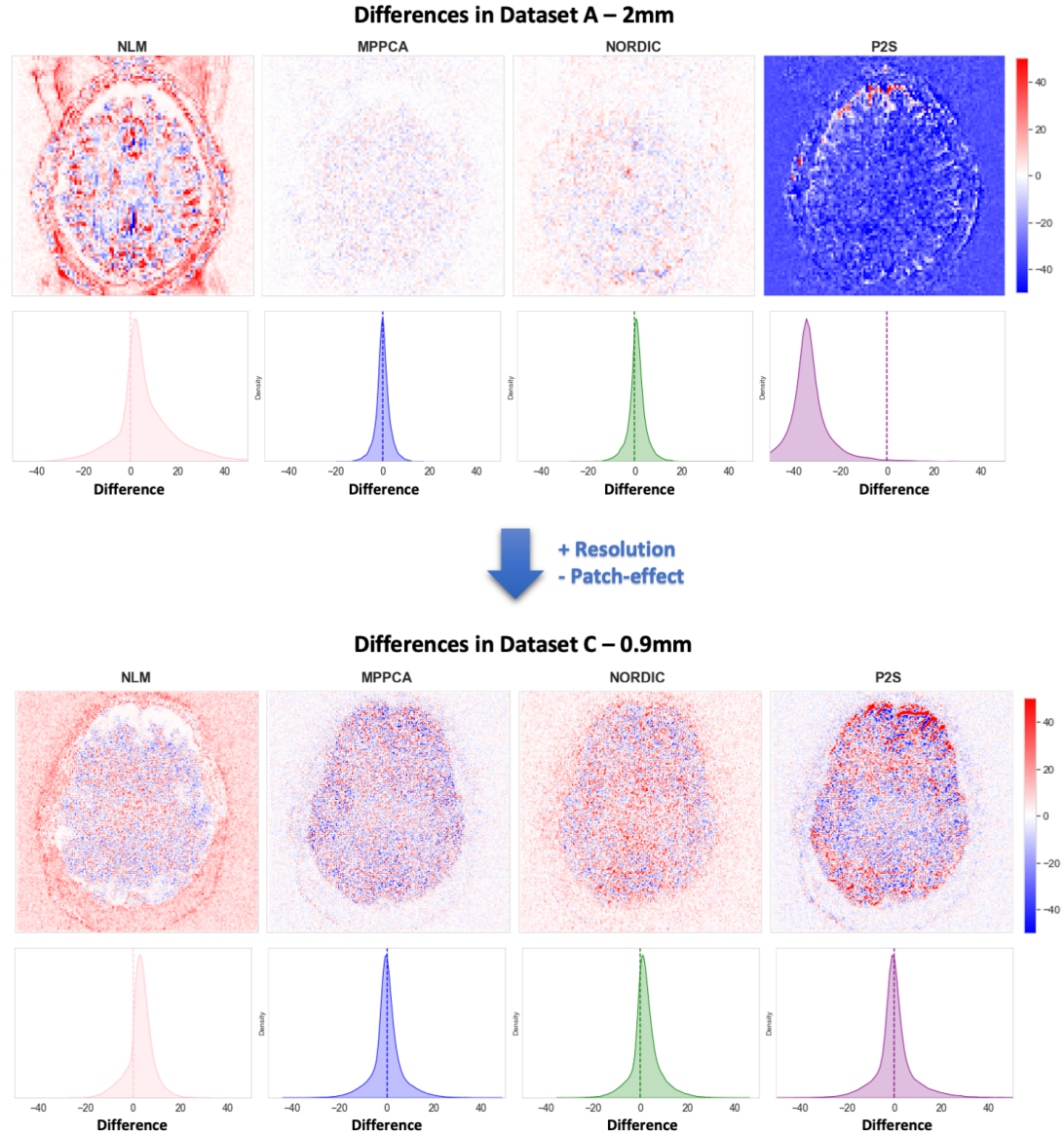


Figure 4.11: *Difference between RAW and denoised data* - The difference d , where $d = S_{\text{RAW}} - S_{\text{denoised}}$, is represented for each method in an axial map for an exemplar slice, and its corresponding histogram of values. If differences arise from the elimination of random thermal noise, they should follow random spatial patterns and zero-mean Gaussian distributions.

How to evaluate whether the removed parts are only noise or they also contain signal features is a hard problem, especially not having access to the ground-truth. The residual maps in Fig.4.11 suggested that most of the removed components can be attributed to random noise, especially in PCA-based methods (e.g., no anatomical structures observed in their residuals).

Another important observation is the preservation of the noise-floor in the majority of approaches. Apart from NORDIC, which was the only one able to reduce noise-floor, all methods practically keep the same noise-floor as the RAW data i.e., they only reduce the variance of the signal (which is translated into SNR and CNR gains). As commented in section 2.3.1, the preservation of the noise-floor can produce a rectification of the signal in the WM and induce potential biases in the estimates. In WM, this rectification can be observed in voxels with high anisotropy that produces high signal attenuation. At high SNR, rectification is not noticeable. However, as the SNR decreases, the chances of hitting the noise-floor when the gradient orientation is closely parallel to the fibre orientation and the signal attenuation is maximum become more likely. As NORDIC has a lower noise-floor, the rectification happens later and a larger dynamic range of the signal is returned.

This behaviour is demonstrated more clearly on the right plot of Fig.4.12 for the signal of a voxel in the mid-body of the Corpus Callosum. The dMRI signal for that voxel has been sorted based on the alignment of the corresponding gradient orientation and the primary fibre orientation in this voxel (given by the DTI primary eigenvector). The signals from orientations parallel to the primary fibre orientation are presented first, followed by signals with increasing perpendicularity to the primary fibre orientation. Hence, the first signals are maximally attenuated, while the last signals are minimally attenuated. As shown in the figure before denoising (orange line), the maximally attenuated signals have been rectified by the noise-floor and appear roughly similar to the minimally attenuated signals (i.e. small dynamic range). Comparing the PCA-based approaches, differences can be found between denoising in the magnitude-data, which reduces the variance but does not make a difference in this rectification effect (blue line); and denoising in the complex domain, which, apart from the variance, also reduces the noise-floor and increases considerably the dynamic range of the signal (green line).

It is important to point out that none of the denoising approaches returns back a Rician noise distribution for this ultra-high-res dataset. Most of them seem to follow a symmetric non-central chi distribution, but NORDIC is the one that has a considerably lower noise-floor.

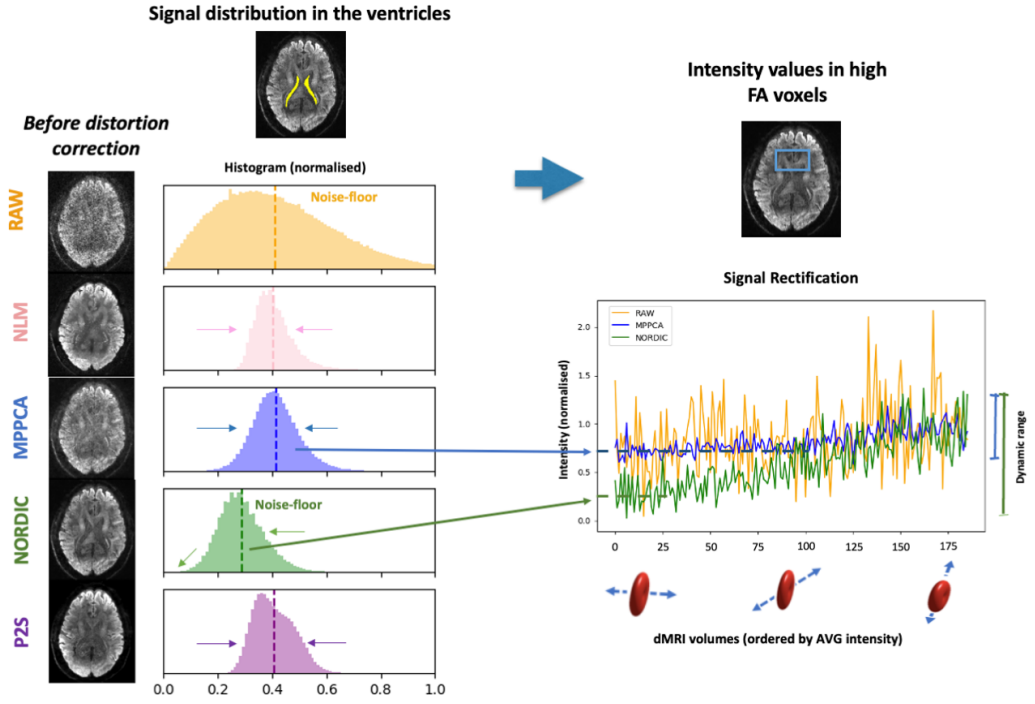


Figure 4.12: Noise-floor preservation and signal rectification (Dataset C) - On the left, the histogram of signal intensity in the ventricles obtained from the different denoised datasets. On the right, signal intensity values in a random high FA voxel ($FA > 0.8$) from the Corpus Callosum (CC) are represented, ordered by the intensity of the reference image (in this case, the reference is the RAW average of the multiple repeats)

4.4.3 Spatial smoothing

Fig.4.13 shows the estimated resolution in the images before and after denoising compared to the nominal resolution (dashed line). An estimation of the % smoothing induced per axis by each method is shown in Table 4.3. As noticed in the RAW data, due to the point-spread-function blurring, resolution along the phase-encoding direction is always smaller than nominal, while in the frequency encoding direction it is very close to nominal, as expected.

All denoising methods introduce an additional loss of spatial resolution compared to the RAW data, which is expected given the nature of patch-based denoising algorithms, with

the percentage of spatial smoothing increasing at higher resolutions (see Table 4.3). NLM is the method that induces the most spatial smoothing, confirming what is already reported elsewhere (e.g. [Mishro et al., 2021]) and observed in previous figures. On the other hand, NORDIC keeps closer to the original resolution overall, closely followed by MPPCA, inducing what could be interpreted as a lower level of co-dependency induced between voxels in both frequency (x-axis) and phase encoding (y-axis) direction. Note that estimates of P2S have not been included here for similar reasons to gains in CNR and SNR from sec.4.4.1, as *Resels* estimation relies on model fit residuals.

	FWHM-x (%)	FWHM-y (%)
Dataset A - RAW	2.16	1.56
Dataset A - NLM	15.49	21.52
Dataset A - MPPCA	11.54	15.29
Dataset A - NORDIC	8.92	6.68
Dataset B - RAW	-4.98	14.19
Dataset B - NLM	27.6	45.7
Dataset B - MPPCA	12.81	24.66
Dataset B - NORDIC	8.67	16.54
Dataset C - RAW	-0.95	16.9
Dataset C - NLM	300.73	360.37
Dataset C - MPPCA	11.96	38.16
Dataset C - NORDIC	5.04	23.14

Table 4.3: Spatial smoothing induced by each method in each dataset, quantified as the percentage ratio between the denoised-data resel size and the nominal resolution .

4.4.4 Convergence in high-SNR regimes

Previous figures have shown that PCA-based methods can still provide some gains in angular CNR (Fig.4.10), even at high-SNR regimes, where noise levels are expected to be minimal. On the other hand, we saw that denoising can induce a partial loss of spatial resolution between 5-10% (see Table 4.3), so one could ask whether denoising is beneficial or detrimental in this high-SNR scenario.

In Fig.4.14 we show the agreement of probabilistic tractography results obtained by each method and the population-average UKB atlas [Warrington et al., 2020]. MPPCA and NORDIC show similar agreement on average as RAW data, with a slightly lower variance, suggesting convergence of these methods at this SNR-regime. Interestingly, NLM caused

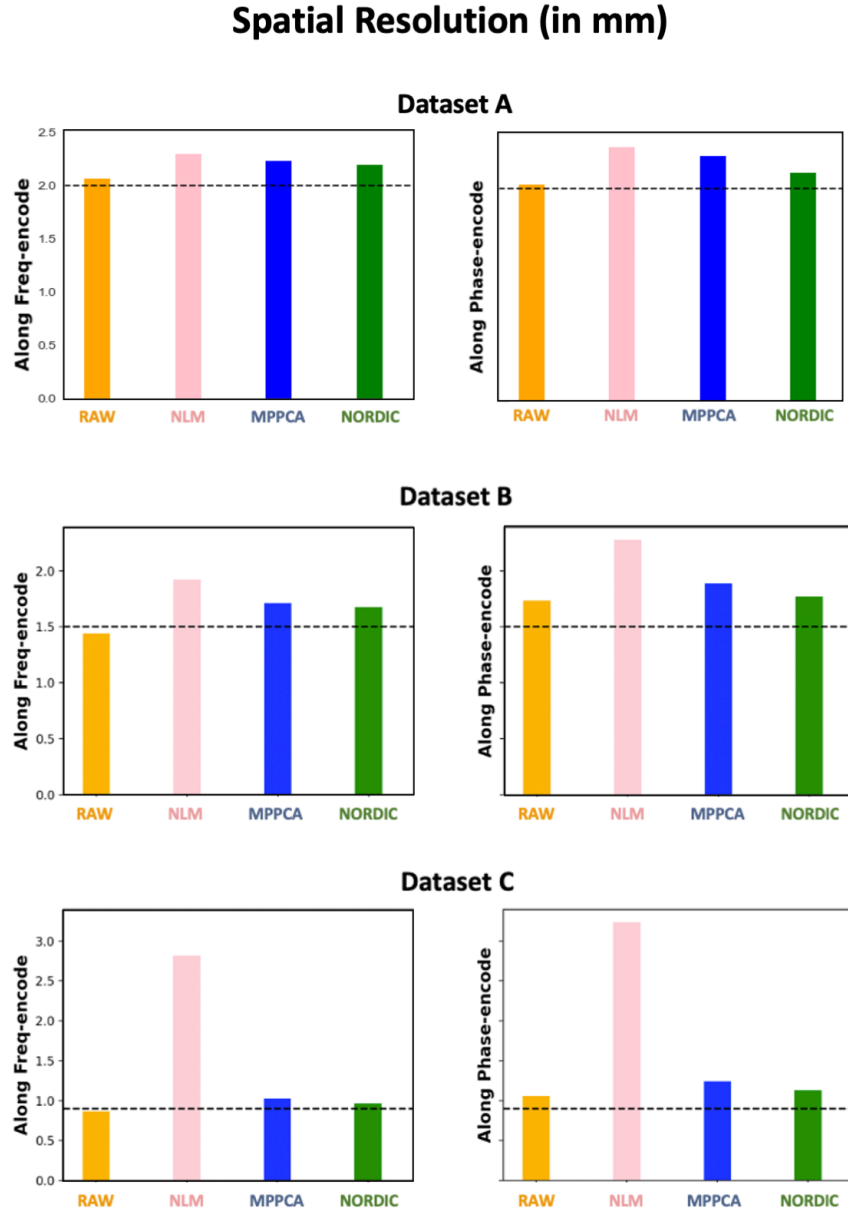


Figure 4.13: *Estimated voxel resolution* (by *FSLsmoothest*) for the denoised and non-denoised datasets, along the frequency (acquisitions along *x*-axis) and encode (acquisitions along *y*-axis) directions - Top row: Dataset A; Middle Row: Dataset B; Bottom row: Dataset C. Dashed line: acquisition nominal resolution

a worse behaviour than RAW data; the small CNR gains we observed before seem to be at cost of a higher loss of spatial specificity (Fig.4.13) and the potentially larger partial volume averaging that may hinder the propagation of the tracts.

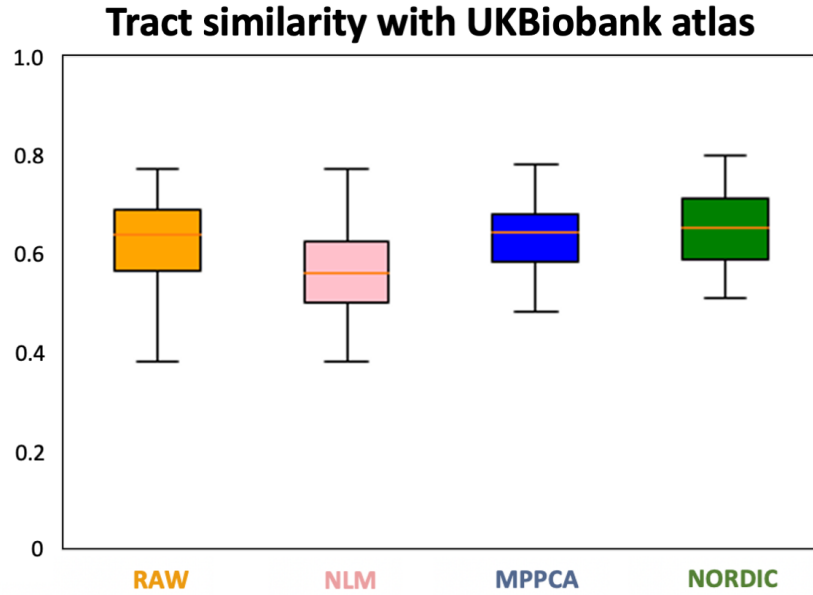


Figure 4.14: *Tract correlations with the UKB population-average atlas (Dataset A, 2mm)* - The boxplots show the spatial tract correlation of each method with the UKB population-average atlas. Non-denoised data and PCA-based methods report very similar results, suggesting convergence, while NLM provide lower correlation, suggesting loss of information.

4.4.5 Improvements in modelling performance

At higher resolutions, denoised datasets are expected to demonstrate better modelling performance, assessed here as an improvement in sensitivity to detect fibre crossings and in precision in fibre orientation estimates. Fig.4.15 demonstrates examples for a crossing-fibre model, where 0.9mm denoised data support more fibre complexity in the Centrum Semiovale, a region where most of the voxels are expected to have at least two fibre bundles crossing. Interestingly, the raw data support false positive fibre crossings in the ventricles, as no crossings or fibres are expected in this CSF-filled region, while the denoised data do not show this behaviour.

Fig.4.15 shows the percentage of voxels with two-way and three-way crossing-fibres in the Centrum Semiovale. In this case, both NLM and NORDIC provide higher rates of detection of secondary and third fibres at 0.9mm resolution, although MPPCA also shows considerable improvements over raw data.

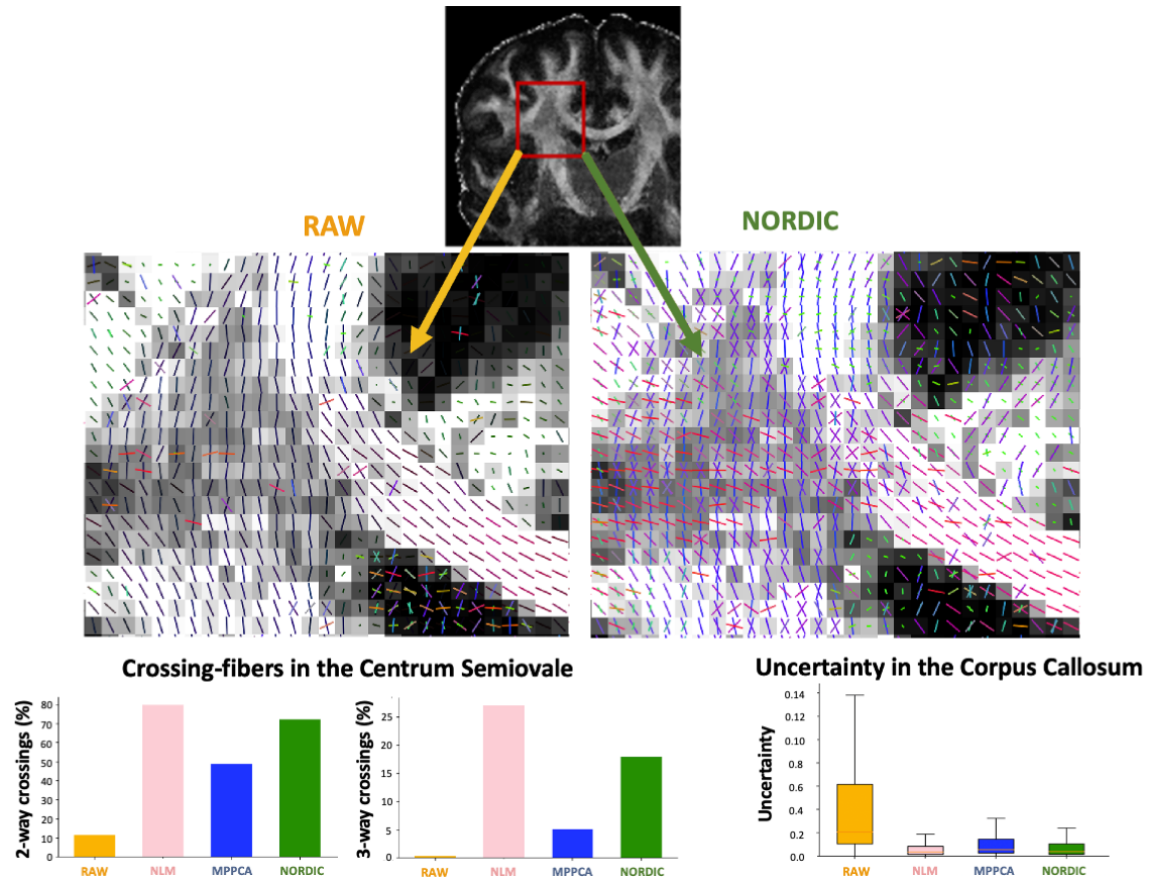


Figure 4.15: *A comparison of modelling performance between RAW and denoised data (Dataset C) - Top: Example of crossings detected in the Centrum Semiovale in the RAW and NORDIC-denoised data. Bottom left: Rate of crossing detection in the Centrum Semiovale. Bottom right: First fibre uncertainty measured in the voxels of the Corpus Callosum.*

Precision in the estimates can be assessed by the MCMC-estimated uncertainty in the fibre orientation. To avoid interpretation complications between model complexity and uncertainty, we focused on a region where most voxels are expected to exhibit a single fibre (i.e. model complexity stays the same), the midbody of the corpus callosum. Overall, as all denoising approaches reduced variance, the precision is considerably increased by all of them as well.

Overall, filtered datasets have demonstrated a considerable improvement in both sensitivity and precision in modelling performance at low SNR levels.

4.4.6 Capitalising on increased SNR to push spatial resolution

Improvements in effective SNR/CNR by denoising open opportunities that only bespoke setups can provide, such as allowing recovering information from barely unusable data at ultra-high spatial resolution. Or reducing scan time, as the increased SNR per volume reduces the need for sampling many volumes or multiple repeats. In this and the following sections, we explore the feasibility of these two applications of denoising.

Fig.4.16 demonstrates an example of doing tractography at an ultra-high resolution dataset (0.9mm isotropic), acquired using a clinical scanner. For reference, the FA image is shown on the left, before and after denoising. While a number of tracts cannot be reconstructed at all using the RAW data (seven bundle tracts in total, namely: left Acoustic Radiation, left and right Corticospinal Tracts, right Fornix, Middle Cerebellar Peduncle, and left and right Superior Longitudinal Fasciculus 1) and even after averaging (it also missed both Cortico-Spinal Tracts, the Middle Cerebellar Peduncle, and the right Superior Longitudinal Fasciculus 1), denoising allows good reconstruction of all considered WM tracts, except the Middle Cerebellar Peduncle in MPPCA and NLM.

To quantify further the improvement in tractography, Fig.4.17 represents the spatial tract correlation across all tracts with the High-resolution HCP atlas [Warrington et al., 2020], used here as an average population reference. All denoising methods improved the agreement of tractography results with the HCP Atlas, compared to the raw data. As it is reported elsewhere [Warrington et al., 2020], the common spatial tract correlation of data

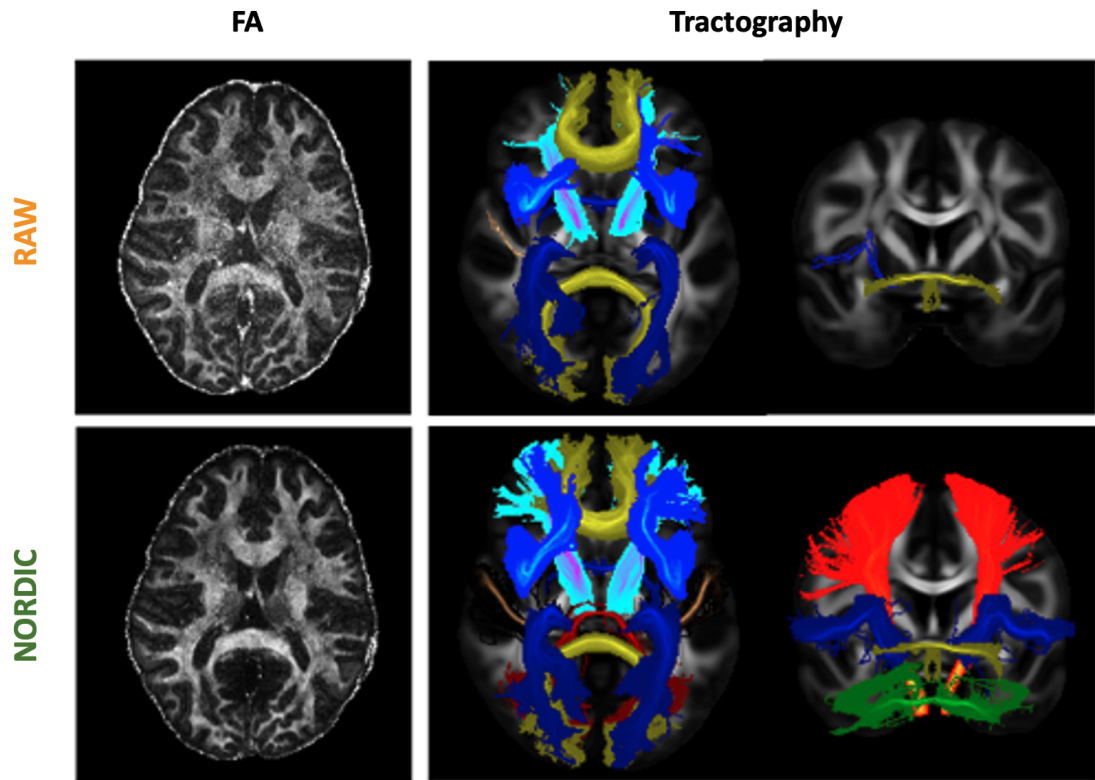


Figure 4.16: Examples of FA maps (left) and Maximum Intensity Projection (MIP) of tracts reconstructed (right) by RAW and NORDIC-denoised data at 0.9mm (Dataset C). At this resolution, denoising allows to recover multiple bundles that were missed in the non-filtered version.

at 1mm with this population average is around 0.5-0.6. NLM and NORDIC provided results around this expected range.

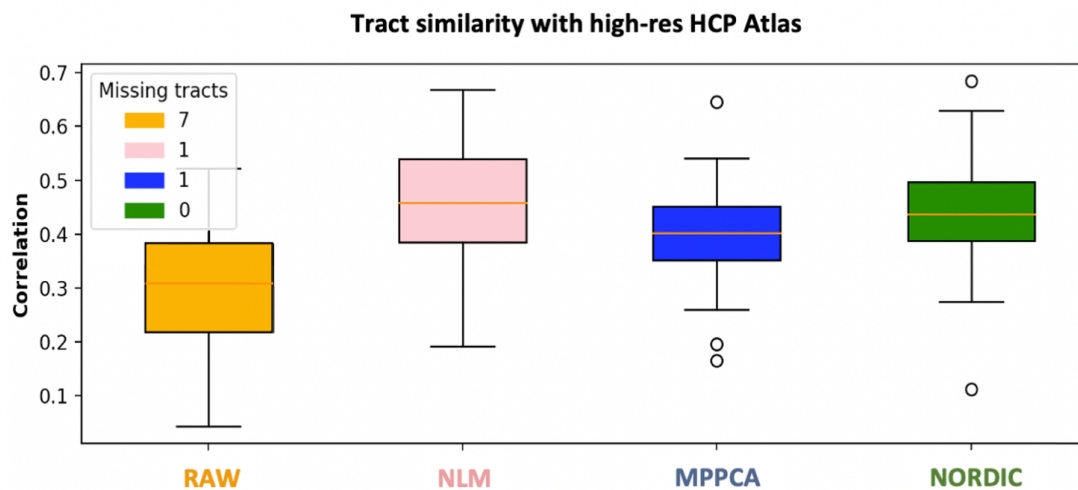


Figure 4.17: Spatial tract correlation of RAW and denoised versions of Dataset C (0.9mm) with the HCP atlas used as a reference for the average population.

4.4.7 Capitalising on increased SNR to reduce scan time

Denoising inherently aims to improve the effective SNR per sample. This opens the question of whether scan time can be reduced, i.e. whether acquiring fewer denoised volumes is equivalent to acquiring many noisy volumes. Fig.4.18 shows a qualitative comparison for fractional anisotropy maps, pre- and post-NORDIC denoising from Dataset C and how these compare against maps from multiple averages Fig.4.18 (top) and from considerably longer scan times in a single-repeat acquisition Fig.4.18 (bottom). In both cases, filtered data provided similar maps with a considerably lower number of volumes.

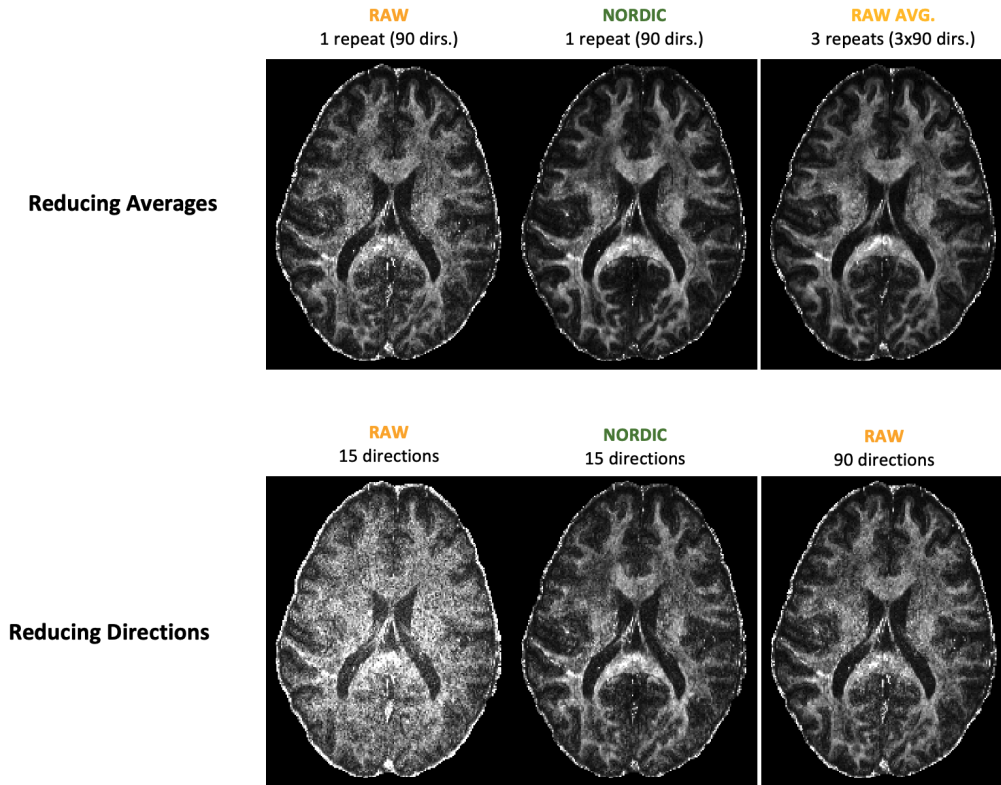


Figure 4.18: Qualitative *FA maps* comparisons (Dataset C). Top row: Denoised single-repeat data vs. multiple-averages. Bottom row: A low number of volumes from denoised data vs a high number of volumes from RAW.

To quantify these improvements, we calculated the spatial correlation between FA and MD scalar maps in WM obtained from subsets of the original data and the FA and MD maps obtained from repeats of the original data. Fig.4.19.A shows the agreement between FA maps in Dataset C (average of the 3 repeats of the dataset used as a reference), where MP-PCA and NORDIC converge faster to the reference than the non-filtered data (in Dataset B, correlation of 0.96 with 1/9 of the data used). However, in this case, NLM deviates

considerably, potentially because of the effects of spatial smoothing. Interestingly, all of them including the RAW data, don't seem to converge to the multiple averages but to a different point (i.e., correlations do not converge at 1). Still, the correlation is high (0.88) and there are potential reasons that could contribute to this deviation, such as geometric misalignments during the image registrations to obtain the average or even the lack of enough volumes to reach the true convergence.

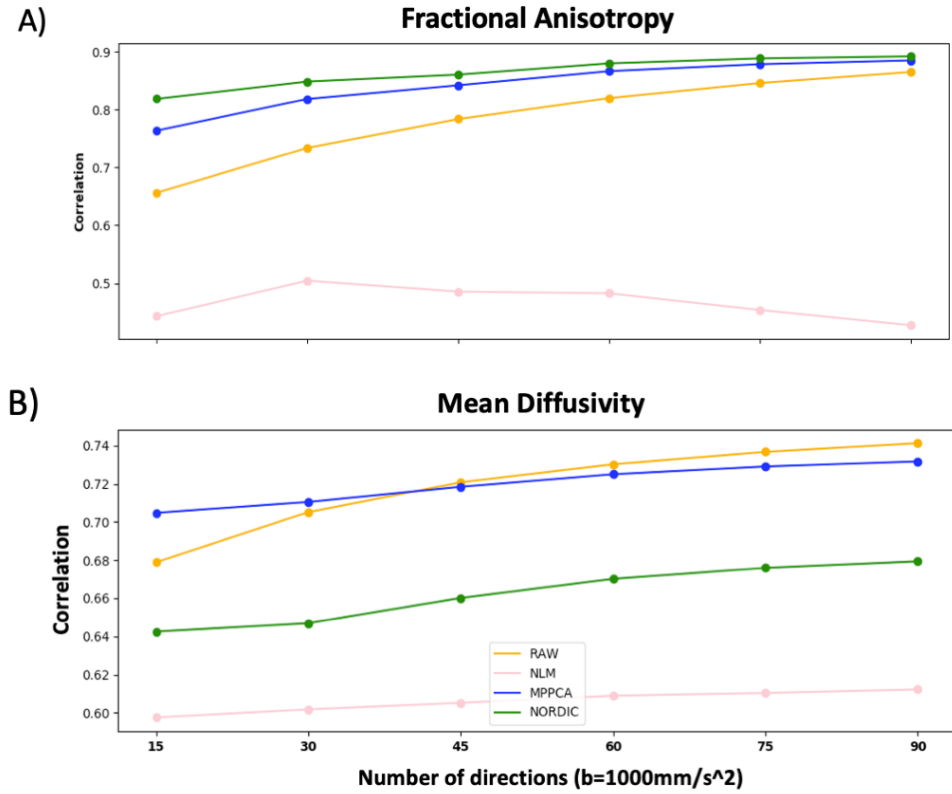


Figure 4.19: Convergence to the multiple-averages (reference) assessed by the correlations in tensor model estimates (DTIFIT) of subsets from Dataset C (0.9mm). Top: Fractional Anisotropy correlations. Bottom: Mean Diffusivity correlations.

However, these deviations from the ground-truth after convergence are more evident and less easily interpretable when looking at the MD (Fig.4.19.B). Particularly, the NORDIC-filtered subsets follow their own trend and converge to a completely different point than RAW and MPPCA-filtered data. This initially puzzling behaviour can be explained by the effect of the noise-floor (which is preserved in MPPCA). As observed in Fig.4.12, the non-filtered data had an elevated noise-floor. As it has been reported before (e.g., [Tax et al., 2021]), averaging images with positive noise-floor will preserve the noise-floor in the averaged image as well (see Fig.4.20 A and B). Averaging is an approach capable to

deal with the random variance of noise but not with the bias introduced by the change of statistical properties (i.e., deviation from Gaussian to non-central-Chi distributions). As a consequence, methods that were not able to address the noise-floor such as MPPCA, reported a similar convergence point to the average as the RAW data in the MD. However, NORDIC reduces the noise-floor, therefore, in regions with very high attenuation, NORDIC returns a larger dynamic range of the signal, capturing higher attenuations and producing higher (and more accurate) diffusivity values (see Fig.4.20.C). As expected, this effect would be much more noticeable at lower SNR scenarios, as in dataset C. This was not observed in the correlations of the FA maps because it is a variance metric, so it is not affected as much by the signal attenuation effects (noise-floor) as the MD.

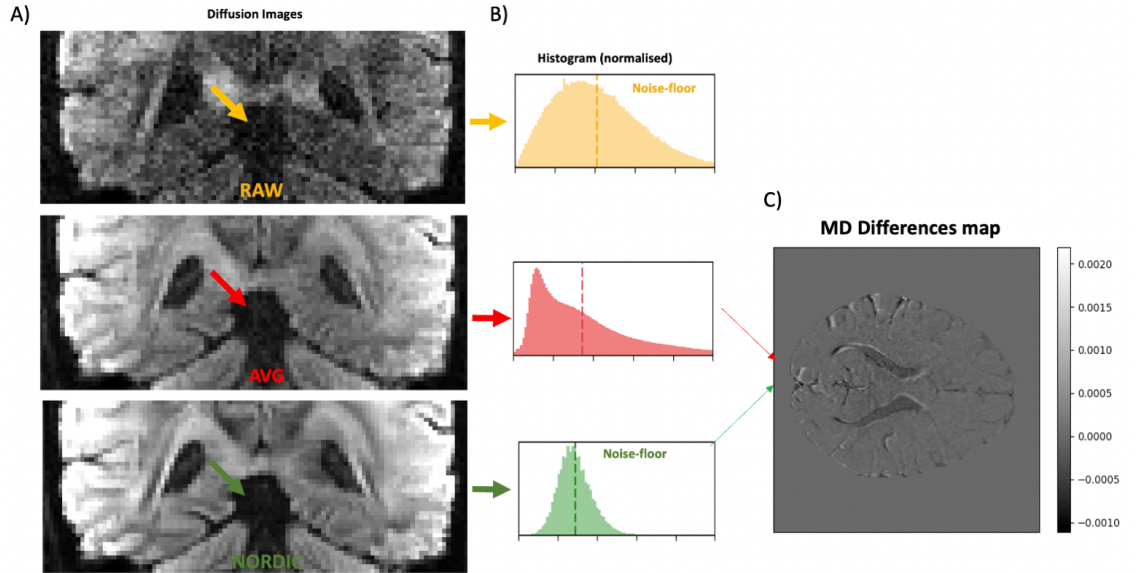


Figure 4.20: Preservation of the noise-floor in the multiple averages - Dataset B. A) Preservation of an elevated noise-floor in the RAW data and in the multiple-averages compared to NORDIC. B) Histograms of the CSF voxel intensities. C) Results from the difference between the MD maps of multiple averages and NORDIC, i.e. $MD_{avg} - MD_{NORDIC}$.

To further explore the above hypothesis (i.e. that differences in convergence are due to differences in the noise-floor), we masked the MD correlation calculation to those regions in the WM that have low FA (i.e., low attenuation). Noise-floor effects are expected to be higher for high FA regions in WM and less noticeable for low FA regions. Doing so for $FA < 0.2-0.3$ (Fig.4.21), we observed a much better behaviour, where NORDIC MD estimates converge to the average more than the other approaches. This provides extra evidence that differences seen before are driven by differences in handling the noise-floor. Taken together, the results in this section suggest that denoising offers the potential to

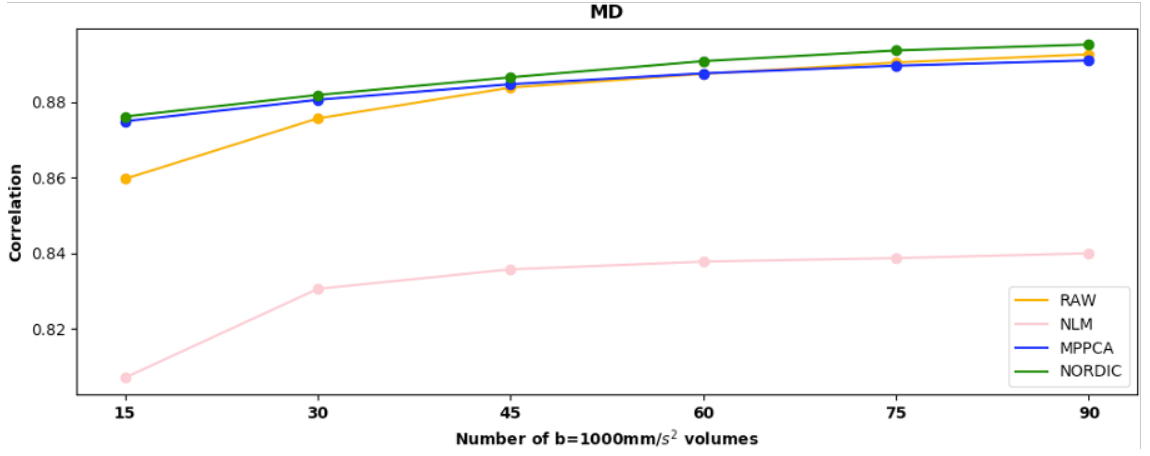


Figure 4.21: Convergence of the MD in denoised datasets to the multiple-averages (reference) assessed by correlations in regions with low attenuation ($FA < 0.2$), where the noise-floor is less likely to rectify the signal.

reduce scan time. The gained SNR can be capitalised to reduce either the number of repeats, or the number of diffusion-sensitising volumes, while maintaining similar performance to full undenoised datasets. Some denoising approaches can offer advantages in terms of reducing scan time up to half or more with similar performance, although this needs to be evaluated for each particular data.

4.5 Discussion

We have devised and presented a novel framework for evaluating diffusion MRI denoising approaches (EDDEN). Although works in structural and functional MRI exist [Heo et al., 2020; Kay, 2022; Mishro et al., 2021], this is the first time to the best of our knowledge that has been comprehensively evaluated how denoising methods in diffusion MRI deal with effects of noise-induced variance (that lead to increased uncertainty), and with effects from the noise-floor (that lead to bias). We have also examined potential loss of resolution, due to the patch-based nature of many denoising methods. We have used bespoke data to evaluate existing approaches at various SNR regimes.

Our results, taken together, suggest **benefits in denoising** dMRI data overall, especially in medium- to low-SNR regimes. However, these benefits seem to increase when denoising **in the complex domain**. For instance, we saw more noise-floor suppression, less spatial smoothing, higher SNR/CNR and better modelling performance when denoising complex data compared to denoising magnitude-only data. This is not surprising; some of the

assumptions on thermal noise made by algorithms (e.g., zero-mean Gaussian distribution) are better fulfilled in the complex domain. Hence, all the algorithms studied managed to reduce the noise-related variance but none of them except NORDIC could address the noise-floor bias. Given that noise variance effects can be mitigated by acquiring more data, but noise-floor effects cannot be mitigated always by longer scan times, this is an important finding on the usability and usefulness of dMRI denoising methods.

Gains in signal quality obtained by denoising can open **opportunities for new applications**. For instance, we have observed how denoising can turn **barely usable data at sub-mm resolution** (very low SNR) to data that high-level analyses can run on, like tractography. Denoising can be potentially applied also to **reduce scan time** needed to achieve a certain SNR-level, either by reducing the need for averages or the need for many dMRI volumes. This advantage can be applied in both clinical routine and research (e.g., post-mortem scans).

Nevertheless, although denoising can be beneficial in a number of ways, caution is needed on how it is applied. Similar to the “No Free Lunch Theorem” in optimization [Wolpert and Macready, 1997], we have observed that the **impact of denoising in dMRI signals is not trivial and the performance can vary considerably** depending on the data, the features being studied, or the algorithm employed. For instance, denoising can result in a loss of spatial resolution and different approaches induce larger or smaller losses. This strengthens the argument that establishing an objective approach is necessary to characterise and evaluate the impact of the above choices. Frameworks such as **EDDEN may help to identify the range of cases within each method performs well or not**.

Very few approaches for objectively evaluating dMRI denoising have been introduced and most of the times they rely on ad-hoc criteria, ranging from qualitative comparisons to indirect model fitting comparisons [Fadnavis et al., 2020; Manjón and Coupe, 2021; Zhang et al., 2017]. Here, we extend previous efforts by considering a more comprehensive set of important criteria that have been missed before. These highlight aspects that have been missed and/or provide a better justification of the source of observed effects. For instance, improvements of denoising in the complex domain over the magnitude domain were al-

ready reported in [Moeller et al., 2021a], especially as the spatial resolution increases. While they attributed these improvements to the *SNR decrease and a more apparent impact of non-zero i.i.d. thermal noise as revealed in a loss of q-space contrast and residual high-spatial frequency modulations*, we here pointed more accurately to the effect of the noise-floor, which also becomes more evident at lower SNR and that variance and residuals provided by MPPCA and NORDIC are similar. We also found detectable spatial smoothing in both PCA-methods (Fig.4.13) despite previous studies suggested the opposite [Fadnavis et al., 2020; Moeller et al., 2021a; Veraart et al., 2016b]. This difference may be due to the evaluation method employed. For instance, in [Fadnavis et al., 2020] the assessment made is just a visual check on the denoised image; in [Veraart et al., 2016b] the evaluation is done in terms of frequency components; in [Moeller et al., 2021a], they used 1) comparison to the multiple-averages (which is inherently smoothed due to averaging residual motion and/or distortions) to estimate the optimal patch size, and 2) loss of angular contrast between dMRI of the same acquisition with same b-values but different gradient orientation. We here rely on covariance estimation within the same volume and comparisons with respect to the nominal resolution in single-repeat non-denoised data measurements, which can avoid biased comparisons. Nevertheless, when assumptions are met, applying the Marchenko-Pastur approach to PCA methods, where the bias-variance trade-off is controlled by the number of dimensions, can provide a threshold good-enough to achieve a considerable reduction of variance without introducing concerning biases (less than 10% for MPPCA and NORDIC compared to RAW data resolution) even operating at very low SNR regimes.

Relatively recent studies have attempted to define a scoring system for image-denoising methods. For instance, [Chow and Rajagopal, 2017] adapted the blind/reference-less image spatial quality evaluator (BRISQUE) [Mittal et al., 2012] to evaluate denoising for structural MRI. In dMRI, [Fadnavis et al., 2022b] recently proposed the Noise Uncertainty Quantification (NUQ) metric, where a Bayesian framework proposed by [Sjölund et al., 2018] is used to develop an uncertainty score for parameter estimates in any model fitted using weighted least squares. Similar to the Ball&Sticks model, the lower the noise in the data, the lower the uncertainty expected; thus, the NUQ is proposed to be used to compare denoising methods. EDDEN aims to provide a considerably more generic

framework, where uncertainty quantification in model fitting is a single component of our approach. Given the challenges in uncertainty quantification (and in separating different uncertainty sources, as seen in the Background chapter), this may be insufficient to fully characterise the effects of denoising in the signal as it only reflects potential noise-related variance changes. Summary scores need to inform about variance, biases and error metrics in different domains of the signal, as the features proposed in this work. EDDEN’s proposed comprehensive evaluation process is compatible with the view expressed previously [Veraart et al., 2016b] that advocates valuations mainly at the signal level instead of studying their accumulated effect on downstream analysis, as the rest of pre-processing and/or modelling stages may be altering/masking the modifications performed purely by denoising. Here, we proposed specific steps to evaluate gains in signal quality, preservation of noise and expected signal properties, and covariance or spatial smoothing induced.

In our study, we have focused on a big source of bias caused by the signal rectification produced by the noise-floor. Previous studies have defined errors or deviations in terms of a so-called **”bias-variance trade-off”**, considering bias sources introduced by the denoising algorithm itself, as suggested in [Kay, 2022] for MRI. That is, methods that focus on reducing the variance may be indirectly introducing algorithmic biases (in addition to the noise-floor signal bias). We saw some of these effects in our study. For instance, we observed that NLM reduced the most variance among the methods studied at the expense of inducing excessive blurring in the image and, therefore, a higher risk of introducing biases (e.g., partial volume effects). The effects of this can be observed in Figures 4.19 and 4.20 where despite reducing variance should push NLM towards the ground-truth, it deviates considerably from it or other methods that preserve the noise-floor, such as MPPCA.

The challenge in defining a ground truth

An important finding from our analyses was the challenge to define a good in-vivo ground-truth. In the context of recovering SNR, a ground-truth can be defined by multiple averages of the same subject with the same acquisition parameter, so the variability originated from noise can be averaged out while the signal information remains. This can work under the assumption of no major motion effects and that a set of measurements corrupted by thermal noise will converge to the true signal across repeated acquisitions because they

are governed by additive, zero-mean, symmetric noise, i.e. measurements can have high variance but are unbiased. However, we have seen that magnitude images do not necessarily follow these properties, due to the noise-floor effects that can be difficult to remove. Particularly at low SNR, noise-floor-contaminated images will result in a noise-floor contaminated average, thus preserving the bias in the powder average, preventing accurate signal modelling [Dietrich et al., 2001; Jones et al., 2013]. This can lead to erroneous validation of denoising methods if such multiple-averages are used as ground-truth, opening the question on how to define a good reference. For instance, for a number of evaluations denoising in the magnitude domain looked closer to the magnitude "gold-standard" than denoising in the complex domain, as both of them preserved noise-floor effects. A potential solution to these discrepancies would be averaging in the complex domain instead of magnitude images. This will preserve the assumptions about noise properties and avoid the noise-floor bias [Eichner et al., 2015]. This approach has not been evaluated here, but it is a natural extension of the work presented in order to avoid the issues we identified with a biased "gold-standard" average.

In fact, this lack of ground-truth forces also the need for supplementary metrics beyond the error or the bias-variance trade-off, as there is no actual reference to compare with and calculate the error. As we have seen in Fig.4.19, the bias in the multiple-averages at very low SNR (i.e., lack of proper reference) was leading to wrong conclusions when purely relying on error metrics. In fact, this lack of "ground-truth" dataset further propagates in defining ground-truth values for quantitative model estimates. For instance, it is challenging to assess results at very high resolution that deviates from standard results obtained at lower resolutions. Findings such as having a higher rate of crossings at higher resolution could mean either increasing sensitivity or a problem of overfitting. All in all, it is not clear that the quality of downstream analysis results necessarily mirrors the performance of these methods on low-level data metrics, such as the amount of rectification on voxel-wise signals. An example of this is the high rates of crossings provided by NLM (4.15), despite presenting clear rectification in the signal. The definition of domain-specific quantitative scores for the evaluation of denoising is needed, as the ones commented on at the beginning of the Discussion section of this chapter (e.g. BRISQUE or NUQ). However, according to all the above, this may be insufficient to fully characterise the effects of de-

noising in the signal as it only reflects potential noise-related variance changes. Summary scores need to inform about variance, biases and error metrics in different domains of the signal, as the features proposed in this chapter. However, again, the lack of ground-truth impedes the design of quantitative metrics for all of them and more work is needed.

Denoising as Filtering vs Deterministic Prediction

Traditionally, denoising can be considered as a filtering operation, where an algorithm identifies components of the measured signal that represent noise and filters these out from the true signal. A more recent view is that denoising can be based on forward parametric or non-parametric model predictions (i.e. effectively deterministic regressions), e.g., P2S [Fadnavis et al., 2020]. In such a scenario, the measured signal can be totally discarded and surrogated instead by a prediction that depends on many other voxels except the one being predicted. This seems a more invasive approach, which produces by definition, "noise-free" model-predicted data. This type of approach caused issues in our evaluations, as existing tools expect input measurement-like data rather than deterministic model predictions. It is very likely that with further development on approaches like FSL's EddyQC will work with P2S-denoised data, but a philosophical question arises: *is this really denoising, given that any deterministic model regression will provide "denoised" data? Why someone would use the P2S-model to denoise, rather than the multi-shell Gaussian Process model in Eddy or even the DTI model predictions that will be equally "denoised"?*

More formally, most of ML-methods are oriented to achieve high predictive performance rather than providing explanatory modelling [Shmueli, 2010]. These models aim to minimize the error (a combination of bias and variance) so unless predictions are perfect, there is a potential that bias exists for denoising techniques trained by predictive performance. An intuitive example can be data imputation (not quite different of what is done in P2S effectively). Let's suppose we delete a small region of a picture and in-painting techniques are used to fill in the missing region. While we are likely to obtain a reasonable-looking image that generally conforms to natural image statistics, it is obvious that this is no substitute for actual measurements. Had there been a specific object in the deleted region, it is likely that the in-painting approach would miss this completely and instead fill the

region with general texture priors by, e.g. drawing a river where there was a small village. The imputation would likely incur a massive bias that can have an enormous impact when modelling medical signals. Arguably, this type of risk could be assumed if the generative approach is achieving levels of performance or addressing a task that any other method can. However, the observed performance of these methods in dMRI is in the same range as the alternatives in the state of the art and they are not addressing any novel challenge (not even the noise-floor bias). Note that similar techniques applied in reconstruction with aims for faster acquisition (and less noise introduced in the final image) may not fall under the same consideration, as they are learning instead how to *select* the optimal (and actually measured) components.

All in all, one could ask whether any dMRI model reviewed in the Background chapters can be seen also as a denoising approach. In fact, these can be seen arguably as more advantageous, as they are voxel-wise (so no patch-based smoothing expected) and include biophysical validated constraints. Nevertheless, sophisticated or not, we have observed that these forward predictions would become tied to a specific model and its underlying assumptions. In that respect approaches like P2S (or Eddy’s Gaussian Processes) that are agnostic to any particular biophysical model may provide a more unbiased forward prediction for data. However, the question still remains on whether they can be considered as denoising ”filters” and we expect interesting discussions around this topic in the future.

Limitations and Future Work

Conclusions from our work have certain limitations that open possibilities for future extensions. Firstly, we have used a specific biophysical model (Ball&Sticks) to evaluate modelling performance. Findings in this work could vary due to the specific model choice; we, however, expect that the main trends remain regardless of the biophysical model used, as some of them are based on evaluations before modelling, i.e. at the signal level. Secondly, the denoising methods used were ”off-the-shelf” (default) implementations. Potential improvements may be found with dedicated fine-tuning of parameters that suit better the datasets or evaluations provided. This is in fact one of the applications of EDDEN, to be used as a framework for denoising optimisation. Hence, there are many possible experiments that can be done with the actual level of development of the ED-

DEN framework, such as evaluating a more comprehensive set of algorithms, different flavours or exploring whether hybrid approaches that combine two or more methods are advantageous (e.g., as suggested in [Mishro et al., 2021]). Thirdly, we applied denoising always before distortion correction and did not explore the potential interaction between the two. It is not clear yet whether denoising should be performed before or after motion and/or distortion correction. While the signal is assumed to be low-rank in methods like MPPCA or NORDIC, the reality is that raw signal is not only mixed with thermal noise but also with random subject motion, volume-to-volume variability, signal dropouts and other Eddy current distortions. For instance, patch-based denoising can have patches comprising of signal piled up/stretched from/to regions outside the considered patch due to susceptibility-induced distortions. Whether distortion correction before or after denoising helps performance is something to be explored. This can be linked also to the question of whether denoising can help in **reproducibility** of results (e.g. test-retest experiments [Schilling et al., 2021]) and assist in data harmonisation.

Amongst these future opportunities, another aspect that can be evaluated is whether signal transformations done to reduce the noise-floor bias after magnitude calculation to meet statistical assumptions (e.g., Koay’s method of moments [Koay and Basser, 2006; Koay et al., 2009] or the Variance Stabilization Transform (VST) approach proposed in [Ma et al., 2020]) or the introduction of terms that compensate for the noise-floor (e.g. modelling assuming Rician noise instead of Gaussian) can provide any advantage at all regarding signal rectification, as variations applying it has been found for some models by previous works (e.g., [Hutchinson et al., 2017]). Preliminary experiments suggest they don’t; most models already take into account the noise, so these approaches only try to improve model fitting by adding some components in the model (and, hence, small shifts in the scalar parameters can exist) but cannot recover the signal already lost under the rectification, strengthen the idea of denoising in the complex space.

All of the above can be evaluated with the current implementation of EDDEN. However, a desired extension is to include the framework in fully **quantifiable scores**, which is not available currently. This would help in the endeavour of comparing the performance of different algorithms or finding the optimal parametrisation of each method. Given the

lack of ground-truth and the complexity of quantifying some parameters, a potential line of research is the application of unsupervised ML to learn no-reference IQA scores (e.g., [Chow and Paramesran, 2016; Lin et al., 2020; Stepień et al., 2021]), such as the Modified-BRISQUE for MRI [Chow and Rajagopal, 2017] commented previously. This type of score can offer a more sophisticated alternative to SNR or CNR, where aspects such as spatial resolution loss or noise-floor could be included. Although the end-goal of EDDEN is to evaluate real measurements in order to reflect the true complexity of the brain and possible interactions during acquisition (anatomical complexity, scanner inhomogeneities, motion, etc.), simulations can be added to the framework to help testing some of these features (e.g., covariance induced by methods, signal and model parameters ground-truth, etc.).

Chapter 5

Mapping uncertainty in dMRI using MCMC: A hybrid approach

5.1 Introduction

In the previous chapter, we saw how denoising can be beneficial for increasing data quality and improving modelling performance. In this chapter, we are concerned with mapping uncertainty (which can be noise-induced or not) while estimating the parameters of biophysical models. We will be relying on a classical Bayesian inference (i.e. MCMC) in this chapter and we will explore in the next chapter whether alternative options based in modern ML approaches can offer advantages.

As we saw before, when we fit a model to dMRI data, the task is to estimate the values of the model parameters ω that best explain the measured images Y . This can be done either deterministically or stochastically. The latter option is an intriguing choice: instead of estimating a single value for each model parameter, we can estimate the conditional distribution $\pi(\omega|Y)$, also known as the **posterior** distribution of the parameters given the data. This distribution provides an inherent characterisation of **uncertainty** for the estimated model.

Bayes theorem provides a simple rule to relate this posterior distribution with the **likelihood** function of the data, $\pi(Y|\omega)$, as well as the prior distribution of the model parameters $\pi(\omega)$. Bayesian inference relies on this rule and offers a suitable framework to

address this estimation problem. However, this also includes the calculation of a quantity $\pi(Y)$, the probability of the data after integrating out all possible model parameter values (see Equation 3.1), also known as model evidence. In the general case, this can be a high-dimensional integral (with dimensions as the number of model parameters) not possible to solve analytically most of the times. To overcome the difficulty of not having access to the posterior distribution, different approaches can be used to provide an approximated estimation of the posterior.

Sampling-based methods are a popular option for the estimation of the posterior and **Markov Chain Monte Carlo (MCMC)** [Gilks et al., 1995], amongst others, is one of the most commonly-used. In MCMC, instead of providing an analytic expression of the posterior density, samples from it are produced without the need of knowing/calculate the evidence. As the evidence is a normalising constant, the target posterior distribution can be sampled up to a proportionality constant. MCMC has been extensively applied in dMRI biophysical modelling (e.g., in [Behrens et al., 2003, 2007; Kaden and Kruggel, 2012; Pisharady et al., 2018; Sotiropoulos et al., 2013a, 2016]). Despite the strength of MCMC, it can be very computationally expensive in its most basic form, (Random-walk MCMC), where independent explorations for each model parameter are performed in the parameter space in an iterative fashion. This can be problematic to the degree that bespoke heavily-parallelised computational frameworks, such as on GPUs ([Hernandez-Fernandez et al., 2019]), are deemed necessary to use these approaches in practice.

Popular alternatives to MCMC include **Variational Bayes (VB)** [Blei et al., 2017] and **Approximate Bayesian Computation (ABC)**. In VB, rather than iterative sampling as in MCMC, a global approximation to the posterior is done under the constraints of an analytical functional form (typically using Normal distributions) [Chappell et al., 2009; Kaden et al., 2008]. Although it is considerably faster than MCMC, the posterior distributions provided may not capture the true uncertainty (they are less specific) and they also impose assumptions on priors that can make it difficult to incorporate improper priors, such as shrinkage Automatic Relevance Determination priors, that provide effective ways for Bayesian model selection [Behrens et al., 2003]. On the other hand, ABC relies on a generative forward process that allows to address problems where even the likelihood

is not tractable, although there exists still a dependency on simulations. ABC-inspired extensions will be the subject of the next chapter.

Here, we revisit the classical Random-walk MCMC paradigm, as initially introduced into the dMRI field by [Behrens et al., 2003] and [Behrens et al., 2007]. We propose and evaluate new algorithmic designs that improve efficiency, both for MCMC sampling and computing. Specifically, we move away from sampling each model parameter independently of each other, to a block-update MCMC paradigm that allows inference of multiple or all model parameters simultaneously. To achieve this, we devise ways to incorporate local approximations (Laplace approximations) and parameter covariances into the MCMC. We demonstrate the utility of this approach in performing more efficient sampling for an exemplar dMRI biophysical model, the Ball&Sticks.

The chapter is organised as follows: the following section overviews the theory on the Markov-Chain Monte Carlo and the Laplace approximation. In Methods, we propose three different algorithmic designs for block updates: 1) an Independence sampler, 2) a Random-Walk Block-proposal MCMC, and 3) a Hybrid design of the block and individual update of parameters. We also provide details about the Ball&Sticks model implementation used for this chapter, the evaluation metrics, and the synthetic and real used for them. Results follow, where we evaluate the different designs in simulated and in-vivo brain data. The chapter ends with a discussion about the results obtained, contributions and limitations, and the potential future steps of this line of research.

5.2 Theory

5.2.1 Markov-Chain Monte-Carlo (MCMC) sampling

The Markov-Chain Monte-Carlo (MCMC) algorithm is the result of combining different concepts to achieve the sampling of a target density distribution $\pi(q)$:

- **Monte-Carlo:** Formally, these are non-deterministic numerical methods used to approximately calculate a definite integral. In Bayesian inference, they discretise the problem of estimating the evidence, converting the integration into a summation (see eq.5.1). For this discretisation, pseudorandom numbers are sampled from a

known probability distribution.

- **Markov-Chain:** This is a stochastic model describing a sequence of events, in which the probability of each event depends only on the state attained in the previous event. To perform a stochastic exploration of the parameter space, the MCMC uses a Markov chain C .

In the problem addressed in this chapter, the target density distribution $\pi(q)$ is the posterior distribution of parameters ω given the observed data Y , i.e. $\pi(q) = \pi(\omega|Y)$. Hence, a Markov-Chain Monte-Carlo method relies on a sampler where the next (ω_{new}) position depends only on the current position (ω). To generate the Markov chain C , a function $f(\omega_{new}|\omega)$ is applied sequentially to move stochastically from the current position ω to the new point ω_{new} . In the context of MCMC samplers, this function is generally known as the *proposal distribution*. In the simplest implementation, with the unique constraint of preserving the target distribution $\pi(\omega|Y)$, C is expected to generate a **random walk** across the regions of high probability, from which we can compute the expectation $E_{\pi(\omega|Y)}$. Hence, the MCMC sampling process can be expressed as:

$$E_{\pi(\omega|Y)} = \int_{\Omega} \pi(\omega_{new}) \cdot f(\omega_{new}|\omega) \cdot d\omega_{new} \rightarrow E_{\pi(\omega|Y)} = \frac{1}{N} \sum_{i=1}^N f(\omega_{i_{new}}) \quad (5.1)$$

The aim is to sample from regions that generate the largest contributions to the expectation $E_{\pi(\omega|Y)}$ rather than computing the target distribution analytically. The sampling should be concentrated in parts of the parameter space that are relevant for the target density $\pi(\omega|Y)$; this is called the **typical set** and denotes the set of parameter space where the stochastic exploration is made by the Markov-Chain.

Monte-Carlo's estimation has an absolute error proportional to $1/\sqrt{N}$ (Central Limit Theorem). This theoretically guarantees that, given *enough* computational resources or time (to do enough iterations), the chain will explore completely any target distribution regardless of where the chain has been initialized: $\lim_{C \rightarrow \infty} E_{\pi(\omega_{new}|Y)} = E_{\pi(\omega|Y)}$. How many are *enough* cannot be easily defined and will depend on many factors, such as the complexity of the distribution, the precision of the result required, the implementation of the MCMC, etc. Therefore, an open problem is to determine how many steps are needed to converge to the target distribution within an acceptable error.

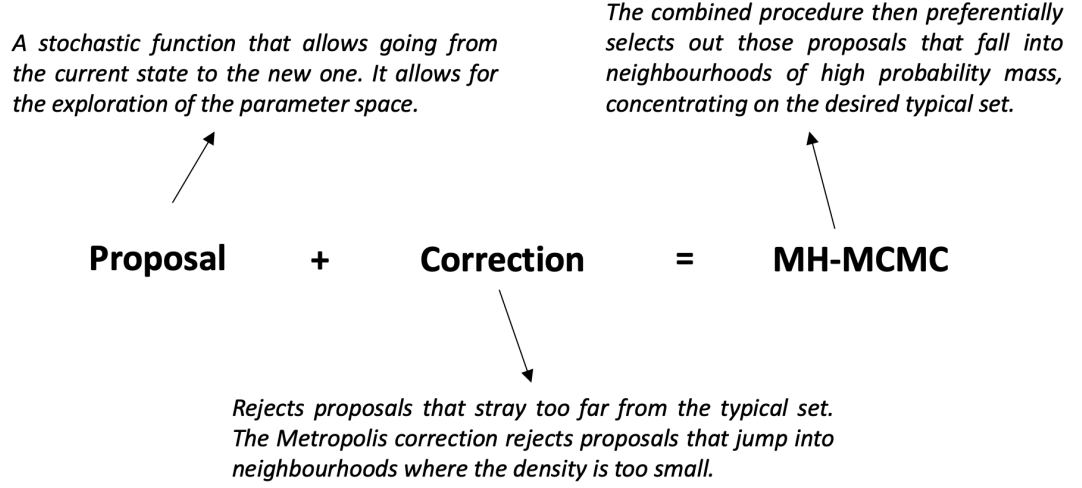


Figure 5.1: *The Metropolis-Hastings MCMC algorithm comprises two main steps: the proposal values and a rejection-acceptance step that corrects such proposals to continue the explorations in the typical set.*

Different MCMC **step sampling methods** follow different approaches to perform the stochastic transitions in parameter space; throughout this chapter, the **Metropolis-Hasting (MH) algorithm** [Hastings, 1970; Metropolis et al., 1953] will be used to sample from posterior distributions of the form $\pi(\omega|Y)$ when estimating the parameters ω given the measurements Y .

The MH algorithm introduces two main steps into the MCMC sampling [Hastings, 1970]: 1) a proposal distribution $\pi(\omega_{new}|\omega)$ to build the stochastic Markov chain, and 2) a sample acceptance/rejection based on the Metropolis-Hasting's criteria. The original MH-MCMC algorithm, also called **Random Walk Metropolis (RWM)**, utilizes a univariate Gaussian distribution as the proposal, centred on the last accepted ω value with random jumps given as a function of the standard deviation σ^2 :

$$\pi(\omega_{new}|\omega) = \mathcal{N}(\omega, \sigma^2) \quad (5.2)$$

Regarding the acceptance criteria, the probability of accepting the new value ω_{new} as a sample of the target distribution depends on the quantity r :

$$r = \min \left(1, \frac{\pi(\omega|\omega_{new}) \cdot \pi(\omega_{new}|Y)}{\pi(\omega_{new}|\omega) \cdot \pi(\omega|Y)} \right) \quad (5.3)$$

Being a Gaussian, $\pi(\omega_{new}|\omega)$ is a symmetric distribution in the random-walk setup (i.e., the likelihood of jumping to ω_{new} from ω is the same as the likelihood of jumping back to ω from ω_{new}). Therefore, the proposal densities in eq.5.3 cancel because $\pi(\omega_{new}|\omega) = \pi(\omega|\omega_{new})$, leaving for the acceptance criteria the compact form of:

$$r = \min\left(1, \frac{\pi(\omega_{new}|Y)}{\pi(\omega|Y)}\right) \quad (5.4)$$

Then, if $u < r$ with $u \sim \mathcal{U}(0, 1)$, the proposed sample ω_{new} is accepted as a sample from the target distribution and the algorithm repeats itself by starting at the new sample ω_{new} ; otherwise, the sample is rejected and the algorithm starts in the next iteration at the last accepted sample ω . The algorithm is self-repeating, so it can be carried out as long as required. In practice, the total number of samples needed from the target distribution is decided in advance.

Finally, it is noteworthy that, for numerical reasons, the target density in eq.3.1 can be also expressed in logarithmic terms. In such case, the negative log-posterior distribution is also known as **energy function**, \mathcal{E} (i.e., the unnormalized negative log-density):

$$\mathcal{E} = -\log(\pi(\omega|Y)) = -(\log(\pi(Y|\omega)) + \log(\pi(\omega))) \quad (5.5)$$

Taking the log allows to turn multiplications of many quantities smaller than 1 (e.g. the priors) to addition, therefore ensuring numerical stability. All in all, the logarithm is a monotonic function and does not change the properties of the density functions.

Dependence on initialization and burn-in

Under ideal conditions, Markov chains explore the target distribution in three distinct phases: 1) First, the Markov chain has to converge to the typical set, i.e. to parts of the parameter space relevant to the target distribution; 2) Once the Markov chain finds the typical set, the bias decreases rapidly and the estimators become much more accurate; 3) As the Markov chain refines its exploration of the typical set, it gradually reduces estimator errors towards zero.

Depending on the initial value of the parameters and the covariance of the proposal distri-

bution, convergence could take a very long time. Parameters can be initialised randomly from the prior distribution or through a deterministic fitting method that helps to start close to an approximate typical set and accelerate convergence.

Finally, if the Markov chain is run for an infinite amount of time, the effect of the initial values decreases to zero. In practice, we only have time for a limited number of samples. Hence, the samples generated by the Markov chain during the initial walk for identifying the typical set can bias the estimators. To ensure convergence to the target distribution, it is common practice to discard a chosen number of early samples (**burn-in period**). This is the time considered necessary for the algorithm to stabilize and converge to the target posterior distribution (i.e. to go from step 1 to 2).

5.2.2 Mixing and sampling efficiency in random-walk MCMC

Because of its simplicity and ease of implementation, Random-walk Metropolis (RWM) is still popular in many applications. Unfortunately, implementation simplicity comes with poor scaling behaviour for both many dimensions and complex target distributions.

When the number of dimensions of the target distribution rises, RWM tends to suffer from the *curse of dimensionality*. The volume of the typical set gets squeezed in the parameter space and the regions of relevance for the target distribution tend to stretch and get lost in an increasing volume of space that gives little contribution to the desired integral. I.e. it becomes more likely to obtain a point outside the typical set in the new RWM proposals, leading to many proposed samples being rejected by the MH criteria. As a consequence, the resulting Markov chain experiences poor mixing, traversing slowly the parameter space and proposing many correlated samples, which results in biased estimations.

This is a problem of mixing and sampling efficiency. Assuming the model is correctly specified (excessive auto-correlation may indicate problems with model specification also) and that we have reached the stationary target distribution (i.e., we are at step 2 described in sec.5.2.1), there are a couple of strategies commonly used to mitigate most of these issues: thinning and Adaptive sampling.

Thinning to reduce sample auto-correlation in Markov Chains

Thinning is useful to reduce the auto-correlation of an already obtained chain. A chain can be thinned by keeping every t -th sampled drawn. E.g., thinning of $t = 20$ means keeping only 1 from every 20 consecutive samples (as they are likely correlated and, therefore, offering similar information).

Note that, although it is a common trick applied to already sampled distributions and it has been found to be convenient, this implies throwing away a $\frac{t-1}{t}$ fraction of all the posterior samples generated. This is a considerable waste of computational time, i.e., it is not improving the efficiency of the sampler nor reducing the number of jumps but only the number of already obtained samples.

Adaptive Metropolis

The acceptance rate and estimator errors are closely related to the sampling efficiency of a Metropolis chain. For RWM, a high acceptance rate means that most new samples occur *close* to the current data point, i.e. the Markov chain is moving rather slowly and not exploring the full parameter space. On the other hand, a low acceptance rate means that the proposed samples are often taken far from the target distribution and, consequently, rejected. In both cases, obtained samples are likely to be autocorrelated, producing biased estimators.

The scale in the proposal distribution effectively controls this acceptance probability. Ideally, an efficient Metropolis sampler has an acceptance rate that is neither too high (it is not moving around the mode only) nor too low (it is not only producing samples from outside of the posterior). With such optimal design, efficiency is achieved by reducing the total number of jumps needed to estimate the posterior distribution. For instance, [Gelman et al., 1997] showed that if both the target and proposal densities are normal, and the width of the proposal distribution is correct, the optimal acceptance probability for the Markov chain should be around 0.45 in a single dimensional problem (for unimodal Gaussian distributions), and asymptotically approaches 0.234 in higher dimensions. If the initialization is not good enough, a strategy to reach an optimal covariance is by introducing an Adaptive Metropolis algorithm, where the proposals are scaled by a correcting

factor λ_L .

$$\pi(\omega_{new}|\omega) = \omega + \lambda \cdot \mathcal{N}(0, \sigma^2) \quad (5.6)$$

There are different strategies to update the value of λ . For instance, for normal proposal distributions, [Haario et al., 2001] proposed to automatically adapt the width of the proposal distributions of each parameter independently, during the burn-in phase, by a λ_L as a function of the observed acceptance rate and the target acceptance rate:

$$\lambda_L = \frac{1 + \text{accepted}}{1 + \text{rejected}} > 0 \quad (5.7)$$

In particular, this λ_L is designed to have an acceptance ratio close to 50% in the RWM. If the acceptance ratio in a given tuning loop L is too low, new samples come from the tails or outside of the acceptance interval. Hence, the scale of λ is reduced, reducing the step-size in new jumps and avoiding new proposals trying to exit the high probability regions. On the other hand, if the acceptance ratio is too high, most of the samples may be obtained from the neighbourhood of the mode. In this situation, the posterior distribution is not being fully sampled (or it will require many steps for accurate results), so λ_L will be increased providing a larger step-size in new jumps to reach regions in the tails. Thus, the advantage of the Adaptive Metropolis is that it adjusts the proposal variances **automatically**. The 1 added in both numerator and denominator can be set to any small value $\delta > 0$; it is included just for numerical reasons, to prevent λ_L becoming 0 if the number of acceptances or rejections is 0.

5.2.3 Model selection and automatic relevance determination

In many cases, it is useful to establish the relevance of some model parameters in a nested model fitting setting, i.e. where a parent model can be reduced to simpler models with a subset of the original parameters. Deciding which of these models is supported by the data is a model selection problem. In dMRI multi-compartment biophysical models, it is often needed to establish whether the addition of a compartment to the model is supported by the data. For instance, when estimating fibre orientations, different voxels may have a different number of fibres. In such scenarios, we need a technique to decide how many model parameters/compartments are relevant and therefore adjust the parameter space

accordingly.

A technique used in Bayesian inference serves this aim automatically and is known as **Automatic Relevance Determination (ARD)** [MacKay, 1995; Wipf and Nagarajan, 2007]. ARD requires the use of a certain type of prior on the parameter whose relevance needs to be determined. This prior distribution can take a number of forms, with one of the simplest being a Gaussian distribution with zero-mean but unknown variance (see appendix B.6).

$$\pi(\omega)_{ARD} \sim \mathcal{N}(0, \sigma_{ARD}^2) \quad (5.8)$$

The variance σ_{ARD} is then estimated as a hyper-parameter. If there is no evidence in the data for the existence of the original parameter ω , this variance term will shrink, giving a very strict prior around zero and forcing ω to zero (hence turning off effectively the part of the model linked to that parameter). However, if ω is supported by the data, σ_{ARD}^2 will obtain high values, making the prior wide (at the extreme a uniform distribution). This will allow freedom for ω to take any value supported by the data.

By using these sparsity-induced priors, we can effectively perform an online model-selection in the presence of multiple compartments in the model, as any unnecessary parameters are automatically forced to zero. By setting ARD priors to the volume fraction parameters of these compartments, we a-priori penalise the existence of multiple compartments and let the data likelihood contribution point towards adding compartments for improved posteriors. As indicated in [Behrens et al., 2007], the expression of the Gaussian distribution based ARD applied to f_i can be simplified to:

$$\pi(\omega)_{ARD} \sim \frac{1}{\omega} \quad (5.9)$$

Using ARD is different from other model selection techniques, as it does not need fitting different models to the data separately, and comparing them on the basis of a metric reflecting data fit and model complexity. As long as the modes are nested, when we use ARD we fit the more complex model, but ensuring that parameters that are not supported by the data do not contribute to the likelihood.

5.2.4 Improving random-walk MCMC: Block proposals and the Laplace approximation

One of the greatest challenges in MCMC sampling is achieving good mixing of the chains. A number of factors determine the performance of a Metropolis sampler. For instance, when sampling each parameter individually with a univariate proposal in multidimensional problems, as in the RWM, the chain may mix too slowly because of marginal target distributions being too narrow and also because of ignoring structure and covariances in the parameters. Hence, it can take a long time for the chain to fully explore each dimension alone. In such multivariate parameter model, if all parameters were proposed with one **joint proposal distribution**, the sampling and the evaluation of the likelihood would occur for all of them once at the same time. Then, if accepted, this could accelerate by xK the MCMC algorithm, where K is the number of parameters in the joint proposal.

The challenge here is to find such a joint proposal (called block-proposal for the rest of the chapter) that resembles the actual posterior distribution of the parameters, as that is the distribution used to sample from in order to approximate the posterior. That is, it should mimic both the mode and the covariance. The Large sample theory states that the posterior distribution of a set of continuous parameters can be well approximated by a multivariate normal distribution (see [Gelman et al., 2015], Appendix B, and Section 7.4 of [Schervish, 1995]). That is why a multivariate normal block proposal often works well in practice.

$$\pi(\Omega_{new}|\omega) = \mathcal{MVN}(\Omega, \Sigma) \quad (5.10)$$

where Ω denotes a vector with the modes of the parameters, $\Omega = (\omega_1, \omega_2, \dots, \omega_k)$, and Σ is the covariance matrix. However, making this multivariate proposal adaptive is not as straightforward as in the univariate case given the interaction between parameters in the covariance matrix. Providing a good initialisation or learning a proper covariance Σ that produces a good ratio of acceptances and rejections can be a challenge, especially when the model parameters vary by more than a few orders of magnitude or if the posterior correlations are high. The simplest approach is to initialize Σ by using an identity matrix multiplied by a factor. However, this can take a number of tuning phases before the

proposal distribution is tuned to be optimal. Numerical optimization routines, such as the quasi-Newton method, can be also employed to find a starting covariance matrix in a more efficient way. Deterministic non-linear optimization is performed on the joint posterior distribution to find its mode, and the covariance matrix is a quadratic approximation at the posterior mode. However, there are cases where the optimization could fail to find a covariance matrix that is positive definite. An alternative for initializing this block covariance is using a Laplace approximation [MacKay, 1992; Tierney and Kadane, 1986].

The Laplace approximation

The Laplace approximation (LA) is an integral approximation method [MacKay, 1992; Tierney and Kadane, 1986]. In a Bayesian framework, it provides a way of locally approximating a density whose normalization constant (evidence) cannot be evaluated. More specifically, the LA entails approximating the joint posterior $\pi(\omega|Y)$ by a multivariate normal distribution, which provides an analytical solution for the integral. Generally, the LA relies on approximating any function $f(x)$ by a second-order Taylor series expansion around the value x^* :

$$f(x) \approx f(x^*) + f'(x^*)(x - x^*) + \frac{1}{2!}f''(x^*)(x - x^*)^2 + \dots \quad (5.11)$$

Therefore, to approximate the log-density, $\text{Log}(\pi(\omega|Y))$, around the Maximum A Posteriori (MAP) estimate $x^* = \hat{\omega}$:

$$\log(\pi(\omega|Y)) \approx f(\hat{\omega}) + f'(\hat{\omega})(\omega - \hat{\omega}) + \frac{1}{2}f''(\hat{\omega})(\omega - \hat{\omega})^2 = f(\hat{\omega}) + \frac{1}{2}f''(\hat{\omega})(\omega - \hat{\omega})^2 \quad (5.12)$$

as $f'(\hat{\omega}) = 0$ (because $\hat{\omega}$ is the maximum). The resultant expression above can be reshaped into the expression of a Gaussian density, which is effectively the Gaussian approximation $q(\cdot)$ to the log-posterior proposed by the LA:

$$\begin{aligned} q(\cdot) \approx \log(\mathcal{N}(\omega|\mu, \Sigma)) &\approx \log(2\pi)^{-1/2} \cdot \Sigma^{-1/2} \cdot \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right) = \\ &\frac{1}{2} \log(\Sigma) - \frac{1}{2} \log(2\pi) - \frac{1}{2} \Sigma^{-1}(\omega - \mu)^2 = \frac{1}{2} \log\left(\frac{\Sigma}{2\pi}\right) - \frac{1}{2\Sigma}(\omega - \mu)^2 \end{aligned} \quad (5.13)$$

Thus, the LA results in a local Gaussian approximation to the log-posterior centred around the estimated mode $\mu = \hat{\omega}$ and with negative curvature at the mode $\Sigma = -f''(\hat{\omega})^{-1}$.

This technique can be used for reasonably well-behaved functions that have most of their mass concentrated around a small area and has very rapidly decreasing tails, i.e. this approximation works well if the posterior distribution is similar to normal distribution but it will suffer if the posterior is not bell-shaped distribution e.g., multimodal or highly skewed (see Fig.5.2).

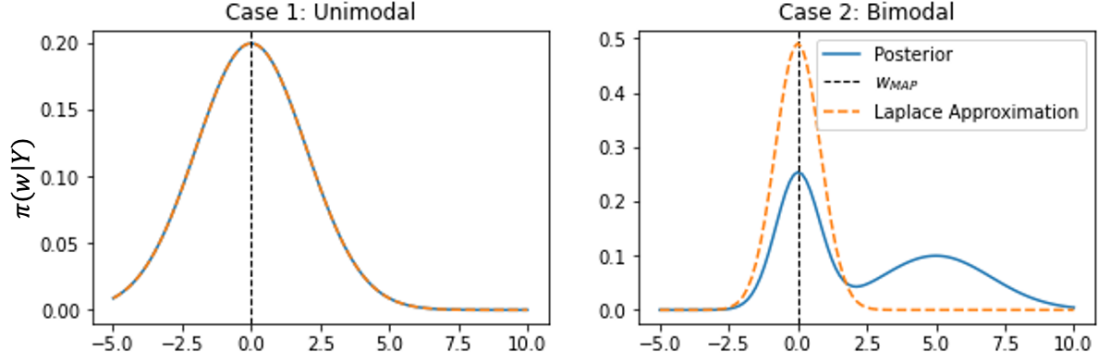


Figure 5.2: Examples of Laplace Approximation (LA) - Left: Approximation to a unimodal bell-shaped distribution, where the LA can capture most of the density. Right: Approximation to a bimodal posterior; as it deviates from expected Gaussianity, there are regions that the LA cannot cover. Figure reproduced from <https://pat.chormai.org/blog/2021-laplace-approximation>

Hence, this approximation $q(\cdot)$ can be used as the multivariate Gaussian proposal distribution of the MCMC, $\pi(\omega_{new})$:

$$\pi(\omega_{new}|\hat{\omega}, -H^{-1}) \sim \mathcal{N}(\hat{\omega}, -H^{-1}) \quad (5.14)$$

where

$$H = \frac{\partial^2 \log(\pi(\omega|Y))}{\partial Y^2} \Big|_{\omega=\hat{\omega}} \quad (5.15)$$

and $\hat{\omega}$ is the MAP value that maximises the logarithm of the posterior distribution

$$\hat{\omega} = \underset{\omega}{\operatorname{argmax}}(\log(\pi(\omega|Y))) \quad (5.16)$$

Note that here the proposal in eq.5.14 is not conditional on previous accepted ω values (as in eq.5.10 or eq.5.6), but fixed. This is the basis of the known **Independence Sampler**, which assumes that the LA approximation is accurate enough and does not require the implementation of a Markov Chain to search in the parameter space. Then, under the assumption of non-informative priors, eq.5.15 and 5.16 reduce to finding the mode and the

inverse Hessian of the likelihood, respectively. This inverse of the Hessian can be actually used as the variance-covariance matrix in a block-RWM, i.e. it is not limited to operating in Independence Sampler only (see Appendix C for demonstrations).

5.3 Methods

As explained above, one of the main challenges of the Random-walk MCMC (RWM) relies on its sampling and computational efficiency. This burden becomes more relevant in data-heavy scenarios like diffusion MRI, where likelihoods need to be computed thousands of times for every model parameter in every voxel. To solve these challenges, we devise new MCMC approaches that i) use block proposals that consider correlations in the parameter space and have more efficient jumps for groups of parameters compared to random-walk independent jumps, ii) use the Laplace approximation to find a proper initialization of the covariance matrix for block proposals, iii) employ optimal subsets of the parameters to accommodate the coexistence of proper and improper priors (such as ARD).

The description of such algorithms is presented in this section. We apply these to an exemplar model, the Ball&Sticks model, and we compare them with the original random-walk MH-MCMC implementation [Behrens et al., 2007]. We start by presenting this RWM and subsequently how this can be improved using the proposed designs. The data and evaluations employed to assess the performance of the proposed methods are also described.

5.3.1 Inference for the Ball&Sticks model using random-walk MH-MCMC

As an exemplar biophysical model for this work, we use the Ball&Sticks model (see eq.5.18). A Bayesian inference approach for this model using RWM is available in FSL [Behrens et al., 2007] and we will use a similar implementation in Python here. Assuming independence between diffusion-sensitising measurements i ($i = 1 : N$), and zero-mean additive Gaussian noise with precision $\tau = 1/\sigma^2$, the **likelihood function** of the data, $\pi(Y|\omega)$, is [Behrens et al., 2003]:

$$\pi(Y|\omega, \tau) \sim \mathcal{N}(S_i, \tau)$$

$$\Rightarrow \pi(Y|\omega, \tau) = \prod_{i=1}^N \left(\frac{\sqrt{\tau}}{\sqrt{2\pi}} \right) \cdot \exp \left(-\frac{\tau}{2} \sum_{i=1}^N (Y_i - S_i)^2 \right) \quad (5.17)$$

where Y_i is the measured signal with gradient i , N is the number of gradients or volumes, ω are model parameters of interest, and S_i is the predicted signal by the forward model, given by the Ball&Sticks model (see 2.2.4):

$$S_i(b_i, g_i, \omega_i) = S_0 \cdot \left[\left(1 - \sum_{j=1}^M f_j \right) \cdot \exp(-b_i \cdot d) + \sum_{j=1}^M f_j \cdot \exp(-b_i \cdot d \cdot (g_i \cdot v_j^T)^2) \right] \quad (5.18)$$

with S_0 the measured signal with no diffusion-sensitising gradients applied, d is the diffusivity, f_j and v_j are the volume fraction and fibre orientation vector of the j -th stick ($j = 1 : M$), respectively, and b_i and g_i are the magnitude and vector direction of gradient i (i.e., b-value and b-vector), $i = 1 : N$. Note that v_j^T is commonly expressed in spherical coordinates, as it allows to reduce the number of unknown parameters (from x, y, z to θ, ϕ):

$$v_j : [x_j, y_j, z_j]_{cartesian} \rightarrow [\cos(\phi_j) \sin(\theta_j), \sin(\theta_j) \sin(\phi_j), \cos(\theta_j)]_{spherical} \quad (5.19)$$

Acquisition parameters b_i and g_i are known, so the unknown parameters are $\{S_0, d, \theta_j, \phi_j, f_j, \tau\}$.

It is possible to marginalize out the precision parameter τ (see appendix B.4), having as a result a likelihood dependent only on model parameters ω :

$$\pi(Y|\omega) \propto \frac{\sum_{i=1}^N (Y_i - S_i)^2 - \frac{N}{2}}{2} \quad (5.20)$$

$$\Rightarrow \log(\pi(Y|\omega)) \propto -\frac{N}{2} \cdot \log \left(\frac{\sum_{i=1}^N (Y_i - S_i)^2}{2} \right) \quad (5.21)$$

Regarding the **prior distributions** of the parameters, $\pi(\omega)$, non-informative distributions can be used [Behrens et al., 2003], besides from a positivity constraint on S_0 , d , and f_j (for $j \geq 1$). For secondary volume fractions, ARD priors are used in order to discern whether they are needed or not:

$$\pi(\theta_j, \phi_j) \sim \mathcal{U}(0, 2\pi)|_{cart} \rightarrow \frac{|\sin \theta_j|}{2}|_{sph}$$

$$\pi(f_1) \sim \mathcal{U}(0, 1)$$

$$\pi(f_j) = -\frac{1}{f_j} \quad (j \geq 2)$$

$$\pi(d) \sim \mathcal{U}(0, 1)$$

$$\pi(S_0) \sim \mathcal{U}(0, \infty)$$

Putting all the above together (the priors and the likelihood), and following the Bayes Rule, the **posterior distribution** $\pi(\omega|Y)$ has the following form:

$$\pi(\omega|Y) \propto \pi(Y|\omega) \cdot \pi(\theta, \phi) \cdot \pi(f) \cdot \pi(d) \cdot \pi(S_0) \quad (5.22)$$

And in terms of energy function \mathcal{E} :

$$\boxed{\mathcal{E} = -\log(\pi(\omega|Y)) \propto \frac{N}{2} \log \frac{\sum_{i=1}^N (Y_i - S_i)^2}{2} - \sum_{j=1}^M \log \frac{|\sin(\theta_j)|}{2} + \sum_{j=2}^M \log(f_j)} \quad (5.23)$$

Model parameters are initialised by a non-linear deterministic fitting (i.e., the Levenberg-Marquadt, (LM)) to the signal measured. For further considerations about the initialization of the parameters, see appendix B.2. The adaptive Metropolis implemented has been detailed in sec.5.2.2. The RWM algorithm implemented can be found in Algorithm 1.

Algorithm 1 Adaptive Univariate RWM**Require:** $Y, \pi(Y|\omega), \pi(\omega)$

```

1: procedure ESTIMATE  $\pi(\omega|Y)$ 
2:   Initialize MCMC parameters ▷ jumps, burn-in,  $\lambda_L$ , etc. defined by user
3:   Initialize forward model and  $x_0$  ▷ See appendix B.2
4:    $\omega^{old} = x_0$ 
5:   Calculate  $\pi(Y|\omega^{old})$  and  $\pi(\omega^{old})$  ▷ ARD applied
6:   Calculate  $\pi(\omega|Y)^{old}$  ▷ Bayes Rule
7:
8:   for  $l = 1$  to  $burn - in + n_{jumps}$  do
9:     for  $k = 1$  to  $n_{params}$  do
10:      Propose  $\omega_{l,k}^{new} \sim \lambda_L \cdot \mathcal{N}(\omega_k^{old}, \sigma)$  ▷ following eq. 5.10
11:      if  $\omega_{l,k}^{new}$  is within boundaries then
12:        Calculate  $\pi(Y|\omega_{l,k}^{new})$  and  $\pi(\omega_{l,k}^{new})$  ▷ ARD applied
13:        Calculate  $\pi(\omega|Y)^{new}$  ▷ Bayes Rule
14:         $r = \min\left(1, \frac{\pi(\omega_{l,k}^{new})}{\pi(\omega_{l,k}^{old})}\right)$ 
15:        if  $r < \mathcal{U}(0, 1)$  then ▷ MH acceptance criteria
16:          Store  $\omega_{l,k}^{new}$  samples ▷ For  $\pi(\hat{\omega}|Y)$ 
17:          Update  $\omega_k^{old} = \omega_{l,k}^{new}$ 
18:          Update  $\pi(\omega|Y)^{old} = \pi(\omega|Y)^{new}$  ▷  $\omega_{l,k}$  is accepted
19:        else
20:          Go to line 23
21:        end if
22:      else
23:        Store  $\omega_k^{old}$  samples ▷ For  $\pi(\hat{\omega}|Y)$ 
24:        Keep  $\omega_k^{old} = \omega_k^{old}$  ▷  $\omega_{l,k}^{new}$  is rejected
25:        Keep  $\pi(\omega|Y)^{old} = \pi(\omega|Y)^{old}$ 
26:        rejected += 1
27:      end if
28:    end for
29:    if  $i = \text{mod}(L)$  then
30:      Update  $\lambda_L$ 
31:      Reset accepted and rejected counters
32:    end if
33:  end for
34:
35:  Discard burn-in samples in  $\pi(\hat{\omega}|Y)$ 
36:  Sample every X samples of  $\pi(\hat{\omega}|Y)$  ▷ Thinning
37:  return  $\pi(\hat{\omega}|Y)$ 

```

5.3.2 Alternative designs to random-walk MH-MCMC

Method 1: Independence Sampler

In the standard RWM implementation, the univariate proposal distribution depends on the current values of the parameters so the next jump starts from that point, and does so independently for each parameter dimension. The most efficient alternative (assuming the proposal distribution is correct) is to define a **fixed joint proposal distribution** (i.e. no adaptive steps nor Markov-Chain), from where new samples are proposed as a block at every iteration, considering all parameters at once. This paradigm is known as **Independence sampler** as there is no dependence on the previous steps of the chain. Hence, it is crucial to achieve a proposal distribution that mimics the target density so the independence sampler will propose values for the parameters that are close to the real values and, therefore, have a sufficient acceptance ratio and allowing for a reduction in the computation needed.

Here, we use the **Laplace approximation** of the joint posterior of all parameters to define such proposal. In order to calculate the covariance matrix, we need to get the second derivatives with respect to the parameters we want to estimate (the Hessian of the energy function). The Hessian is then evaluated at an optimum of the parameter space and we find these values for the model parameters by a non-linear fit to the observations. The analytical derivatives to compute the Hessian for the Ball&Sticks model with 3 sticks can be found in appendix C.2. This analytical derivation of the Hessian will be used for all the methods proposed in this chapter (except the univariate RWM, as it does not need it).

It's important to note that with this proposal the symmetric assumption used in Metropolis of $\pi(\omega_{new}|\omega) \equiv \pi(\omega|\omega_{new})$ is not valid anymore. It is therefore necessary to revert the MH acceptance criteria to its original form of eq.5.3.

A limitation of this implementation is that only proper priors can be considered (i.e., no ARD priors are used), as the Laplace Approximation works under the assumption well-behaved density functions. We do not expect particularly efficient performance from this algorithm, but it is considered here as a building block for the following designs.

Algorithm 2 Independence Sampler**Require:** Y , $\pi(Y|\Omega)$, $\pi(\Omega)$, Hessian H

```

1: procedure ESTIMATE  $\pi(\Omega|Y)$ 
2:   Initialize MCMC parameters ▷ jumps, burn-in,  $\lambda_L$ , etc. defined by user
3:   Initialize forward model  $x_0$  and covariance  $\Sigma$  ▷ See appendices B.2 and C.2
4:   Calculate  $\pi(Y|x_0)$  and  $\pi(x_0)$  ▷ No ARD priors
5:   Calculate  $\pi(\Omega|Y)^{old} = \pi(Y|x_0) \cdot \pi(x_0)$  ▷ Bayes Rule
6:
7:   for  $l = 1$  to  $burn - in + n_{jumps}$  do
8:     Propose  $\Omega_l^{new} \sim \mathcal{N}(x_0, \Sigma)$  ▷ Independence Sampler
9:     if all  $\omega_k$  in  $\Omega_l^{new}$  are within boundaries then
10:      Calculate  $\pi(Y|\Omega_l^{new})$  and  $\pi(\Omega_l^{new})$  ▷ No ARD applied
11:      Calculate  $\pi(\Omega|Y)^{new}$  ▷ Bayes Rule
12:       $r = \min\left(1, \frac{\pi(x_0|\Omega_l^{new}) \cdot \pi(\Omega_l^{new})}{\pi(\Omega_l^{new}|x_0) \cdot \pi(x_0)}\right)$ 
13:      if  $r < \mathcal{U}(0, 1)$  then ▷ MH acceptance criteria
14:        Store  $\Omega_l^{new}$  samples ▷ For  $\pi(\hat{\Omega}|Y)$ 
15:        Update  $\pi(\Omega|Y)^{old} = \pi(\Omega|Y)^{new}$  ▷  $\omega_{l,k}$  is accepted
16:      else
17:        Go to line 19
18:      end if
19:    else
20:      Store  $x_0$  samples ▷ For  $\pi(\hat{\Omega}|Y)$ 
21:      Keep  $\pi(\Omega|Y)^{old} = \pi(\Omega|Y)^{old}$  ▷  $\Omega_l^{new}$  is rejected
22:    end if
23:  end for
24:
25:  Discard burn-in samples in  $\pi(\hat{\Omega}|Y)$ 
26:  Sample every  $X$  samples of  $\pi(\hat{\Omega}|Y)$  ▷ Thinning
27:  return  $\pi(\hat{\Omega}|Y)$ 

```

Method 2: Adaptive Block-RWM

An alternative to the simple Independence Sampler can be obtained by introducing two modifications: 1) A random walk paradigm within the joint proposals, i.e. the joint proposal of all parameters depends on the current values of the chain, rather than being a fixed proposal. 2) Fine-tune the covariance of the joint proposal in an adaptive style. Hence, we are still updating all the parameters at once (in a block) and the Laplace approximation is still used to initialize the joint proposal distribution (modes and covariances). However, modification 1) helps to reach the multivariate target posterior if the initialization is not good enough; and modification 2) allows the covariance to be updated along the procedure, which can add flexibility to the new proposals and increase the acceptance ratios,

especially if the covariance provided by the LA is not good enough (e.g., when working with not bell-shaped target distributions).

As we have seen in the theory section, there are multiple options to adapt the covariance Σ . For instance, the tuning phase can be used to find an optimal scale for the covariance matrix Σ and, again, there are multiple ways to define this optimal scale. Here, as acceptance/rejections happen for all parameters simultaneously, we implement the same adaptive Metropolis scheme used in the univariate RWM, where a factor λ_L is updated every L MCMC iteration. Finally, as we allow for adaptive covariances and the LA is used only to initialize the proposal, we can use ARD priors, such as $ARD \sim \frac{1}{f_j}$ for f_j (for $j > 1$).

Algorithm 3 Adaptive Block-RWM

H]

Require: $Y, \pi(Y|\Omega), \pi(\Omega), \text{Hessian } H$

```

1: procedure ESTIMATE  $\pi(\Omega|Y)$ 
2:   Initialize MCMC parameters ▷ jumps, burn-in,  $\lambda_L$ , etc. defined by user
3:   Initialize forward model  $x_0$  and covariance  $\Sigma$  ▷ See appendices B.2 and C.2
4:    $\Omega^{old} = x_0$ 
5:   Calculate  $\pi(Y|\Omega^{old})$  and  $\pi(\Omega^{old})$  ▷ ARD optional
6:   Calculate  $\pi(\Omega|Y)^{old}$  ▷ Bayes Rule
7:
8:   for  $l = 1$  to  $burn - in + n_{jumps}$  do
9:     Propose  $\Omega_l^{new} \sim \lambda_L \cdot \mathcal{MVN}(\Omega_l^{old}, \Sigma)$  ▷ following eq.??
10:    if all  $\omega_k$  in  $\Omega_l^{new}$  are within boundaries then
11:      Calculate  $\pi(Y|\Omega_l^{new})$  and  $\pi(\Omega_l^{new})$  ▷ ARD optional
12:      Calculate  $\pi(\Omega|Y)^{new}$  ▷ Bayes Rule
13:       $r = \min\left(1, \frac{\pi(\Omega_l^{new})}{\pi(\Omega_l^{old})}\right)$ 
14:      if  $r < \mathcal{U}(0, 1)$  then ▷ MH acceptance criteria
15:        Store  $\Omega_l^{new}$  samples ▷ For  $\pi(\hat{\Omega}|Y)$ 
16:        Update  $\Omega^{old} = \Omega_l^{new}$ 
17:        Update  $\pi(\Omega|Y)^{old} = \pi(\Omega|Y)^{new}$  ▷  $\Omega_l^{new}$  is accepted
18:        accepted += 1
19:      else
20:        Go to line 23
21:      end if
22:    else
23:      Store  $\Omega^{old}$  samples ▷ For  $\pi(\hat{\Omega}|Y)$ 
24:      Keep  $\Omega^{old} = \Omega^{old}$  ▷  $\Omega_l^{new}$  is rejected.
25:      Keep  $\pi(\Omega|Y)^{old} = \pi(\Omega|Y)^{old}$ 
26:      rejected += 1
27:    end if
28:    if  $i = \text{mod}(L)$  & Adaptive=True then
29:      Update  $\lambda_L$ 
30:      Reset accepted and rejected counters
31:    end if
32:  end for
33:
34:  Discard burn-in samples in  $\pi(\hat{\Omega}|Y)$ 
35:  Sample every X samples of  $\pi(\hat{\Omega}|Y)$  ▷ Thinning
36:  return  $\pi(\hat{\Omega}|Y)$ 

```

Method 3: Hybrid-MCMC

The advantages of treating all model parameters jointly, as done in the previous algorithms, are linked to the challenges of finding an appropriate joint proposal. In practice, grouping all parameters in the same set can be rather inefficient, especially when parameters have vastly different scales as in the Ball&Sticks model or can become degenerate due to ARD priors. Also, as the number of parameters gets large, it is much more likely to have (proposal) samples that fall well into the tails of the target distribution, producing too small acceptance ratios.

An intermediate alternative to alleviate these difficulties is to allocate a group of parameters into smaller blocks and update them sequentially. To the best of our knowledge, there is no theory to determine an optimal "blocking" for an arbitrary parametric model. A rule of thumb followed in practice is to form small groups of correlated parameters that belong to the same context in the formulation of the model.

We therefore implemented an approach that performs block-updates on a subset of parameters, while leaving parameters with improper priors to be updated independently. We called it **Hybrid-MCMC**. More specifically, we evaluated two potential blocking configurations in preliminary experiments:

- **Hybrid-MCMC configuration 1 (or Hybrid1):** We split the parameters based on potential *degeneracy* due to ARD priors. We considered block $\Omega_1 = \{s_0, d, \theta_1, \phi_1, f_1\}$, being the parameters that are always relevant for the model and block $\Omega_2 = \{\theta_2, \phi_2, \theta_3, \phi_3\}$, being that parameters that can become degenerate, as they reflect compartments with an ARD on the respective volume fractions. Parameters in each of these blocks are treated jointly. We perform independent jumps for parameters f_2 and f_3 , where ARD priors are applied, i.e. $\Omega_3 = \{f_2\}$ and $\Omega_4 = \{f_3\}$.
- **Hybrid-MCMC configuration 2 (or Hybrid2):** We split based on *contextual similarity*, with blocks $\Omega_1 = \{s_0, d, f_1\}$, containing non-orientational parameters and $\Omega_2 = \{\theta_1, \phi_1, \theta_2, \phi_2, \theta_3, \phi_3\}$ being the parameters that represent orientations in a wrapped $[0, 2\pi]$ space. Parameters in each of these blocks are treated jointly. We perform independent jumps for parameters f_2 and f_3 , where ARD priors are applied,

i.e. $\Omega_3 = \{f_2\}$ and $\Omega_4 = \{f_3\}$.

Both hybrid-MCMC configurations can still accelerate the MCMC process updating most of the parameters in a block-manner while leaving freedom to use sparsity priors for model selection on the fly.

Algorithm 4 Hybrid MCMC

Require: Y , $\pi(Y|\omega)$, $\pi(\omega)$, Hessian H

```

1: procedure ESTIMATE  $\pi(\omega|Y)$ 
2:   Initialize MCMC parameters                                ▷ jumps, burn-in,  $\lambda_L$ , etc. defined by user
3:   Initialize forward model  $x_0$  and covariance  $\Sigma$           ▷ See appendices B.2 and C.2
4:    $\omega^{old} = x_0$ 
5:   Define  $K$  blocks of parameters  $\omega_k$  and covariance  $\Sigma_k$ 
6:   Calculate  $\pi(Y|\omega^{old})$  and  $\pi(\omega^{old})$                     ▷ ARD applied
7:   Calculate  $\pi(\omega|Y)^{old}$                                     ▷ Bayes Rule
8:
9:   for  $l = 1$  to  $burn_{in} + n_{jumps}$  do
10:    for  $k = 1$  to  $n_{blocks}$  do
11:      Propose  $\Omega_{l,k}^{new} \sim \Omega_k^{old} + \lambda_L \cdot \mathcal{N}(0, \Sigma_j)$     ▷ following eq. ??
12:      if all  $\omega$  in  $\Omega_{l,k}^{new}$  are within boundaries then
13:        Calculate  $\pi(Y|\Omega_{l,k}^{new})$  and  $\pi(\Omega_{l,k}^{new})$           ▷ ARD applied
14:        Calculate  $\pi(\Omega|Y)^{new}$                                 ▷ Bayes Rule
15:         $r = \min\left(1, \frac{\pi(\Omega_{l,k}^{new})}{\pi(\Omega_{l,k}^{old})}\right)$ 
16:        if  $r < \mathcal{U}(0, 1)$  then                                ▷ MH acceptance criteria
17:          Store  $\Omega_{l,k}^{new}$  samples                                ▷ For  $\pi(\hat{\Omega}|Y)$ 
18:          Update  $\Omega_k^{old} = \Omega_{l,k}^{new}$ 
19:          Update  $\pi(\Omega|Y)^{old} = \pi(\Omega|Y)^{new}$               ▷  $\Omega_{l,k}^{new}$  is accepted
20:          accepted += 1
21:        else
22:          Go to line 26
23:        end if
24:      else
25:        Store  $\Omega_k^{old}$  samples                                ▷ For  $\pi(\hat{\Omega}|Y)$ 
26:        Keep  $\Omega_k^{old} = \Omega_k^{old}$                             ▷  $\Omega_{l,k}^{new}$  is rejected.
27:        Keep  $\pi(\Omega|Y)^{old} = \pi(\Omega|Y)^{old}$ 
28:        rejected += 1
29:      end if
30:    end for
31:    if  $i = \text{mod}(L)$  then
32:      Update  $\lambda_L$ 
33:      Reset accepted and rejected counters
34:    end if
35:  end for
36:
37:  Discard burn-in samples in  $\pi(\hat{\omega}|Y)$ 
38:  Sample every  $X$  samples of  $\pi(\hat{\omega}|Y)$                                 ▷ Thinning
39:  return  $\pi(\hat{\omega}|Y)$ 

```

5.3.3 Data

Simulated data

First evaluations of the proposed algorithms were done in simulated data. The dMRI signal was simulated using the Ball&Sticks model and a single-shell ($b = 1000 \text{ s/mm}^2$) sampling scheme with 64 diffusion sensitizing directions and 5 $b_0 = 0$ volumes. Gaussian

noise was added to the signal in quadrature (i.e. resulting to Rician noise) for the $b = 0$ s/mm^2 signal. Using this parametrisation, we simulated three datasets (see also Fig.5.3):

- **Dataset A:** dataset simulated with a **single fibre orientation per voxel**. We kept fixed $S_0 = 1000$, $d = 0.0012$ [mm^2/s]. We vary the volume fraction in steps of $\Delta f_1 = 0.1$ from $f_1 : [0.1 - 0.9]$, and the orientation v_1 keeping fixed $\phi_1 = \pi/6$ [rad] and varying $\theta_1 : [0, \pi/2]$ [rad] with increasing $\Delta\theta_1 = \pi/10$ [rad]. A total of 5x8 voxels were simulated per realisation. The aim of this dataset is to evaluate whether the algorithms can perform online model selection and return accurate estimates in a fairly simple fibre configuration with different values of volume fraction and fibre orientation.
- **Dataset B:** dataset simulated with **two fibres per voxel**. We kept fixed $S_0 = 1000$, $d = 0.0012$ [mm^2/s] and $v_1 = [1, 0, 0]_{cart}$. Volume fractions of both fibres were kept as equal $f_1 = f_2$, varying in a range of $f_i : [0.2 - 0.45]$ with steps of $\Delta f_i = 0.05$. Simultaneously, we vary the crossing angle X_{angle} between v_1 and v_2 such as $X_{angle} : [40 - 90](^\circ)$ with $\Delta X_{angle} = 5^\circ$. A total of 5x10 voxels per realisation. Similar to Dataset A, here we evaluate whether they can perform online model selection and the accuracy of estimates in voxels with a more complex fibre configuration.
- **Dataset C:** dataset simulated with **three fibres per voxel**. We kept fixed $S_0 = 1000$, $d = 0.0012$ [mm^2/s], $f_1 = 0.3$, and the fibre orientations perpendicular such as $v_1 = [1, 0, 0]_{cart}$, $v_2 = [0, 1, 0]_{cart}$, and $v_3 = [0, 0, 1]_{cart}$. We vary the volume fraction of the second fibre as $f_2 : [0.2 - 0.3]$ with $\Delta f_2 = 0.01$, and $f_3 : [0.2 - 0.3]$ with $\Delta f_3 = 0.02$. A total of 5x10 voxels per realisation. The aim of this dataset is to evaluate the accuracy and precision of identifying multiple fibres where volume fractions have similar (low) values.

For each dataset, we simulated $z = 100$ noisy realisations to evaluate median errors. In all cases, we fit the Ball&Sticks model with (up to) 3 fibre compartments. To assess the effect of noise, we repeated these experiments for SNR=15 and SNR=30. Given the similar volume fraction values in datasets B and C, a fibre orientation matching between ground-truth and estimates is performed prior to assessing the results, based on the lowest crossing angle between fibre orientations (see eq.5.26).

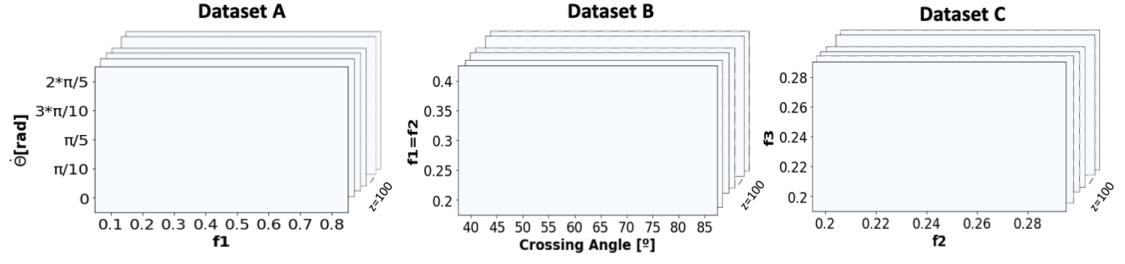


Figure 5.3: *Scheme of the simulated datasets.*

Note that the data in this study is simulated using the identical biophysical model that is later employed in the likelihood computation during the MCMC procedure. It is essential to acknowledge that employing the same model for both simulation and evaluation could introduce biases when assessing the accuracy and precision of a given biophysical model (i.e. whether the model accurately captures and reproduces the biophysical properties of the studied anatomy). However, in this chapter, our focus is on comparing the performance of different inference strategies, namely component-wise MCMC versus block-approaches MCMC, using the same dataset and biophysical model. Therefore, any potential biases resulting from using the same model for simulation and evaluation would be consistent across all approaches and hence would not affect the validity of the comparisons. Similarly, we expect conclusions to remain valid if a different model were used for simulation and inference.

In-vivo brain data

We used diffusion MRI data from 15 subjects randomly selected from the UKBiobank dataset, accessed through UKBiobank project number 43822 (PI: Sotiropoulos). All data were already pre-processed from the UKBiobank for distortion and motion correction [Alfaro-Almagro et al., 2018; Miller et al., 2016]. We extracted the lower $b = 0$ and $b = 1000 \text{ s/mm}^2$ shells, as these are compatible with the assumptions of the Ball&Sticks model used here [Jbabdi et al., 2012].

Algorithms and MCMC parametrisation

The proposed MCMC algorithms were evaluated: the Independence Sampler, the adaptive Block-proposal Random Walk MH-MCMC (Block-RWM), and the two Hybrid-MCMC configurations (Hybrid1 and Hybrid2). For all algorithms except the Independence Sam-

pler, the default configuration included the adaptive joint proposal with the scalar factor λ_L (see sec.5.2.2) and the ARD priors applied to f_2 and f_3 . All these designs were compared to the original univariate random-walk MH-MCMC (RWM) as proposed in [Behrens et al., 2007].

For all the algorithms, we used a standard set of MCMC hyper-parameters (e.g., used in the FSL's BedpostX implementation as default): Burn-in=1000, jumps=1250, thinning=25. This setup returns 50 already-thinned samples with the burn-in samples period discarded. These 50 samples are used to construct the posterior density estimates. Common to all configurations are the initialization of parameters x_0 (mean of the proposal) done by the Lavenberg-Marquadt non-linear fitting and the non-informative priors described in the previous section. The λ_L factor is set to be automatically updated every $L = 50$ loops (for the RWM, Block-RWM and the Hybrid-MCMC configurations).

5.3.4 Evaluations

Accuracy

To evaluate **accuracy**, we calculated the errors in the mean estimates. Once posterior model parameters densities are sampled, their mean posterior is extracted, i.e., $\langle \pi(\hat{\omega}|Y) \rangle$ for scalar parameters $\{S_0, d, f_j\}$. For the fibre orientations v_j , the principal eigenvector of the dyadic tensor was used to get the average fibre orientations across θ_j and ϕ_j MCMC samples. In **simulations**, the ground-truth value of each parameter ω is available. For scalar parameters, the voxel-wise error e can be calculated directly as the difference between the ground-truth and the mean estimates:

$$e_\omega = \omega - \langle \pi(\hat{\omega}|Y) \rangle \quad (5.24)$$

This can be also expressed in percentage rates:

$$e_{\omega|\%} = 100 * \frac{\omega - \langle \pi(\hat{\omega}|Y) \rangle}{\omega} \quad [\%] \quad (5.25)$$

For fibre orientations, we calculated the crossing-angle X_{angle} (in degrees) between the true v_j and the estimated one such as:

$$X_{angle} = \arccos \langle v_j \cdot \hat{v}_j \rangle \quad [^\circ] \quad (5.26)$$

where v_j is the fibre orientation considered as truth, \hat{v}_j is the estimated dyadic vector, and the $\langle \cdot, \cdot \rangle$ denotes the scalar product between both vectors.

For the **in-vivo brain data**, we used the mean estimates from the univariate RWM as the reference values for the evaluations, as RWM is the implementation currently used in many tested and validated software packages, such as FSL. Hence, the error in scalar parameters was calculated as the difference between mean estimates, and the % error as:

$$e_{\omega|\%} = \left| 100 * \frac{\langle \pi(\hat{\omega}|Y) \rangle_{RWM} - \langle \pi(\hat{\omega}_j|Y) \rangle}{\langle \pi(\hat{\omega}|Y) \rangle_{RWM}} \right| \quad [\%] \quad (5.27)$$

Regarding fibre orientations, the error was calculated as in eq.5.26, where v_j is now the dyadic vector provided by the RWM.

Precision

In both simulations and in-vivo brain data, **precision** was evaluated by comparing the width of the posterior distributions provided by the RWM (used as a reference) and the rest of the methods. For the fibre orientation, by getting the principal eigenvalue of the dyadic tensor we could obtain the dispersion of the direction samples. Here, secondary fibres are expected to have higher dispersion where such fibres are not supported by the data.

Performance

Regarding the **performance** of the algorithm, there are different features of the Markov chain that can be evaluated, such as **convergence** to the target posterior distribution, **mixing** quality and **sampling efficiency**. For convergence, there are no conclusive tests on whether a Markov chain has converged, especially in multivariate chain [Roy, 2020]; we assumed it reached stationarity if the distribution of points was not changing as the chain progressed (e.g. relatively constant mean and variance). We analysed traceplots of the chains and the posterior density estimates for that purpose.

Mixing quality can be assessed by calculating the autocorrelation function (ACF), where the estimated sample autocorrelation $\hat{\rho}_h$ at lag h (i.e. time difference between observations) is defined in terms of the sample autocovariance function:

$$\hat{\rho}(h) = \frac{\text{Cov}(X_t, X_{t+h})}{\text{Var}(X_t)} \quad (5.28)$$

If we assume the chain starts at $t = 0$, the autocorrelation reduces to:

$$\hat{\rho}(h) = \frac{\hat{\psi}(h)}{\hat{\psi}(0)} \quad (5.29)$$

where $\hat{\psi}(h)$ is the sample autocovariance function of lag h :

$$\hat{\psi}(h) = \frac{1}{n} \sum_{i=1}^{n-h} (\omega_i^{t+h} - \tilde{\omega}_i)(\omega_i^t - \tilde{\omega}_i) \quad (5.30)$$

where n is the total sample size such $h < n$, ω_i^t denotes the value of the variable i at time t , and $\tilde{\omega}_i$ represents the sample mean of the variable i . However, given that we use $n = 50$ samples to build the posterior, the ACF needs to be computed with a limited number of lags to keep reliable correlation estimates. We fix it to $h = 5$ lags.

As a complementary measure to examine the mixing/sampling efficiency of the Markov Chain, we also provided the Effective Sample Size (ESS) [Kass et al., 1998]:

$$ESS = \frac{n}{\tau} = \frac{n}{1 + 2 \sum_{j=1}^{\infty} \rho_h(\omega)} \quad (5.31)$$

where n is the total sample size, $\rho_h(\omega)$ is the estimated autocorrelation at lag h for parameter ω (h defines a threshold point after which the autocorrelations are very close to 0) and τ is the autocorrelation time.

All the evaluations explained above are voxel-wise and are calculated after burn-in and thinning. For **global performance**, we compared the **computational time** required to complete the different MCMC procedures. This time was measured by running the in-house Python implementations in a dedicated single-core CPU (see sec.1.2 for detailed specs).

5.4 Results

In this section, we will show results obtained from evaluating the different MCMC approaches presented above. For the representation of the results, the following colour code will be used: RWM in black, Independence Sampler in red, Block-RWM in blue, Hybrid1 in green, and Hybrid2 in magenta. We provide representative examples that help to understand the behaviour of the samplers. Note that the results shown in this section (e.g., traceplots, error calculations, etc.) have been already thinned (thinning=25) and had burn-in samples discarded.

5.4.1 Simulations

In this subsection, we evaluate the sampling performance, accuracy and precision of the different MCMC samplers with simulated data where ground-truth values are known. From these results, we select the samplers with the best performance to compare them in in-vivo brain data.

Convergence to the stationary distribution and sampling efficiency

As a first step, we confirmed whether the behaviour of the chains is as desired after thinning and burn-in period. Fig.5.4 shows the traceplots provided by each MCMC from exemplar voxels in Dataset A (only one fibre), Dataset B (two fibre-crossing), and Dataset C (three fibre-crossings) for the diffusivity, $\pi(d|Y)$, and the volume fractions, $\pi(f_1|Y)$, $\pi(f_2|Y)$, and $\pi(f_3|Y)$. These figures already provide insights about differences in the behaviour of the different algorithms. For instance, the Independence Sampler is barely able to move away from the non-linear initialization values in the parameter space (e.g., Dataset A). Given the complexity of the model fitted, the Laplace approximation is too smooth to cover appropriately the multimodal posterior density. As a consequence, almost 100% of the samples were rejected by the MH acceptance criteria, being incapable of performing a search in the parameter space. For this reason, we do not consider it any further and we focus on the other approaches.

The adaptive Metropolis and the Markov Chain changes introduced in the Block-RWM and Hybrid-MCMC configurations allow them to reshape the Laplace approximation and achieve better mixing, especially the Hybrid-MCMC approaches which are mostly over-

lapped with the univariate RWM. The Block-RWM shows unstable behaviour in some cases where the number of compartments in the model fitted is different than the true underlying fibre configuration (e.g., see example of 2 fibres in Fig.5.4); it provides good mixing and similar estimates to the univariate RWM or the Hybrid-MCMC otherwise (see examples of 3 fibres in Fig.5.4).

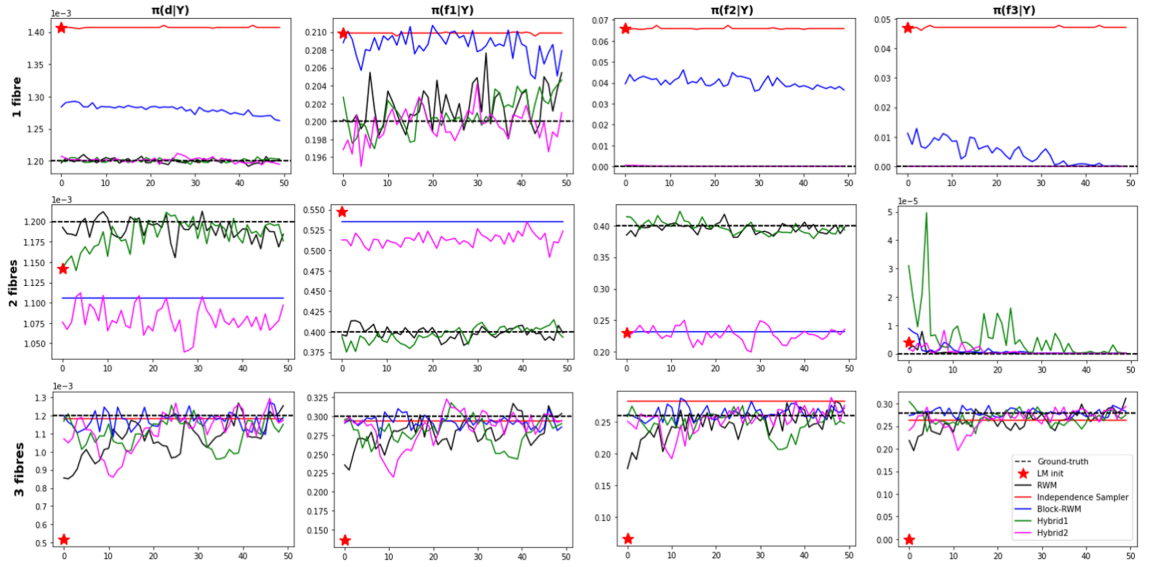


Figure 5.4: *Median traceplots (across realisations) of the diffusivity and volume fractions (columns) obtained by each MCMC approach (denoted by colors) for three different exemplar voxels in dataset A (top row), dataset B (middle row) and dataset C (bottom row) with SNR=30. The ground-truth value is indicated with a black dashed line and the LM initialization value with a red star.*

Fig.5.5 shows a histogram of the median posterior estimates across realisations for the same exemplar voxels showed in Fig.5.4. The univariate RWM and the Hybrid1 are the methods with better agreement, with similar variability in the mean estimates (width of the distribution) and modes close to the ground-truth (black dashed lines) for all datasets. The Hybrid2 deviates from them in Dataset B. The Block-RWM shows higher variability in terms of accuracy and precision, especially for configurations with 1 and 2 fibres; on the other hand, it performs best in Dataset C.

Similarly, boxplots in Fig.5.6 show a histogram of the width of the posteriors (given by the standard deviation) obtained across the 100 noisy realisations. Although there is no ground-truth for this metric, the Block-RWM shows much lower uncertainty quantification

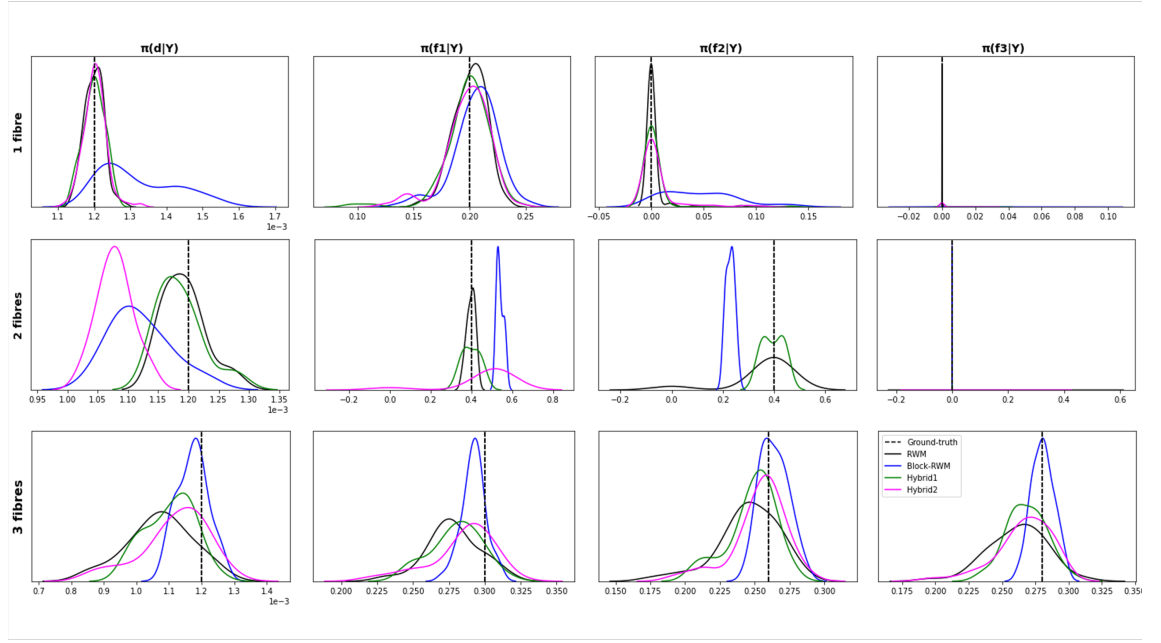


Figure 5.5: Histogram of the median posterior estimates across realisations for the same exemplar voxels shown in Fig.5.4. For visualizations reasons, the Independence Sampler is not represented (it would be a mass point on the initial values of the MCMC).

compared to the rest of the algorithms, especially in the cases of one and two fibres. On the other hand, it provides higher widths in cases where the ARD drives the parameter to 0, i.e. in $f2$ and $f3$ for the case of 1 fibre dataset, and in $f3$ in the case of 2 fibres dataset. Both Hybrid approaches provided similar results, being the Hybrid1 a bit closer to the behaviour observed in the RWM.

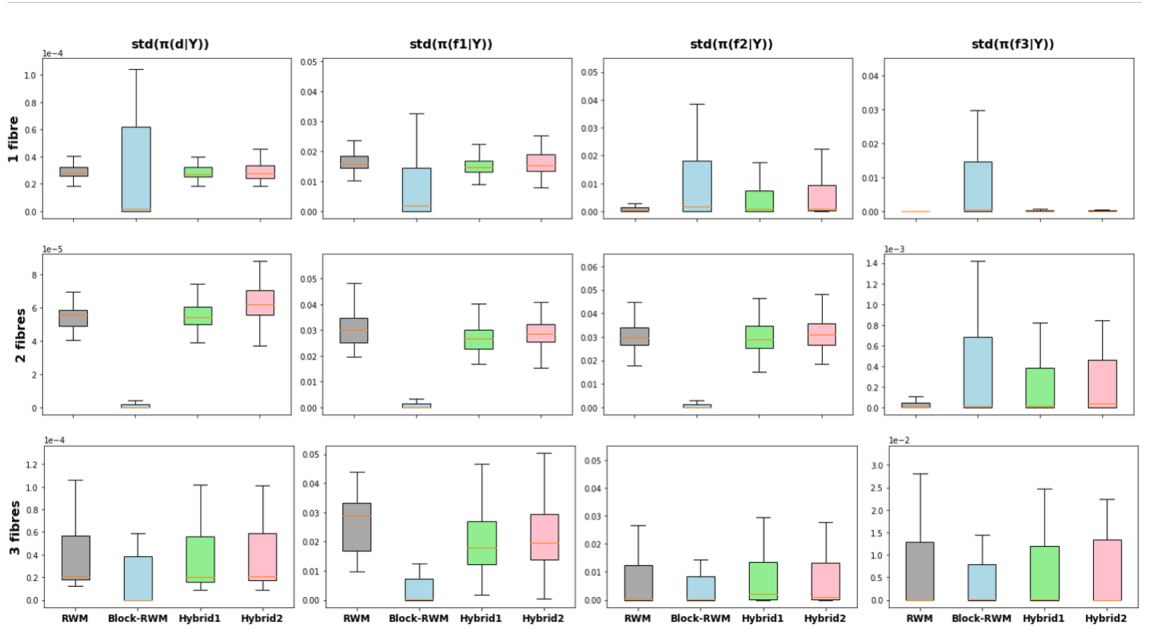


Figure 5.6: Boxplots of the posterior width (given by their standard deviation) across realisations for the same exemplar voxels showed in Fig.5.4. For visualizations reasons, the Independence Sampler is not represented.

Regarding sampling efficiency, Fig.5.7 shows representative examples of the autocorrelation in the chains, and the ESS obtained for f_1 . As the posterior is built with 50 samples (after thinning and burn-in), the ideal Effective Sample Size (ESS) should be close to 50, i.e., 50 non-autocorrelated samples. The univariate RWM performs consistently the best for all parameters, followed by the Hybrid1 and the Hybrid2, and finally the Block-RWM with the highest autocorrelation. Nevertheless, according to what was observed in the traceplots, any of them showed signs of concerning autocorrelation issues (except those cases where the Block-RWM could not move).

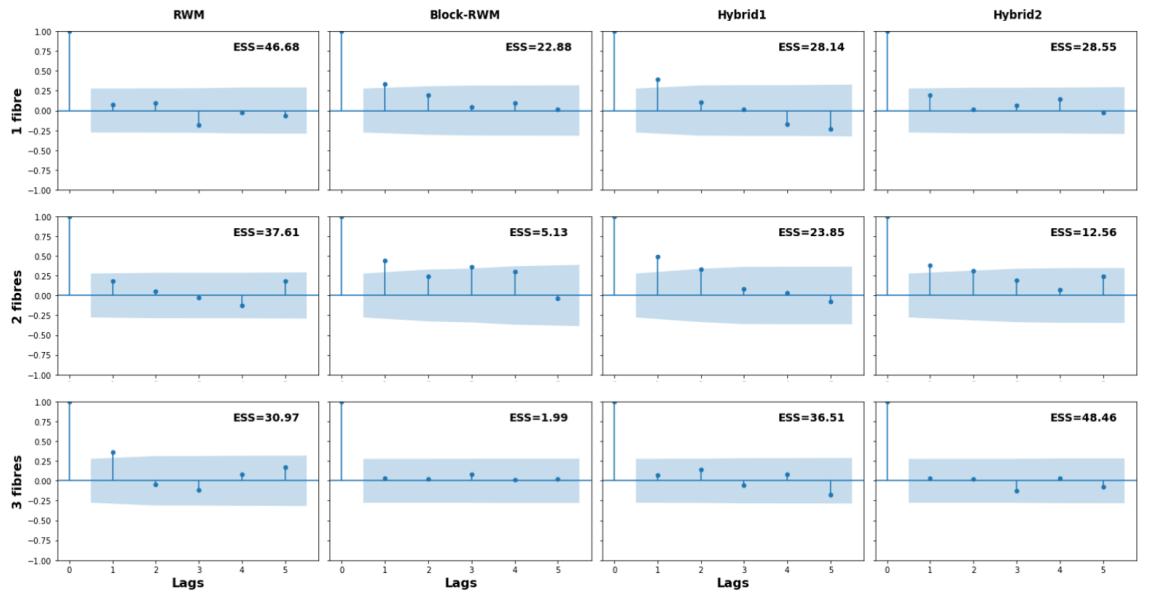


Figure 5.7: *Autocorrelation Function (ACF) and Effective Sample Size (ESS) for the samples of the volume fraction of the first fibre in the same voxels shown in Fig.5.4 and Fig.5.5.*

Analysing the accepted and rejected proposed samples, the univariate RWM and the Hybrid-MCMC configurations showed a homogeneous behaviour, independently of the type of pattern simulated, where $\approx 50\%$ of the samples were accepted and the other half were rejected by the MH acceptance criteria, as designed to do. The Block-RWM showed more variability. Nevertheless, the above corroborates the advantages of the adaptive Metropolis even in high-dimensional problems with block-proposals. Furthermore, the behaviour observed confirms that the adaptive Metropolis can be run only during the burn-in period to reduce the computational cost, as it quickly converges to the desired behaviour (in less than 10 loops, see Fig.5.8). Hence, despite the use of block configurations, and contrary to what has been suggested in the Theory section, both Hybrid-MCMC ap-

proaches have shown the capability to obtain acceptance ratios around 50% consistently, comparable to what is expected for the component-wise RWM (and, therefore, allowing to reduce computational time).

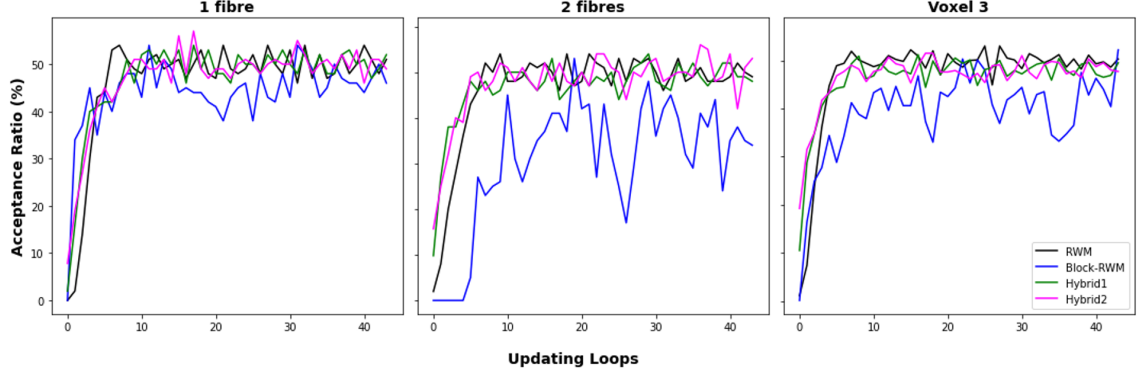


Figure 5.8: *Median acceptance ratio* (across realisations) obtained by each MCMC approach (in colors) for an exemplar random voxel in the 3 different datasets (in columns). It includes burn-in period loops (i.e., up to updating loop=20).

Accuracy and Precision

Based on the above, we can assume that all algorithms reached the stationary distribution and we can evaluate the accuracy and precision of the estimates. The values shown in this subsection represent the median values across the $z = 100$ realisations.

Fig.5.9 shows the median number of fibre compartments detected by each method in each dataset. For each voxel, compartment j is considered as supported by the data if $\langle \pi(f_j|Y) \rangle > 0.05$ [Behrens et al., 2007]. Given the low angular contrast of the simulated data (only $b = 1000s/mm^2$ shell), multi-way crossings can be wrongly estimated (e.g., RWM and Hybrid-MCMC in Dataset C at SNR=15). Increasing the SNR (or b-value), all methods provided a good estimation of the number of fibres, as shown in Fig.5.9 (bottom row).

The following figures show the median estimation errors for each algorithm and for each dataset, as indicated in eq.5.25 (for scalar parameters) and eq.5.26 (for fibre orientations). In the case of **single-fibre patterns** (Fig.5.10), errors decreased (or slightly changed) with increasing SNR. The performance of the univariate RWM and the Hybrid-MCMC is pretty similar, while the Block-RWM showed the highest error rates overall. Nevertheless,

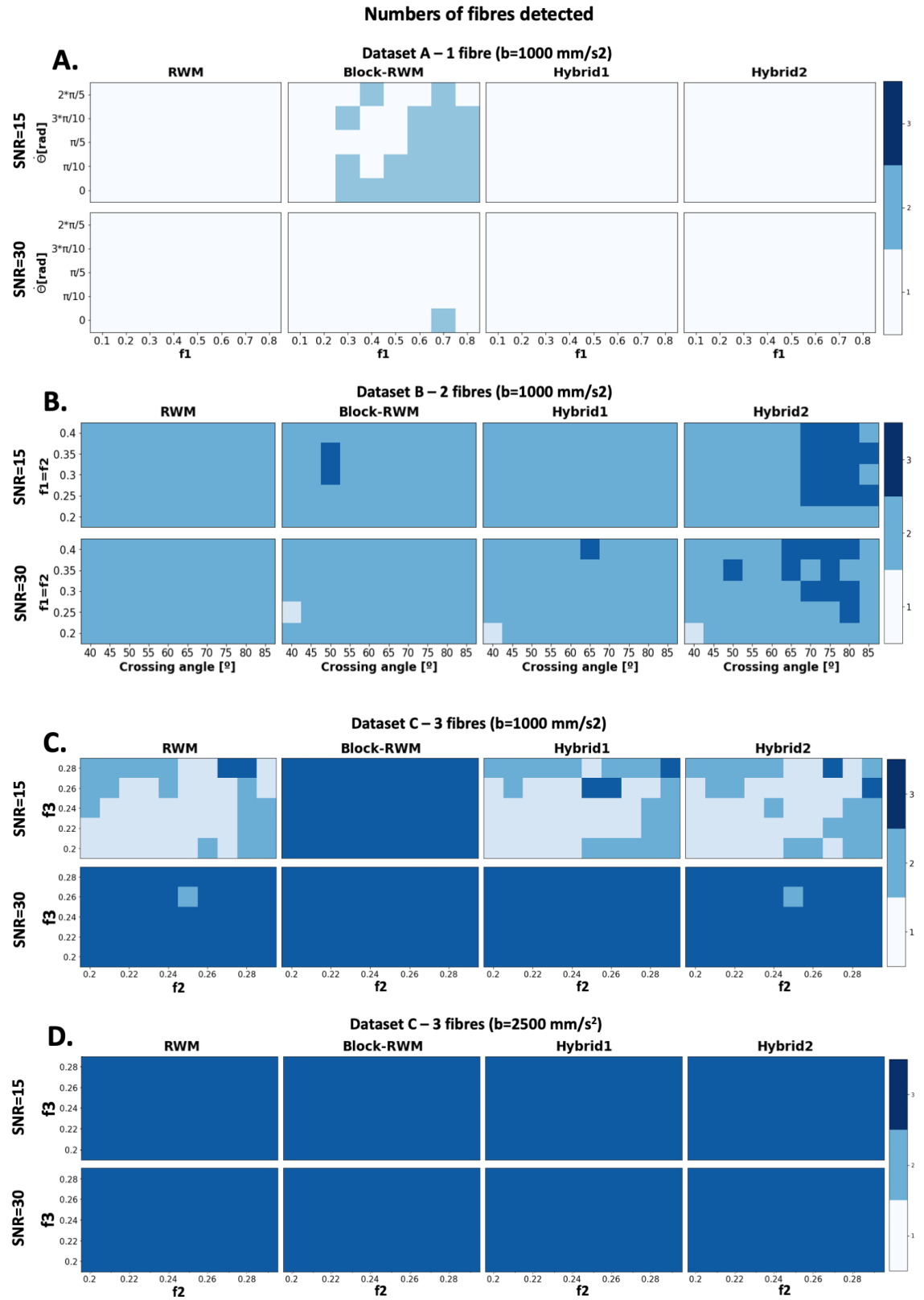


Figure 5.9: *Number of fibres detected at different SNR levels (rows) for each algorithm (columns) for Datasets A (A.), Dataset B (B.), Dataset C (C.), and Dataset C simulated with higher angular contrast ($b = 2500 \text{ s/mm}^2$).*

as the fibre configuration in Dataset A is fairly simple, errors were low for all methods (maximum errors are around 5% for mean diffusivity of volume fraction, and 5 degrees of difference in the fibre orientation estimates). Slight overestimation of the diffusivity was generally compensated with small underestimation of the volume fraction for all methods.

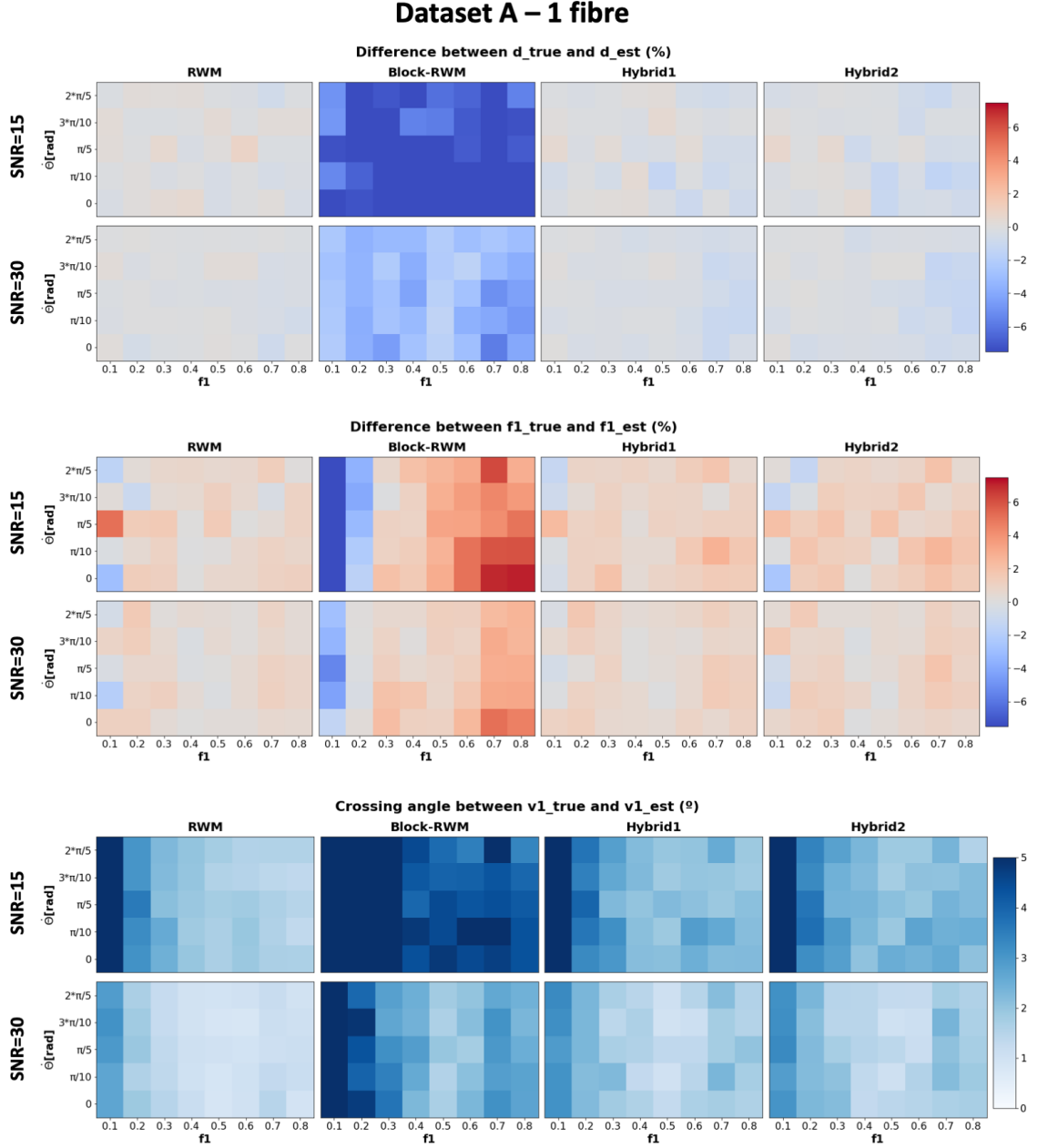


Figure 5.10: *Error maps obtained by each method in Dataset A at different SNR levels.*

Fig.5.11 and Fig.5.12 present the difference between approaches in Dataset B, where **2-fibre crossings** are simulated. As reported in Fig.8 of [Behrens et al., 2007], for relatively low b values and low SNR, the univariate RWM is known to provide a reliable estimate

for fibre configurations with angles between compartments above 40-45 degrees. This is observed also here. All methods tend to overestimate diffusivity and the volume fraction of the second fibre, while underestimating the f_1 . This intercoupling between parameters is much more compensated in the univariate RWM and the Hybrid1 than in the others, as shown in the maps of the f_{sum} (Fig.5.11). The most similar accuracy compared to the RWM is provided by the Hybrid1 although it needs larger crossing angles to provide reliable estimates when $f_1 = f_2 \leq 0.25$. This is also the region where it failed in detecting the correct number of fibres (Fig. 5.9). In these cases, the Hybrid1 returns only 1 fibre with a volume fraction similar to $f_1 + f_2$ and fibre orientation similar to v_1 . Out of that region, the error rates are drastically lower independently of the noise level, similar to the RWM (e.g., less than 10-15 degrees error in fibre orientation estimates).

While the Block-RWM showed good performance detecting the number of fibres, it reported the highest median errors, e.g., up to 40% higher error rates and orientation errors of up to 50° more compared to the univariate RWM. In many crossings scenarios, Block-RWM and Hybrid2, which use a joint-proposal for all the fibre orientations, provided completely perpendicular (and wrong) fibre orientation estimates; updating all the fibre orientations in the same block may be imposing too much structure and inflexibility in the proposal. This would explain the high errors obtained despite returning the correct number of fibres (determined either by a good initialization or by the action of the ARD).

There are also differences in the precision of the estimates. Fig.5.13 shows for different model parameters the dispersion across MCMC samples (median calculated across the noisy realisations). Again, although there are no ground-truth values for the dispersion or the standard deviation here, we used the RWM estimates as a reference, as they have been extensively employed in real brain data, e.g., to guide probabilistic tractography. In this context, the Block-RWM suffers from a heavy underestimation of uncertainty, while the Hybrid-MCMC approaches returned similar ranges of uncertainty to the RWM, with uncertainty decreasing with increasing SNR and higher dispersion in secondary fibres, as expected. This agrees with what has been observed in Fig.5.6. Two reasons can explain this behaviour: 1) a unique joint-proposal can lead to a lack of flexibility that makes it difficult to cover appropriately the multivariate target posterior, and 2) when improper

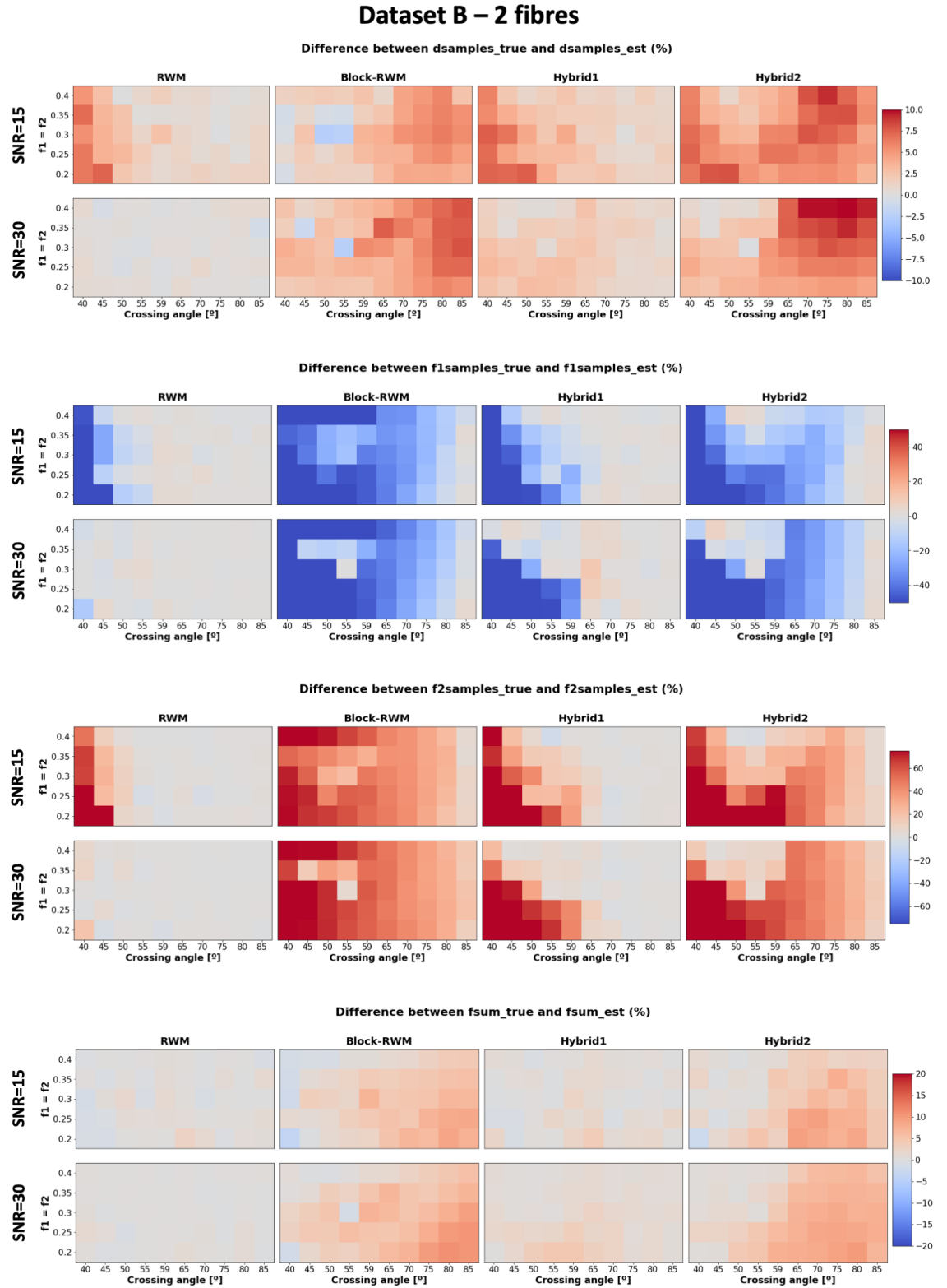


Figure 5.11: *Error maps* obtained by each method in Dataset B for the scalar parameters at different SNR levels.



Figure 5.12: *Difference maps* obtained by each method in Dataset B for the fibre orientations at different SNR levels.

priors (i.e. ARD) are not updated independently, this can lead to a severe underestimation of uncertainty in the posterior parameter estimates. For the above reasons, and taking all previous results together, we do not consider the Block-RWM nor the Hybrid2 any further and we focus on the Hybrid1 approach as an alternative to the univariate RWM.

Fig.5.14 shows exemplar results for Dataset C, where **perpendicular 3-fibre crossings** were simulated. As stated before, at the given low contrast in the images, 3-way crossings are likely to be missed; therefore, median error estimates here are only calculated on realisations where the number of fibres was correctly estimated (across all noisy realisations, around 50% and 75% of the voxels for low and high SNR, respectively). Doing so, differences between RWM and Hybrid1 algorithms are barely noticeable for all parameters and errors returned by both methods are low (less than 5% in scalar parameters and less than 15 degrees for fibre orientations at high SNR). Dispersion maps (see Appendix D.0.1), show similar behaviour in both approaches.

In summary, the Hybrid1 configuration has shown greater flexibility and performance

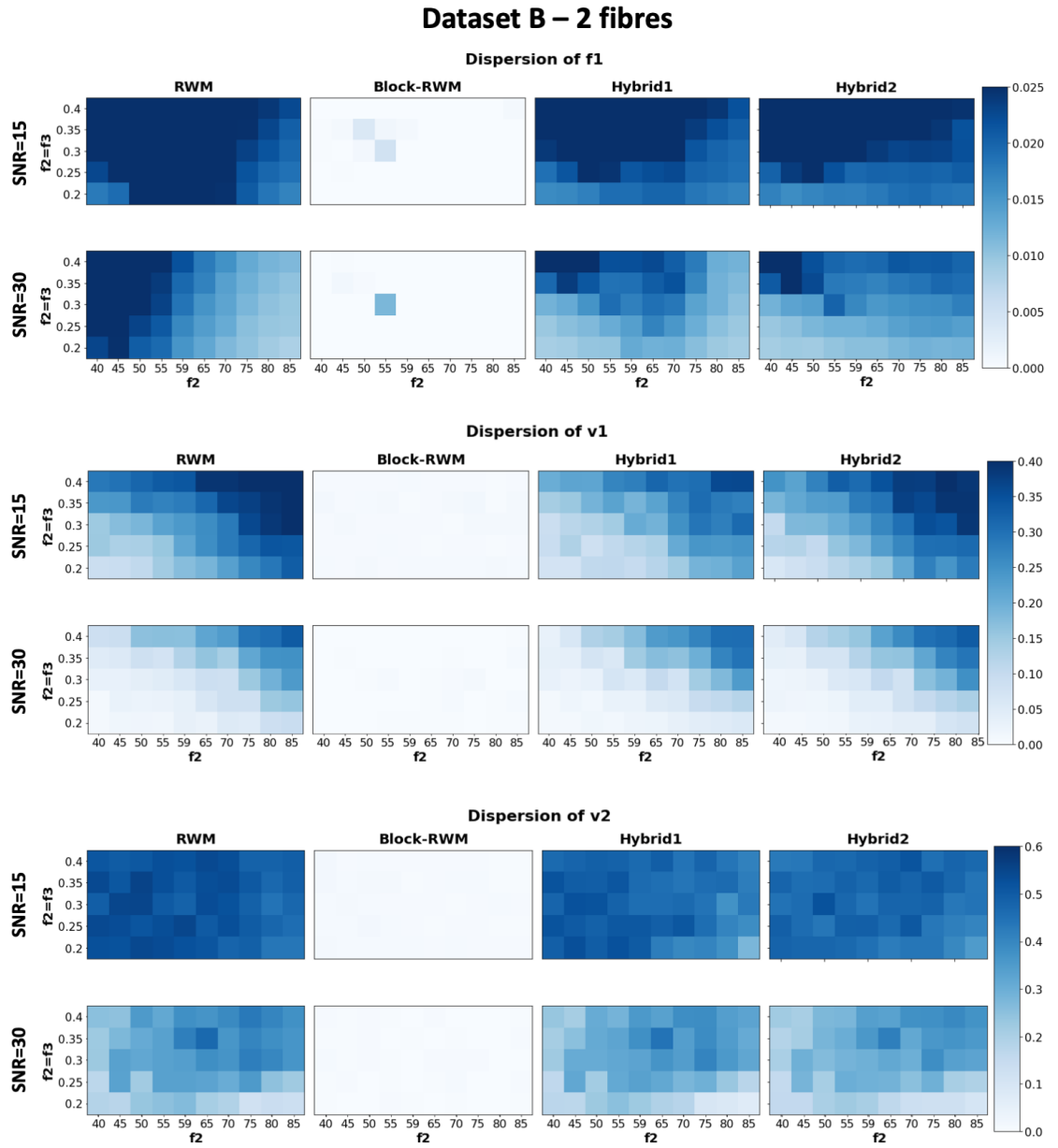


Figure 5.13: *Standard deviation or dispersion* obtained by each method for f_1 and the fibre orientations in Dataset B at different SNR levels.

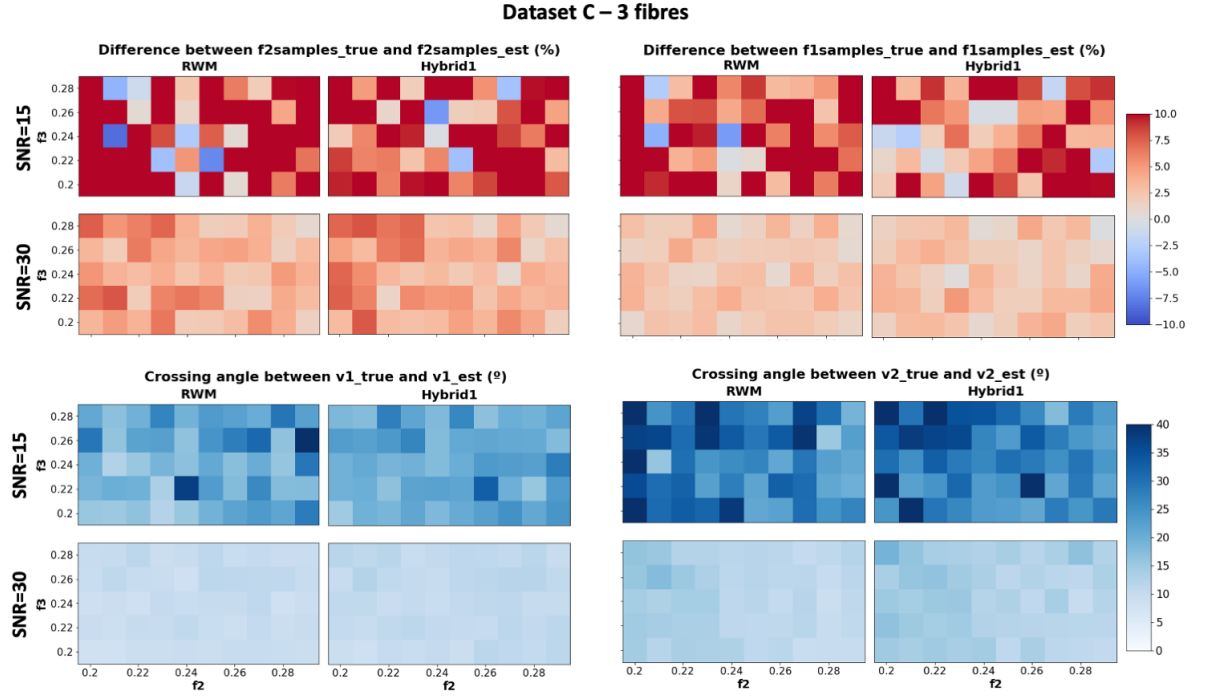


Figure 5.14: *Error maps of the first and second fibres estimates obtained (see eq.5.25 and eq.5.26) by the RWM and Hybrid1 methods in Dataset C at different SNR levels.*

than the other alternatives. It allows for an adaptive algorithm, leaves parameters with improper priors to operate independently and breaks the joint-proposal constraint imposed on the fibre orientations. It has shown similar behaviour to the classical univariate RWM for different noise levels in terms of mixing, acceptance rates, number of fibres correctly detected (i.e. compatible with the ARD), and error estimates. This suggests that similar levels of accuracy and precision can be achieved in a more efficient sampling scheme, but further validation in in-vivo brain data needs to be performed.

Computational speed-up

The theoretical speed-up at voxel-wise inference provided by the Hybrid-MCMC designs against the univariate RWM can be calculated. For instance, for a model with three fibre compartments (11 parameters), Hybrid-MCMC configurations update the parameters in 4 sub-blocks, as explained in Methods. This should provide a theoretical speedup of $11/4 = 2.75$, i.e. the Hybrid1 should need only 36% of the RWM time to infer the posterior density given the same MCMC parametrisation (jumps, thinning, etc.).

Fig.5.15 shows the experimental computation time in a single-core CPU obtained for the different algorithms implemented and different numbers of iterations (*Hybrid1* actually

represents also the Hybrid2, as they require the same computational times). The dashed line indicates the ideal improvement, using the experimental time in the RWM as the reference.

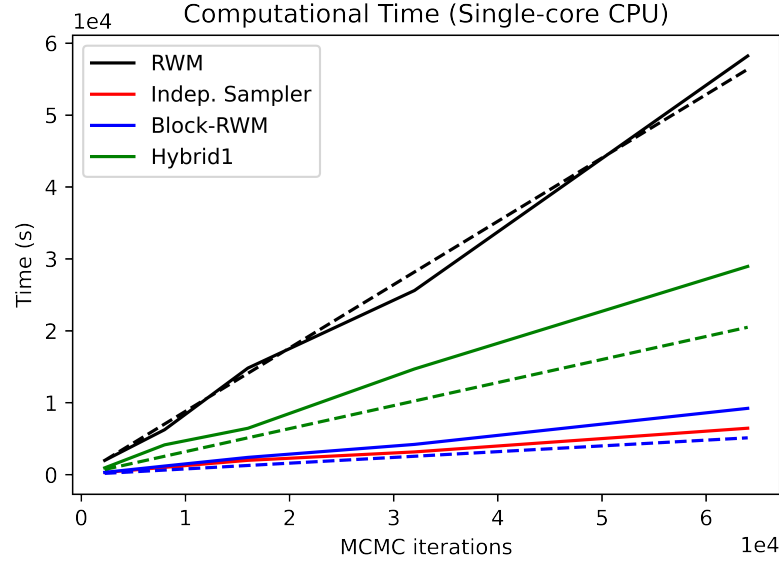


Figure 5.15: Comparison of the computational time required by the different MCMC approaches implemented in function of the number of MCMC iterations. The dashed indicates the ideal behaviour.

The speed-up obtained by the proposed approaches has been found to be lower than the theoretical results. In the case of the Hybrid1, it is around 2-fold (experimental) vs 2.75-fold (theoretical) for the particular model considered here. The deviation can be due to the fact that some of the evaluations carried on in the Hybrid approach, such as the multivariate proposals, are more expensive to compute than in the univariate RWM. Nevertheless, the code used here was not particularly optimised for performance and further algorithmic optimizations may contribute to getting closer to the theoretical speed-ups.

5.4.2 In-vivo data

In this section, we will compare results from the univariate Random-walk MCMC and the Hybrid1 MCMC algorithms applied to in-vivo brain data, as this was found to be the optimal alternative in simulations. First, we evaluate voxel-wise estimates from the Ball&Sticks model. Then, we use these model estimates to perform tractography and assess the agreement between both methods.

Voxel-wise estimates and uncertainty mapping

As a first evaluation, we assessed the mixing and convergence of the chains in randomly-selected voxels from areas where different fibre complexity is anticipated: one from the middle of the Corpus Callosum (one fibre orientation expected), and two from the Centrum Semiovale (2-way and 3-way crossings expected). Fig.5.16 shows the already thinned and burn-in discarded traceplots for both MCMC approaches. With the selected burn-in and thinning, samples show convergence with no sign of autocorrelation and good mixing. This convergence can be also confirmed by 1) the ARD effect, which keeps f_j at 0 in those voxels that don't have secondary or third crossing fibres, and 2) the overlapping between the conditional posterior densities observed in the RWM and the Hybrid1 (Fig.5.17). Hence, proper mixing of the chains seems to be achieved by both MCMC approaches.

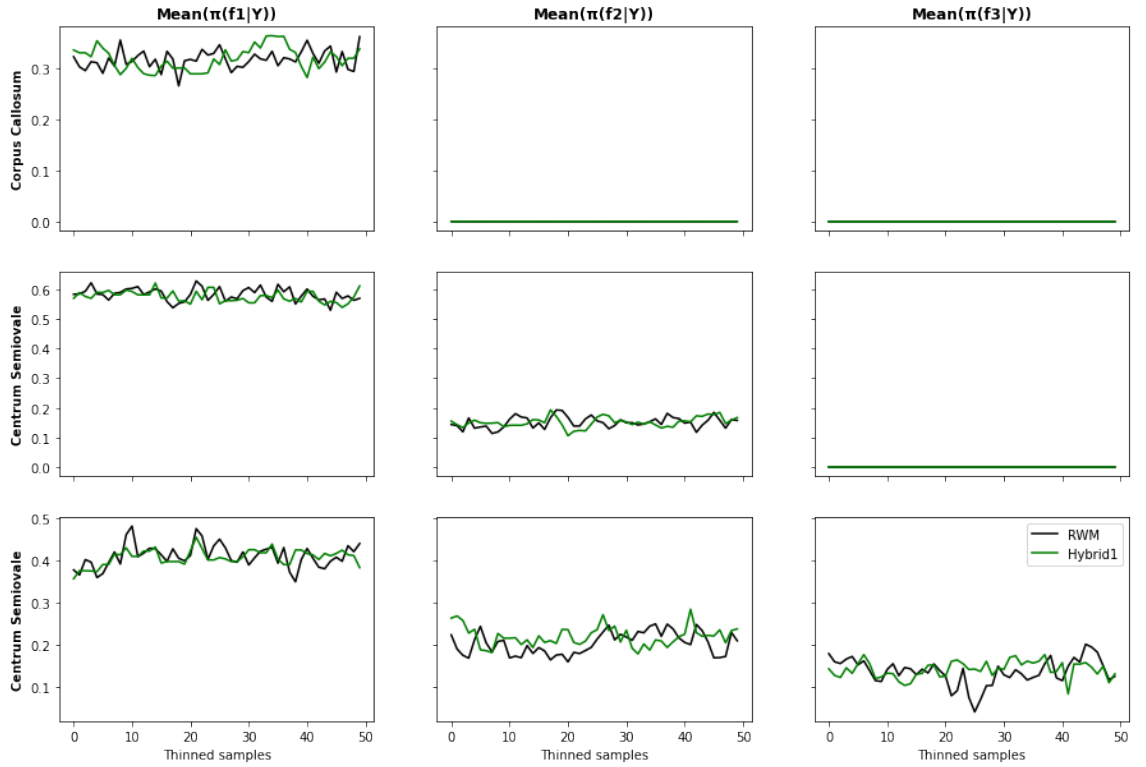


Figure 5.16: Comparison of MCMC Markov Chain traceplots in the RWM and the Hybrid-MCMC for 3 representative voxels in white matter.

Fig.5.18 shows a comparison between **posterior mean maps of scalar parameters** obtained for an exemplar axial slice. For volume fractions, we report differences and scatter plots where $\langle \pi(f_j|Y) \rangle > 0.05$ (i.e. where these compartments are supported by the data) in both the RWM and the Hybrid1 approaches; for the mean diffusivity, we report values

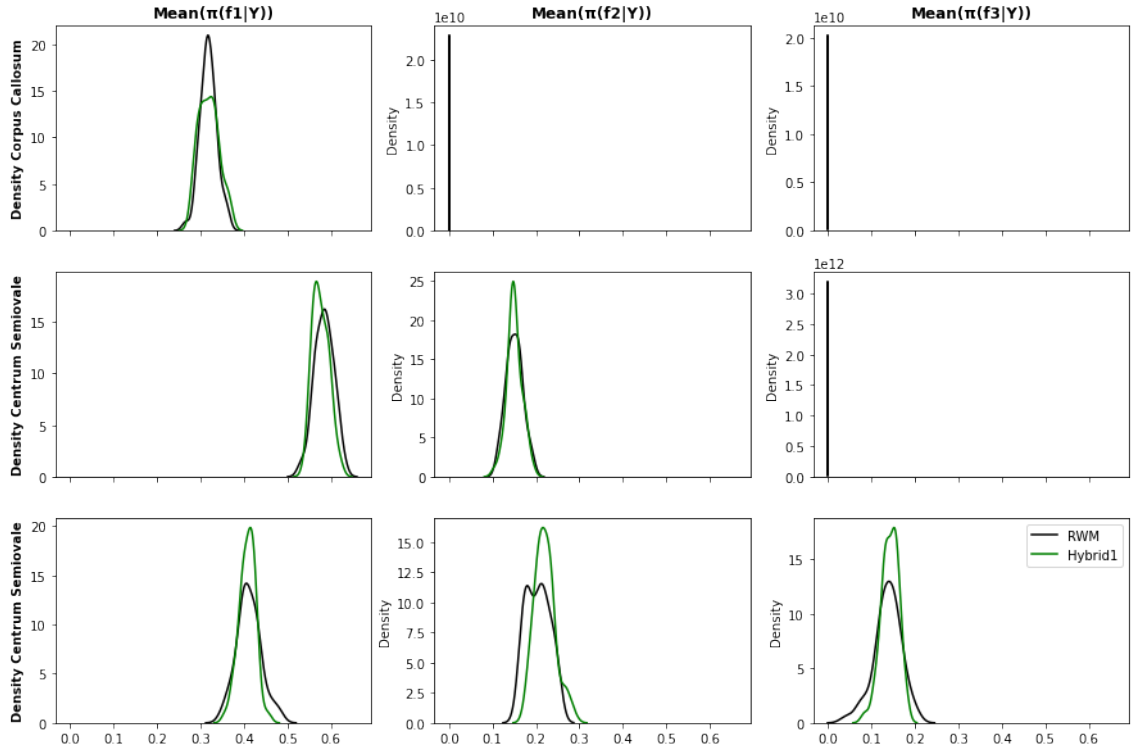


Figure 5.17: *Posterior density for the same voxels as in Fig.5.16*

in the whole WM. We can qualitatively observe that there are no noticeable differences between the map contrasts obtained for d and f_1 . Differences are normally distributed and there is a high correlation r between mean estimates of both approaches ($r_d = 0.92$ and $r_{f_1} = 0.96$). Higher differences are observed in secondary fibres. The RWM seems to provide slightly higher f_2 values (see small right tail in the histograms or the predominant *red* colour in the residual map); higher (and more accurate) f_2 values in RWM were also observed in some simulated patterns (e.g., see small crossing angles Fig.5.11). Nevertheless, the correlation is still high ($r_{f_2} = 0.86$) and mean maps look considerably similar. These slightly higher f_2 values compared to the Hybrid1 are accompanied by a slightly lower number of 3-way crossings. The Hybrid1 approach seems to have a higher sensitivity, capturing a higher number of 3-way crossings within regions like the Centrum Semiovale.

Examples of **fibre orientation estimates** are shown in 5.19. Qualitatively, both approaches return plausible and similar fibre-crossing patterns. They both depict complex orientations in the Centrum Semiovale crossing from left to right, inferior to superior and anterior to posterior, corresponding to callosal, pyramidal and longitudinal tracts respectively. Differences (in degrees) are quantified in Fig.5.20 between the dyadic orientation

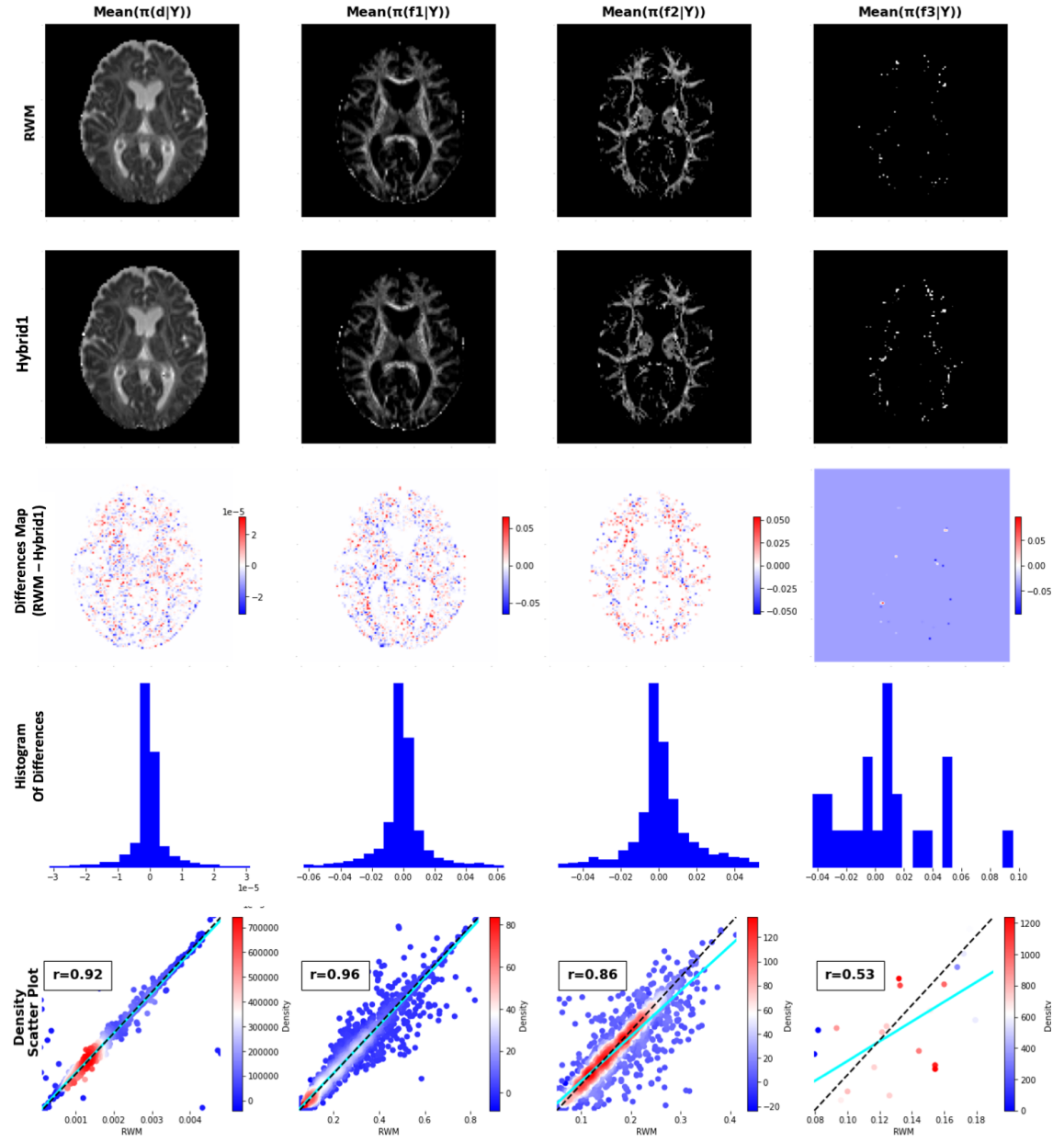


Figure 5.18: Comparison of mean estimates between the RWM and the Hybrid-MCMC. Differences and scatter plots are calculated only in WM. In the scatter plot: the dashed line shows the ideal perfect correlation (diagonal); the light blue line indicates the actual linear relationship between estimates.

vectors v_j provided by each approach. Differences are in general small (95th percentile of differences: 17 degrees for v_1 , and 36 degrees for v_2). They only become more substantial in cases of 3-way crossings, due to the Hybrid1 approach returning slightly more crossings than the RWM. Overall, however, both methods return a considerably smaller degree of 3-way crossings, compared to 2-way crossings, as expected.

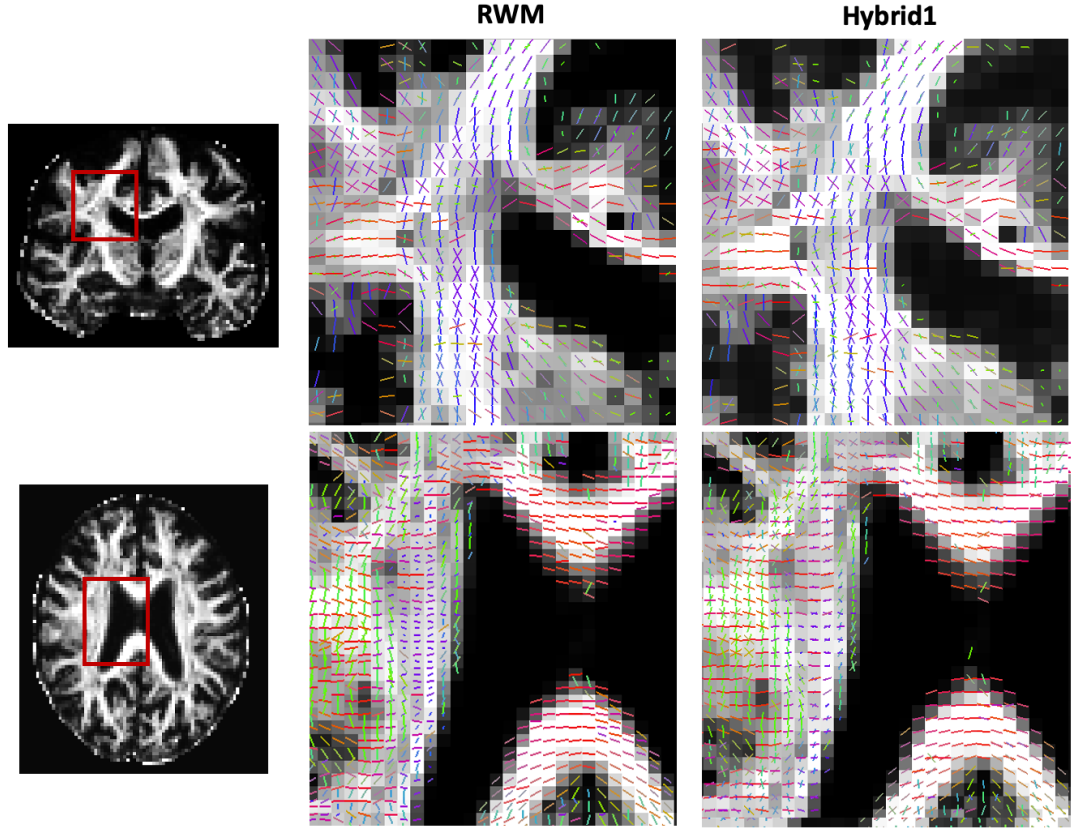


Figure 5.19: Fibre orientation estimates (mean) from the RWM and the Hybrid-MCMC in the Centrum Semiovale (coronal view - top row, axial view - bottom row), overlaid on the mean anisotropy map (mean of the sum of all anisotropic volume fractions f_j).

Finally, results for the **fibre orientation dispersion** are shown in Fig.5.21. Differences are presented in white matter, as more relevant for tractography. The Hybrid1 tends to return smaller uncertainties, particularly for secondary fibres (see circled area in the Density Scatter Plot of Fig.5.21). Nevertheless, dispersion maps in the Hybrid algorithm still follow similar patterns as the RWM, especially for v_1 and v_2 (i.e., low uncertainty in WM and high uncertainty in CSF and GM). The underestimated uncertainties are randomly scattered throughout the whole brain (e.g., see *Differences map* in Fig.5.21) and differences between both approaches are 0 on average (see *Histogram of Differences* centered at 0). To further explore the impact of these differences in tract propagation, we performed

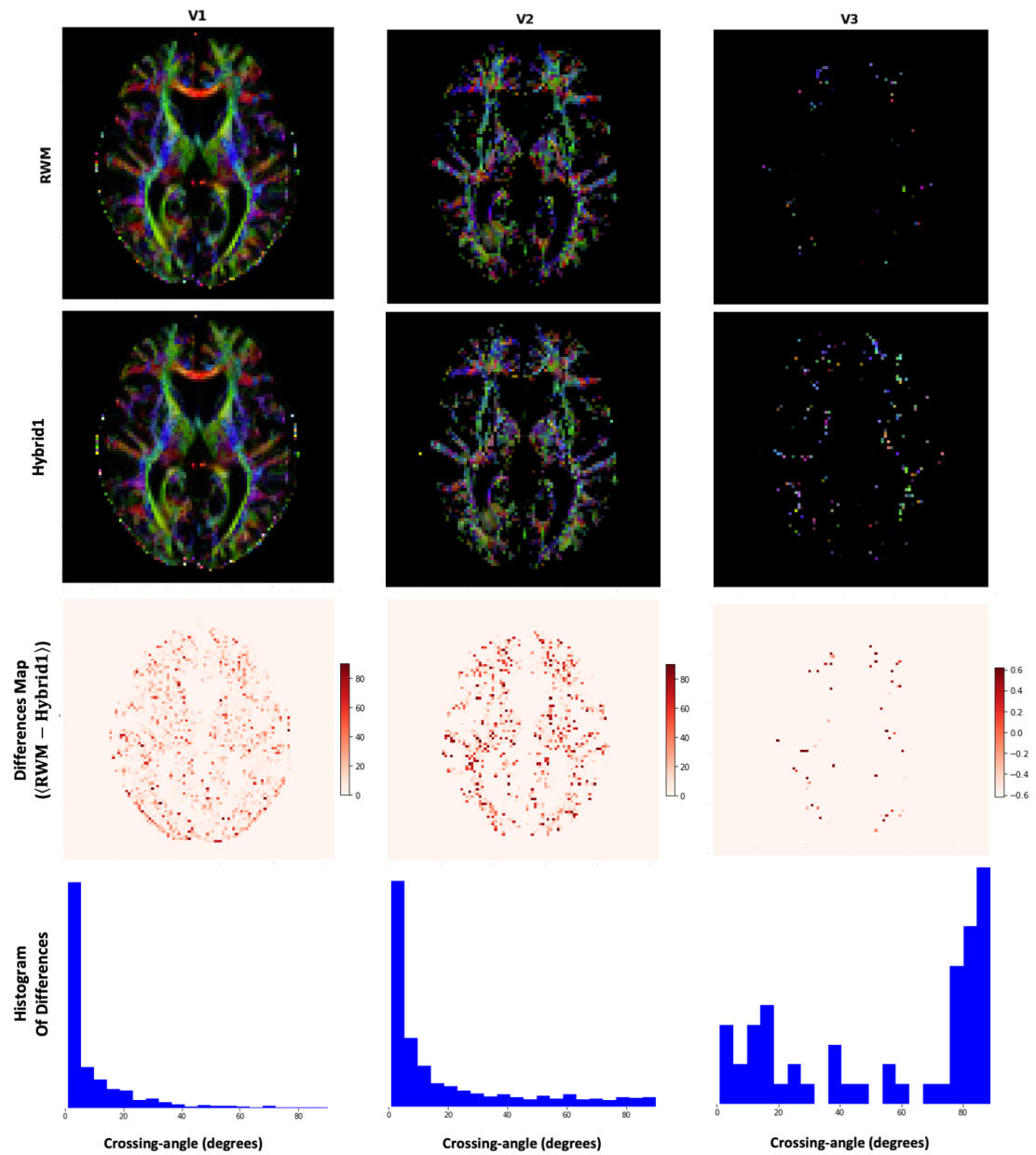


Figure 5.20: Comparison of fibre orientation estimates between the RWM and the Hybrid-MCMC.

tractography and compared results obtained using the orientation estimates from RWM and Hybrid1.

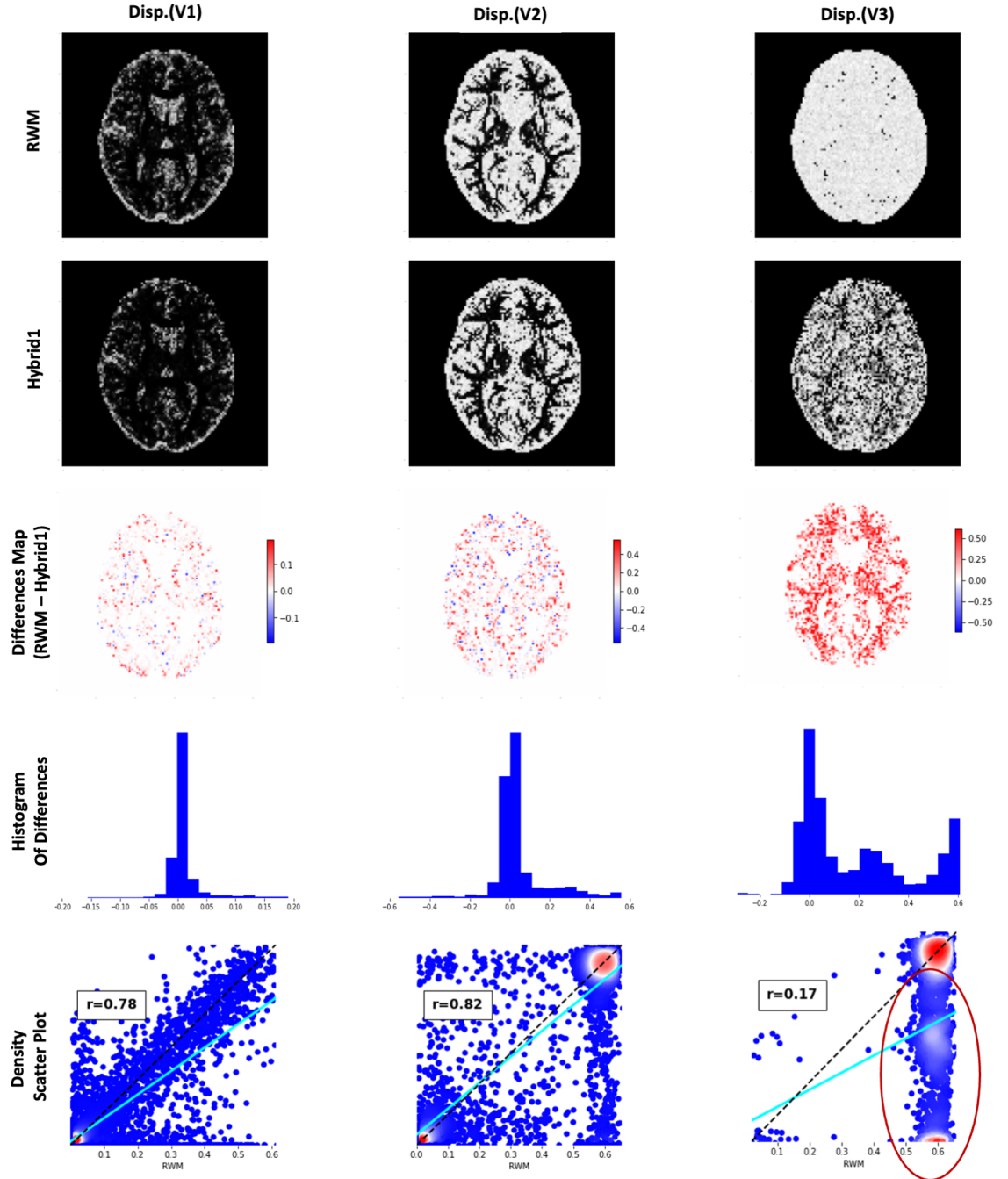


Figure 5.21: Comparison of dispersion estimates between the RWM and the Hybrid-MCMC. Differences and scatter plots are calculated only in WM. In the scatter plot: the dashed line shows the ideal perfect correlation (diagonal); the light blue line indicates the actual linear relationship between estimates. The circled area refers to the points where the uncertainty was underestimated in the Hybrid1.

Probabilistic Tractography

We compared tracts reconstructed using XTRACT [Warrington et al., 2020] and the voxel-wise orientation estimates obtained from the RWM and the different Hybrid-MCMC configurations. Fig.5.22 shows the Maximum Intensity Projection (MIP) of a set of tracts from a single subject as an exemplar case for the RWM (left column) and the Hybrid-MCMC configuration 1 (right column). All MCMC parameters were kept as default. Visually, both sets of tracts seem to be almost identical. To quantify the similarity between them, Fig.5.23 represent the spatial tract correlation between both MCMC approaches across 15 subjects of the UKBiobank dataset. Comparing tract correlations by subject (top row), we can see a high level of agreement across subjects, with all median correlations ≥ 0.90 . There are few outlier tracts with low correlation, that correspond to tracts with known complexities, e.g., thin tracts such as the Anterior Commissure (*ac*), with complex geometries such as the Fornix (*fx*), or tracts with multiple crossings such as the Superior Longitudinal Fasciculus 2 (*slf2*). In fact, these are the set of tracts that also demonstrate lower correlations (bottom row of Fig.5.23), as expected.

An example of outliers tracts detected in the results above is shown in Fig.5.24 with the Fornix (both left and right) from the subject with the lowest correlation in this tract. It can be seen that tracts in fact are not reconstructed successfully with any of the approaches (even if the Hybrid1 looks slightly better), as the single-shell version of the UKBiobank dataset used (50 directions, $b = 1000 \text{ s/mm}^2$) is not good enough to allow good reconstruction of this thin pathway. Nevertheless, similar to the subject's correlations, most of the group of tracts returned median correlations above 0.9.

In the absence of ground-truth for these tracts, we used the HCP WM population atlas [Warrington et al., 2020] as gold-standard and compared the performance of the RWM and the Hybrid1 against it. Similar to the above, Fig.5.25 shows the spatial tract correlation grouped by subjects (top row) and by tracts (bottom row) of each MCMC approach with respect to the atlas. In both cases, the spatial correlation distributions are very similar and follow similar patterns, i.e. harder to reconstruct tracts agree less with the atlas for both methods. The Hybrid1 returns on average almost identical median correlations to the atlas compared to the RWM: 0.56 (std. 0.11) vs 0.55 (std. 0.11), by averaging subjects

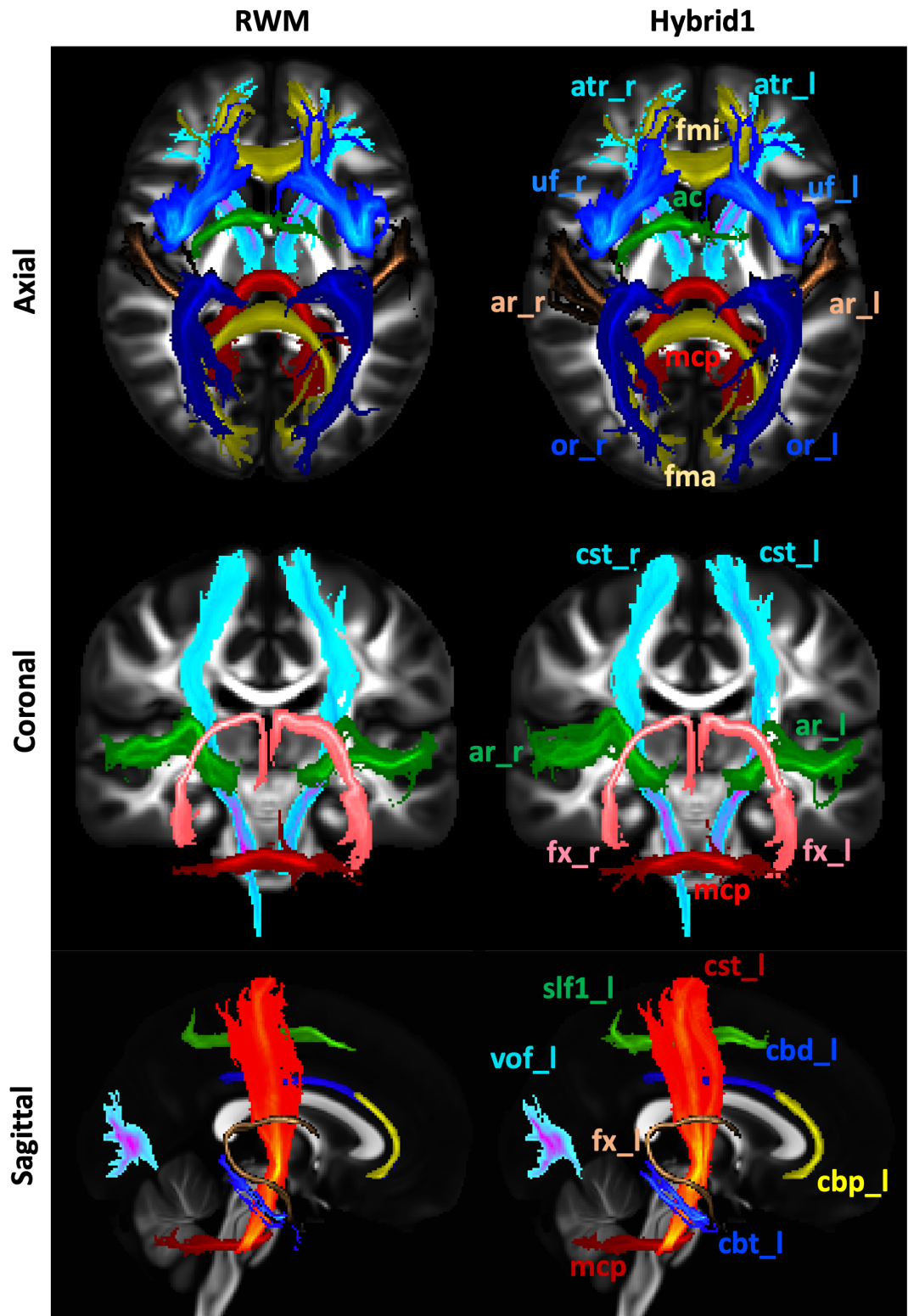


Figure 5.22: *Maximum Intensity Projection (MIP) of different tracts obtained by using estimates from the RWM and the Hybrid1. The abbreviation definition for each tract can be found in [Warrington et al., 2020]*

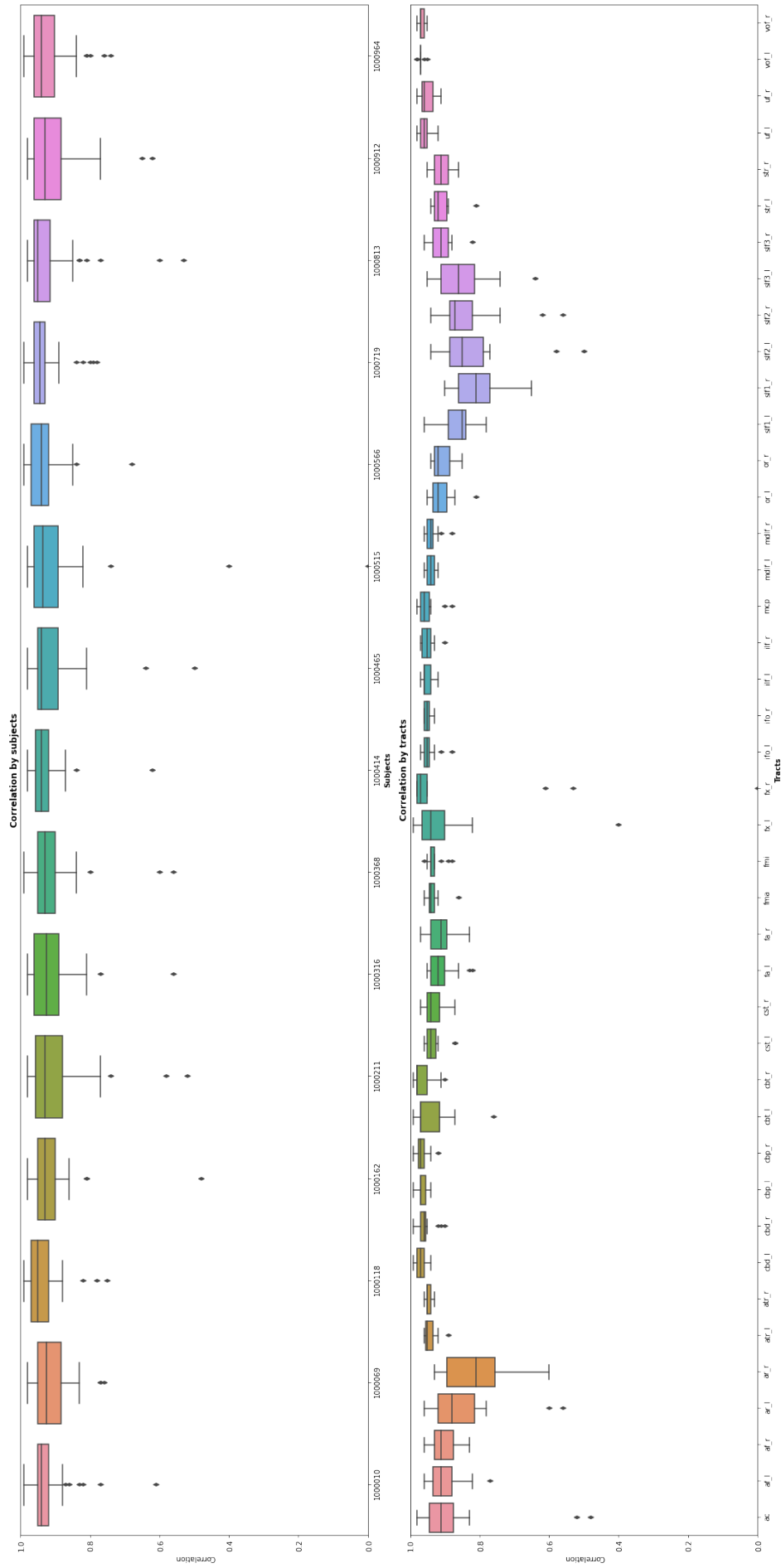


Figure 5.23: Spatial Tract Correlations between the RWM and the Hybrid1 - Correlations are grouped by: tracts belonging to each of the selected subjects from the UKBiobank dataset (Top row); subject correlations obtained for each tract defined in the XTRACT protocol (Bottom row). Colours are assigned randomly and do not contain information. The abbreviation definition for each tract can be found in [Warrington et al., 2020]

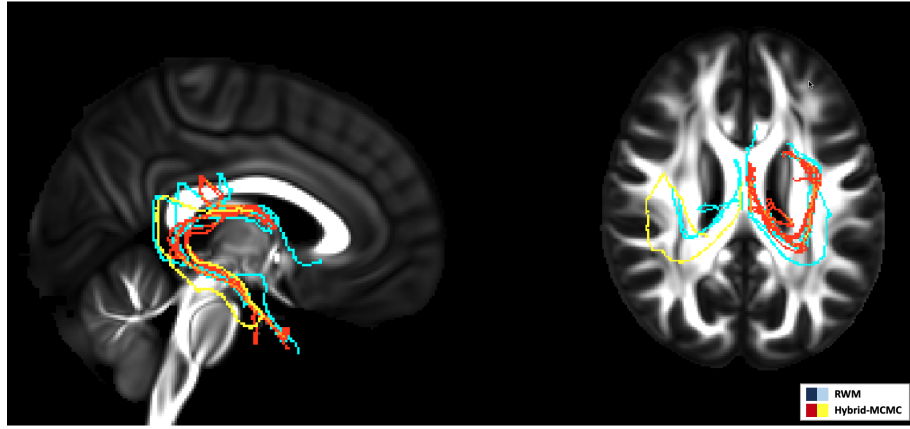


Figure 5.24: Example of outlier tract - Comparison of the MIP in the Fornix tract (subject 1000515) obtained from the RWM and the Hybrid-MCMC.

per tract, respectively; and 0.56 (std. 0.023) vs 0.55 (std. 0.024) by averaging tracts per subject.

Results obtained from a longer run of the Hybrid1 (burn-in=3000, thinning=100, jumps=5000) can be found in the appendix D.1. Increasing the number of MCMC iterations provides slightly higher correlations overall (see Fig.D.2). However, it does not recover the outliers observed while increases computational time, in this case, by 5-fold (i.e. this is dependent on other factors, such insufficient as angular contrast). Hence, keeping the Hybrid-MCMC configuration 1 with the *default* MCMC parameters provided performance and results that are in a high level of agreement with current univariate MCMC implementations while providing the speed-up commented in sec.5.4.1.

5.5 Discussion

The adaptive component-wise Random-walk MH-MCMC (RWM) is an established method heavily used in dMRI over the last 15 years. However, the idea of using a random walk process to sample the parameter space, independently for each model parameter, can be inherently inefficient and does not scale well with larger dimensionality. In this work, we have designed, implemented and evaluated different MCMC alternative modifications based on the idea of proposing and sampling multiple parameters at the same time (in blocks) to improve the algorithmic efficiency of the RWM while preserving accuracy. More specifically, we have built upon the idea of an Independence Sampler, and devised an adaptive *all-at-once* Block-RWM, and two Hybrid-MCMC configurations. All of them relied

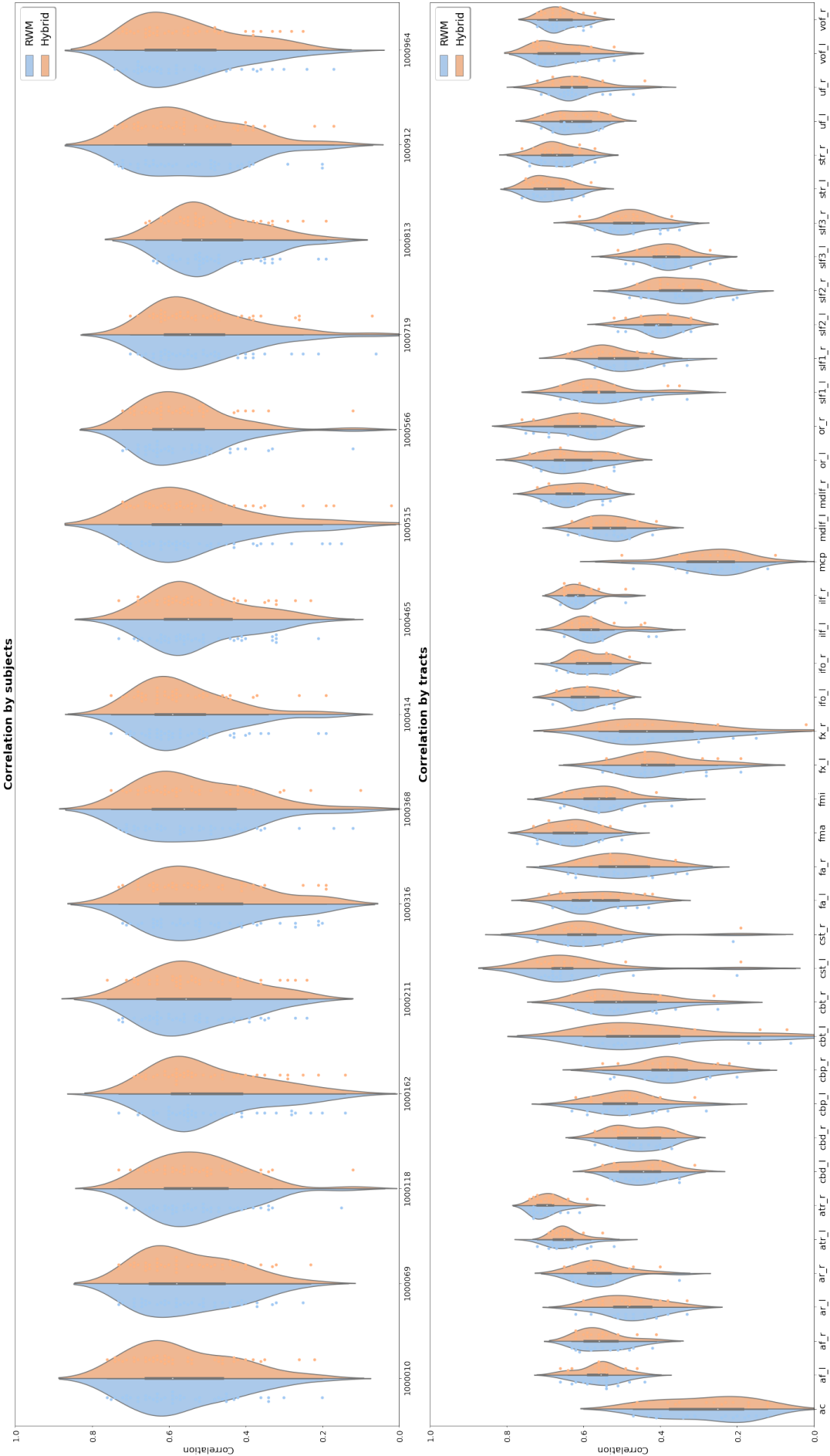


Figure 5.25: Spatial Tract Correlations with the HCP population-averaged atlas. MCMC approaches compared are the RWM (blue) and Hybrid-MCMC (orange) with default MCMC parameters. Correlations are grouped by: multiple subjects from the UKBiobank dataset (Top row); Defined tracts from the XTRACT protocol (Bottom row)

on the Laplace approximation as an efficient and principled manner to define the initialization and block-covariance of parameters.

A novelty in our designs is that we addressed the challenge of including improper priors in the model, such as shrinkage priors, in the form of automatic relevance determination (ARD). And at the same time, we allowed for adaptive proposals and therefore efficient mixing. The more conventional Independence Sampler was quickly discarded due to poor performance but provided an important building block for the subsequent approaches. The **adaptive Block-RWM** considers all parameters in a single block and introduces 1) iterative adaptation for the block proposal covariance matrix Σ depending on the acceptance ratio, constraining/expanding the Laplace approximation when it becomes too smooth/tight in higher dimensions and reducing the number of rejections; and 2) the multivariate proposal is re-centred to the last accepted sample, which helps to converge to the target distribution (i.e., introduces the Markov-Chain). However, it showed unstable behaviour in cases where the model fitted does not agree with the underlying fibre configuration and also a significant underestimation of the width of the posterior distribution (i.e., the uncertainty of parameters). This can be due to the inflexibility of the proposal and too much structure imposed by having a unique block, making it challenging to mix properly in the presence of improper priors. In fact, updating all fibre orientations in a joint proposal showed similarly challenging behaviour, as observed in the Hybrid2.

One potential reason for these issues is that all the parameters in the Ball&Sticks model are not necessarily correlated. E.g., the existence of one compartment does not determine the existence of the other. Despite that, only one covariance matrix Σ is considered for all of them in the Block-RWM, which is globally modified by λ_L based on the average acceptance ratios of the parameters. Therefore, the covariance of the parameters is not optimally tuned and the results may vary depending on the fibre configuration complexity found in each voxel. This can lead to acceptance of samples that are only close to the modes, returning as a consequence very narrow posteriors as the ones observed.

The **Hybrid1** design takes these issues into consideration and allows for greater flexibility; it allows for an adaptive algorithm, using optimal subsets of the parameters in order to

accommodate for the coexistence of proper and improper priors and the multiple compartments being updated independently on each other. In doing so, it showed comparable accuracy to the established RWM in both simulated experiments and in-vivo brain data (with correlations of almost 1 for some parameters). We observed some differences at voxel-wise level estimates of the secondary fibres, with smaller volume fractions returned by the Hybrid1 compared to RWM in in-vivo brain data. These were typically accompanied by more 3-way crossings. Nevertheless, these observed deviations showed a spatially-random distribution and did not affect higher-level analysis that uses these orientations, such as tractography. Both the RWM and the Hybrid1 reported a very high level of agreement in tractography results. This has been achieved while providing a computational speed-up of 2x, which can be particularly helpful when processing large datasets (e.g., UKBiobank). But more importantly, this paves the way for Bayesian inference or more complex biophysical multi-compartment models [Alexander et al., 2019; Sotiropoulos et al., 2016], where dimensionality becomes a limiting factor in practice for the random-walk MCMC.

Recently, a number of parallelisation approaches have been proposed to accelerate random-walk MCMC [Hernández et al., 2013; Hernandez-Fernandez et al., 2019; Kim et al., 2022; Madhyastha et al., 2017]. These are based on the power of GPUs and parallelise effectively the model fitting across image voxels. Our approach is compatible with such parallelisation techniques, but also offers a design improvement rather than only a computational speed-up, which offers the potential for even larger scalability.

Limitations and Future Work

Despite the extensive number of experiments performed there are yet certain evaluations that could be assessed to validate the Hybrid1 as a generalisable alternative of the established RWM. For instance, we have used a specific single-shell biophysical model (Ball&Sticks) to evaluate MCMC inference performance, but other models should be evaluated. Furthermore, factors like tractability of the likelihood, model complexity, or the lack of analytic Hessian have not been tested and can influence the results.

There is also an inherent limitation in this approach given trade-off n_{jumps} vs speed-up: the more jumps we used in the Hybrid-MCMC, the more similar it can become to the

RWM results, but the lower the speed-up. On the other hand, the Laplace approximation used here requires the calculation of second derivatives to initialize the covariance. In this case, it was possible to do it analytically. However, there are multiple scenarios where this is not possible and numerical differentiation is needed, reducing the computational speed-up and introducing potential inaccuracies.

Nevertheless, the Hybrid-MCMC allows also for further algorithmic optimization. For instance, GPU tree-like parallelization at two levels, across volumes and across likelihood calculations, such as in [Hernandez-Fernandez et al., 2019] can be investigated in future work. Also in this direction, new Machine Learning approaches have been proposed in the last years to learn the covariance and perform model inversion in an unsupervised way [Cranmer et al., 2020]. We will explore more about this line of research in the next chapter.

We observed a general trend of underestimation of the fibre orientation uncertainties at voxel-wise level in the block approaches. Even if the effect was less noticeable for Hybrid1 and we did not see demonstrations in probabilistic tractography results, it is unclear how this can affect other scenarios. One potential reason might reside in the incapability of the block-approach to cover appropriately multimodal posterior distributions. As we start from the LA, which is a local approximation to the modes, there may be too much shrinkage around them. Nevertheless, it is worth noting that there is no ground-truth for uncertainty and these conclusions are obtained from comparisons with the univariate RWM. As commented above, in the next chapter we will explore alternative inference frameworks relying on ML approaches that also allow us to skip the definition of the block proposals and covariance matrices.

Chapter 6

Amortised simulation-based inference in diffusion MRI

6.1 Introduction

Inference and uncertainty quantification in imaging-based modelling provides a principled way of estimating non-linear model parameters, but also assessing confidence on results [Jones, 2003], quantifying noise effects [Behrens et al., 2003], and aiding experimental design [Alexander, 2008]. In the past chapter, we explored algorithmic improvements on classical Bayesian inference approaches (MCMC) to improve their sampling efficiency. In this chapter, we devise and evaluate a different -in principle- framework, based on the idea of generative modelling using neural networks.

Machine learning has been used before for model fitting in the context of dMRI (see overview in 3.3.4). However, most of the approaches commented on provide point-estimate fitting and do not express uncertainty over model parameters. In the approach presented in this chapter, we devise an inference approach using neural networks and apply it to a dMRI microstructural model. The assumption is that using enough complexity in a considered network trained on synthetic data, we can learn an arbitrary complex mapping between the posterior distribution of model parameters and observations [Cranmer et al., 2020]. We can then use this mapping to perform model estimation and inference in the parameters using new data. Given the high data dimensionality with modern dMRI protocols (e.g., [Sotiropoulos et al., 2013b]), the large space of parameters, and/or the

complexity, or lack of tractable likelihoods in some biophysical models, this approach can be an attractive alternative to traditional inference approaches.

The presented framework builds upon the explosive growth of deep learning over recent years, which has made possible network-based inference. Advances in **deep generative methods** have enabled frameworks that use data to learn *representations* of any target latent density, such as the posterior distribution of model parameters. This can be cast as a problem of density estimation and there exist several interconnected approaches in the current literature [Bond-Taylor et al., 2022]. One of the most popular is Generative Adversarial Networks (GANs), which learn to represent latent, high-dimensional distributions in an adversarial manner [Goodfellow et al., 2014]. Another class of generative approaches, termed "likelihood-based", seeks to learn a model represented by, mainly, neural networks that assign a high likelihood to the observed data samples. This includes *Auto-Regressive models* (AR) [Bengio et al., 2003], *Normalizing Flows* (NF) [Tabak and Turner, 2013; Tabak and Vanden-Eijnden, 2010], and *Variational Autoencoders* (VAEs) [Kingma and Welling, 2014]. New capabilities have been opened recently by *diffusion models*, which have both likelihood-based and score-based interpretations and are considered to be as powerful as GANS, without requiring the complex adversarial training [Dhariwal and Nichol, 2021; Ramesh et al., 2022; Saharia et al., 2022]; and Generative Flow Networks (GFlowNets) [Bengio et al., 2021], which combine principles from reinforcement learning, deep generative models and energy-based probabilistic modelling to allow for non-parametric Bayesian inference. All in all, network-based inference is a relatively new field with a lot of synergies between methods and a very quick pace of development.

In this work, we will focus on likelihood-based approaches. More specifically, we explore Simulation-Based Inference (SBI) frameworks [Cranmer et al., 2020], which allow us to combine learning latent features of importance from networks (e.g., Deep Nets), the interpretability of mechanistic models (e.g., multi-compartment models) and Bayesian principles. SBI uses forward models to provide a data-driven alternative to classical inference based on training. Generally, an artificial neural network (ANN) learns a Bayesian model inversion, allowing inference for unseen data without the need for repeating the model inversion (amortised inference). We present here one of the first applications of

SBI in voxel-wise microstructural modelling for dMRI data. Importantly, we also directly compare SBI with classical Bayesian inference for the same model using MCMC. This new perspective can open new opportunities for solving likelihood-free problems in neuroimaging and/or provide massive computational speedups in the inference process.

The chapter is organised as follows: the following section overviews the theory on Simulation-Based Inference and modern density estimation methods. In *Methods*, we describe the design and implementation of the SBI approach for the inference of voxel-wise parameters, using the Ball&Sticks model as a proof-of-concept. Results follow, where we evaluate the novel approach in simulated and in-vivo brain data and compare it against MCMC. The chapter ends with a discussion about the results obtained, contributions, limitations, and the potential future steps of this line of research.

6.2 Theory

6.2.1 Simulation-Based Inference

Classical Bayesian inference methods, such as MCMC, requires a likelihood function, $\pi(Y|\omega)$, to provide samples from the posterior distribution of parameters $\pi(\omega|Y)$. However, this approximation typically needs a non-linear iterative process, which can be challenging and expensive. Furthermore, there may also be cases where simulations can be readily obtained from a modelling process, but inversion becomes analytically and/or computationally intractable (e.g. due to complex likelihoods). A group of methods have been proposed to overcome such challenges by making use of simulations from a forward generative process $f_Y(\omega)$ [Diggle and Gratton, 1984; Gourioux et al., 1993], which implicitly defines a likelihood $\pi(Y|\omega)$ (that may be tractable or not). These can be used to learn a mapping (and its inversion) between parameters and observations.

This principle was first followed by **Approximate Bayesian Computation (ABC)** [Beaumont, 2019; Beaumont et al., 2002]. ABC are simulation-based methods that aim to approximate the posterior density by generating large amounts of data Y_{sim} from a forward generative process $f_Y(\omega)$, with parameters ω drawn from defined prior distributions $\pi(\omega)$. A set of summary statistics $S(Y)$ can be calculated for each simulated sample so

the dimensionality of the problem is reduced. The problem relies then on finding the set of parameters ω that best mimics the observed data Y_{obs} (or $S(Y_{obs})$), without the necessity to compute or invert a likelihood. One of the initial and simplest ABC approaches is Rejection-ABC [Pritchard et al., 1999; Tavaré et al., 1997]. To decide what parameter proposals to keep, an ad-hoc distance function ρ and a tolerance threshold value ϵ are defined, such that ω that provide $\rho(Y_{sim}, Y_{obs}) < \epsilon$ are accepted as samples from the posterior distribution; they are rejected otherwise. This process is repeated as long as needed to obtain the density posterior. As Rejection-ABC samples from the prior, and the posterior is typically much narrower than the prior, most of the proposals are rejected for a given observation, i.e. there is a trade-off between the number of accepted samples and how conservative ϵ can be. Algorithms to improve sampling efficiency have been proposed on top of the rejection scheme, such as MCMC-ABC [Marjoram et al., 2003] or Sequential Monte Carlo (SMC) [Sisson et al., 2007], which iteratively re-define proposals by perturbing parameters that have been already accepted to explore more often the areas where parameter values have shown higher posterior density.

However, ABC methods have important **shortcomings**, including the sensitivity to ad-hoc decisions (namely, $S(Y)$, ρ and ϵ), and the *curse of dimensionality*. Given the dimensionality of some problems and the poor scalability of most of these methods, ABC often needs the definition of summary statistics that inherently leads to information loss and worsens the approximation. At the same time, this makes it harder to define *similarity* conditions between observed and simulated data. E.g., the smaller the ϵ , the better the approximation but the more samples are likely to be rejected; and this trade-off becomes more critical when using high-dimensional or multimodal data.

Regression Adjustment techniques for ABC introduced a new perspective by assuming that there exists some structure that relates the set of plausible values ω and the observed data. For instance, [Beaumont et al., 2002] hypothesized that such structure can be defined by a linear regression model with fixed variance. The parametric model is trained on simulated data in order to learn a mapping from Y_{sim} to ω . This learnt mapping is then used to correct the values of posterior samples gathered by e.g., Rejection-ABC to reduce discrepancies between the simulated and observed data. This adjustment increased the

number of samples accepted from 5% in Rejection-ABC to 20%, approximately. Later, [Blum and François, 2009] relaxed the constraints and let non-linear Regression Adjustments be calculated using feed-forward neural networks (NNs), increasing now the rate of accepted samples up to 90% and, therefore, allowing for much higher dimensionalities and ϵ values.

These developments have led to a shift of perspective: if making these adjustments allows us to accept almost all proposed samples, the inference problem can then be recast: from using iterative sampling approaches that rely on model inversion to finding the latent underlying relationship between observed data and model parameters in a data-driven manner using an abstract descriptor, such as a neural network. This in principle allows to approximate the exact posterior distribution (i.e. $\epsilon = 0$) of parameters conditioned on observed data with the aid of a prior over parameters and a stochastic simulator, without computing potentially intractable log-likelihood nor defining distance metrics or tolerance thresholds. This is a pattern recognition problem where Machine Learning algorithms excel and high-dimensionality problems can be tackled. This combination of Bayesian principles and Machine Learning to approximate not only the posterior but any target density distribution is what currently defines the field of **Simulation-Based Inference (SBI)** [Brehmer, 2021; Brehmer et al., 2020; Cranmer et al., 2020]. In the next section, a more detailed description of the SBI frameworks is provided.

6.2.2 Density Estimation using Neural Networks

Methods that rely on neural networks to directly learn the mapping from $f_Y(\omega)$ to the posterior, known as **neural density estimators (NDE)**, have offered promising results in neuroscience and neuroimaging problems [Gonçalves et al., 2020; Hashemi et al., 2022; Jallais et al., 2022]. Neural networks are considered universal approximators, i.e. given sufficient conditions, they can map any arbitrarily complex function [Cybenko, 1989; Hornik et al., 1989], although common feed-forward networks are designed to provide point-estimates. Instead, the strategy in NDE is to obtain a probability distribution as output from the neural network; this can be achieved by embedding a conditional density estimator into them so the point-estimates in the output layers effectively define the

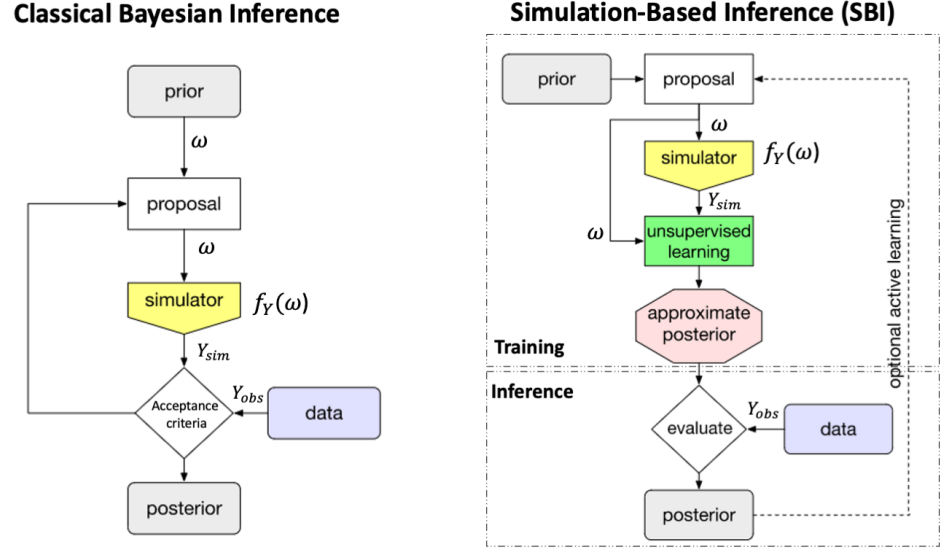


Figure 6.1: Comparison between Classical Bayesian Inference and Simulation-based Inference - Both approaches rely on the use of a forward simulator $f_Y(\omega)$. However, methods like MCMC use these to compare with the observed data, accepting or rejecting samples in an iterative way for each new case, while SBI uses the simulations to train a neural network that approximates the exact posterior via a parametric model. Using this approximation, samples from the posterior can be obtained with a single forward pass to the network for any new case. Figure adapted from [Cranmer et al., 2020]

hyperparameters of the target distribution (e.g. the mean and variance of a Gaussian distribution).

Formally, an NDE is a parametric model that takes as inputs pairs of datapoints (u, v) (which can be simulated) and returns a conditional probability density $Q_\psi(U|V)$ that is an approximation to the true conditional density $\pi(U|V)$. A family of flexible conditional distributions, parametrised with parameters ψ and assumed to resemble the target density, is proposed. The problem relies then upon estimating the optimal parameters ψ of such distributions. This general formulation can be used to approximate any conditional density distribution: an approximated likelihood $Q_\psi(Y|\omega)$, termed as Neural Likelihood Estimation (NLE) [Durkan et al., 2018; Papamakarios et al., 2018b]; the posterior distribution of parameters $Q_\psi(\omega|Y)$, termed as **Neural Posterior Estimation (NPE)** [Greenberg et al., 2019; Lueckmann et al., 2017; Papamakarios and Murray, 2018]; or a ratio of likelihoods $\frac{Q_\psi(Y_0|\omega_0)}{Q_\psi(Y_1|\omega_1)}$, termed as Neural Ratio Estimation (NRE) [Durkan et al., 2020; Hermans et al., 2020]. In all cases, the learnt model subsequently allows direct inference of the posterior given new data from the trained network, without model inversion.

This property is known as **amortised inference**. Note that, despite training with pairs of simulated data (u, v) , this framework is essentially unsupervised in nature as the posterior is reconstructed without the need of labels in real input data. Also, it is Bayesian in the general sense (getting the posterior) although the density is obtained by fitting in a point-estimate way the parameters ψ of $Q_\psi(\omega|Y)$ using typical feed-forward networks $\mathcal{F}(Y, \omega)$, i.e. $\psi = \mathcal{F}(Y, \omega)$.

Deciding what conditional density to approximate depends on the problem being tackled. For instance, in both NLE and NRE the model learnt by the networks provides a surrogate likelihood, so they require a further sampling step to build the posterior (e.g., MCMC). On the other hand, the NPE can sample the posterior directly from the NDE but does not provide a surrogate likelihood that can be reused across inference tasks.

Given the core of the methodology relies on forward simulations, it is crucial that the breadth and depth of simulated data capture all the important aspects of real data for the relevant range of parameter values. Two techniques are commonly used to address this challenge: 1) To introduce **inductive bias** via the parameter priors or model specification, hence simulate more around parameter values that are more strongly supported a-priori. 2) To perform **active learning** and focus on more "important" regions of the parameter space. A simple scheme of active learning is the sequential variant (SNPE, SNLE, and SNRE), where after each round of training the estimated posterior becomes the new prior. This is a similar concept to what MCMC-ABC and SMC introduced in Rejection-ABC. However, although the sequential approach is usually more simulation-efficient (e.g., see an empirical comparison in [Lueckmann et al., 2021]), amortisation is partially lost after each iteration as it focuses on regions of high probability for a particular observation. This is particularly problematic in multimodal posterior distributions, as in each round of the SNPE you deviate a bit more from the true posterior, going to the region of maximum likelihood where the posterior will ultimately collapse to a delta or mass point at the limit of infinite iterations.

The aim is to maximize the log-likelihood of the observed data with respect to model parameters, which is usually expressed in terms of a minimization problem such as $\underset{\psi}{\operatorname{argmin}} \{-\log Q_\psi(Y|\omega)\}$.

For NPE, the training adjusts the hyper-parameters of the network ψ (which would provide the parameters of the posterior distribution in the output layer). If the density estimator is flexible enough and sufficient simulated training data (ω_{sim}, Y_{sim}) is given, this loss function leads to perfect recovery of the exact posterior [Papamakarios and Murray, 2018]. Given that infinite data is not feasible, it is crucial to at least have flexible density estimators to approximate the posteriors as best as possible.

In this chapter, we focus on the estimation of the posterior density by using Neural Posterior Estimation (NPE), detailed in the next section.

6.2.3 Neural Posterior Estimation (NPE)

The idea implemented in the NPE dates back to the regression adjustment approaches [Blum and François, 2009]; these were not flexible enough to accurately estimate the posterior, and they were only used within some other ABC method to allow for a larger ϵ . By making use of the flexibility of NDEs, [Papamakarios and Murray, 2018] extended the concept of learning a mapping between data and model parameters to directly approximate the *exact* posterior density.

The active learning scheme used in the sequential NPE (SNPE) approach also introduced in [Papamakarios and Murray, 2018] consists in using the posterior learnt in the iteration i as prior for the iteration $i + 1$. As a consequence, the approximated posterior $Q_\psi(\omega|Y)$ deviates from the true posterior after each iteration, as $Q_\psi(\omega|Y)_{i+1} \propto \pi(\omega)_{i+1} \cdot \pi(Y|\omega)_{i+1}$, where $\pi(\omega)_{i+1} = Q_\psi(\omega|Y)_i$ (i.e if learning exists, the posterior will contract from the prior definition, leaving out regions of the parameter space). Hence, the posterior needs a post-hoc correction factor of $\frac{\pi(\omega)_i}{\pi(\omega)_{i+1}}$ to avoid over-contraction. The SNPE-A proposed by [Papamakarios and Murray, 2018] restricted the priors to Uniform or Gaussians, and the density estimator to a Mixture Density Network (MDN) capable of representing a mixture of Gaussians (see below). By applying these restrictive conditions of application, analytical corrections were possible. In SNPE-B [Lueckmann et al., 2017], the correction is directly introduced in the loss function, extending the application to any NDE and not only MDNs. Although SNPE-B seemed to solve the contraction issue, the changes in the loss made training much more difficult. Finally, [Greenberg et al., 2019] manipulated again

the mathematical formulation of the problem and the loss to dynamically reparametrise the proposals and provide a final solution: the Automatic Posterior Transform (APT), or just SNPE-C. SNPE-C allows for the use of any type of NDE with a low training cost. Once the networks are trained, amortised inference allows sampling from the posterior given an observation with a single forward pass through the network, without model inversion.

For completeness, a description is provided for some of the common NDE architectures compatible with SNPE-C that will be used here.

Mixture Density Networks

Mixture Density Networks (MDN) were proposed by [Bishop, 1994] and use neural networks to learn the coefficients (means and standard deviations) of a Gaussian Mixture model (GMM), $Q_\psi(\hat{\omega}|Y)$, which density approximates the target distribution $\pi(\omega|Y)$ (see Fig.6.2):

$$Q_\psi(\hat{\omega}|Y) = \sum_{j=0}^{m-1} \alpha_j \cdot \pi(Y|\mu_j, \sigma_j), \quad (6.1)$$

with m the number of Gaussian components, α_j the mixing coefficients ($\sum \alpha = 1$) and $\pi(Y|\mu_j, \sigma_j)$ the density function of a univariate Gaussian distribution. In principle, any probability density function can be approximated to arbitrary accuracy.

Auto-regressive Likelihood models

The term auto-regressive (AR) originates from time-series models where observations from previous steps are used to predict the value of the current one. Similarly, AR likelihood models [Bengio et al., 2003] are based on the chain rule of probability (eq.6.2), where the probability of a random vector variable $X=(x_1, ..., x_n)$ can be decomposed as:

$$\pi(\mathbf{X}) = \prod_{i=1, k < i}^N p(\hat{x}_i | x_{1:k}) \quad (6.2)$$

i.e., the auto-regressive condition relies on x_i being explained by the $i - 1$ components.

This formulation can lead to impractical learning of the joint distribution given the high

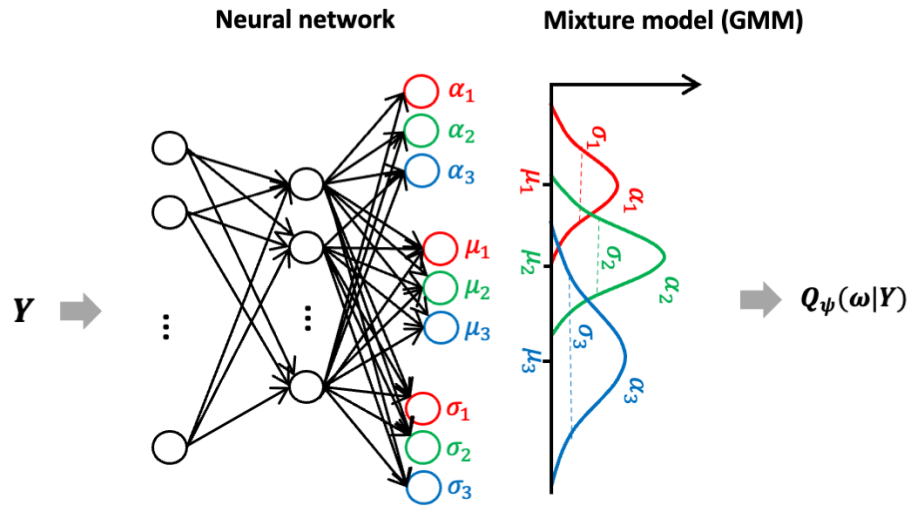


Figure 6.2: Mixture Density Network - The output of the neural network provides the parameters (means and standard deviations) of a Gaussian Mixture Model (GMM).

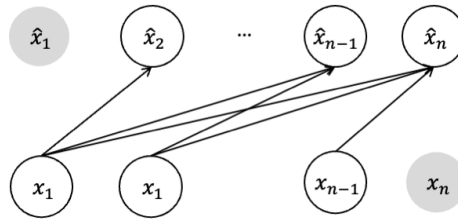


Figure 6.3: Auto-regressive property - Being X a random variable, the estimate \hat{x}_i can be only explained by the x_k components, being $k < i$. Image adapted from [Bond-Taylor et al., 2022]

number of conditionals that may need to be estimated if N is sufficiently high. Instead, each conditional distribution can be specified as parameterised functions with a fixed number of parameters, e.g. assuming each follows a given distribution and using NNs to estimate the parameters of that distribution. Among different techniques existing to build AR models (e.g., Recurrent NNs such as Long Short-Term Memories, Self-attention, etc.), *masking* is one of the most popular [Larochelle and Murray, 2011; Uria et al., 2016]. Here, pre-defined binary masks are applied to the weights of a neural network of any type. These masks are used to decide which connections to keep and which connections to drop in order to satisfy the AR property, while enabling density estimation and without relying on sequential loops that are common in AR models.

This approach is used in [Germain et al., 2015] to build a **Masked Autoencoder for Distribution Estimation (MADE)** (see Fig.6.4). In this case, to satisfy the AR property for conditional distributions (i.e., that \hat{x}_i only depends on the preceding inputs $x_{<i}$), the idea is that there should be no network path with the next layer, i.e., at least one connection must be 0 (see eq.8 and eq.9 from [Germain et al., 2015]). Masking here essentially is equivalent to a logical removal of some connections or the input data. To make easier the learning by the network, they propose a list of possible masks to choose between (i.e., selecting the mask becomes a hyperparameter).

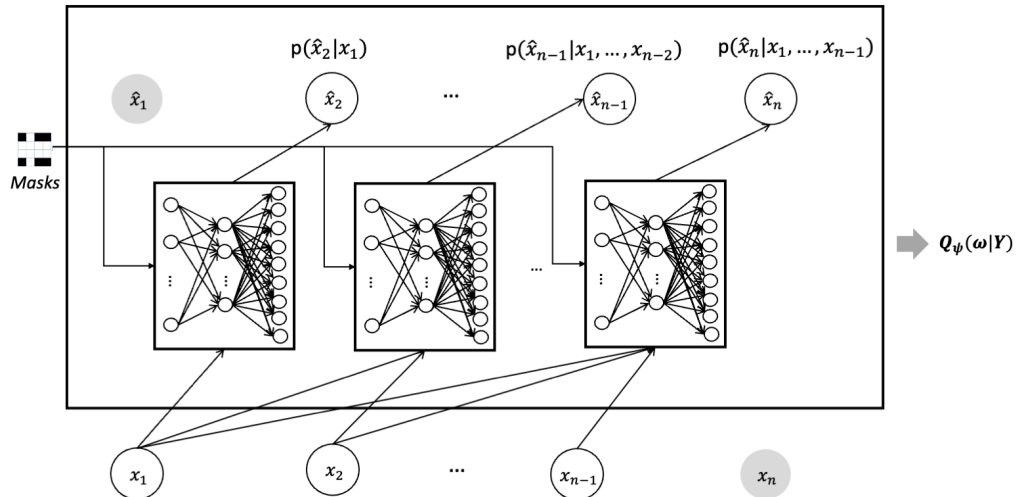


Figure 6.4: Masked Auto-regressive Density Estimation (MADE) - By masking a standard NN architecture (e.g. an autoencoder) to accomplish for AR properties (see Fig.6.4), MADE can provide conditional density estimation in the outputs.

Normalizing Flows

Normalizing Flows (NF) is a class of generative models that transform a simple base distribution $x_0 \sim \pi(x_0)$ into a complex target density $x_k \sim \pi(x_k)$, where both sampling and density evaluation can be efficient and exact. The framework was defined in [Rippel and Adams, 2013; Tabak and Turner, 2013; Tabak and Vanden-Eijnden, 2010], and later popularised by [Rezende and Mohamed, 2015] in the context of variational inference, and by [Dinh et al., 2015] for density estimation, rising as an efficient alternative to e.g., MDNs. NF have shown also good scalability and high expressivity for density estimation [Kobyzev et al., 2021; Papamakarios et al., 2019].

To obtain the target density, a *flow* of concatenated non-linear and *flexible* deterministic transformations f_k are applied to a base distribution x_0 , such as $x_K = f_K(x_{K-1}) \circ \dots \circ f_2(x_1) \circ f_1(x_0)$ (see Fig.6.5). Essential conditions are that the defined transformations f_k are 1) invertible (i.e., bijective functions that have an inverse function f_k^{-1}), and 2) smooth differentiable functions (diffeomorphisms).

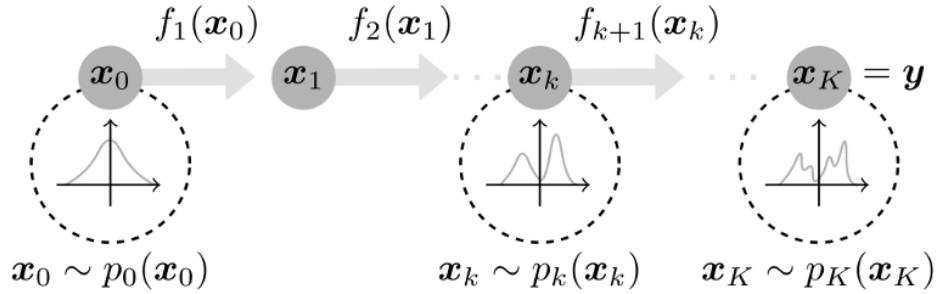


Figure 6.5: *Normalising Flows (NFs) can build complex multi-modal distributions by mapping a simple distribution x_0 through invertible functions $f_k(x_{k-1})$. Image extracted from [Bond-Taylor et al., 2022]*

The mathematical formulation is grounded in the assumption of volume preservation (total probability sums to 1) and the *change of variable* formula:

$$p_k(x_k) = p_0(x_0) \prod_{k=1}^K \left| \det \left(\frac{\partial f_k(x_{k-1})^{-1}}{\partial x_{k-1}} \right) \right| = p_0(x_0) \prod_{k=1}^K \left| \det \left(\frac{\partial f_k(x_{k-1})}{\partial x_{k-1}} \right) \right|^{-1}, \quad (6.3)$$

i.e., the base and the target distributions are related by a scaling factor given by the prod-

uct of the Jacobian determinant of the successive invertible mappings, $\frac{\partial f_k(x_{k-1})}{\partial x_{k-1}}$, that tells how much each ∂x_{k-1} is stretched into ∂x_k . In the case of estimating the conditional posterior distribution, $p_k(x_k) \rightarrow Q_\psi(\hat{\omega}|Y)$. As transformations f_k can be arbitrarily complex, the NF can be embedded into a neural network and the problem resides then in finding the optimal hyper-parameters of f_k (the weights of the networks), that will provide in return the parameters ψ of $Q_\psi(\hat{\omega}|Y)$ in the output layer. Effectively, it is not needed to know the target distribution. The inverse transformation enables to sample from the posterior by simply extracting values from the base distribution and pushing forward these samples through the sequence of learned transformations f_k . This allows for fast evaluation as it only needs a single forwards pass through the flow to generate samples from the posterior conditioned on some observed data.

Commonly, the bottleneck is the computation of the Jacobian determinants, so most of the works in recent years have focused on how to achieve an efficient computation of it while preserving enough flexibility in the transformations (see [Kobyzev et al., 2021; Papamakarios et al., 2019] for reviews). In this chapter, we will make use of two of the state-of-the-art NFs: the **Masked Auto-regressive Flow (MAF)** [Papamakarios et al., 2018a] and the **Neural Spline Flow (NSF)** [Durkan et al., 2019].

Masked Auto-regressive Flow (MAF): MAF [Papamakarios et al., 2018a] uses MADE as a building block for each layer, i.e. each layer in MAF is a feed-forward neural network with masked weight matrices, such that the auto-regressive property holds. Thanks to the probability chain rule, these properties enable us to factorize the target distribution as a sequence of 1-dimensional conditional densities, where the parameters of each of these conditional density estimators (i.e., MADEs) are estimated by a neural network. The transformations f_k are linked to auto-regressive conditioning, providing a triangular Jacobian by design that allows for low-cost computation of the absolute determinant.

Neural Spline Flows (NSF): Another method of creating invertible functions is by implementing monotonic splines as transformations f_k , i.e. a piece-wise function consisting of P segments, where each segment is a simple function that is easy to invert (e.g. a low-degree polynomial). Outside of such interval, the transformation can default to

a simple identity function. Different functional forms have been explored (in order of increasing flexibility): linear and quadratic splines [Müller et al., 2019], linear-rational splines [Dolatabadi et al., 2020], and rational-quadratic splines, which are the ones used in Neural Spline Flows [Durkan et al., 2019]. With this formulation, any arbitrary monotonic function can be approximated in a given interval by a neural network; having enough intervals, any function can be approximated while using a differentiable closed-form function that can be inverted analytically (quick to compute).

6.3 Methods

Similarly to the previous chapter, our goal is to estimate the posterior distribution $\pi(\omega|Y)$ of a dMRI biophysical model parameters ω given measured data Y , where $\pi(\omega|Y) \propto \pi(Y|\omega)\pi(\omega)$. For that, we designed a SBI framework and compared it against a classical Bayesian inference approach, the univariate Metropolis-Hastings MCMC, referred to from now on just as MCMC. We have explored how the posterior in MCMC is obtained by inverting the forward model using approximate/iterative approaches. In our SBI implementation, this is achieved by training an NPE to learn the mapping between parameters and the data. For that, two inputs are needed to be specified: 1) a prior distribution $\pi(\omega)$ defining the plausible range of parameters ω , and 2) a generative forward process $f_Y(\omega)$ to create a simulated training dataset $\{\omega_{sim}, Y_{sim}\}$. These are used to train a neural density estimator as a parametric model, $Q_\psi(\omega|Y)$, by maximizing the total log-probability, $\sum_N \log Q_\psi(\omega_N|Y_N)$ (or minimizing the negative log-probability), with respect to hyperparameters ψ (see Fig.6.6).

To the best of our knowledge, this is the first time SBI is implemented to solve such a problem. Hence, as a first approximation and without loss of generality, we evaluated the above framework using a single-shell single-fibre Ball&Sticks model as an exemplar for $f_Y(\omega)$, as it is a model with tractable likelihood where MCMC can be also applied for comparisons.

$$f_Y(\omega) \rightarrow A(d, f_1, v_1, \sigma^2) = (1 - f_1) \cdot e^{-b_i d} + f_1 \cdot e^{-b_i d (g_i v_1^T)^2} + \epsilon \quad (6.4)$$

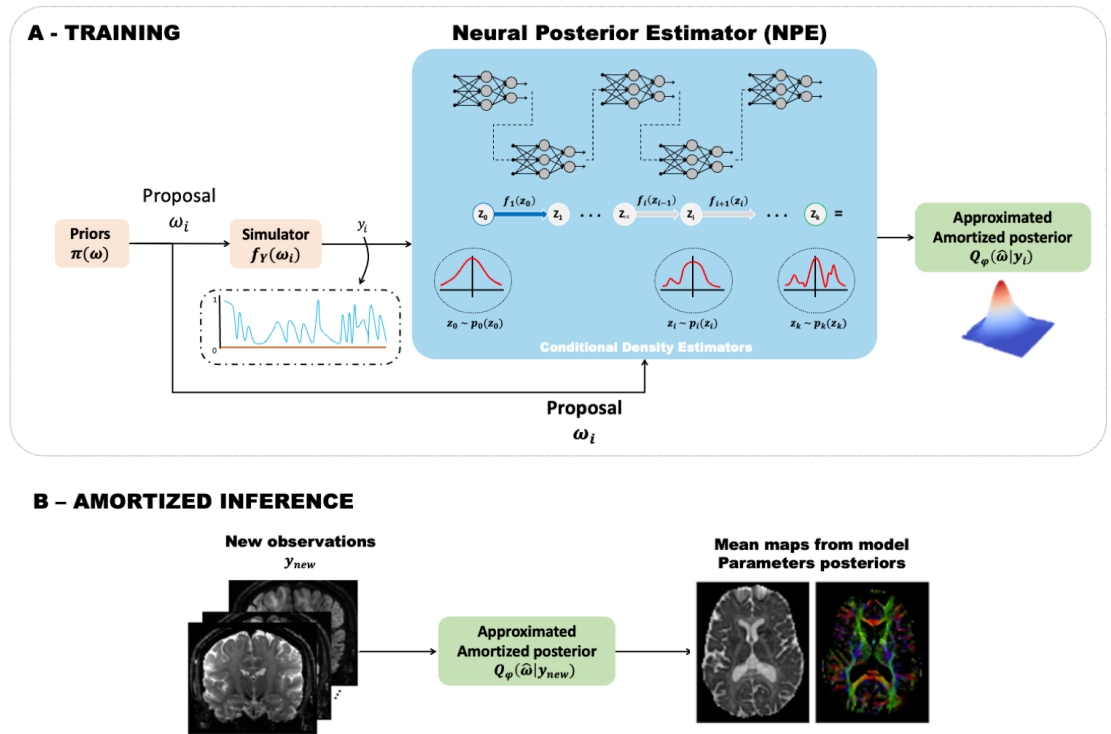


Figure 6.6: NPE scheme implementation - A. Training phase: Samples can be obtained from the prior distribution of parameters and used to produce forward predictions Y_i . The pairs $\{\omega_i, Y_i\}$ are passed to a neural density estimator to learn the mapping between, which will provide the approximate posterior distribution $Q_\psi(\tilde{\omega}|Y)$. B. Inference: Once the model is learnt, new observations can be passed to it to obtain the samples from the posterior distribution.

with ϵ indicating a constant level noise of SNR=30. Note that we learn the mapping between the signal attenuation $A = S/S_0$ for a given dMRI acquisition scheme (64 directions, $b = 1000 \text{ s/mm}^2$) and the parameters, rather than the signal S directly. S_0 is a scaling factor that is directly measured over multiple repeats in a dMRI experiment, it can therefore be fixed to the mean value of these repeats (i.e. the mean of the $b = 0 \text{ s/mm}^2$ measurements in each voxel). Including such an arbitrary-scaled factor in the training could increase substantially the size of the training dataset, so we omitted it for simplicity. For the implementation of the NPEs, we have used the Pytorch-based **sbi** Python package [Tejero-Cantero et al., 2020].

6.3.1 Training

We focused on **Neural Posterior Estimators (NPE)**, as they bypass the need for further sampling as in NLE methods. Also, we performed one round of training instead of following a sequential NPE (SNPE) approach, in order to be able to perform amortised inference.

A critical component of the process is the training data. We employed Eq. 6.4 to generate N training sets $\{\omega_{sim}, Y_{sim}\}$. We kept the dMRI sampling scheme constant (64 directions, $b = 1000 \text{ s/mm}^2$). We then obtained millions of training samples by sweeping through the four model parameters of interest in ways that generated uniform coverage over specific ranges of parameter values.

Algorithm 5 Neural Posterior Estimation as in [Papamakarios and Murray, 2018]

Require: $f_Y(\omega)$, $\pi(\omega)$

1: **procedure** ESTIMATE $\hat{\pi}(\omega|Y)$

2: **for** $n = 1$ to N **do**

3: $\omega_n \sim \pi(\omega)$

▷ Sample a proposed value from the prior

4: $Y_n = f_Y(\omega_n)$

▷ Simulate data Y_n using the forward model and the proposal sampled

5: **end for**

6: $\psi \leftarrow \text{argmin} \sum_n^N -\log Q_\psi(\omega_n|Y_n)$

▷ Learn the parameters ϕ of the conditional density estimator

7: Set $\hat{\pi}(\omega|Y_0) = Q_\psi(\omega|Y)$

8: **return** Sample from $\hat{\pi}(\omega|Y_0)$

▷ Sample from $\hat{\pi}(\omega|Y_0)$ for new observations Y_0

In order to achieve optimal performance, we evaluated different aspects of the framework. We kept the network hyper-parameters (number of layers, activation function, etc) at their

defaults provided in the *sbi* package and subsequently evaluated:

Density estimators: Although NSF has been shown to be the most flexible flow that generally outperforms its alternatives in the (S)NPE (see appendix H.5. in [Lueckmann et al., 2021]), the appropriate density estimator is problem specific. For instance, MDNs have been proven to be advantageous in SNPE if the posterior is simple enough since they allow for analytical corrections of the posterior. In [Jallais et al., 2022], MAF is used instead of NSF. Hence, we evaluated a range of density estimators: Mixture Density Networks (MDN), Masked Autoencoder Density Estimator (MADE), Masked Auto-regressive Flow (MAF), and Neural Spline Flow (NSF).

Size of training data: There is no rule of thumb to pre-define the size of the training dataset in SBI. [Lueckmann et al., 2021] evaluated the performance of the networks in different tasks using up to 10^6 of training samples, while [Jallais et al., 2021] addressed an inference problem using a grey matter biophysical model and used $N = 10^5$ samples for training. Here we generated 10^7 simulations and evaluated the convergence as a function of the training data size (i.e. by taking random subsets of the full set). All NDEs were trained using the same pre-simulated datasets. We avoided summary statistics and mapped the simulated data directly to model parameters.

Parameter priors: Priors on the model parameters are directly linked to the accuracy of the model learned (the closer the simulated ranges are to the true parameters, the less interpolation the model needs). Using established knowledge of the problem (inductive bias), we defined positivity constraints and a known plausible range of values. For scalar parameters, we used uniform prior distributions over the following ranges: $d \sim U(10^{-5}, 7.5 \cdot 10^{-3})$ [mm^2/s], $f_1 \sim U(0.001, 0.999)$. Regarding the diffusivity, it is also common to use a Gamma distribution as a prior such as $d \sim \Gamma(\alpha, \beta)$, with plausible values $\alpha = 3.5$ and $\beta = 1500$ [Behrens et al., 2003]. We compared both options to see which one is more convenient.

For spherical parameters linked to fibre orientations, using uniform distributions such as $\theta \sim U(0, \pi)$ [rads] and $\phi \sim U(0, 2\pi)$ [rads] leads to oversampling of the poles in spherical

coordinates (i.e., parts of the sphere with ϕ close to either 0 or π) and too much sparsity around the equator ($\phi = \pi/2$), biasing the training dataset and the performance of the network [Devroye, 1986]. Therefore, we used the correction in eq.6.6 that ensures uniformly distributed points on the sphere (see Fig.6.7):

$$\theta \sim 2\pi \cdot \mathcal{U}(0, 1) \quad (6.5)$$

$$\phi \sim \arccos(1 - 2 \cdot \mathcal{U}(0, 1)) \quad (6.6)$$

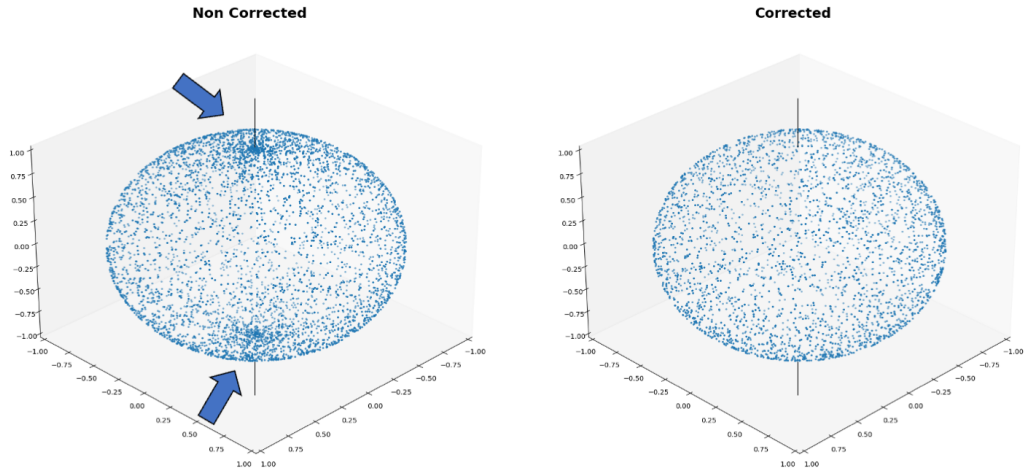


Figure 6.7: Uniform sampling in the sphere - On the left, oversampling of the poles due when using uniform distribution for spherical angles θ and ϕ . On the right, uniform sampling on the sphere was obtained after applying the correction in eq.6.6.

6.3.2 Evaluations

Experiments and Metrics

To evaluate the implemented SBI framework, we performed different experiments. We first explored what architecture and hyper-parameters provide the best results. For that, we explored a range of possible density estimators, the convergence depending on training data size, and parameter priors using synthetic data where ground-truth values are known.

Based on accuracy metrics using ground-truth values from the synthetic test datasets, we selected the best NPE configuration. Then, we compared the best NPE candidate against

results provided by classical MCMC in both synthetic and in-vivo brain data. In the former, we have ground-truth values; in the latter, we assume the MCMC estimates as the reference values for comparisons, as these can be considered to be samples from the exact/true posterior distribution. Results from these comparisons against MCMC will help also to assess the predictive uncertainty estimates from SBI, as there is no ground-truth for the posterior density.

In all these experiments, we evaluated the accuracy, precision and performance of the novel SBI framework. For **accuracy**, we used the same metrics employed in Chapter 5.3.4. We calculated the errors in the median posterior estimates. For scalar parameters, the error e for each synthetic realisation can be calculated directly as the difference between the known ground truth value and the median posterior estimates (see eq.5.24 and 5.25). For fibre orientations, we calculated the crossing-angle X_{angle} (in degrees) between the ground truth v_j and the mean estimated one (extracted from the dyadic vectors), as indicated in eq.5.26. In the in-vivo data, where no ground truth exists, we computed the errors against the MCMC mean estimates.

Precision was evaluated in both synthetic and in-vivo brain data. We evaluated the behaviour of uncertainty quantification as a function of noise level and training data size; uncertainty is expected to reduce as the SNR and/or the training data size increase. Further sensitivity analysis of the posterior can be useful for identifying pathologies in the model inversion process. For instance, using the synthetic test data, we calculated the posterior z-scores and the posterior shrinkage for each parameter ω , as defined in [Betancourt, 2018]:

$$z - score = \left| \frac{\hat{\omega} - \omega}{\sigma_{post}} \right| \quad (6.7)$$

$$shrinkage = 1 - \frac{\sigma_{post}^2}{\sigma_{prior}^2} \quad (6.8)$$

where $\hat{\omega}$ indicates the estimated posterior mean, ω is the ground-truth value used for simulations, and σ_{prior}^2 and σ_{post}^2 indicate the variance of the prior and posterior, respectively. The posterior z-score provides a quantification of how well the posterior distribution en-

velops the ground-truth, with a small z-score being an indicator of true values accurately encompassed in the posteriors. The shrinkage quantifies how much the posterior distribution contracts from the initial prior distribution for a given observation, so large contractions (large shrinkages) are indicators of a learning process; if in addition the z-score is small, the learning process is likely to be successful. Hence, representing one measurement against the other allows us to define regions of confidence for the model inversion (e.g., bottom right in Fig.6.8). Finally, we compared the width of the approximated posterior distributions provided by MCMC (used as a reference) and SBI in both synthetic and in-vivo brain data.

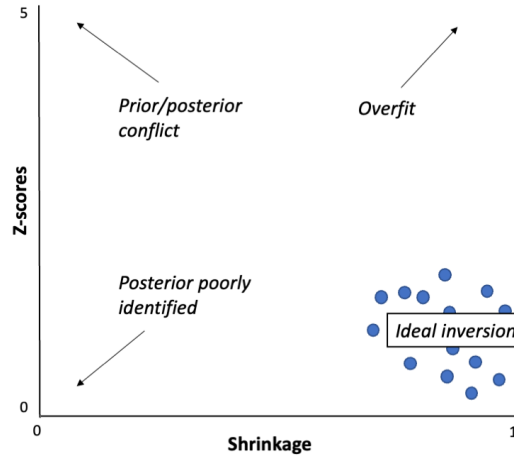


Figure 6.8: Sensitivity analysis of the posterior - An ideal experiment should be informative (large shrinkage) and accurate (small z-scores) for every observation and parameter. Deviations from here may indicate different biases. For instance, small posterior shrinkage indicates an experiment that poorly identifies the given parameter component, while large z-scores indicate inferences biased away from the true data-generating process. Large values in both shrinkage and z-score indicate overfitting, while concentration to the top left indicates a poorly chosen prior that biases the model configuration space away from the presumed true data-generating process.

Regarding **performance**, if the neural network is not properly trained, it won't be able to recognise and operate correctly with new observations. In such scenarios, sampling from the network may take relatively long. We evaluated global performance by comparing the **computational runtime** required for training and inference in SBI and MCMC. Runtime was measured from running the in-house Python implementations in a dedicated single-core CPU of a MacbookAir M1 2020.

Data

For the evaluation of the inference frameworks, we firstly used synthetic data to explore how well the different SBI designs performed (note that these are different simulated sets than the simulated data we used to train the NPEs). For these synthetic *test* data we used the Ball&Sticks model to generate noisy realisations of the dMRI signal attenuations, given some ground truth parameter values, and we tested how well each SBI design could estimate these ground truth values. Then, we further explored the performance of SBI against classical MCMC in synthetic and real brain data.

Synthetic test data: We generated a set of 1000 synthetic single-shell (64 directions, $b = 1000 \text{ s/mm}^2$) single-fibre dMRI signals covering the whole range of parameter values used in the training sets. Parameter values were sampled from the priors to generate new synthetic observations for which we knew the ground truth parameter values. These covered uniformly the possible range of parameters the network has been trained on. These data were also used to compare the different options of training (type of NPE, size of training data, priors) and select the best configurations.

For comparisons against the MCMC, we generated 100 noisy realisations for a given SNR level of the same sets of underlying fibre patterns. This is a 5x8 single-shell (64 directions, $b = 1000 \text{ s/mm}^2$) single-fibre set of dMRI signals, which allowed easier visualisation of performance in a specific WM setup, where the dependency on volume fraction and fibre orientation was assessed.

Unless specified otherwise, both datasets have a constant level noise of SNR=30.

In-vivo brain data: We used standard 3T data acquired in a Siemens Prisma scanner from a healthy volunteer, with $2\text{x}2\text{x}2 \text{ mm}^3$ voxels (104x104x56 voxels) that had dMRI acquisition parameters compatible with the ones used in the training: 64 directions, $b = 1000 \text{ s/mm}^2$. We estimated fibre orientations and performed probabilistic tractography using XTRACT [Warrington et al., 2020] and compared results with the ones obtained using conventional MCMC.

6.4 Results

6.4.1 Design, training and evaluation of the SBI network

All results in this section are obtained from evaluating the networks with the *synthetic test data* for the inference. We first evaluated the **accuracy** obtained using the different density estimators (i.e., MDN, MADE, MAF and NSF) in recovering the ground-truth values used for simulations. Fig.6.9 shows exemplar posteriors (top row in each subplot) and the % error obtained (bottom row) for each parameter by the different density estimators trained with a subset of 50.000 samples. Overall, density estimators based on normalising flows (MAF and NSF) have shown the best results, with NSF being the one with the best performance in recovering the ground-truth, and MDN the worst. As a sanity check, it is expected to see the posterior contracted with respect to the priors. The priors (uniform within a bounded range) are not represented here for visualization reasons, but it is straightforward to observe that all density estimators, but MADE, demonstrate this contraction around the ground truth value, for all the exemplar cases shown.

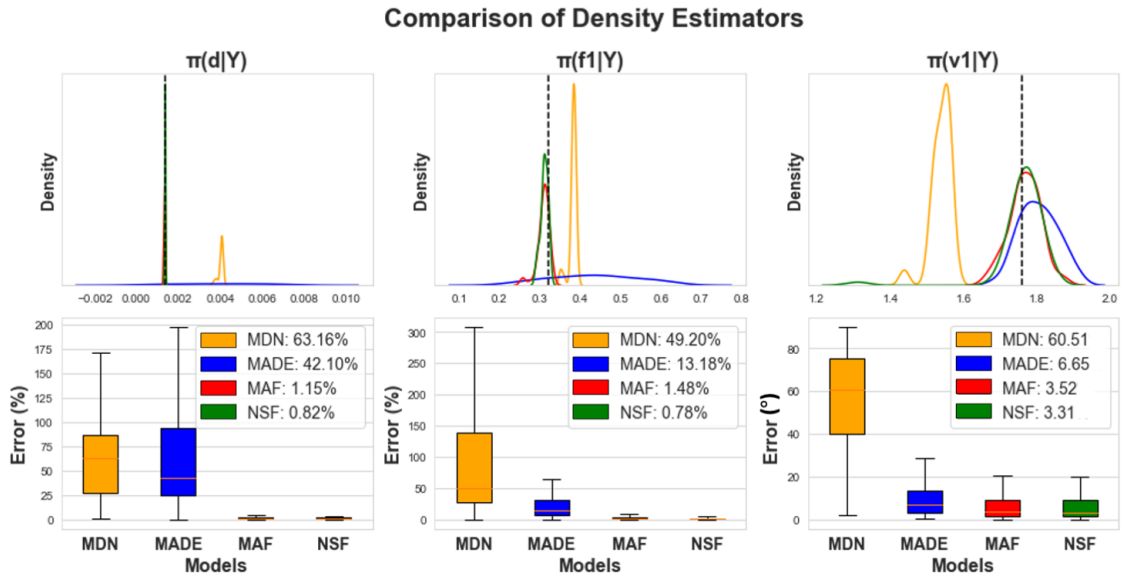


Figure 6.9: Comparison of density estimators. The figure shows results obtained from different density estimators trained with a dataset 50,000 samples. Top: posterior density parameters obtained from each density estimator in an exemplar dMRI signal from the synthetic test data. The ground-truth value is indicated with a black dashed line. Bottom: Median error estimates obtained from the mean parameters in the synthetic test data (1,000 realisations).

Using the NSF as a density estimator, a comparison of the test errors obtained when using different priors for the diffusivity and different sizes of training data can be found in

Fig.6.10. Using a uniform prior for d showed more stable results and faster convergence than using a Gamma distribution, although for training data sizes larger than 500.000 samples, the results converge to similar behaviour for all parameters.

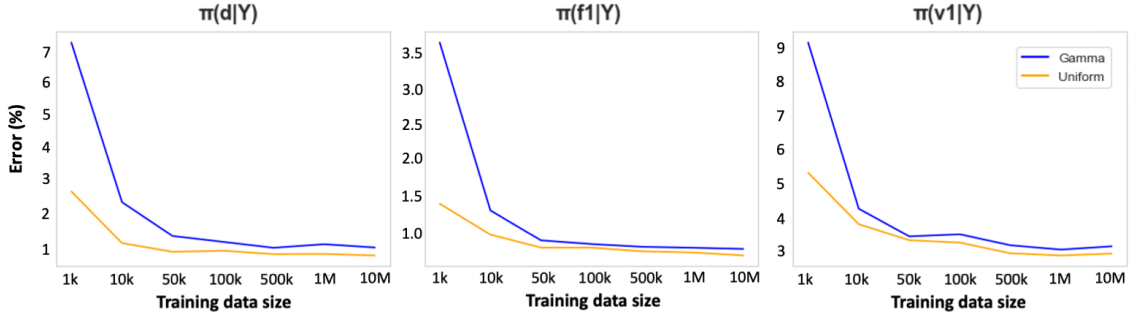


Figure 6.10: Mean error in parameter estimates for different training data sizes and different diffusivity prior distributions - The figure shows a comparison of test error obtained from using $\pi(d) \sim U(10^{-5}, 7.5 \cdot 10^{-3})$ [mm²/s] (in orange) or $\pi(d) \sim \Gamma(3.5, 1500)$ (in blue) for different size of training data and using a Neural Spline Flow as density estimator.

So far we have seen that an NSF density estimator with uniform priors and 10^6 training samples provides very good performance in terms of accuracy in the synthetic data for all the parameters. We kept these features fixed (unless specified otherwise) and we subsequently evaluated performance in terms of the **precision** of the inference. Fig.6.11 represents the standard deviation of the posteriors for increasing SNR levels and different sizes of training data. As expected, the uncertainty decreases as the SNR increases for all models; also, models trained with more simulations returned higher precision.

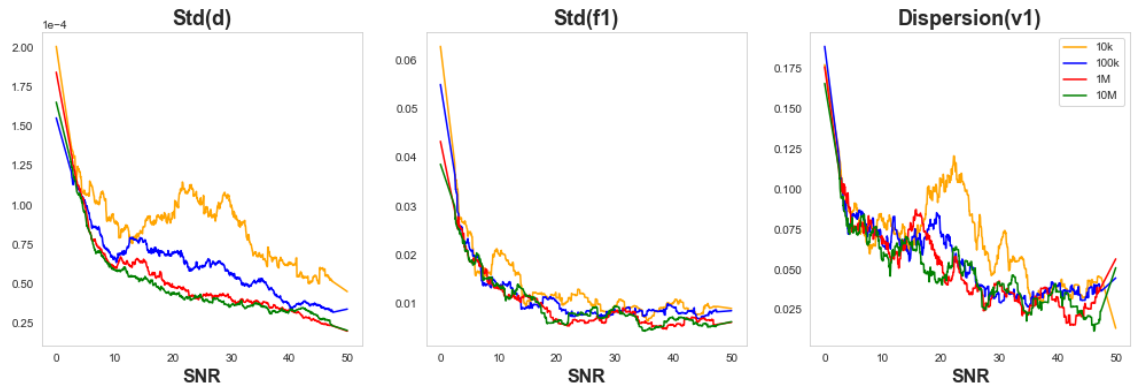


Figure 6.11: Variations in the width of the posteriors - Each line shows the std (for d and f_1) and the dispersion (for v_1) obtained by models trained with a different number of simulations (see legend) respect to increasing SNR level. The inference is done in the synthetic test data.

Fig.6.12 shows the changes in the standard deviation of the posterior samples when a different number of posterior samples are considered. Results suggest that the width of the marginal posteriors can be assumed relatively stable when at least 50-100 samples are

used to build the posterior distribution. We used 50 samples in subsequent comparisons against MCMC (where we also use 50 samples drawn from the posterior).

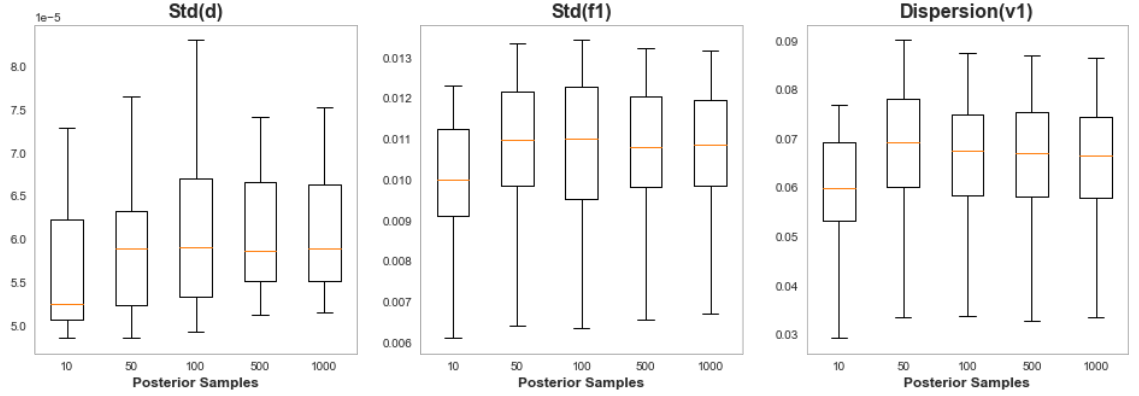


Figure 6.12: *Variations in the width of the posteriors as a function of the number of posterior samples.*

As a further assessment for the precision of the estimates, the density scatter plot in Fig.6.13 shows the relationship between the posterior z-scores and the posterior shrinkage (reliability plot) [Betancourt, 2018], where the combination of high shrinkage and low z-scores represents an ideal anticipated behaviour from a well-trained network. As expected from previous figures (see Fig.6.9), high shrinkage is observed for all parameters in the NSF, i.e. the posteriors are well-identified and considerably contracted compared to the initial uniform priors. A higher density of points around small z-scores indicates that the mean estimates are concentrated around the ground-truth values. Taken together, these results provide evidence of learning by the networks. Some cases exist that suggest behaviour towards the overfitting region (high z-scores and shrinkage), however, the largest portion of cases lie in the ideal region.

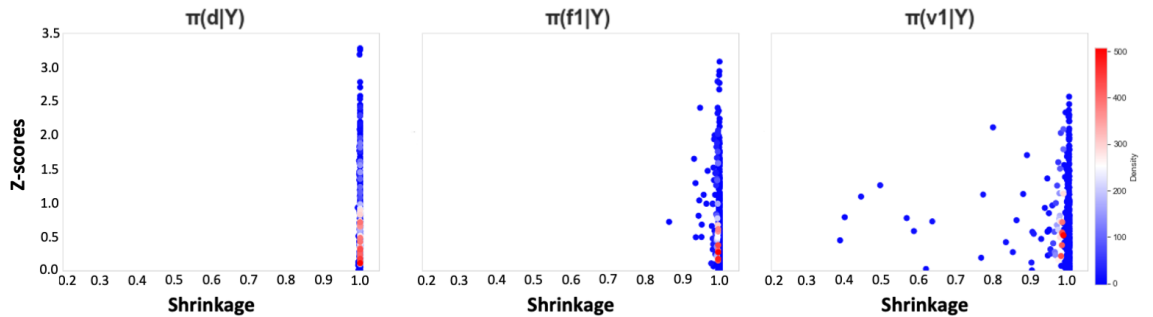


Figure 6.13: *Posterior fit evaluation* - The density scatter plots show a measure of the reliability of the inference using synthetic data, plotting the posterior z-scores against the shrinkage of the posteriors as defined in [Betancourt, 2018] and sec.6.3.2

Finally, Fig.6.14 shows the relationship between the **computational runtime** in a single-CPU (in hours) with training data size and the training error measured as the log-probability in the validation subset. Runtime shows a linear relationship with the size of training data. In terms of error, the validation log probability curve seems to converge at 1 million samples. This agrees with what has been observed before, e.g. if the posterior barely changes in terms of mean and std (i.e. variations can be attributed to the randomness of sampling) for varying parameters such as training data size, it is a sign of convergence; otherwise, if posterior still gets tighter, more simulations are likely needed. In Fig.6.11 we observed how the uncertainty remains pretty much the same in the model trained with 1M or 10M samples. Hence, based on the results so far, we can assume convergence (in terms of accuracy and precision) for the given biophysical model for an NSF trained with 1 million samples. This requires only around 2 hours; furthermore, this training needs to be done just once for a model (amortised inference).

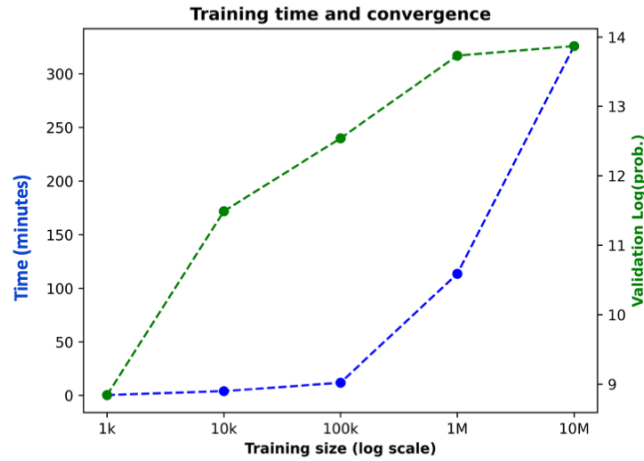


Figure 6.14: Training size vs training time - The figure shows the Log-probability obtained (in green) in the test set obtained by an NSF with uniform priors for training dataset size. Similarly, in blue, it is indicated the time needed to train the network for a given training dataset size measured in a single-core CPU.

In terms of generating the simulated training data, creating a simulated dataset of $N = 10^6$ realisations took around 20 minutes. In terms of inference, obtaining 50 samples from the posterior distribution from the already trained NSF is almost instantaneous and takes ~ 1 msec. Still, as both simulation and inference runtime are linear with N , the runtime can be reduced by C by parallelising at the data level, with C being the number of CPU-threads used.

6.4.2 Comparisons of SBI to MCMC – Synthetic data

After exploring a range of possible NPE hyper-parameters experimentally in the previous subsection, we found the configuration that provides best results is using a Neural Spline Flow (NSF) as density estimator, uniform priors that cover the whole plausible range of the parameters ω ($d \sim \mathcal{U}(0, 5e-3) \text{ mm}^2/\text{s}$, $f \sim \mathcal{U}(0, 1)$, orientations: $[\theta \sim \mathcal{U}(0, \pi)$; $\phi \sim \mathcal{U}(0 - 2\pi) \text{ rads}]$, and a training data size of $N = 10^6$ samples. Using a network trained with this configuration, we first compared results in a specific WM setup using a synthetic dMRI dataset (see sec.6.3.2) against a more classical inference approach, the random walk MCMC.

Fig.6.15 shows two exemplar cases of the posterior estimates given by each approach. In terms of signal prediction, both provide accurate reconstruction. For a normalised signal, MCMC returns an RMSE of 0.03 and 0.06, for the first and second exemplars of Fig.6.15, respectively; RMSE in SBI is 0.05 for both cases. However, there exist differences in the width of the posterior of the diffusivity.

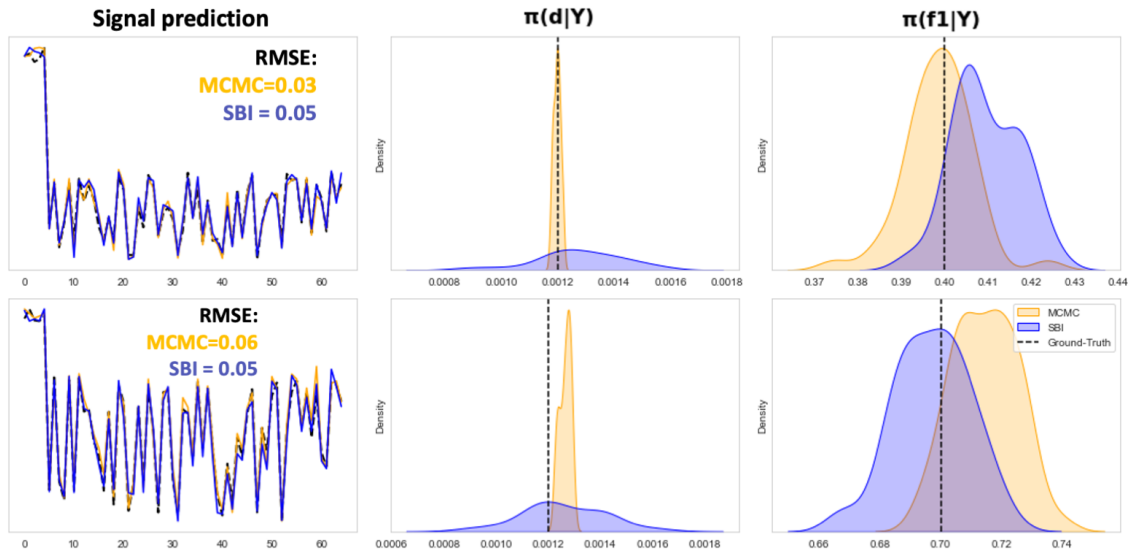


Figure 6.15: Posterior Predictive Check - Comparisons between the ground-truth values (black dashed line) and the estimates given by MCMC (orange) and SBI (blue) for two different, randomly-chosen synthetic data realisations (in rows). Left column: Signal prediction. Middle column: posterior distribution obtained for the diffusivity. Right column. posterior distribution obtained for the volume fraction.

Fig.6.16 represents the median error across the 100 realisations for MCMC and SBI. SBI returns slightly higher median errors than MCMC, although these are still low (below 5% for scalar parameters, and 10 degrees for the fibre orientation). In SBI, there is a certain

tendency to overestimate the mean parameter, especially in the diffusivity.

Regarding precision, Fig.6.17 shows the uncertainty estimation of each method quantified by the standard deviation of the posterior across the 100 realisations. Uncertainty mapping in the volume fraction and fibre orientation is similar in both approaches, while in the diffusivity SBI provides uncertainty in the order of 10 times higher, as anticipated by 6.15. In relative terms, the standard deviation of the posterior distributions is in the order of 2-5% of the estimated value for both SBI and MCMC, except in the diffusivity model parameter in the SBI that is in the order of 40%, suggesting an overestimation in this case.

6.4.3 Comparisons of SBI to MCMC – In-vivo brain data

Using the same trained network as in the subsection above, a comparison between SBI estimates and MCMC outputs in in-vivo brain data is shown in Figs. 6.18 and 6.19. In terms of accuracy, SBI shows a high level of agreement with the MCMC. Although the difference maps and the *histogram of differences* suggest slightly higher values in the diffusivity and the volume fraction compared to MCMC. This is in agreement to the observed in synthetic data. Nevertheless, the differences are always below 5% with an overall correlation of $r = 0.84$ for the diffusivity and $r = 0.94$ for the anisotropic volume fraction. There is also a high agreement in the fibre orientation estimates, with a median difference of 4.3° (IQR= 6.21°) between SBI and MCMC mean posterior estimates.

Regarding precision, the uncertainty of the posterior parameter distributions provided by SBI is lower compared to the MCMC, especially in non-WM regions. However, these differences were not observed in synthetic datasets; in fact, there was an overestimation of the uncertainty in the diffusivity parameter. Given that the uncertainty mapping agreed with the expected behaviour in terms of SNR and data size (see Figs.6.12 and 6.11), a potential reason for this might be differences between SBI and MCMC in capturing signal complexity that is unexplained by the fitted model. In the synthetic data, we simulated one fibre compartment and fitted a model with one compartment. But in real data we also have signals arising from multiple crossing fibre compartments. While MCMC seems to capture this unexplained signal from the extra compartments as the source of uncertainty

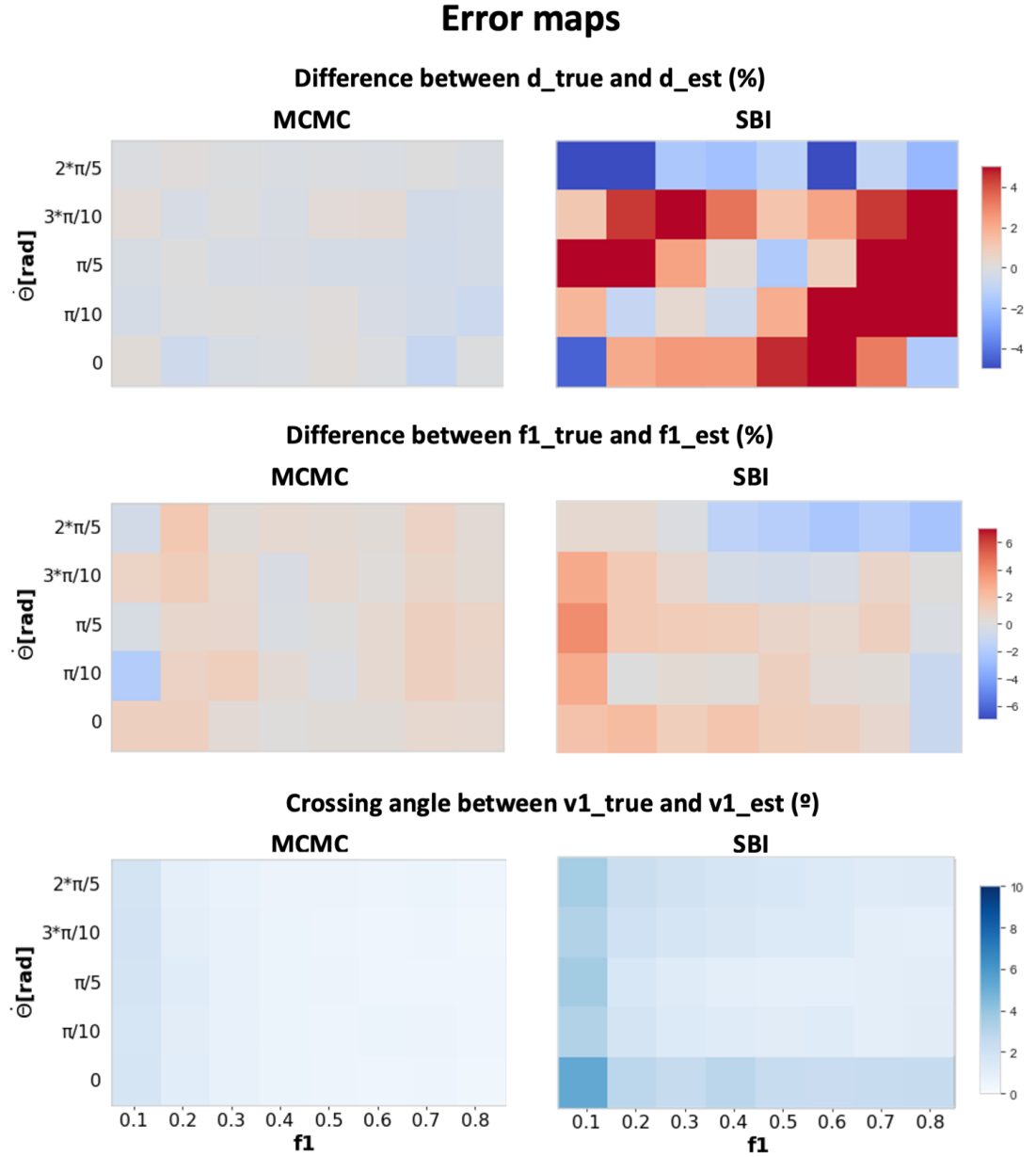


Figure 6.16: *Error maps in synthetic data - MCMC vs SBI* - Comparisons of the error obtained in the MCMC and SBI estimates across 100 noisy realisations ($SNR=30$) of the same WM setup.

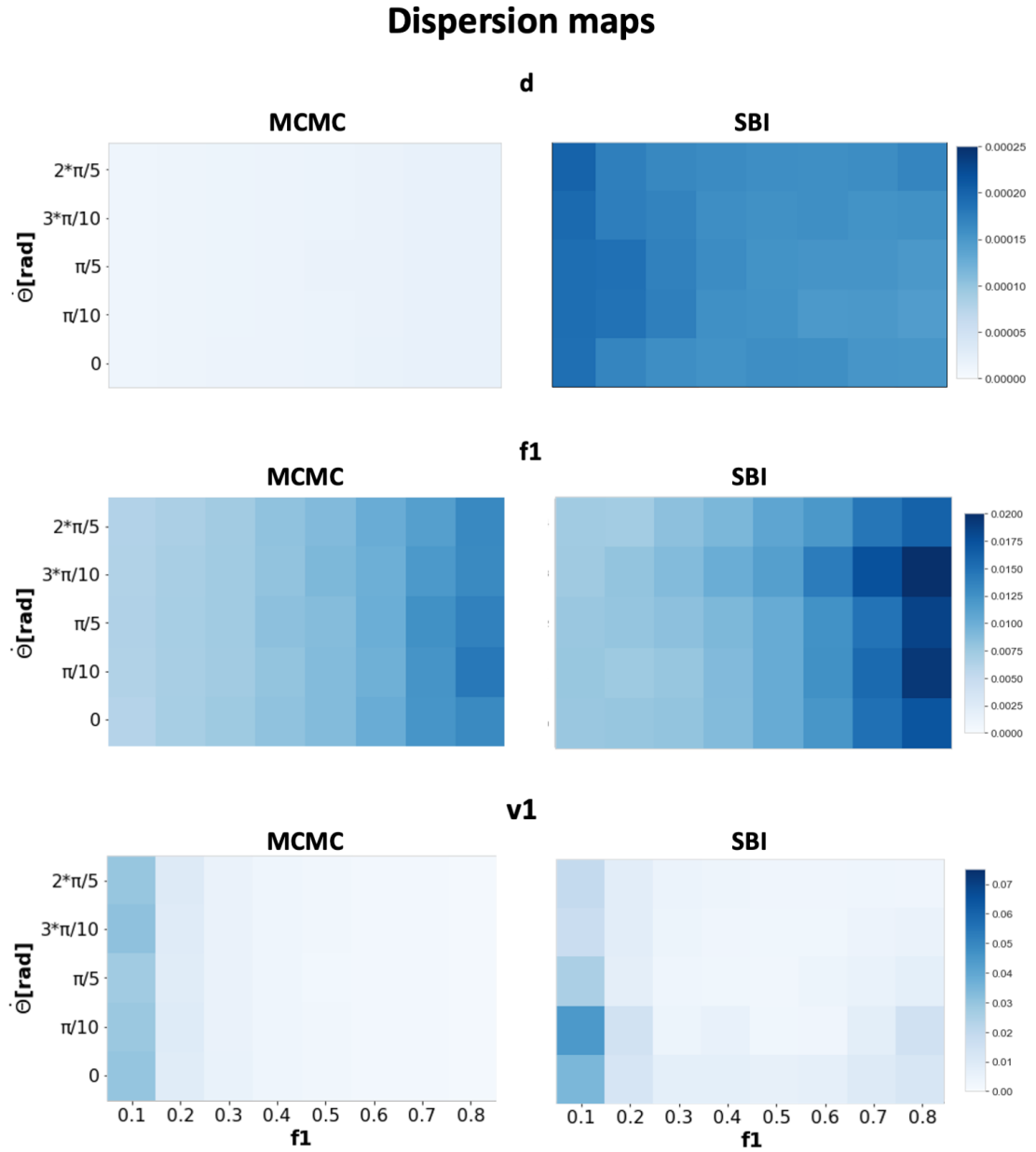


Figure 6.17: *Dispersion maps in synthetic data - MCMC vs SBI* - Comparisons of the dispersion maps obtained in the MCMC and SBI estimates across 100 noisy realisations (SNR=30) of the same WM setup. Dispersion is measured as the standard deviations of the posterior distribution.

returning a higher dispersion in the estimates, the SBI seems to return overconfident posteriors.

We explored whether these differences could have an effect on the tract bundles obtained using probabilistic tractography [Warrington et al., 2020] and the fibre orientation posteriors provided by SBI and MCMC. Given that no crossing fibres are considered in these experiments, a number of tracts that go through complex white matter patterns could not be resolved successfully with these single-fibre vector fields (specifically the bilateral Acoustic Radiations (AR) and superior Longitudinal Fasciculi (SLF) I and II). Nevertheless, reasonable estimates were obtained from the rest of the tracts with a relatively high correlation (median correlation $r = 0.74$) between the SBI-obtained tracts and those obtained from the MCMC approach (see Fig.6.20). Despite the agreement in the main core of the tracts, we can observe how the difference in orientation uncertainties between SBI and MCMC are reflected in spatially more confined distributions for some of the SBI tracts (e.g. the Frontal Aslant, pointed with yellow arrows in Fig.6.20).

Finally, we evaluated the computational performance in this in-vivo brain dataset. Using the in-house Python implementations and using a single-CPU core, obtaining the posteriors for a whole-brain unseen dataset (104x104x56 voxels) took 20 minutes, while running MCMC took more than 100 hours. This reflects a speed-up of 2-3 orders of magnitude in parameter inference once the network is trained, as found also in other studies (e.g., [Hashemi et al., 2022]).

6.5 Discussion

Simulation-Based Inference takes advantage of the capability of Machine Learning methods to handle large dimensionality problems while preserving Bayesian principles to approximate the posterior distribution. Furthermore, untackled challenges can be addressed within this framework as it can provide amortised inference without the necessity of an explicit likelihood; the only requirement is to have access to a forward process that can simulate data. In this chapter, we have designed and evaluated the application of SBI in voxel-wise microstructural modelling of the brain for dMRI data for the first time. As a proof-of-concept, we used a single-shell single-fibre Ball&Sticks model, studied the op-

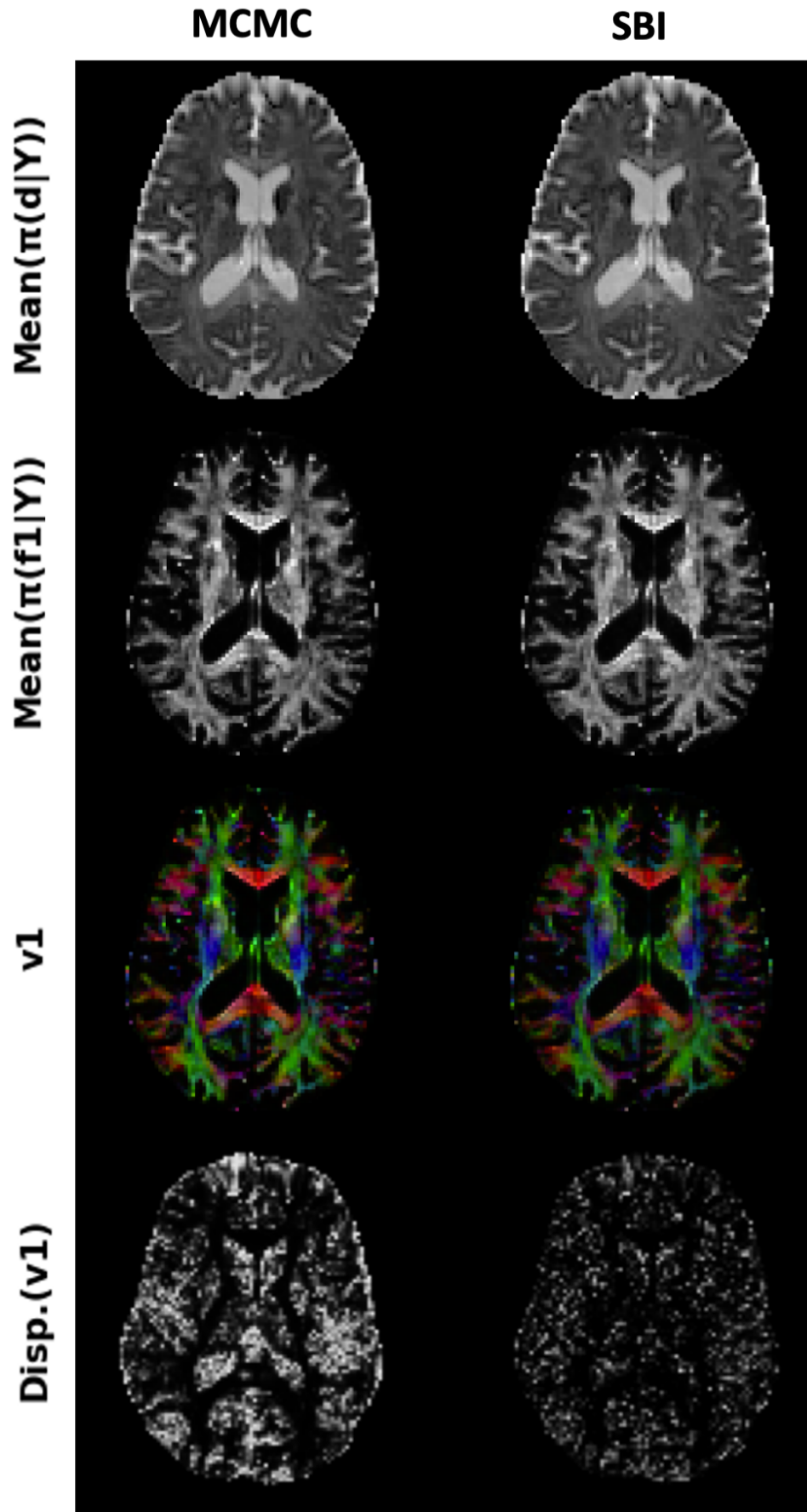


Figure 6.18: *Qualitative comparison between MCMC and SBI estimates in in-vivo brain data* - Each column shows the mean maps obtained for the diffusivity and volume fraction, the fibre orientation, and the dispersion of the fibre orientation estimates (as exemplar of precision maps).

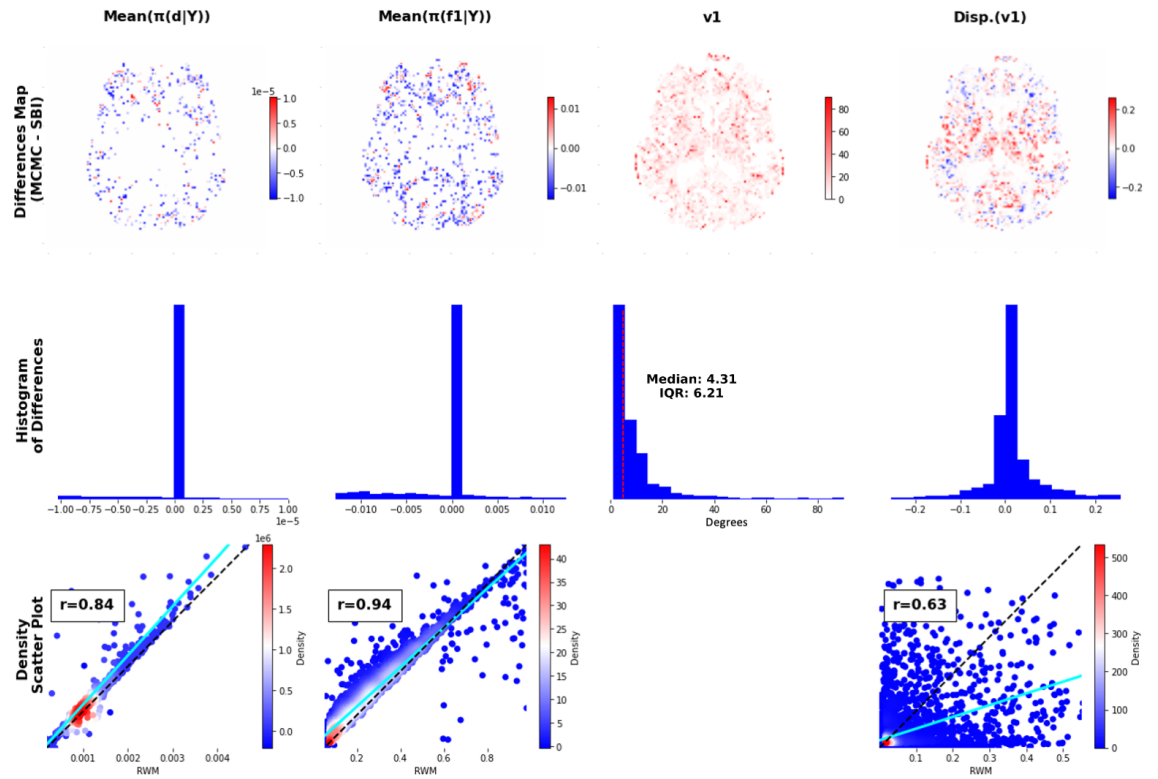


Figure 6.19: *Quantitative comparison of MCMC and SBI estimates in in-vivo brain data* - First row: difference maps between estimates from MCMC and SBI. Second row: histogram of these differences. Third row: a correlation plot between scalar parameters, respectively. Differences are calculated in the WM only using a $FA > 0.1$ mask.

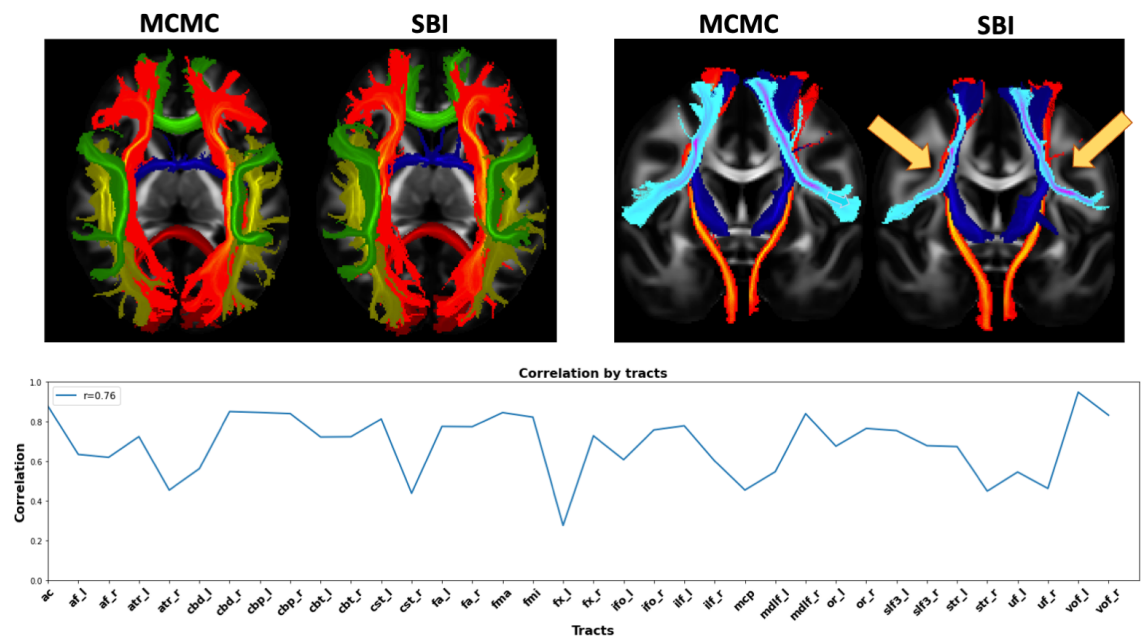


Figure 6.20: *Comparison of probabilistic tractography obtained from MCMC and SBI estimates in in-vivo brain data* - Top: Comparison of Maximum Intensity Projection of some exemplar tracts obtained by using MCMC and SBI model estimates. Bottom: Correlation between MCMC and SBI tracts. Legend indicates the median correlation between all tracts (discarding missed tracts).

timal design of the network, and compared results against classical inference approaches (MCMC) in both synthetic and in-vivo brain data.

The Neural Spline Flow (NSF) was the density estimator that performed the best, closely followed by the Masked Auto-regressive Flow (MAF). This agrees with what has been found in other SBI studies [Lueckmann et al., 2021], also in neuroimaging [Hashemi et al., 2022]). In our study, we utilized the NSF as the density estimator along with synthetic data to determine the most stable and accurate results for all model parameters. We discovered that employing uniform priors with biophysical-informed ranges for all parameters, with appropriate corrections for angular parameters to ensure uniformity in spherical space rather than Cartesian space, yielded the best outcomes. Bayesian inference can benefit from inductive biases, which can be introduced by defining more specific prior distributions for the parameters. To explore this, we utilized the Gamma distribution, previously employed in other works for diffusivity [Behrens et al., 2003]. However, we found that using a uniform prior for diffusivity outperformed in terms of accuracy and stability. This may initially appear counterintuitive since inductive biases are typically employed to enhance convergence during training (e.g., enforcing biophysical-informed ranges). Nevertheless, similar findings have been reported in other simulation-based inference studies [Lueckmann et al., 2021]. The reasons for these architectures favouring uniform priors are not yet fully understood. It remains unclear whether defining overly specific priors might overly constrain the parameter space during training, leading to undersampling/oversampling in certain regions, whether there is any dependency on the geometry of the parameter space (e.g. we have commented on the consequences of this type of bias when the poles of the sphere were undersampled), or it is just lack of training. Further investigation is required to shed light on this matter. Interestingly, recent developments in the field, such as Active Learning [Rezende et al.] and the sequential approaches mentioned in section 6.2.1, offer intriguing alternatives for leveraging more accurate and complex priors specific to the data. These avenues provide opportunities for future research and warrant exploration.

Regarding precision, uncertainty in ML models is expected to be reduced with higher SNRs and/or larger and more diverse datasets [Hermans et al., 2021; Stickland and Murray, 2020], and we have observed this in synthetic data. We have also verified that using

only 50 samples to build the posteriors (allowing for faster inference and direct comparisons against MCMC) is not biasing the results and that the posteriors obtained are mostly located in reliable regions as defined in [Betancourt, 2018]. Regarding training data size, a set of $N = 10^6$ samples was shown to be close to convergence in terms of error, while keeping a very low training time of the network. This size of the training set is in a similar range obtained in other works inferring a similar number of model parameters with non-linear relationships [Jallais et al., 2022; Lueckmann et al., 2021].

Using this configuration (NSF trained with $N = 10^6$ samples simulated with parameters drawn from uniform priors), we compared model parameter estimates obtained by the novel SBI implementation and a classical univariate-MCMC approach in synthetic data. Accuracy of estimates was overall high in both methods, with a slight overestimation in the SBI. In terms of precision, SBI returned similar widths of the posteriors for all parameters except the diffusivity, which showed up to 10 times broader posterior distributions.

A similar comparison was performed for **in-vivo brain data**. The only requirement for this data is to be obtained with the same acquisition parameters as the ones used for the training dataset. Again, a high level of agreement was found in the mean parameters estimates, confirming the accuracy of the SBI approaches. However, we found differences in the precision of the estimates. Compared to MCMC estimates, SBI has shown lower uncertainty values for model parameter estimates. These differences with respect to the MCMC were not observed in simulated data. Two potential hypotheses can be explored to explain these differences: 1) There is a model miss-specification and the simulated data used for training were not realistic enough; however, this should have led to an incorrect training of the network and, as a consequence, very bad accuracy in mean parameter estimation, which was not the case. 2) SBI and MCMC behave differently when there is a residual signal unexplained by the complexity of the considered model. The model fitted was a single-fibre model; simulated data were drawn from single-fibre patterns, but in-vivo brain data can contain multiple compartments. As a consequence, the variance explained by the predicted signal is lower in general and the uncertainty of the estimates is higher, as happened in the MCMC. However, the SBI has not been able to map this uncertainty correctly. This seems to be a more plausible explanation, which points to a limitation of

the SBI framework (see sec.6.5.1 below for a more extensive discussion on this matter). Nevertheless, probabilistic tractography (which is built upon uncertainty estimates of the fibre orientations) using SBI model estimates has provided plausible results and a relatively high level of agreement with those obtained using MCMC (median correlations above 0.7).

We further evaluated SBI in terms of its computational performance. Given the amortised implementation, the computational cost of SBI largely resides in the training phase. Once the network is trained, inference for new unobserved data can be done almost interactively, with a simple forward pass on the trained network, providing experimental speed-ups of 2 to 3 orders of magnitude compared to classical inference approaches like MCMC.

The idea of applying ML to learn the mapping between data and model parameters in dMRI is relatively recent but has been demonstrated in previous studies (e.g., [Golkov et al., 2016; Gyori et al., 2022; Ye et al., 2020]). However, to the best of our knowledge, only [Hashemi et al., 2022; Jallais et al., 2022] provide a similar approach for inference at whole-brain scale including uncertainty mapping, instead of point-estimate predictions.

Of particular interest is [Jallais et al., 2022], where the authors fitted a 3-compartment biophysical model (neurites, somas and extra-cellular space (ECs)) to dMRI signal, in order to characterize cortical cytoarchitecture [Palombo et al., 2020]. They designed an NPE, with a MADE as density estimator with a similar training set-up to ours (e.g., 10^5 simulations, ADAM optimizer and default parameters in the *sbi* Python package) and validation using both synthetic and in-vivo brain data. However, relevant differences can be found in our design. For instance, they relied on a vector of 7 rotationally invariant summary features used to describe the dMRI signal [Jallais et al., 2022]. They used this approach to allow inference for data from any dMRI sampling scheme (i.e. for any b-value and b-vector), while in our design the sampling scheme was fixed. Although the suitability of using summary statistics is highly dependent on the data and model used, the above suggests this could be a future aspect to investigate for both making training easier by having a lower dimensionality and to allow using different sampling schemes or models for the same trained network. Regarding precision, they evaluated the variance of the posteriors by quantifying the reduction of the variance as a function of different

simulated parameter values, but not by comparing against other methods. In our study, we evaluated the effect of the SNR and training data size in the dispersion of the parameters and confirmed the expected behaviour that uncertainty reduces as SNR or training data size increases. Furthermore, we showed that sharper distributions are not always equivalent to accurate precision mapping, but a potential symptom of overconfident estimates, so studying the variance of SBI estimates alone (as done in [Jallais et al., 2022]) can be misleading and further diagnostics are needed (see sec.6.5.1 below for a discussion).

Taken together, results from both works provide complementary observations of how SBI has shown potential to be applied to a variety of biophysical models while providing ground-breaking speed-ups thanks to the amortised inference. This makes SBI frameworks a promising approach with a potential impact on neuroimaging. However, there are open questions in multiple directions that need to be addressed, such as the generalisability and applicability of the frameworks and, more crucially, the uncertainty mapping done by SBI.

6.5.1 Uncertainty mapping in SBI

We have observed situations where the uncertainty quantification obtained by the MCMC and the novel SBI approach differed. Exploring which approach provides the right width for the posterior distributions is challenging as there is no *ground-truth*. However, classical Bayesian approaches, such as MCMC, have been extensively used in the past, while there are recent studies that highlight the potential of overconfident behaviours in NNs [Guo et al., 2017]; therefore, cautious use of methods that rely on them, such as SBI frameworks, may be needed [Hermans et al., 2021].

Recent studies have looked into how these deviations in uncertainty mapping can arise and have pointed to a complex mix of multiple sources. For instance, while vanilla neural networks are proven to provide well-calibrated probabilities [Niculescu-Mizil and Caruana, 2005], techniques designed to boost accuracy in current deep learning approaches (weights decay, batch normalization, active learning, etc.) affect output probability calibration [Guo et al., 2017] and coverage of the posterior [Hermans et al., 2021], leading to potential overconfidence. In simple terms, if the probabilities are not calibrated they cannot be directly interpreted as a degree of confidence. Training data size, imbalance in

the training set or data distribution shifts in new observations with respect to the training data (also known as out-of-distribution data) are also common sources that contribute to badly calibrated output probabilities [Bjorck and Gomes, 2021; Gyori et al., 2022; Ovadia et al., 2019]. Model complexity and model miss-specification is especially relevant as well in an SBI context as simulations are used to learn about real-world observations [Schmitt et al., 2022]. Some of these sources could have been captured in our experiments with SBI, such as the flexibility of density estimators, and the effects of training data size, noise, or imbalanced training data if corrections are not made to angular parameters; however, others such as increased uncertainty by model miss-specification have not been evidenced.

To alleviate the impact of these factors in uncertainty quantification some approaches have been very recently proposed. For instance, post-processing recalibration techniques of the output probabilities can be applied (e.g. see [Cranmer et al., 2016; Kuleshov and Deshpande, 2022; Macêdo et al., 2022]). However, these posterior corrections are not always accurate nor solve the bias in the learnt surrogate model. Hence, there is an active field of research looking for well-calibrated algorithmic designs capable of preserving the accuracy of the current ones, such as: defining new objective functions for correct uncertainty mapping in NNs [Delaunoy et al., 2022; Glöckler et al., 2022; Macêdo et al., 2022], or new alternative methods based in Bayesian-NNs [Blundell et al., 2015] or use ensembles that perform model averaging that have been proved experimentally to improve generalisation [Hansen and Salamon, 1990; Lakshminarayanan et al., 2017; Ovadia et al., 2019; Stickland and Murray, 2020] and increase the expected coverage in SBI frameworks [Hermans et al., 2021].

All in all, SBI has shown great potential for being used as a probabilistic framework and addressing untackled problems by classical inference methods. However, while improvements in modern ML methods in recent years have boosted performance in predictive accuracy they might be showing symptoms of falling short in giving correct estimates of their predictive uncertainty. In this thesis, we have stressed on multiple occasions that correct uncertainty mapping is crucial and how it can be useful in a number of dMRI contexts, such as identifying regions in the parameter space or in the brain where it is harder to fit a given model; or assessing whether the acquisition scheme is sufficiently informative

[Jallais et al., 2022]. Hence, further research in this direction is a promising opportunity to extend the reliability and applications of simulation-based inference frameworks in neuroimaging.

6.5.2 Limitations and Future Work

Given the novelty of SBI, a limitation is that the theoretical underpinnings are much less understood than in classical Bayesian approaches. In general, more research needs to be done regarding guarantees of the approximations and, especially, uncertainty mapping, as we discussed above.

Specific to the problem addressed in this chapter, and despite the promising results, we used a toy example here to provide proof of principle. More evaluations can be done and further modeling complexity can be considered. For instance, we have only evaluated a model with a single fibre compartment, so the generalisation of the problem to complex fibre patterns is still unclear from these experiments. We observed challenges regarding uncertainty quantification. Hence, a direct extension would be to improve on this using some of the techniques described above (e.g., recalibration of output probabilities, modification of loss functions or alternative methods to NNs). Particularly contrasting with classical approaches, as we did here, can also provide deeper insights into the generalisation of the new approaches with different parameter acquisitions (b-shells, resolution, etc.) or the compatibility with complex parametrization, such as angular parameters, or improper priors, such as the ARD.

Furthermore, one of the main motivations for SBI is the capability to address problems where there is a forward process (but not necessarily a single likelihood) to invert. Hence, the suitability to perform inference in models where there is not an explicit likelihood, but a series of iterative reconstruction and regularisation steps, like the Constrained Spherical Deconvolution (CSD) [Tournier et al., 2004, 2007] is an appealing future direction.

Following that line of research, many applications in neuroimaging can be envisioned in the longer term. Theoretically, if a neural network is appropriately chosen, tuned and trained,

it can represent any relationship between inputs and outputs if such a relationship exists. Framed into an SBI framework, a probabilistic mapping can be learnt between raw dMRI signals to the desired outputs (even directly from k-space data), e.g., biophysical model parameters or connectivity indices. Current pipelines barely provide feedback from previous steps despite the information being partly lost or smoothed at every pre-processing stage (registration, motion correction, denoising, modelling fibre orientations, etc.). However, by providing a direct mapping using a neural network, 1) all these modelling and assumptions taken during pre-processing steps are avoided, and 2) the effects of each layer on the final result are propagated back to adjust preceding layers, preventing the loss of information during intermediate steps.

Chapter 7

Conclusions and future directions

7.1 Main Contributions

Advanced imaging techniques, such as diffusion MRI, have had explosive growth and have shown great potential in studying brain architecture and structure. However, the measurements are indirect, in the sense that quantities of interest cannot be observed directly from images and have to be estimated. Over the last years a number of models have been proposed to link measurements from the scanner to different biophysical properties of brain tissue. This inverse mapping process creates challenges from estimating parameters for potentially highly-complex models to mapping uncertainty in these estimates. Hence, approaches that can address this inverse problem in an efficient manner are of crucial importance. To contribute towards that direction, we have devised and evaluated inference frameworks for computational diffusion MRI that deal with uncertainty mapping. We also explored one of the largest sources of uncertainty, thermal noise, and how denoising approaches can be beneficial.

Chapter 4 presented a novel framework for evaluating diffusion MRI denoising approaches (EDDEN). By setting objective criteria based on what a well-behaved denoising algorithm should offer, we evaluated the performance of commonly-used denoising approaches. Importantly, we captured both effects that noise can introduce: a) increasing variance of signal, but also b) causing signal biases. We showed that magnitude-based denoising approaches deal with the first challenge, but not with the second, leaving noise-floor effects in the "denoised" signal. They also cause larger loss of spatial resolution compared to

complex space-based denoising. The latter are more successful in dealing with noise-floor effects as well. All in all, we demonstrated that dMRI denoising, when done properly, can provide gains in signal quality and model estimability, and this can open opportunities for new applications, such as reducing scan time or pushing the boundaries for spatial resolution.

After demonstrating how some sources of uncertainty can be reduced, we considered in the following chapters the problem of mapping this uncertainty when inferring models from data. In Chapter 5 we tackled this challenge by proposing and evaluating new algorithmic designs of the Metropolis-Hastings MCMC in order to improve efficiency in both sampling and computing. We proposed moving away from the conventional univariate random-walk sampling scheme to a block-update MCMC paradigm that allows inference of multiple model parameters simultaneously. Among the different designs proposed, the Hybrid-MCMC algorithm showed the best performance. This new algorithm finds the appropriated covariance matrix of parameters via a Laplace approximation followed by an adaptive learning procedure, while permitted parameters with improper priors to operate independently. Hybrid-MCMC achieved similar accuracy in biophysical model parameter estimates to the univariate random-walk approach, in both synthetic and in-vivo brain data, while reducing the inference time by more than two-fold.

Finally, in Chapter 6 we approached the inverse problem from a different perspective. We designed and evaluated a novel simulation-based inference framework that relies on a generative forward process and the use of neural networks to learn an implicit mapping between data and parameters of interest. Using neural density estimators, these mappings are not simply considering point estimates but whole distributions, such as likelihoods and posterior densities of parameters given the data, allowing Bayesian inference in a data-driven way. Using a dMRI model as a proof-of-principle, we evaluated the framework in both synthetic and in-vivo brain data. Results showed a high level of accuracy in the mean estimates of model parameters, while achieving orders of magnitude faster inference than classical methods such as MCMC. Moreover, this type of framework allowed for inference amortisation even in problems that cannot be explicitly modelled with a tractable likelihood, paving the way for exploring new opportunities in neuroimaging-based models.

7.2 Future Directions

The methods and explorations presented in this thesis can be envisioned as complementary advances in the field. These could be translated into new research pathways and direct applications into clinical routine settings. For example, the utilization of denoising techniques has been discussed as a means to enable shorter scanning acquisitions, reducing the burden on patients undergoing scanning. Moreover, the introduction of high-resolution protocols holds promise for more precise and personalized diagnoses. It is worth noting that these methods extend beyond diffusion MRI, as various inherently noisy MRI modalities can benefit from principled denoising approaches, both in research (e.g., post-mortem imaging [Roebroek et al., 2019a]) and different targets in clinical applications (such as non-proton MRI [Weber et al., 2020] or DW-spectroscopy [Palombo et al., 2018]). However, to validate the efficacy of denoising, it is crucial to establish a principled manner for its evaluation, which necessitates a deep understanding of how the signal is modified by denoising approaches. Further work on developing standardised benchmarking for denoising is necessary, and we outline potential next steps below in section 7.2.1.

When combined with the new amortised inference approaches explored in this thesis, which also benefit from higher signal-to-noise ratios, these avenues become even more accessible. In clinical practice, they can remove the need for extensive computational resources traditionally associated with neuroimaging inference. For instance, by leveraging trained models, advanced and specific biophysical parameters could be displayed alongside the readily available FA or MD maps, generated in near real-time and aiding diagnostic purposes.

While exploring alternative avenues of research, such as hardware improvements, accelerated image acquisition, and GPU implementations, can enhance computational performance as well, the availability of high-quality ultra-high-resolution data (achieved through denoising) and methodologies like the SBI framework examined in this thesis offer exciting prospects for addressing unexplored research questions. One area of exploration involves investigating more complex crossing-fibre models. As observed, SBI has the potential to revolutionise inference speed (e.g., for models with multiple crossings like the Ball&Sticks model) but, more significantly, provide a novel inference approach for complex models.

For example, current methods are unable to perform inference in spherical deconvolution models such as Constrained Spherical Deconvolution (CSD) [Tournier et al., 2004, 2007]. Typically, these intricate microstructural models are solved using iterative heuristics that do not provide information about parameter correlations or find global minima [Daducci et al., 2015]. However, data can still be simulated from these models as a mixture of tensors, rendering them suitable for SBI exploration. Looking into the long term, validating these tools may unlock exciting opportunities to tackle some of the field’s current challenges, such as multimodal fusion of data or direct mapping from raw measurements to relevant features, bypassing intermediate pre-processing steps.

7.2.1 Benchmarking for dMRI denoising

The interest in pushing spatial resolution boundaries, the convenience of reducing scan times, or achieving higher diffusion-sensitisation to increase angular contrast [Moeller et al., 2021b; Roebroek et al., 2019b] have motivated research for algorithmically reducing thermal noise effects in the dMRI signal [Buades et al., 2005; Fadnavis et al., 2020; Manjón and Coupe, 2021; Moeller et al., 2021a; Pal and Rathi, 2022; Veraart et al., 2016b; Zeng et al., 2021]. However, many questions remain unanswered: *how should one denoise dMRI data? Are there any undesired effects caused by denoising? Are theoretical properties of the dMRI signal preserved?* or even whether “*to denoise or not*”.

These type of concerns are generally addressed in other fields by a benchmark [Kay et al., 2013; Plotz and Roth, 2017], which is lacking currently in dMRI. By implementing a controlled environment formed by a set of principled evaluations designed for a given type of data, and a fixed dataset representing different challenging scenarios, it is possible to characterise the performance of any algorithm. This process is required in order to answer the questions above in a standard and reliable manner.

This is basically the idea behind our framework EDDEN and turning it to a benchmark is the aim for the future. At the moment, some of the evaluations are empirically-driven. Quantitative summary metrics are needed to allow automated performance comparison of existing and new algorithms, or to find the optimal parametrisation of each method in an objective manner. Given the lack of ground-truth for in-vivo brain data and the complexity

of quantifying some parameters, one could work in two directions: 1) incorporating to the framework synthetic data where ground-truth is known, and 2) exploring Machine Learning options to learn no-reference quantifiable scores [Chow and Rajagopal, 2017; Lin et al., 2020; Mittal et al., 2012; Stepień et al., 2021], especially for their application to in-vivo brain data. In fact, this type of score can offer a more sophisticated alternative to SNR or CNR, where aspects such as spatial resolution loss or noise-floor could be included.

7.2.2 Direct mapping from raw signal to model estimates

Machine Learning provides powerful ways to learn arbitrarily-complex transfer functions without the necessity of any model or a priori knowledge of the data. Their application to perform analyses in dMRI, even beyond biophysical model parameters fitting, is expanding [Knoll et al., 2020; Poulin et al., 2019; Ravi et al., 2019]. For instance, [Wasserthal et al., 2018] used a CNN to directly segment tracts from the fODF peaks without using tractography, image registration or parcellation.

This raises the question of whether we could extend this idea to directly map noisy raw data to downstream model parameters, such as biophysical model parameters or connectivity indices, without the need of intermediate steps; by integrating those into one forward process. For instance, pre-processing data is usually needed to reduce the impact in modeling of artifacts during acquisition, such as noise or motion; registration to a common space for comparisons; or even just for human visualization and interpretation. However, no new information is produced during this process, i.e. all the information should be already contained in the acquired k-space measurements. Pre-processing may rely on inaccurate statistical assumptions, model simplifications and/or signal alterations to *smooth* the data, that can affect the final estimates. A direct mapping, as for instance allowed in an SBI framework, may provide crucial advantages: 1) Avoid signal modifications across various intermediate steps, thus reducing the sources of interpolation artifacts and/or uncertainty; 2) Use uncertainty in model estimates for more efficient acquisition and experimental design; 3) Massive reductions of computational time given amortised inference; 4) Reduce differences arising from different pre-processing and analysis pipelines in favour of standardisation.

Nevertheless, this entails important challenges and considerations. For instance, how the outcome may be affected when images are not distortion-corrected or in a common space has not been investigated, so it is unclear what steps can be really avoided or not. How much of the processing can be reversed and be identifiable may also depend on how data are acquired. Furthermore, although SBI operates via unsupervised learning, training relies on forward simulations. If the generative models employed are simplifications of the reality, the surrogate model learnt by SBI will be inherently limited as well. Nevertheless, this is an exciting stream of research that now seems to be tangible, as these methods are capable of handle high-dimensionalities while imposing little assumptions on data.

7.2.3 Fusing information from multiple sources

The fact that a flexible density estimator can be used to learn any arbitrarily complex function also paves the way for an avenue extensively desired in the neuroimaging community: information fusion from different sources. Different MRI modalities provide indirect and complementary approaches to probe on same occasions the same underlying brain properties. For instance, brain connectivity and networks can be probed using different MRI modalities, including anatomical MRI (anatomical covariance), resting-state functional MRI (functional connectivity) and diffusion MRI (structural connectivity) [Jbabdi et al., 2015]. However, the current paradigm treats them independently in ad-hoc modality-specific pipelines, ignoring complementary information they convey during modelling. This results in errors and ambiguities to a degree that hinders the ability to map individual brains reliably, affecting and possibly biasing [Maier-Hein et al., 2017], and affects interpretability of findings [Cole, 2010; Eickhoff et al., 2015]. Furthermore, these errors reduce reproducibility in neuroscience research and impede its translation into clinical applications [Griffanti et al., 2016].

Attempts to fuse information and reduce these errors have been implemented in the past to integrate cross-modalities data [Cabral et al., 2011; Honey et al., 2009; Kasabov, 2012], complementary within-modality data [Sotiropoulos et al., 2013a, 2016], or different models for the same data modality [Assaf et al., 2013; Howard et al., 2019]. In fact, the concept of fusing modalities to provide a unified generative model linked by common biophysical characteristics has been proposed a while ago [Jbabdi et al., 2007; Woolrich and Stephan,

2013] although it has not been comprehensively evaluated and adapted, due to challenges posed by classical inference approaches. Nowadays, there is not actual fusion of information so each individual modality will carry its own sources of error after the individual processing/modelling. As a workaround, a common approach is to use one modality to restrict the other, but in a unidirectional manner, e.g. assume anatomical connectivity obtained from experimental data as a true scaffold (which is not) to study parameters from generative models of large-scale neural populations coupled to that structure [Cabral et al., 2011; Deco et al., 2012; Honey et al., 2009].

The approaches that truly allow data fusion are generally limited by the computational feasibility of inverting the model or just the lack of well-established joint forward models that can be written as a likelihood. While the Hybrid-MCMC showed advantages by allowing for block-conditioning during inference, it would still need the definition of tractable likelihood. Hence, SBI could provide a potential framework to fuse information without the need of explicit joint forward models while 1) being able to address highly dimensional parameters spaces, 2) introducing naturally prior knowledge and biological constraints, and 3) assessing the uncertainty of the estimations. Ideally, constraining the conditional density estimation using information from multiple sources should unlock the potential for more accurate brain mapping, even at the individual level.

Bibliography

- T. Abraham and J. Feng. Evolution of Brain Imaging Instrumentation. *Seminars in Nuclear Medicine*, 41(3):202–219, May 2011. ISSN 00012998. doi: 10.1053/j.semnuclmed.2010.12.001.
- S. Aja-Fernández, A. Tristán-Vega, and W. S. Hoge. Statistical noise analysis in GRAPPA using a parametrized noncentral Chi approximation model. *Magnetic resonance in medicine*, 65(4):1195–1206, Apr. 2011. ISSN 1522-2594. doi: 10.1002/mrm.22701.
- S. Aja-Fernández, V. Brion, and A. Tristán-Vega. Effective noise estimation and filtering from correlated multiple-coil MR data. *Magnetic Resonance Imaging*, 31(2):272–285, Feb. 2013. ISSN 0730725X. doi: 10.1016/j.mri.2012.07.006.
- S. Aja-Fernández, G. Vegas-Sánchez-Ferrero, and A. Tristán-Vega. Noise estimation in parallel MRI: GRAPPA and SENSE. *Magnetic Resonance Imaging*, 32(3):281–290, Apr. 2014. ISSN 1873-5894. doi: 10.1016/j.mri.2013.12.001.
- D. C. Alexander. A general framework for experiment design in diffusion MRI and its application in measuring direct tissue-microstructure features. *Magnetic Resonance in Medicine*, 60(2):439–448, Aug. 2008. ISSN 07403194, 15222594. doi: 10.1002/mrm.21646.
- D. C. Alexander, D. Zikic, A. Ghosh, R. Tanno, V. Wottschel, J. Zhang, E. Kaden, T. B. Dyrby, S. N. Sotiropoulos, H. Zhang, and A. Criminisi. Image quality transfer and applications in diffusion MRI. *NeuroImage*, 152:283–298, May 2017. ISSN 10538119. doi: 10.1016/j.neuroimage.2017.02.089.
- D. C. Alexander, T. B. Dyrby, M. Nilsson, and H. Zhang. Imaging brain microstructure with diffusion MRI: Practicality and applications. *NMR in Biomedicine*, 32(4), Apr. 2019. ISSN 0952-3480, 1099-1492. doi: 10.1002/nbm.3841.
- F. Alfaro-Almagro, M. Jenkinson, N. K. Bangerter, J. L. Andersson, L. Griffanti, G. Douaud, S. N. Sotiropoulos, S. Jbabdi, M. Hernandez-Fernandez, E. Vallee, D. Vidaurre, M. Webster, P. McCarthy, C. Rorden, A. Daducci, D. C. Alexander, H. Zhang, I. Dragonu, P. M. Matthews, K. L. Miller, and S. M. Smith. Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank. *NeuroImage*, 166:400–424, Feb. 2018. ISSN 10538119. doi: 10.1016/j.neuroimage.2017.10.034.
- E. Aliotta, H. Nourzadeh, J. Sanders, D. Muller, and D. B. Ennis. Highly accelerated, model-free diffusion tensor MRI reconstruction using neural networks. *Medical physics*, 46(4):1581–1591, Apr. 2019. ISSN 2473-4209. doi: 10.1002/mp.13400.
- M. H. Alkinani and M. R. El-Sakka. Patch-based models and algorithms for image denoising: A comparative review between patch-based images denoising methods for additive noise reduction. *J Image Video Proc.*, 2017(1):58, Dec. 2017. ISSN 1687-5281. doi: 10.1186/s13640-017-0203-4.

- J. L. R. Andersson and S. N. Sotiropoulos. An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. *NeuroImage*, 125: 1063–1078, Jan. 2016. ISSN 1095-9572. doi: 10.1016/j.neuroimage.2015.10.019.
- A. Annavarapu and S. Borra. Development of magnetic resonance image de-noising methodologies: A comprehensive overview of the state-of-the-art. *Smart Health*, 18: 100138, Nov. 2020. ISSN 23526483. doi: 10.1016/j.smhl.2020.100138.
- Y. Assaf, D. C. Alexander, D. K. Jones, A. Bizzi, T. E. Behrens, C. A. Clark, Y. Cohen, T. B. Dyrby, P. S. Huppi, T. R. Knösche, D. LeBihan, G. J. Parker, and C. Poupon. The CONNECT project: Combining macro- and micro-structure. *NeuroImage*, 80: 273–282, Oct. 2013. ISSN 10538119. doi: 10.1016/j.neuroimage.2013.05.055.
- P. J. Basser. Inferring microstructural features and the physiological state of tissues from diffusion-weighted images. *NMR in biomedicine*, 8(7-8):333–344, 1995. ISSN 0952-3480. doi: 10.1002/nbm.1940080707.
- P. J. Basser, J. Mattiello, and D. LeBihan. Estimation of the Effective Self-Diffusion Tensor from the NMR Spin Echo. *Journal of Magnetic Resonance, Series B*, 103(3):247–254, Mar. 1994. ISSN 1064-1866. doi: 10.1006/jmrb.1994.1037.
- M. Bastiani, M. Cottaar, K. Dikranian, A. Ghosh, H. Zhang, D. C. Alexander, T. E. Behrens, S. Jbabdi, and S. N. Sotiropoulos. Improved tractography using asymmetric fibre orientation distributions. *NeuroImage*, 158:205–218, Sept. 2017. ISSN 10538119. doi: 10.1016/j.neuroimage.2017.06.050.
- M. Bastiani, M. Cottaar, S. P. Fitzgibbon, S. Suri, F. Alfaro-Almagro, S. N. Sotiropoulos, S. Jbabdi, and J. L. Andersson. Automated quality control for within and between studies diffusion MRI data using a non-parametric framework for movement and distortion correction. *NeuroImage*, 184:801–812, Jan. 2019. ISSN 10538119. doi: 10.1016/j.neuroimage.2018.09.073.
- S. Basu, T. Fletcher, and R. Whitaker. Rician Noise Removal in Diffusion Tensor MRI. In D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, R. Larsen, M. Nielsen, and J. Spurring, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2006*, volume 4190, pages 117–125. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006. ISBN 978-3-540-44707-8 978-3-540-44708-5. doi: 10.1007/11866565_15.
- J. Batson and L. Royer. Noise2Self: Blind Denoising by Self-Supervision. *arXiv:1901.11365 [cs, stat]*, June 2019.
- Y. Beauferris, J. Teuwen, D. Karkalousos, N. Moriakov, M. Caan, G. Yiasemis, L. Rodrigues, A. Lopes, H. Pedrini, L. Rittner, M. Dannecker, V. Studenyak, F. Gröger, D. Vyas, S. Faghih-Roohi, A. K. Jethi, J. C. Raju, M. Sivaprakasam, M. Lasby, N. Nogovitsyn, W. Loos, R. Frayne, and R. Souza. Multi-Coil MRI Reconstruction Challenge – Assessing Brain MRI Reconstruction Models and their Generalizability to Varying Coil Configurations, Dec. 2021.
- M. A. Beaumont. Approximate Bayesian Computation. page 27, 2019.
- M. A. Beaumont, W. Zhang, and D. J. Balding. Approximate Bayesian Computation in Population Genetics. *Genetics*, 162(4):2025–2035, Dec. 2002. ISSN 1943-2631. doi: 10.1093/genetics/162.4.2025.

- T. Behrens, M. Woolrich, M. Jenkinson, H. Johansen-Berg, R. Nunes, S. Clare, P. Matthews, J. Brady, and S. Smith. Characterization and propagation of uncertainty in diffusion-weighted MR imaging. *Magnetic Resonance in Medicine*, 50(5):1077–1088, Nov. 2003. ISSN 0740-3194, 1522-2594. doi: 10.1002/mrm.10609.
- T. Behrens, H. J. Berg, S. Jbabdi, M. Rushworth, and M. Woolrich. Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? *NeuroImage*, 34(1):144–155, Jan. 2007. ISSN 10538119. doi: 10.1016/j.neuroimage.2006.09.018.
- T. E. Behrens and O. Sporns. Human connectomics. *Current Opinion in Neurobiology*, 22(1):144–153, Feb. 2012. ISSN 09594388. doi: 10.1016/j.conb.2011.08.005.
- E. Bengio, M. Jain, M. Korablyov, D. Precup, and Y. Bengio. Flow Network based Generative Models for Non-Iterative Diverse Candidate Generation. In *Advances in Neural Information Processing Systems*, volume 34, pages 27381–27394. Curran Associates, Inc., 2021.
- Y. Bengio, R. Ducharme, P. Vincent, and C. Janvin. A neural probabilistic language model. *The Journal of Machine Learning Research*, 3(null):1137–1155, Mar. 2003. ISSN 1532-4435.
- J. I. Berman, S. Chung, P. Mukherjee, C. P. Hess, E. T. Han, and R. G. Henry. Probabilistic streamline Q-ball tractography using the residual bootstrap. *NeuroImage*, 39(1):215–222, Jan. 2008. ISSN 10538119. doi: 10.1016/j.neuroimage.2007.08.021.
- A. S. Bernstein, N.-k. Chen, and T. P. Trouard. Bootstrap analysis of diffusion tensor and mean apparent propagator parameters derived from multiband diffusion MRI. *Magnetic resonance in medicine*, 82(5):1796–1803, Nov. 2019. ISSN 0740-3194. doi: 10.1002/mrm.27833.
- M. Betancourt. Calibrating Model-Based Inferences and Decisions, Mar. 2018.
- D. L. 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. *Radiology*, Nov. 1986. doi: 10.1148/radiology.161.2.3763909.
- C. M. Bishop. Mixture density networks. <https://publications.aston.ac.uk/id/eprint/373/>, 1994.
- B. Biswal, F. Z. Yetkin, V. M. Haughton, and J. S. Hyde. Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magnetic Resonance in Medicine*, 34(4):537–541, Oct. 1995. ISSN 0740-3194. doi: 10.1002/mrm.1910340409.
- J. Bjorck and C. P. Gomes. Dataset Curation Beyond Accuracy. Mar. 2021.
- D. M. Blei, A. Kucukelbir, and J. D. McAuliffe. Variational Inference: A Review for Statisticians. *Journal of the American Statistical Association*, 112(518):859–877, Apr. 2017. ISSN 0162-1459, 1537-274X. doi: 10.1080/01621459.2017.1285773.
- M. G. B. Blum and O. François. Non-linear regression models for Approximate Bayesian Computation. *Statistics and Computing*, 20(1):63–73, Mar. 2009. ISSN 1573-1375. doi: 10.1007/s11222-009-9116-0.
- C. Blundell, J. Cornebise, K. Kavukcuoglu, and D. Wierstra. Weight Uncertainty in Neural Networks, May 2015.

- S. Bond-Taylor, A. Leach, Y. Long, and C. G. Willcocks. Deep Generative Modelling: A Comparative Review of VAEs, GANs, Normalizing Flows, Energy-Based and Autoregressive Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11):7327–7347, Nov. 2022. ISSN 0162-8828, 2160-9292, 1939-3539. doi: 10.1109/TPAMI.2021.3116668.
- J. Brehmer. Simulation-based inference in particle physics. *Nature Reviews Physics*, 3(5): 305–305, May 2021. ISSN 2522-5820. doi: 10.1038/s42254-021-00305-6.
- J. Brehmer, K. Cranmer, S. Mishra-Sharma, F. Kling, and G. Louppe. Mining gold: Improving simulation-based inference with latent information. page 7, 2020.
- W. Brun, G. Keren, G. Kirkebøen, and H. Montgomery. *Perspectives on Thinking, Judging, and Decisions Making: A Tribute to Karl Halvor Teigen*. Universitetsforlaget, Oslo, 2011. ISBN 978-82-15-01878-2.
- A. Buades, B. Coll, and J.-M. Morel. A non-local algorithm for image denoising. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05)*, volume 2, pages 60–65 vol. 2, June 2005. doi: 10.1109/CVPR.2005.38.
- J. Cabral, E. Hugues, O. Sporns, and G. Deco. Role of local network oscillations in resting-state functional connectivity. *NeuroImage*, 57(1):130–139, July 2011. ISSN 1095-9572. doi: 10.1016/j.neuroimage.2011.04.010.
- J. Cabral, E. Hugues, M. L. Kringelbach, and G. Deco. Modeling the outcome of structural disconnection on resting-state functional connectivity. *NeuroImage*, 62(3):1342–1353, Sept. 2012. ISSN 10538119. doi: 10.1016/j.neuroimage.2012.06.007.
- P. T. Callaghan. The measurement of motion using spin echoes. *Principles of nuclear magnetic resonance microscopy*, pages 328–370, 1991.
- P. T. Callaghan, C. D. Eccles, and Y. Xia. NMR microscopy of dynamic displacements: K-space and q-Space imaging. *Journal of Physics E: Scientific Instruments*, 21(8):820–822, Aug. 1988. ISSN 0022-3735. doi: 10.1088/0022-3735/21/8/017.
- E. J. Canales-Rodríguez, J. H. Legarreta, M. Pizzolato, G. Renzonnet, G. Girard, J. R. Patino, M. Barakovic, D. Romascano, Y. Alemán-Gómez, J. Radua, E. Pomarol-Clotet, R. Salvador, J.-P. Thiran, and A. Daducci. Sparse wars: A survey and comparative study of spherical deconvolution algorithms for diffusion MRI. *NeuroImage*, 184:140–160, Jan. 2019. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2018.08.071.
- E. Caruyer, C. Lenglet, G. Sapiro, and R. Deriche. Design of multishell sampling schemes with uniform coverage in diffusion MRI. *Magnetic Resonance in Medicine*, 69(6):1534–1540, 2013. ISSN 1522-2594. doi: 10.1002/mrm.24736.
- M. Catani, R. J. Howard, S. Pajevic, and D. K. Jones. Virtual in Vivo Interactive Dissection of White Matter Fasciculi in the Human Brain. *NeuroImage*, 17(1):77–94, Sept. 2002. ISSN 10538119. doi: 10.1006/nimg.2002.1136.
- M. A. Chappell, A. R. Groves, B. Whitcher, and M. W. Woolrich. Variational Bayesian Inference for a Nonlinear Forward Model. *IEEE Transactions on Signal Processing*, 57(1):223–236, Jan. 2009. ISSN 1941-0476. doi: 10.1109/TSP.2008.2005752.
- M. A. Chappell, M. W. Woolrich, S. Kazan, P. Jezzard, S. J. Payne, and B. J. MacIntosh. Modeling dispersion in arterial spin labeling: Validation using dynamic angiographic measurements. *Magnetic Resonance in Medicine*, 69(2):563–570, 2013. ISSN 1522-2594. doi: 10.1002/mrm.24260.

- G. Chen, Y. Wu, D. Shen, and P.-T. Yap. Noise reduction in diffusion MRI using non-local self-similar information in joint $x - q$ space. *Medical Image Analysis*, 53:79–94, Apr. 2019. ISSN 13618415. doi: 10.1016/j.media.2019.01.006.
- Y. Chen, C.-B. Schönlieb, P. Liò, T. Leiner, P. L. Dragotti, G. Wang, D. Rueckert, D. Firmin, and G. Yang. AI-Based Reconstruction for Fast MRI—A Systematic Review and Meta-Analysis. *Proceedings of the IEEE*, 110(2):224–245, Feb. 2022. ISSN 1558-2256. doi: 10.1109/JPROC.2022.3141367.
- J. Cheng, A. Ghosh, T. Jiang, and R. Deriche. Model-free and analytical EAP reconstruction via spherical polar Fourier diffusion MRI. *Med Image Comput Comput Assist Interv*, 13(Pt 1):590–597, 2010. doi: 10.1007/978-3-642-15705-9_72.
- L. S. Chow and R. Paramesran. Review of medical image quality assessment. *Biomedical Signal Processing and Control*, C(27):145–154, 2016. ISSN 1746-8094. doi: 10.1016/j.bspc.2016.02.006.
- L. S. Chow and H. Rajagopal. Modified-BRISQUE as no reference image quality assessment for structural MR images. *Magnetic Resonance Imaging*, 43:74–87, Nov. 2017. ISSN 0730-725X. doi: 10.1016/j.mri.2017.07.016.
- D. Christiaens, M. Reisert, T. Dhollander, S. Sunaert, P. Suetens, and F. Maes. Global tractography of multi-shell diffusion-weighted imaging data using a multi-tissue model. *NeuroImage*, 123:89–101, Dec. 2015. ISSN 10538119. doi: 10.1016/j.neuroimage.2015.08.008.
- D. Christiaens, J. Veraart, L. Cordero-Grande, A. N. Price, J. Hutter, J. V. Hajnal, and J.-D. Tournier. On the need for bundle-specific microstructure kernels in diffusion MRI. *NeuroImage*, 208:116460, Mar. 2020. ISSN 10538119. doi: 10.1016/j.neuroimage.2019.116460.
- S. Chung, Y. Lu, and R. G. Henry. Comparison of bootstrap approaches for estimation of uncertainties of DTI parameters. 2006. doi: 10.1016/j.neuroimage.2006.07.001.
- W. T. Clarke, C. J. Stagg, and S. Jbabdi. FSL-MRS: An end-to-end spectroscopy analysis package. *Magnetic Resonance in Medicine*, 85(6):2950–2964, 2021. ISSN 1522-2594. doi: 10.1002/mrm.28630.
- Cole. Advances and pitfalls in the analysis and interpretation of resting-state FMRI data. *Frontiers in Systems Neuroscience*, 2010. ISSN 16625137. doi: 10.3389/fnsys.2010.00008.
- D. L. Collins, A. P. Zijdenbos, V. Kollokian, J. G. Sled, N. J. Kabani, C. J. Holmes, and A. C. Evans. Design and construction of a realistic digital brain phantom. *IEEE transactions on medical imaging*, 17(3):463–468, June 1998. ISSN 0278-0062. doi: 10.1109/42.712135.
- C. D. Constantinides, E. Atalar, and E. R. McVeigh. Signal-to-noise measurements in magnitude images from NMR phased arrays. *Magnetic resonance in medicine*, 38(5): 852–857, Nov. 1997. ISSN 0740-3194. doi: 10.1002/mrm.1910380524.
- T. E. Conturo, N. F. Lori, T. S. Cull, E. Akbudak, A. Z. Snyder, J. S. Shimony, R. C. McKinstry, H. Burton, and M. E. Raichle. Tracking neuronal fiber pathways in the living human brain. *Proceedings of the National Academy of Sciences of the United States of America*, 96(18):10422–10427, Aug. 1999. ISSN 0027-8424. doi: 10.1073/pnas.96.18.10422.

- K. Cranmer, J. Pavez, and G. Louppe. Approximating Likelihood Ratios with Calibrated Discriminative Classifiers, Mar. 2016.
- K. Cranmer, J. Brehmer, and G. Louppe. The frontier of simulation-based inference. *Proc Natl Acad Sci USA*, 117(48):30055–30062, Dec. 2020. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1912789117.
- G. Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, 2(4):303–314, Dec. 1989. ISSN 1435-568X. doi: 10.1007/BF02551274.
- A. Daducci, E. J. Canales-Rodríguez, H. Zhang, T. B. Dyrby, D. C. Alexander, and J.-P. Thiran. Accelerated Microstructure Imaging via Convex Optimization (AMICO) from diffusion MRI data. *NeuroImage*, 105:32–44, Jan. 2015. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2014.10.026.
- A. D. Davis, S. Hassel, S. R. Arnott, G. B. Hall, J. K. Harris, M. Zamyadi, J. Downar, B. N. Frey, R. W. Lam, S. H. Kennedy, and S. C. Strother. Biophysical compartment models for single-shell diffusion MRI in the human brain: A model fitting comparison. *Physics in Medicine & Biology*, 67(5):055009, Feb. 2022. ISSN 0031-9155. doi: 10.1088/1361-6560/ac46de.
- A. C. Davison and D. V. Hinkley. *Bootstrap Methods and Their Application*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 1997. ISBN 978-0-521-57471-6. doi: 10.1017/CBO9780511802843.
- A. De Luca, A. Ianus, A. Leemans, M. Palombo, N. Shemesh, H. Zhang, D. C. Alexander, M. Nilsson, M. Froeling, G.-J. Biessels, M. Zucchelli, M. Frigo, E. Albay, S. Sedlar, A. Alimi, S. Deslauriers-Gauthier, R. Deriche, R. Fick, M. Afzali, T. Pieciak, F. Bogusz, S. Aja-Fernández, E. Özarslan, D. K. Jones, H. Chen, M. Jin, Z. Zhang, F. Wang, V. Nath, P. Parvathaneni, J. Morez, J. Sijbers, B. Jeurissen, S. Fadnavis, S. Endres, A. Rokem, E. Garyfallidis, I. Sanchez, V. Prchkovska, P. Rodrigues, B. A. Landman, and K. G. Schilling. On the generalizability of diffusion MRI signal representations across acquisition parameters, sequences and tissue types: Chronicles of the MEMENTO challenge. *NeuroImage*, 240:118367, Oct. 2021. ISSN 10538119. doi: 10.1016/j.neuroimage.2021.118367.
- G. Deco, M. Senden, and V. Jirsa. How anatomy shapes dynamics: A semi-analytical study of the brain at rest by a simple spin model. *Frontiers in Computational Neuroscience*, 6, 2012. ISSN 1662-5188. doi: 10.3389/fncom.2012.00068.
- A. Delaunoy, J. Hermans, F. Rozet, A. Wehenkel, and G. Louppe. Towards Reliable Simulation-Based Inference with Balanced Neural Ratio Estimation. Aug. 2022.
- 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. *IEEE Transactions on Biomedical Engineering*, 54(3):462–472, Mar. 2007. ISSN 0018-9294. doi: 10.1109/TBME.2006.888830.
- M. Descoteaux, R. Deriche, T. R. Knösche, and A. Anwander. Deterministic and probabilistic tractography based on complex fibre orientation distributions. *IEEE transactions on medical imaging*, 28(2):269–286, Feb. 2009a. ISSN 1558-254X. doi: 10.1109/TMI.2008.2004424.
- M. Descoteaux, R. Deriche, D. Le Bihan, J.-F. Mangin, and C. Poupon. Diffusion propagator imaging: Using Laplace’s equation and multiple shell acquisitions to reconstruct

- the diffusion propagator. *Information Processing in Medical Imaging: Proceedings of the ... Conference*, 21:1–13, 2009b. ISSN 1011-2499. doi: 10.1007/978-3-642-02498-6_1.
- A. Deshmane, V. Gulani, M. A. Griswold, and N. Seiberlich. Parallel MR Imaging. *Journal of magnetic resonance imaging : JMRI*, 36(1):55–72, July 2012. ISSN 1053-1807. doi: 10.1002/jmri.23639.
- L. Devroye. *Non-Uniform Random Variate Generation*. Springer, New York, NY, 1986. ISBN 978-1-4613-8645-2 978-1-4613-8643-8. doi: 10.1007/978-1-4613-8643-8.
- P. Dhariwal and A. Nichol. Diffusion Models Beat GANs on Image Synthesis. In *Advances in Neural Information Processing Systems*, volume 34, pages 8780–8794. Curran Associates, Inc., 2021.
- O. Dietrich, S. Heiland, and K. Sartor. Noise correction for the exact determination of apparent diffusion coefficients at low SNR. *Magnetic Resonance in Medicine*, 45(3): 448–453, Mar. 2001. ISSN 0740-3194, 1522-2594. doi: 10.1002/1522-2594(200103)45:3{\${\$}448::AID-MRM1059{\${\$}}3.0.CO;2-W.
- O. Dietrich, J. G. Raya, S. B. Reeder, M. F. Reiser, and S. O. Schoenberg. Measurement of signal-to-noise ratios in MR images: Influence of multichannel coils, parallel imaging, and reconstruction filters. *Journal of Magnetic Resonance Imaging*, 26(2):375–385, Aug. 2007. ISSN 10531807, 15222586. doi: 10.1002/jmri.20969.
- 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. *Magnetic Resonance Imaging*, 26(6):754–762, July 2008. ISSN 0730725X. doi: 10.1016/j.mri.2008.02.001.
- O. Dietrich, A. Biffar, A. Baur-Melnyk, and M. F. Reiser. Technical aspects of MR diffusion imaging of the body. *European Journal of Radiology*, 76(3):314–322, Dec. 2010. ISSN 0720-048X, 1872-7727. doi: 10.1016/j.ejrad.2010.02.018.
- P. J. Diggle and R. J. Gratton. Monte Carlo Methods of Inference for Implicit Statistical Models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 46(2):193–227, 1984. ISSN 0035-9246.
- L. Dinh, D. Krueger, and Y. Bengio. NICE: Non-linear Independent Components Estimation, Apr. 2015.
- H. M. Dolatabadi, S. Erfani, and C. Leckie. Invertible Generative Modeling using Linear Rational Splines, Apr. 2020.
- A. T. Drysdale, L. Grosenick, J. Downar, K. Dunlop, F. Mansouri, Y. Meng, R. N. Fetcho, B. Zebley, D. J. Oathes, A. Etkin, A. F. Schatzberg, K. Sudheimer, J. Keller, H. S. Mayberg, F. M. Gunning, G. S. Alexopoulos, M. D. Fox, A. Pascual-Leone, H. U. Voss, B. Casey, M. J. Dubin, and C. Liston. Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nature medicine*, 23(1):28–38, Jan. 2017. ISSN 1078-8956, 1546-170X. doi: 10.1038/nm.4246.
- J. Dubois and R. Adolphs. Building a Science of Individual Differences from fMRI. *Trends in Cognitive Sciences*, 20(6):425–443, June 2016. ISSN 13646613. doi: 10.1016/j.tics.2016.03.014.
- C. Durkan, G. Papamakarios, and I. Murray. Sequential Neural Methods for Likelihood-free Inference, Nov. 2018.

- C. Durkan, A. Bekasov, I. Murray, and G. Papamakarios. Neural Spline Flows. *arXiv:1906.04032 [cs, stat]*, Dec. 2019.
- C. Durkan, I. Murray, and G. Papamakarios. On Contrastive Learning for Likelihood-free Inference. *arXiv:2002.03712 [cs, stat]*, Dec. 2020.
- W. A. Edelstein, P. A. Bottomley, and L. M. Pfeifer. A signal-to-noise calibration procedure for NMR imaging systems: Signal-to-noise calibration for NMR imaging systems. *Medical Physics*, 11(2):180–185, Mar. 1984. ISSN 00942405. doi: 10.1118/1.595484.
- C. Eichner, S. F. Cauley, J. Cohen-Adad, H. E. Möller, R. Turner, K. Setsompop, and L. L. Wald. Real diffusion-weighted MRI enabling true signal averaging and increased diffusion contrast. *NeuroImage*, 122:373–384, Nov. 2015. ISSN 10538119. doi: 10.1016/j.neuroimage.2015.07.074.
- S. B. Eickhoff, B. Thirion, G. Varoquaux, and D. Bzdok. Connectivity-based parcellation: Critique and implications. *Human Brain Mapping*, 36(12):4771–4792, Dec. 2015. ISSN 1065-9471, 1097-0193. doi: 10.1002/hbm.22933.
- A. Einstein. *Investigations on the Theory of the Brownian Movement*. Courier Corporation, Jan. 1956. ISBN 978-0-486-60304-9.
- N. Evangelou. Regional axonal loss in the corpus callosum correlates with cerebral white matter lesion volume and distribution in multiple sclerosis. *Brain : a journal of neurology*, 123(9):1845–1849, Sept. 2000. ISSN 14602156. doi: 10.1093/brain/123.9.1845.
- S. Fadnavis, J. Batson, and E. Garyfallidis. Patch2Self: Denoising Diffusion MRI with Self-Supervised Learning. page 11, 2020.
- S. Fadnavis, A. Chowdhury, J. Batson, P. Drineas, and E. Garyfallidis. Patch2Self denoising of Diffusion MRI with Self-Supervision and Matrix Sketching, Mar. 2022a.
- S. Fadnavis, J. Sjölund, A. Eklund, and E. Garyfallidis. NUQ: A Noise Metric for Diffusion MRI via Uncertainty Discrepancy Quantification, Mar. 2022b.
- L. Fan, F. Zhang, H. Fan, and C. Zhang. Brief review of image denoising techniques. 2 (1):7, Dec. 2019. ISSN 2524-4442. doi: 10.1186/s42492-019-0016-7.
- S. Farquharson, J.-D. Tournier, F. Calamante, G. Fabinyi, M. Schneider-Kolsky, G. D. Jackson, and A. Connelly. White matter fiber tractography: Why we need to move beyond DTI: Clinical article. *Journal of Neurosurgery*, 118(6):1367–1377, June 2013. ISSN 1933-0693, 0022-3085. doi: 10.3171/2013.2.JNS121294.
- U. Ferizi, T. Schneider, E. Panagiotaki, G. Nedjati-Gilani, H. Zhang, C. A. M. Wheeler-Kingshott, and D. C. Alexander. A ranking of diffusion MRI compartment models with in vivo human brain data. *Magnetic Resonance in Medicine*, 72(6):1785–1792, Dec. 2014. ISSN 0740-3194. doi: 10.1002/mrm.25080.
- A. Fick. Ueber Diffusion. *Annalen der Physik*, 170(1):59–86, 1855. ISSN 1521-3889. doi: 10.1002/andp.18551700105.
- R. H. Fick, N. Sepasian, M. Pizzolato, A. Ianus, and R. Deriche. Assessing the feasibility of estimating axon diameter using diffusion models and machine learning. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pages 766–769, Apr. 2017. doi: 10.1109/ISBI.2017.7950631.
- B. Fischl. FreeSurfer. *NeuroImage*, 62(2):774–781, Aug. 2012. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2012.01.021.

- E. Garyfallidis, M. Brett, B. Amirbekian, A. Rokem, S. van der Walt, M. Descoteaux, I. Nimmo-Smith, and Dipy Contributors. Dipy, a library for the analysis of diffusion MRI data. *Frontiers in Neuroinformatics*, 8, Feb. 2014. ISSN 1662-5196. doi: 10.3389/fninf.2014.00008.
- A. Gelman, W. R. Gilks, and G. O. Roberts. Weak convergence and optimal scaling of random walk Metropolis algorithms. *The Annals of Applied Probability*, 7(1):110–120, Feb. 1997. ISSN 1050-5164, 2168-8737. doi: 10.1214/aoap/1034625254.
- A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC, New York, 3 edition, July 2015. ISBN 978-0-429-11307-9. doi: 10.1201/b16018.
- M. Germain, K. Gregor, I. Murray, and H. Larochelle. MADE: Masked Autoencoder for Distribution Estimation. *arXiv:1502.03509 [cs, stat]*, June 2015.
- E. K. Gibbons, K. K. Hodgson, A. S. Chaudhari, L. G. Richards, J. J. Majersik, G. Adluru, and E. V. R. DiBella. Simultaneous NODDI and GFA parameter map generation from subsampled Q-space imaging using deep learning. *Magnetic resonance in medicine*, 81(4):2399–2411, Apr. 2019. ISSN 1522-2594. doi: 10.1002/mrm.27568.
- W. R. Gilks, S. Richardson, and D. Spiegelhalter. *Markov Chain Monte Carlo in Practice*. CRC Press, Dec. 1995. ISBN 978-0-412-05551-5.
- M. Glöckler, M. Deistler, and J. H. Macke. Variational methods for simulation-based inference, Mar. 2022.
- A. S. Goldberger. *Econometric Theory*. J. Wiley, 1964. ISBN 978-0-471-31101-0.
- V. Golkov, A. Dosovitskiy, J. I. Sperl, M. I. Menzel, M. Czisch, P. Sämann, T. Brox, and D. Cremers. Q-Space Deep Learning: Twelve-Fold Shorter and Model-Free Diffusion MRI Scans. *IEEE Transactions on Medical Imaging*, 35(5):1344–1351, May 2016. ISSN 1558-254X. doi: 10.1109/TMI.2016.2551324.
- P. J. Gonçalves, J.-M. Lueckmann, M. Deistler, M. Nonnenmacher, K. Öcal, G. Bassetto, C. Chintaluri, W. F. Podlaski, S. A. Haddad, T. P. Vogels, D. S. Greenberg, and J. H. Macke. Training deep neural density estimators to identify mechanistic models of neural dynamics. *eLife*, 9:e56261, Sept. 2020. ISSN 2050-084X. doi: 10.7554/eLife.56261.
- I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative Adversarial Networks. *arXiv:1406.2661 [cs, stat]*, June 2014.
- C. Gourieroux, A. Monfort, and E. Renault. Indirect inference. *Journal of Applied Econometrics*, 8(S1):S85–S118, 1993. ISSN 1099-1255. doi: 10.1002/jae.3950080507.
- B. Goyal, A. Dogra, S. Agrawal, B. Sohi, and A. Sharma. Image denoising review: From classical to state-of-the-art approaches. *Information Fusion*, 55:220–244, Mar. 2020. ISSN 15662535. doi: 10.1016/j.inffus.2019.09.003.
- D. S. Greenberg, M. Nonnenmacher, and J. H. Macke. Automatic Posterior Transformation for Likelihood-Free Inference, May 2019.
- L. Griffanti, M. Rolinski, K. Szewczyk-Krolikowski, R. A. Menke, N. Filippini, G. Zamboni, M. Jenkinson, M. T. Hu, and C. E. Mackay. Challenges in the reproducibility of clinical studies with resting state fMRI: An example in early Parkinson’s disease. *NeuroImage*, 124:704–713, Jan. 2016. ISSN 10538119. doi: 10.1016/j.neuroimage.2015.09.021.

- M. A. Griswold, P. M. Jakob, R. M. Heidemann, M. Nittka, V. Jellus, J. Wang, B. Kiefer, and A. Haase. Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magnetic Resonance in Medicine*, 47(6):1202–1210, June 2002. ISSN 0740-3194. doi: 10.1002/mrm.10171.
- X. Gu, A. Eklund, E. Özarslan, and H. Knutsson. Using the Wild Bootstrap to Quantify Uncertainty in Mean Apparent Propagator MRI. *Frontiers in Neuroinformatics*, 13, 2019. ISSN 1662-5196.
- H. Gudbjartsson and S. Patz. The rician distribution of noisy mri data. *Magnetic Resonance in Medicine*, 34(6):910–914, Dec. 1995. ISSN 07403194, 15222594. doi: 10.1002/mrm.1910340618.
- C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger. On Calibration of Modern Neural Networks, Aug. 2017.
- O. J. Gurney-Champion, D. J. Collins, A. Wetscherek, M. Rata, R. Klaassen, H. W. M. van Laarhoven, K. J. Harrington, U. Oelfke, and M. R. Orton. Principal component analysis for fast and model-free denoising of multi b-Value diffusion-weighted MR images. *Physics in medicine and biology*, 64(10):105015, May 2019. ISSN 1361-6560. doi: 10.1088/1361-6560/ab1786.
- N. G. Gyori, M. Palombo, C. A. Clark, H. Zhang, and D. C. Alexander. Training data distribution significantly impacts the estimation of tissue microstructure with machine learning. *Magnetic Resonance in Medicine*, 87(2):932–947, 2022. ISSN 1522-2594. doi: 10.1002/mrm.29014.
- H. Haario, E. Saksman, and J. Tamminen. An Adaptive Metropolis Algorithm. *Bernoulli*, 7(2):223–242, 2001. ISSN 1350-7265. doi: 10.2307/3318737.
- J. Hamilton, D. Franson, and N. Seiberlich. Recent advances in parallel imaging for MRI. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 101:71–95, Aug. 2017. ISSN 0079-6565. doi: 10.1016/j.pnmrs.2017.04.002.
- L. Hansen and P. Salamon. Neural network ensembles. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(10):993–1001, Oct. 1990. ISSN 1939-3539. doi: 10.1109/34.58871.
- M. P. Harms, L. H. Somerville, B. M. Ances, J. Andersson, D. M. Barch, M. Bastiani, S. Y. Bookheimer, T. B. Brown, R. L. Buckner, G. C. Burgess, T. S. Coalson, M. A. Chappell, M. Dapretto, G. Douaud, B. Fischl, M. F. Glasser, D. N. Greve, C. Hodge, K. W. Jamison, S. Jbabdi, S. Kandala, X. Li, R. W. Mair, S. Mangia, D. Marcus, D. Mascali, S. Moeller, T. E. Nichols, E. C. Robinson, D. H. Salat, S. M. Smith, S. N. Sotiropoulos, M. Terpstra, K. M. Thomas, M. D. Tisdall, K. Ugurbil, A. van der Kouwe, R. P. Woods, L. Zöllei, D. C. Van Essen, and E. Yacoub. Extending the Human Connectome Project across ages: Imaging protocols for the Lifespan Development and Aging projects. *NeuroImage*, 183:972–984, Dec. 2018. ISSN 10538119. doi: 10.1016/j.neuroimage.2018.09.060.
- R. Harms, F. Fritz, A. Tobisch, R. Goebel, and A. Roebroeck. Robust and fast nonlinear optimization of diffusion MRI microstructure models. *NeuroImage*, 155:82–96, July 2017. ISSN 10538119. doi: 10.1016/j.neuroimage.2017.04.064.
- H. A. Haroon, D. M. Morris, K. V. Embleton, D. C. Alexander, and G. J. M. Parker. Using the Model-Based Residual Bootstrap to Quantify Uncertainty in Fiber Orientations From \mathbb{S}^2 -Ball Analysis. *IEEE Transactions on Medical Imaging*, 28(4):535–550, Apr. 2009. ISSN 1558-254X. doi: 10.1109/TMI.2008.2006528.

- M. Hashemi, A. N. Vattikonda, J. Jha, V. Sip, M. M. Woodman, F. Bartolomei, and V. K. Jirsa. Simulation-Based Inference for Whole-Brain Network Modeling of Epilepsy using Deep Neural Density Estimators. Preprint, Neurology, June 2022.
- W. K. Hastings. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika*, 57(1):97–109, 1970. ISSN 0006-3444. doi: 10.2307/2334940.
- S. Hayasaka and T. E. Nichols. Validating cluster size inference: Random field and permutation methods. *NeuroImage*, 20(4):2343–2356, Dec. 2003. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2003.08.003.
- Y.-C. Heo, K. Kim, and Y. Lee. Image Denoising Using Non-Local Means (NLM) Approach in Magnetic Resonance (MR) Imaging: A Systematic Review. *Applied Sciences*, 10(20):7028, Oct. 2020. ISSN 2076-3417. doi: 10.3390/app10207028.
- J. Hermans, V. Begy, and G. Louppe. Likelihood-free MCMC with amortized approximate ratio estimators. In *Proceedings of the 37th International Conference on Machine Learning*, ICML’20, pages 4239–4248. JMLR.org, July 2020.
- J. Hermans, A. Delaunoy, F. Rozet, A. Wehenkel, and G. Louppe. Averting A Crisis In Simulation-Based Inference, Oct. 2021.
- M. Hernández, G. D. Guerrero, J. M. Cecilia, J. M. García, A. Inuggi, S. Jbabdi, T. E. J. Behrens, and S. N. Sotiropoulos. Accelerating Fibre Orientation Estimation from Diffusion Weighted Magnetic Resonance Imaging Using GPUs. *PLOS ONE*, 8(4):e61892, Apr. 2013. ISSN 1932-6203. doi: 10.1371/journal.pone.0061892.
- M. Hernandez-Fernandez, I. Regul, S. Jbabdi, M. Giles, S. Smith, and S. N. Sotiropoulos. Using GPUs to accelerate computational diffusion MRI: From microstructure estimation to tractography and connectomes. *NeuroImage*, 188:598–615, Mar. 2019. ISSN 10538119. doi: 10.1016/j.neuroimage.2018.12.015.
- I. Hill, M. Palombo, M. Santin, F. Branzoli, A.-C. Philippe, D. Wassermann, M.-S. Aigrot, B. Stankoff, A. Baron-Van Evercooren, M. Felfli, D. Langui, H. Zhang, S. Lehericy, A. Petiet, D. C. Alexander, O. Ciccarelli, and I. Drobnjak. Machine learning based white matter models with permeability: An experimental study in cuprizone treated in-Vivo mouse model of axonal demyelination. *NeuroImage*, 224:117425, Jan. 2021. ISSN 10538119. doi: 10.1016/j.neuroimage.2020.117425.
- R. K. Hobbie and B. J. Roth. Magnetic Resonance Imaging. In R. K. Hobbie and B. J. Roth, editors, *Intermediate Physics for Medicine and Biology*, pages 515–542. Springer, New York, NY, 2007. ISBN 978-0-387-49885-0. doi: 10.1007/978-0-387-49885-0_18.
- C. J. Honey, O. Sporns, L. Cammoun, X. Gigandet, J. P. Thiran, R. Meuli, and P. Hagmann. Predicting human resting-state functional connectivity from structural connectivity. *Proceedings of the National Academy of Sciences*, 106(6):2035–2040, Feb. 2009. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.0811168106.
- K. Hornik, M. Stinchcombe, and H. White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366, Jan. 1989. ISSN 08936080. doi: 10.1016/0893-6080(89)90020-8.
- H. Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6):417–441, 1933. ISSN 1939-2176. doi: 10.1037/h0071325.

- A. F. Howard, J. Mollink, M. Kleinnijenhuis, M. Pallegage-Gamarallage, M. Bastiani, M. Cottaar, K. L. Miller, and S. Jbabdi. Joint modelling of diffusion MRI and microscopy. *NeuroImage*, 201:116014, Nov. 2019. ISSN 10538119. doi: 10.1016/j.neuroimage.2019.116014.
- P. Huang, M. M. Correia, C. Rua, C. T. Rodgers, R. N. Henson, and J. D. Carlin. Correcting for Superficial Bias in 7T Gradient Echo fMRI. *Frontiers in Neuroscience*, 15, 2021. ISSN 1662-453X.
- E. Hüllermeier and W. Waegeman. Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. *Machine Learning*, 110(3):457–506, Mar. 2021. ISSN 0885-6125, 1573-0565. doi: 10.1007/s10994-021-05946-3.
- E. B. Hutchinson, A. V. Avram, M. O. Irfanoglu, C. G. Koay, A. S. Barnett, M. E. Komlosh, E. Özarslan, S. C. Schwerin, S. L. Juliano, and C. Pierpaoli. Analysis of the effects of noise, DWI sampling, and value of assumed parameters in diffusion MRI models. *Magnetic Resonance in Medicine*, 78(5):1767–1780, 2017. ISSN 1522-2594. doi: 10.1002/mrm.26575.
- A. E. Ilesanmi and T. O. Ilesanmi. Methods for image denoising using convolutional neural network: A review. *Complex Intell. Syst.*, 7(5):2179–2198, Oct. 2021. ISSN 2199-4536, 2198-6053. doi: 10.1007/s40747-021-00428-4.
- M. Jallais, P. Rodrigues, A. Gramfort, and D. Wassermann. Inverting brain grey matter models with likelihood-free inference: A tool for trustable cytoarchitecture measurements. *arXiv:2111.08693 [cs, q-bio]*, Nov. 2021.
- M. Jallais, P. L. C. Rodrigues, A. Gramfort, and D. Wassermann. Inverting brain grey matter models with likelihood-free inference: A tool for trustable cytoarchitecture measurements, May 2022.
- S. Jbabdi, M. Woolrich, J. Andersson, and T. Behrens. A Bayesian framework for global tractography. *NeuroImage*, 37(1):116–129, Aug. 2007. ISSN 10538119. doi: 10.1016/j.neuroimage.2007.04.039.
- S. Jbabdi, S. N. Sotiropoulos, A. M. Savio, M. Graña, and T. E. J. Behrens. Model-based analysis of multishell diffusion MR data for tractography: How to get over fitting problems. *Magnetic Resonance in Medicine*, 68(6):1846–1855, 2012. ISSN 1522-2594. doi: 10.1002/mrm.24204.
- S. Jbabdi, S. N. Sotiropoulos, S. N. Haber, D. C. Van Essen, and T. E. Behrens. Measuring macroscopic brain connections in vivo. *Nature neuroscience*, 18(11):1546–1555, Nov. 2015. ISSN 1097-6256, 1546-1726. doi: 10.1038/nn.4134.
- I. O. Jelescu, J. Veraart, V. Adisetiyo, S. S. Milla, D. S. Novikov, and E. Fieremans. One diffusion acquisition and different white matter models: How does microstructure change in human early development based on WMTI and NODDI? *NeuroImage*, 107: 242–256, Feb. 2015. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2014.12.009.
- I. O. Jelescu, M. Palombo, F. Bagnato, and K. G. Schilling. Challenges for biophysical modeling of microstructure. *Journal of Neuroscience Methods*, 344:108861, Oct. 2020. ISSN 01650270. doi: 10.1016/j.jneumeth.2020.108861.
- M. Jenkinson, C. F. Beckmann, T. E. J. Behrens, M. W. Woolrich, and S. M. Smith. FSL. *NeuroImage*, 62(2):782–790, Aug. 2012. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2011.09.015.

- S. N. Jespersen, C. D. Kroenke, L. Østergaard, J. J. Ackerman, and D. A. Yablonskiy. Modeling dendrite density from magnetic resonance diffusion measurements. *NeuroImage*, 34(4):1473–1486, Feb. 2007. ISSN 10538119. doi: 10.1016/j.neuroimage.2006.10.037.
- S. N. Jespersen, C. R. Bjarkam, J. R. Nyengaard, M. M. Chakravarty, B. Hansen, T. Vosegaard, L. Østergaard, D. Yablonskiy, N. C. Nielsen, and P. Vestergaard-Poulsen. Neurite density from magnetic resonance diffusion measurements at ultrahigh field: Comparison with light microscopy and electron microscopy. *Neuroimage*, 49(1):205–216, Jan. 2010. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2009.08.053.
- 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: Prevalence of Multifiber Voxels in WM. *Hum. Brain Mapp*, 34(11):2747–2766, Nov. 2013. ISSN 10659471. doi: 10.1002/hbm.22099.
- 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. *NeuroImage*, 103:411–426, Dec. 2014. ISSN 1095-9572. doi: 10.1016/j.neuroimage.2014.07.061.
- B. Jeurissen, M. Descoteaux, S. Mori, and A. Leemans. Diffusion MRI fiber tractography of the brain. *NMR in Biomedicine*, 32(4), Apr. 2019. ISSN 0952-3480, 1099-1492. doi: 10.1002/nbm.3785.
- J. B. Johnson. Thermal Agitation of Electricity in Conductors. *Physical Review*, 32(1):97–109, July 1928. doi: 10.1103/PhysRev.32.97.
- D. Jones, M. Horsfield, and A. Simmons. Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging. *Magnetic Resonance in Medicine*, 42(3):515–525, 1999. ISSN 1522-2594. doi: 10.1002/(SICI)1522-2594(199909)42:3{1515::AID-MRM14{3.0.CO;2-Q.
- D. K. Jones. Determining and visualizing uncertainty in estimates of fiber orientation from diffusion tensor MRI. *Magnetic Resonance in Medicine*, 49(1):7–12, Jan. 2003. ISSN 0740-3194, 1522-2594. doi: 10.1002/mrm.10331.
- D. K. Jones. The effect of gradient sampling schemes on measures derived from diffusion tensor MRI: A Monte Carlo study. *Magnetic Resonance in Medicine*, 51(4):807–815, Apr. 2004. ISSN 0740-3194. doi: 10.1002/mrm.20033.
- D. K. Jones. Tractography Gone Wild: Probabilistic Fibre Tracking Using the Wild Bootstrap With Diffusion Tensor MRI. *IEEE Transactions on Medical Imaging*, 27(9):1268–1274, Sept. 2008. ISSN 1558-254X. doi: 10.1109/TMI.2008.922191.
- D. K. Jones and P. J. Basser. Squashing peanuts and smashing pumpkins?: How noise distorts diffusion-weighted MR data. *Magnetic Resonance in Medicine*, 52(5):979–993, Nov. 2004. ISSN 0740-3194, 1522-2594. doi: 10.1002/mrm.20283.
- D. K. Jones and C. Pierpaoli. Confidence mapping in diffusion tensor magnetic resonance imaging tractography using a bootstrap approach. *Magnetic Resonance in Medicine*, 53(5):1143–1149, 2005. ISSN 1522-2594. doi: 10.1002/mrm.20466.
- 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. *NeuroImage*, 73:239–254, June 2013. ISSN 10538119. doi: 10.1016/j.neuroimage.2012.06.081.

- E. Kaden and F. Kruggel. Nonparametric Bayesian inference of the fiber orientation distribution from diffusion-weighted MR images. *Medical Image Analysis*, 16(4):876–888, May 2012. ISSN 13618415. doi: 10.1016/j.media.2012.01.004.
- E. Kaden, A. Anwander, and T. R. Knösche. Variational inference of the fiber orientation density using diffusion MR imaging. *NeuroImage*, 42(4):1366–1380, Oct. 2008. ISSN 10538119. doi: 10.1016/j.neuroimage.2008.06.004.
- E. R. Kandel, J. H. Schwartz, and T. M. Jessell. *Principles of Neural Science*. McGraw-Hill, Health Professions Division, New York, 2000. ISBN 978-0-8385-7701-1 978-0-07-112000-5.
- D. Karimi, L. Vasung, C. Jaimes, F. Machado-Rivas, S. K. Warfield, and A. Gholipour. Learning to estimate the fiber orientation distribution function from diffusion-weighted MRI. *NeuroImage*, 239:118316, Oct. 2021. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2021.118316.
- N. Kasabov. NeuCube EvoSpike architecture for spatio-temporal modelling and pattern recognition of brain signals. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7477 LNAI: 225–243, 2012. ISSN 03029743. doi: 10.1007/978-3-642-33212-8{\textbackslash}_}21.
- R. E. Kass, B. P. Carlin, A. Gelman, and R. M. Neal. Markov Chain Monte Carlo in Practice: A Roundtable Discussion. *The American Statistician*, 52(2):93–100, 1998. ISSN 0003-1305. doi: 10.2307/2685466.
- K. Kay. The risk of bias in denoising methods: Examples from neuroimaging. *PLOS ONE*, 17(7):e0270895, July 2022. ISSN 1932-6203. doi: 10.1371/journal.pone.0270895.
- K. Kay, A. Rokem, J. Winawer, R. Dougherty, and B. Wandell. GLMdenoise: A fast, automated technique for denoising task-based fMRI data. *Frontiers in Neuroscience*, 7, 2013. ISSN 1662-453X.
- P. Kellman and E. R. McVeigh. Image reconstruction in SNR units: A general method for SNR measurement. *Magnetic resonance in medicine*, 54(6):1439–1447, Dec. 2005. ISSN 0740-3194. doi: 10.1002/mrm.20713.
- D. H. C. Kim, L. J. Williams, M. Hernandez-Fernandez, and B. H. Bjornson. Comparison of CPU and GPU bayesian estimates of fibre orientations from diffusion MRI. *PLOS ONE*, 17(4):e0252736, Apr. 2022. ISSN 1932-6203. doi: 10.1371/journal.pone.0252736.
- D. P. Kingma and M. Welling. Auto-Encoding Variational Bayes, May 2014.
- F. Knoll, K. Hammernik, C. Zhang, S. Moeller, T. Pock, D. K. Sodickson, and M. Akcakaya. Deep-Learning Methods for Parallel Magnetic Resonance Imaging Reconstruction: A Survey of the Current Approaches, Trends, and Issues. *IEEE Signal Processing Magazine*, 37(1):128–140, Jan. 2020. ISSN 1558-0792. doi: 10.1109/MSP.2019.2950640.
- C. G. Koay and P. J. Basser. Analytically exact correction scheme for signal extraction from noisy magnitude MR signals. *Journal of Magnetic Resonance*, 179(2):317–322, Apr. 2006. ISSN 10907807. doi: 10.1016/j.jmr.2006.01.016.
- C. G. Koay, E. Özarslan, and P. J. Basser. A Signal-Transformational Framework for Breaking the Noise Floor and Its Applications in MRI. *Journal of magnetic resonance (San Diego, Calif. : 1997)*, 197(2):108–119, Apr. 2009. ISSN 1090-7807. doi: 10.1016/j.jmr.2008.11.015.

- I. Kobyzev, S. J. D. Prince, and M. A. Brubaker. Normalizing Flows: An Introduction and Review of Current Methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(11):3964–3979, Nov. 2021. ISSN 0162-8828, 2160-9292, 1939-3539. doi: 10.1109/TPAMI.2020.2992934.
- M. A. Koch, D. G. Norris, and M. Hund-Georgiadis. An Investigation of Functional and Anatomical Connectivity Using Magnetic Resonance Imaging. *NeuroImage*, 16(1): 241–250, May 2002. ISSN 10538119. doi: 10.1006/nimg.2001.1052.
- S. Koppers and D. Merhof. Direct Estimation of Fiber Orientations Using Deep Learning in Diffusion Imaging. In L. Wang, E. Adeli, Q. Wang, Y. Shi, and H.-I. Suk, editors, *Machine Learning in Medical Imaging*, Lecture Notes in Computer Science, pages 53–60, Cham, 2016. Springer International Publishing. ISBN 978-3-319-47157-0. doi: 10.1007/978-3-319-47157-0_7.
- S. Koppers, M. Friedrichs, and D. Merhof. Reconstruction of Diffusion Anisotropies Using 3D Deep Convolutional Neural Networks in Diffusion Imaging. In T. Schultz, E. Özarslan, and I. Hotz, editors, *Modeling, Analysis, and Visualization of Anisotropy*, pages 393–404. Springer International Publishing, Cham, 2017. ISBN 978-3-319-61357-4 978-3-319-61358-1. doi: 10.1007/978-3-319-61358-1_17.
- B. Kreher, I. Mader, and V. Kiselev. Gibbs tracking: A novel approach for the reconstruction of neuronal pathways. *Magnetic Resonance in Medicine*, 60(4):953–963, Oct. 2008. ISSN 07403194, 15222594. doi: 10.1002/mrm.21749.
- V. Kuleshov and S. Deshpande. Calibrated and Sharp Uncertainties in Deep Learning via Density Estimation. In *Proceedings of the 39th International Conference on Machine Learning*, pages 11683–11693. PMLR, June 2022.
- B. Lakshminarayanan, A. Pritzel, and C. Blundell. Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles, Nov. 2017.
- H. Larochelle and I. Murray. The Neural Autoregressive Distribution Estimator. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pages 29–37. JMLR Workshop and Conference Proceedings, June 2011.
- E. G. Larsson, D. Erdogmus, R. Yan, J. C. Principe, and J. R. Fitzsimmons. SNR-optimality of sum-of-squares reconstruction for phased-array magnetic resonance imaging. *Journal of Magnetic Resonance (San Diego, Calif.: 1997)*, 163(1):121–123, July 2003. ISSN 1090-7807. doi: 10.1016/s1090-7807(03)00132-0.
- P. C. Lauterbur. Image Formation by Induced Local Interactions: Examples Employing Nuclear Magnetic Resonance. *Nature*, 242(5394):190–191, Mar. 1973. ISSN 1476-4687. doi: 10.1038/242190a0.
- M. Lazar and A. L. Alexander. An error analysis of white matter tractography methods: Synthetic diffusion tensor field simulations. *NeuroImage*, 20(2):1140–1153, Oct. 2003. ISSN 10538119. doi: 10.1016/S1053-8119(03)00277-5.
- M. Lazar and A. L. Alexander. Bootstrap white matter tractography (BOOT-TRAC). 2004. doi: 10.1016/j.neuroimage.2004.08.050.
- D. Le Bihan, C. Poupon, A. Amadon, and F. Lethimonnier. Artifacts and pitfalls in diffusion MRI. *Journal of Magnetic Resonance Imaging*, 24(3):478–488, Sept. 2006. ISSN 1053-1807, 1522-2586. doi: 10.1002/jmri.20683.

- N. E. Leeds and S. A. Kieffer. Evolution of Diagnostic Neuroradiology from 1904 to 1999. *Radiology*, 217(2):309–318, Nov. 2000. ISSN 0033-8419, 1527-1315. doi: 10.1148/radiology.217.2.r00nv45309.
- J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila. Noise2Noise: Learning Image Restoration without Clean Data. *ICML*, 2018.
- H. Lin, V. Hosu, and D. Saupe. DeepFL-IQA: Weak Supervision for Deep IQA Feature Learning. *arXiv:2001.08113 [cs, eess]*, Jan. 2020.
- Z. Lin, T. Gong, K. Wang, Z. Li, H. He, Q. Tong, F. Yu, and J. Zhong. Fast learning of fiber orientation distribution function for MR tractography using convolutional neural network. *Medical Physics*, 46(7):3101–3116, 2019. ISSN 2473-4209. doi: 10.1002/mp.13555.
- R. Y. Liu. Bootstrap Procedures under some Non-I.I.D. Models. *The Annals of Statistics*, 16(4):1696–1708, Dec. 1988. ISSN 0090-5364, 2168-8966. doi: 10.1214/aos/1176351062.
- J.-M. Lueckmann, P. J. Goncalves, G. Bassetto, K. Öcal, M. Nonnenmacher, and J. H. Macke. Flexible statistical inference for mechanistic models of neural dynamics. *arXiv:1711.01861 [stat]*, Nov. 2017.
- J.-M. Lueckmann, J. Boelts, D. Greenberg, P. Goncalves, and J. Macke. Benchmarking Simulation-Based Inference. In *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, pages 343–351. PMLR, Mar. 2021.
- A. S. Lundervold and A. Lundervold. An overview of deep learning in medical imaging focusing on MRI. *Zeitschrift für Medizinische Physik*, 29(2):102–127, May 2019. ISSN 09393889. doi: 10.1016/j.zemedi.2018.11.002.
- M. Lustig, D. Donoho, and J. M. Pauly. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine*, 58(6):1182–1195, 2007. ISSN 1522-2594. doi: 10.1002/mrm.21391.
- X. Ma, K. Uğurbil, and X. Wu. Denoise magnitude diffusion magnetic resonance images via variance-stabilizing transformation and optimal singular-value manipulation. *NeuroImage*, 215:116852, July 2020. ISSN 10538119. doi: 10.1016/j.neuroimage.2020.116852.
- D. Macêdo, C. Zanchettin, and T. Ludermir. Distinction Maximization Loss: Efficiently Improving Out-of-Distribution Detection and Uncertainty Estimation by Replacing the Loss and Calibrating. May 2022.
- D. J. C. MacKay. *Bayesian Methods for Adaptive Models*. PhD thesis, California Institute of Technology, 1992.
- D. J. C. MacKay. Probable networks and plausible predictions - a review of practical Bayesian methods for supervised neural networks. page 37, 1995.
- D. J. C. MacKay, D. J. C. M. Kay, and v. J. C. MacKay. *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, Sept. 2003. ISBN 978-0-521-64298-9.
- T. M. Madhyastha, N. Koh, T. K. M. Day, M. Hernández-Fernández, A. Kelley, D. J. Peterson, S. Rajan, K. A. Woelfer, J. Wolf, and T. J. Grabowski. Running Neuroimaging Applications on Amazon Web Services: How, When, and at What Cost? *Frontiers in Neuroinformatics*, 11:63, Nov. 2017. ISSN 1662-5196. doi: 10.3389/fninf.2017.00063.

- 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. Auría, O. Esteban, A. Lemkaddem, J.-P. Thiran, H. E. Cetinçul, 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. *Nature Communications*, 8(1):1349, Dec. 2017. ISSN 2041-1723. doi: 10.1038/s41467-017-01285-x.
- J. V. Manjón and P. Coupe. MRI denoising using Deep Learning and Non-local averaging. page 23, 2021.
- J. V. Manjón, J. Carbonell-Caballero, J. J. Lull, G. García-Martí, L. Martí-Bonmatí, and M. Robles. MRI denoising using non-local means. *Medical image analysis*, 12(4): 514–523, Aug. 2008. ISSN 1361-8423. doi: 10.1016/j.media.2008.02.004.
- 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: Spatially Adaptive Nonlocal Denoising. *Journal of Magnetic Resonance Imaging*, 31(1):192–203, Jan. 2010. ISSN 10531807. doi: 10.1002/jmri.22003.
- J. V. Manjón, P. Coupé, L. Concha, A. Buades, D. L. Collins, and M. Robles. Diffusion Weighted Image Denoising Using Overcomplete Local PCA. *PLoS ONE*, 8(9):e73021, Sept. 2013. ISSN 1932-6203. doi: 10.1371/journal.pone.0073021.
- P. Mansfield and A. A. Maudsley. Medical imaging by NMR. *The British journal of radiology*, 50(591):188–194, Mar. 1977. ISSN 0007-1285. doi: 10.1259/0007-1285-50-591-188.
- V. A. Marčenko and L. A. Pastur. Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik*, 1(4):457, Apr. 1967. ISSN 0025-5734. doi: 10.1070/SM1967v001n04ABEH001994.
- P. Marjoram, J. Molitor, V. Plagnol, and S. Tavaré. Markov chain Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences*, 100(26):15324–15328, Dec. 2003. doi: 10.1073/pnas.0306899100.
- N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, 21(6):1087–1092, June 1953. ISSN 0021-9606. doi: 10.1063/1.1699114.
- K. L. Miller, F. Alfaro-Almagro, N. K. Bangerter, D. L. Thomas, E. Yacoub, J. Xu, A. J. Bartsch, S. Jbabdi, S. N. Sotiropoulos, J. L. R. Andersson, L. Griffanti, G. Douaud, T. W. Okell, P. Weale, I. Dragonu, S. Garratt, S. Hudson, R. Collins, M. Jenkinson, P. M. Matthews, and S. M. Smith. Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nature neuroscience*, 19(11):1523–1536, Nov. 2016. ISSN 1097-6256, 1546-1726. doi: 10.1038/nn.4393.
- P. K. Mishro, S. Agrawal, R. Panda, and A. Abraham. A Survey on State-of-the-art Denoising Techniques for Brain Magnetic Resonance Images. *IEEE Reviews in Biomedical Engineering*, pages 1–1, 2021. ISSN 1937-3333, 1941-1189. doi: 10.1109/RBME.2021.3055556.

- A. Mittal, A. K. Moorthy, and A. C. Bovik. No-Reference Image Quality Assessment in the Spatial Domain. *IEEE Transactions on Image Processing*, 21(12):4695–4708, Dec. 2012. ISSN 1057-7149, 1941-0042. doi: 10.1109/TIP.2012.2214050.
- S. Moeller, E. Yacoub, C. A. Olman, E. Auerbach, J. Strupp, N. Harel, and K. Uğurbil. 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. *Magnetic Resonance in Medicine*, 63(5):1144–1153, May 2010. ISSN 1522-2594. doi: 10.1002/mrm.22361.
- S. Moeller, P. K. Pisharady, S. Ramanna, C. Lenglet, X. Wu, L. Dowdle, E. Yacoub, K. Uğurbil, and M. Akçakaya. NOise reduction with DIstribution Corrected (NORDIC) PCA in dMRI with complex-valued parameter-free locally low-rank processing. *NeuroImage*, 226:117539, Feb. 2021a. ISSN 10538119. doi: 10.1016/j.neuroimage.2020.117539.
- S. Moeller, P. Pisharady Kumar, J. Andersson, M. Akcakaya, N. Harel, R. Ma, X. Wu, E. Yacoub, C. Lenglet, and K. Ugurbil. Diffusion Imaging in the Post HCP Era. *Journal of magnetic resonance imaging : JMRI*, 54(1):36–57, July 2021b. ISSN 1053-1807, 1522-2586. doi: 10.1002/jmri.27247.
- A.-R. Mohammadi-Nejad, S. Pszczolkowski, D. Auer, and S. Sotiropoulos. Multi-modal neuroimaging pipelines for data preprocessing. Zenodo, Feb. 2019.
- S. V. Mohd Sagheer and S. N. George. A review on medical image denoising algorithms. *Biomedical Signal Processing and Control*, 61:102036, Aug. 2020. ISSN 17468094. doi: 10.1016/j.bspc.2020.102036.
- R. J. Morecraft, G. Ugolini, J. L. Lanciego, F. G. Wouterlood, and D. N. Pandya. Classic and Contemporary Neural Tract-Tracing Techniques. In *Diffusion MRI*, pages 359–399. Elsevier, 2014. ISBN 978-0-12-396460-1. doi: 10.1016/B978-0-12-396460-1.00017-2.
- M. Moreno López, J. M. Frederick, and J. Ventura. Evaluation of MRI Denoising Methods Using Unsupervised Learning. 4:642731, June 2021. ISSN 2624-8212. doi: 10.3389/frai.2021.642731.
- S. Mori, B. J. Crain, and P. C. M. V. Zijl. 3D brain fiber reconstruction from diffusion MRI. *NeuroImage*, 7(4 PART II), 1998. ISSN 1053-8119. doi: 10.1016/s1053-8119(18)31543-x.
- M. J. Muckley, B. Ades-Aron, A. Papaioannou, G. Lemberskiy, E. Solomon, Y. W. Lui, D. K. Sodickson, E. Fieremans, D. S. Novikov, and F. Knoll. Training a neural network for Gibbs and noise removal in diffusion MRI. *Magnetic Resonance in Medicine*, 85(1): 413–428, Jan. 2021. ISSN 0740-3194, 1522-2594. doi: 10.1002/mrm.28395.
- T. Müller, B. McWilliams, F. Rousselle, M. Gross, and J. Novák. Neural Importance Sampling, Sept. 2019.
- D. Muresan and T. Parks. Adaptive principal components and image denoising. In *Proceedings 2003 International Conference on Image Processing (Cat. No.03CH37429)*, volume 1, pages I–101–4, Barcelona, Spain, 2003. IEEE. ISBN 978-0-7803-7750-9. doi: 10.1109/ICIP.2003.1246908.
- J. D. Murray, M. Demirtaş, and A. Anticevic. Biophysical Modeling of Large-Scale Brain Dynamics and Applications for Computational Psychiatry. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 3(9):777–787, Sept. 2018. ISSN 24519022. doi: 10.1016/j.bpsc.2018.07.004.

- V. Nath, K. G. Schilling, P. Parvathaneni, C. B. Hansen, A. E. Hainline, Y. Huo, J. A. Blaber, I. Lyu, V. Janve, Y. Gao, I. Stepniewska, A. W. Anderson, and B. A. Landman. Deep learning reveals untapped information for local white-matter fiber reconstruction in diffusion-weighted MRI. *Magnetic Resonance Imaging*, 62:220–227, Oct. 2019. ISSN 0730-725X. doi: 10.1016/j.mri.2019.07.012.
- R. J. M. Navest, S. Mandija, A. Andreychenko, A. J. E. Raaijmakers, J. J. W. Lagendijk, and C. A. T. van den Berg. Understanding the physical relations governing the noise navigator. *Magnetic Resonance in Medicine*, 82(6):2236–2247, Dec. 2019. ISSN 0740-3194. doi: 10.1002/mrm.27906.
- G. L. Nedjati-Gilani, T. Schneider, M. G. Hall, N. Cawley, I. Hill, O. Ciccarelli, I. Drobnjak, C. A. G. Wheeler-Kingshott, and D. C. Alexander. Machine learning based compartment models with permeability for white matter microstructure imaging. *NeuroImage*, 150:119–135, Apr. 2017. ISSN 10538119. doi: 10.1016/j.neuroimage.2017.02.013.
- A. Niculescu-Mizil and R. Caruana. Predicting good probabilities with supervised learning. In *Proceedings of the 22nd International Conference on Machine Learning - ICML '05*, pages 625–632, Bonn, Germany, 2005. ACM Press. ISBN 978-1-59593-180-1. doi: 10.1145/1102351.1102430.
- L. Ning, F. Szczepankiewicz, M. Nilsson, Y. Rathi, and C.-F. Westin. Probing tissue microstructure by diffusion skewness tensor imaging. *Scientific reports*, 11(1):135, Dec. 2021. ISSN 2045-2322. doi: 10.1038/s41598-020-79748-3.
- D. S. Novikov. The present and the future of microstructure MRI: From a paradigm shift to normal science. *Journal of Neuroscience Methods*, 351:108947, Mar. 2021. ISSN 0165-0270. doi: 10.1016/j.jneumeth.2020.108947.
- D. S. Novikov, V. G. Kiselev, and S. N. Jespersen. On modeling. *Magnetic Resonance in Medicine*, 79(6):3172–3193, 2018. ISSN 1522-2594. doi: 10.1002/mrm.27101.
- D. S. Novikov, E. Fieremans, S. N. Jespersen, and V. G. Kiselev. Quantifying brain microstructure with diffusion MRI: Theory and parameter estimation. *NMR in Biomedicine*, 32(4), Apr. 2019. ISSN 0952-3480, 1099-1492. doi: 10.1002/nbm.3998.
- H. Nyquist. Thermal Agitation of Electric Charge in Conductors. *Physical Review*, 32(1): 110–113, July 1928. doi: 10.1103/PhysRev.32.110.
- R. L. O’Gorman and D. K. Jones. Just how much data need to be collected for reliable bootstrap DT-MRI? *Magnetic Resonance in Medicine*, 56(4):884–890, 2006. ISSN 1522-2594. doi: 10.1002/mrm.21014.
- Y. Ovadia, E. Fertig, J. Ren, Z. Nado, D. Sculley, S. Nowozin, J. V. Dillon, B. Lakshminarayanan, and J. Snoek. Can You Trust Your Model’s Uncertainty? Evaluating Predictive Uncertainty Under Dataset Shift, Dec. 2019.
- 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 T. D. Andrews, R. Balan, J. J. Benedetto, W. Czaja, and K. A. Okoudjou, editors, *Excursions in Harmonic Analysis, Volume 2: The February Fourier Talks at the Norbert Wiener Center*, Applied and Numerical Harmonic Analysis, pages 373–399. Birkhäuser, Boston, 2013a. ISBN 978-0-8176-8379-5. doi: 10.1007/978-0-8176-8379-5\19.
- E. Özarslan, C. G. Koay, T. M. Shepherd, M. E. Komlosh, M. O. İrfanoğlu, C. Pierpaoli, and P. J. Basser. Mean Apparent Propagator (MAP) MRI: A novel diffusion imaging

- method for mapping tissue microstructure. *NeuroImage*, 78:16–32, Sept. 2013b. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2013.04.016.
- S. Pajevic and P. J. Basser. Parametric and non-parametric statistical analysis of DT-MRI data. *Journal of Magnetic Resonance*, 161(1):1–14, Mar. 2003. ISSN 10907807. doi: 10.1016/S1090-7807(02)00178-7.
- A. Pal and Y. Rathi. A review and experimental evaluation of deep learning methods for MRI reconstruction, Mar. 2022.
- M. Palombo, N. Shemesh, I. Ronen, and J. Valette. Insights into brain microstructure from in vivo DW-MRS. *NeuroImage*, 182:97–116, Nov. 2018. ISSN 1095-9572. doi: 10.1016/j.neuroimage.2017.11.028.
- M. Palombo, A. Ianus, M. Guerreri, D. Nunes, D. C. Alexander, N. Shemesh, and H. Zhang. SANDI: A compartment-based model for non-invasive apparent soma and neurite imaging by diffusion MRI. *NeuroImage*, 215:116835, July 2020. ISSN 1095-9572. doi: 10.1016/j.neuroimage.2020.116835.
- G. Papamakarios and I. Murray. Fast ϵ -Free Inference of Simulation Models with Bayesian Conditional Density Estimation. *arXiv:1605.06376 [cs, stat]*, Apr. 2018.
- G. Papamakarios, T. Pavlakou, and I. Murray. Masked Autoregressive Flow for Density Estimation. *arXiv:1705.07057 [cs, stat]*, June 2018a.
- G. Papamakarios, D. C. Sterratt, and I. Murray. Sequential Neural Likelihood: Fast Likelihood-free Inference with Autoregressive Flows. page 12, 2018b.
- G. Papamakarios, E. Nalisnick, D. J. Rezende, S. Mohamed, and B. Lakshminarayanan. Normalizing Flows for Probabilistic Modeling and Inference. *arXiv:1912.02762 [cs, stat]*, Dec. 2019.
- G. J. Parker, J. A. Schnabel, M. R. Symms, D. J. Werring, and G. J. Barker. Non-linear smoothing for reduction of systematic and random errors in diffusion tensor imaging. *Journal of Magnetic Resonance Imaging*, 11(6):702–710, June 2000. ISSN 1053-1807, 1522-2586. doi: 10.1002/1522-2586(200006)11:6<702::AID-JMRI18>3.0.CO;2-A.
- G. J. Parker, H. A. Haroon, and C. A. Wheeler-Kingshott. A framework for a streamline-based probabilistic index of connectivity (PICO) using a structural interpretation of MRI diffusion measurements. *Journal of Magnetic Resonance Imaging*, 18(2):242–254, Aug. 2003. ISSN 1053-1807, 1522-2586. doi: 10.1002/jmri.10350.
- C. Pierpaoli and P. J. Basser. Toward a quantitative assessment of diffusion anisotropy. *Magnetic resonance in medicine*, 36(6):893–906, Dec. 1996. ISSN 0740-3194. doi: 10.1002/mrm.1910360612.
- P. K. Pisharady, S. N. Sotiropoulos, J. M. Duarte-Carvajalino, G. Sapiro, and C. Lenglet. Estimation of white matter fiber parameters from compressed multiresolution diffusion MRI using sparse Bayesian learning. *NeuroImage*, 167:488–503, Feb. 2018. ISSN 1095-9572. doi: 10.1016/j.neuroimage.2017.06.052.
- T. Plotz and S. Roth. Benchmarking Denoising Algorithms with Real Photographs. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2750–2759, Honolulu, HI, July 2017. IEEE. ISBN 978-1-5386-0457-1. doi: 10.1109/CVPR.2017.294.

- P. Poulin, D. Jörgens, P.-M. Jodoin, and M. Descoteaux. Tractography and machine learning: Current state and open challenges. *Magnetic Resonance Imaging*, 64:37–48, Dec. 2019. ISSN 0730-725X. doi: 10.1016/j.mri.2019.04.013.
- J. K. Pritchard, M. T. Seielstad, A. Perez-Lezaun, and M. W. Feldman. Population growth of human Y chromosomes: A study of Y chromosome microsatellites. *Molecular Biology and Evolution*, 16(12):1791–1798, Dec. 1999. ISSN 0737-4038. doi: 10.1093/oxfordjournals.molbev.a026091.
- K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger. SENSE: Sensitivity encoding for fast MRI. *Magnetic Resonance in Medicine*, 42(5):952–962, Nov. 1999. ISSN 0740-3194.
- A. Ramesh, P. Dhariwal, A. Nichol, C. Chu, and M. Chen. Hierarchical Text-Conditional Image Generation with CLIP Latents, Apr. 2022.
- S. Ramon y Cajal. Histologie du système nerveux de l’Homme et des vertébrés. *Maloine (Paris)*, 2:891–942, 1911.
- D. Ravi, N. Ghavami, D. C. Alexander, and A. Ianus. Current Applications and Future Promises of Machine Learning in Diffusion MRI. In E. Bonet-Carne, F. Grussu, L. Ning, F. Sepehrband, and C. M. W. Tax, editors, *Computational Diffusion MRI*, pages 105–121, Cham, 2019. Springer International Publishing. ISBN 978-3-030-05831-9.
- T. W. Redpath. Signal-to-noise ratio in MRI. *The British Journal of Radiology*, 71(847):704–707, July 1998. ISSN 0007-1285. doi: 10.1259/bjr.71.847.9771379.
- D. J. Rezende and S. Mohamed. Variational Inference with Normalizing Flows. page 9, 2015.
- D. J. Rezende, S. Mohamed, I. Danihelka, K. Gregor, and D. Wierstra. One-Shot Generalization in Deep Generative Models. page 9.
- S. O. Rice. Mathematical Analysis of Random Noise. *Bell System Technical Journal*, 24(1):46–156, 1945. ISSN 1538-7305. doi: 10.1002/j.1538-7305.1945.tb00453.x.
- O. Rippel and R. P. Adams. High-Dimensional Probability Estimation with Deep Density Models, Feb. 2013.
- A. Roebroek, K. L. Miller, and M. Aggarwal. Ex vivo diffusion MRI of the human brain: Technical challenges and recent advances. *NMR in Biomedicine*, 32(4), Apr. 2019a. ISSN 0952-3480, 1099-1492. doi: 10.1002/nbm.3941.
- A. Roebroek, K. L. Miller, and M. Aggarwal. Ex vivo diffusion MRI of the human brain: Technical challenges and recent advances. *NMR in Biomedicine*, 32(4):e3941, Apr. 2019b. ISSN 0952-3480, 1099-1492. doi: 10.1002/nbm.3941.
- P. B. Roemer, W. A. Edelstein, C. E. Hayes, S. P. Souza, and O. M. Mueller. The NMR phased array. *Magnetic Resonance in Medicine*, 16(2):192–225, Nov. 1990. ISSN 0740-3194. doi: 10.1002/mrm.1910160203.
- V. Roy. Convergence Diagnostics for Markov Chain Monte Carlo. *Annual Review of Statistics and Its Application*, 7(1):387–412, 2020. doi: 10.1146/annurev-statistics-031219-041300.
- C. Saharia, W. Chan, S. Saxena, L. Li, J. Whang, E. Denton, S. K. S. Ghasemipour, B. K. Ayan, S. S. Mahdavi, R. G. Lopes, T. Salimans, J. Ho, D. J. Fleet, and M. Norouzi. Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding, May 2022.

- R. Salvador, A. Peña, D. K. Menon, T. A. Carpenter, J. D. Pickard, and E. T. Bullmore. Formal characterization and extension of the linearized diffusion tensor model: Linearized Diffusion Tensor Model. *Human Brain Mapping*, 24(2):144–155, Feb. 2005. ISSN 10659471. doi: 10.1002/hbm.20076.
- M. J. Schervish. *Theory of Statistics*. Springer Series in Statistics. Springer, New York, NY, 1995. ISBN 978-1-4612-8708-7 978-1-4612-4250-5. doi: 10.1007/978-1-4612-4250-5.
- K. Schilling, Y. Gao, V. Janve, I. Stepniewska, B. A. Landman, and A. W. Anderson. Can increased spatial resolution solve the crossing fiber problem for diffusion MRI? *NMR in biomedicine*, 30(12):10.1002/nbm.3787, Dec. 2017. ISSN 0952-3480. doi: 10.1002/nbm.3787.
- K. G. Schilling, V. Janve, Y. Gao, I. Stepniewska, B. A. Landman, and A. W. Anderson. Histological validation of diffusion MRI fiber orientation distributions and dispersion. *NeuroImage*, 165:200–221, Jan. 2018. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2017.10.046.
- K. G. Schilling, A. Daducci, K. Maier-Hein, C. Poupon, J.-C. Houde, V. Nath, A. W. Anderson, B. A. Landman, and M. Descoteaux. Challenges in diffusion MRI tractography – Lessons learned from international benchmark competitions. *Magnetic Resonance Imaging*, 57:194–209, Apr. 2019. ISSN 0730725X. doi: 10.1016/j.mri.2018.11.014.
- K. G. Schilling, S. Fadnavis, J. Batson, M. Visagie, A. J. Combes, C. D. McKnight, F. Bagnato, E. Garyfallidis, B. A. Landman, S. A. Smith, and K. P. O’Grady. Patch2Self denoising of diffusion MRI in the cervical spinal cord improves intra-cord contrast, signal modelling, repeatability, and feature conspicuity. Preprint, Radiology and Imaging, Oct. 2021.
- M. Schmitt, P.-C. Bürkner, U. Köthe, and S. T. Radev. Detecting Model Misspecification in Amortized Bayesian Inference with Neural Networks, May 2022.
- D. M. Schnyer, P. C. Clasen, C. Gonzalez, and C. G. Bevers. Evaluating the diagnostic utility of applying a machine learning algorithm to diffusion tensor MRI measures in individuals with major depressive disorder. *Psychiatry Research: Neuroimaging*, 264: 1–9, June 2017. ISSN 0925-4927. doi: 10.1016/j.psychresns.2017.03.003.
- K. K. Seunarine and D. C. Alexander. Multiple Fibers: Beyond the Diffusion Tensor. page 18, 2009.
- G. Shmueli. To Explain or to Predict? *Statistical Science*, 25(3), Aug. 2010. ISSN 0883-4237. doi: 10.1214/10-STS330.
- S. A. Sisson, Y. Fan, and M. M. Tanaka. Sequential Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences*, 104(6):1760–1765, Feb. 2007. doi: 10.1073/pnas.0607208104.
- S. A. Sisson, Y. Fan, and M. A. Beaumont. Overview of Approximate Bayesian Computation. *arXiv:1802.09720 [stat]*, Feb. 2018.
- J. Sjölund, A. Eklund, E. Özarslan, M. Herberthson, M. Bånkestad, and H. Knutsson. Bayesian uncertainty quantification in linear models for diffusion MRI. *NeuroImage*, 175:272–285, July 2018. ISSN 10538119. doi: 10.1016/j.neuroimage.2018.03.059.
- S. M. Smith, D. Vidaurre, C. F. Beckmann, M. F. Glasser, M. Jenkinson, K. L. Miller, T. E. Nichols, E. C. Robinson, G. Salimi-Khorshidi, M. W. Woolrich, D. M. Barch, K. Uğurbil, and D. C. Van Essen. Functional connectomics from resting-state fMRI.

- Trends in Cognitive Sciences*, 17(12):666–682, Dec. 2013. ISSN 13646613. doi: 10.1016/j.tics.2013.09.016.
- S. M. Smith, T. E. Nichols, D. Vidaurre, A. M. Winkler, T. E. J. Behrens, M. F. Glasser, K. Ugurbil, D. M. Barch, D. C. Van Essen, and K. L. Miller. A positive-negative mode of population covariation links brain connectivity, demographics and behavior. *Nature neuroscience*, 18(11):1565–1567, Nov. 2015. ISSN 1097-6256, 1546-1726. doi: 10.1038/nn.4125.
- S. N. Sotiropoulos and A. Zalesky. Building connectomes using diffusion MRI: Why, how and but. *NMR in Biomedicine*, 32(4):e3752, Apr. 2019. ISSN 0952-3480, 1099-1492. doi: 10.1002/nbm.3752.
- S. N. Sotiropoulos, D. E. Jones, L. Bai, and T. Kypraios. Exact and analytic bayesian inference for orientation distribution functions. In *2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 1189–1192, Apr. 2010. doi: 10.1109/ISBI.2010.5490207.
- S. N. Sotiropoulos, T. E. Behrens, and S. Jbabdi. Ball and rackets: Inferring fiber fanning from diffusion-weighted MRI. *NeuroImage*, 60(2):1412–1425, Apr. 2012. ISSN 10538119. doi: 10.1016/j.neuroimage.2012.01.056.
- S. N. Sotiropoulos, S. Jbabdi, J. L. Andersson, M. W. Woolrich, K. Ugurbil, and T. E. J. Behrens. RubiX: Combining Spatial Resolutions for Bayesian Inference of Crossing Fibers in Diffusion MRI. *IEEE Transactions on Medical Imaging*, 32(6):969–982, June 2013a. ISSN 0278-0062, 1558-254X. doi: 10.1109/TMI.2012.2231873.
- S. N. Sotiropoulos, S. Jbabdi, J. Xu, J. L. Andersson, S. Moeller, E. J. Auerbach, M. F. Glasser, M. Hernandez, G. Sapiro, M. Jenkinson, D. A. Feinberg, E. Yacoub, C. Lenglet, D. C. Van Essen, K. Ugurbil, and T. E. Behrens. Advances in diffusion MRI acquisition and processing in the Human Connectome Project. *NeuroImage*, 80:125–143, Oct. 2013b. ISSN 10538119. doi: 10.1016/j.neuroimage.2013.05.057.
- S. N. Sotiropoulos, S. Moeller, S. Jbabdi, J. Xu, J. L. Andersson, E. J. Auerbach, E. Yacoub, D. Feinberg, K. Setsompop, L. L. Wald, T. E. J. Behrens, K. Ugurbil, and C. Lenglet. Effects of image reconstruction on fiber orientation mapping from multichannel diffusion MRI: Reducing the noise floor using SENSE: Effects of dMRI Image Reconstruction on Fiber Orientations. *Magnetic Resonance in Medicine*, 70(6):1682–1689, Dec. 2013c. ISSN 07403194. doi: 10.1002/mrm.24623.
- S. N. Sotiropoulos, M. Hernández-Fernández, A. T. Vu, J. L. Andersson, S. Moeller, E. Yacoub, C. Lenglet, K. Ugurbil, T. E. Behrens, and S. Jbabdi. Fusion in diffusion MRI for improved fibre orientation estimation: An application to the 3T and 7T data of the Human Connectome Project. *NeuroImage*, 134:396–409, July 2016. ISSN 10538119. doi: 10.1016/j.neuroimage.2016.04.014.
- O. Sporns, G. Tononi, and R. Kötter. The human connectome: A structural description of the human brain. *PLoS computational biology*, 1(4):e42, Sept. 2005. ISSN 1553-7358. doi: 10.1371/journal.pcbi.0010042.
- E. O. Stejskal and J. E. Tanner. Spin Diffusion Measurements: Spin Echoes in the Presence of a Time-Dependent Field Gradient. *The Journal of Chemical Physics*, 42(1):288, 1965. ISSN 0021-9606. doi: 10.1063/1.1695690.
- I. Stepień, R. Obuchowicz, A. Piórkowski, and M. Oszust. Fusion of Deep Convolutional Neural Networks for No-Reference Magnetic Resonance Image Quality Assessment. *Sensors*, 21(4):1043, Feb. 2021. ISSN 1424-8220. doi: 10.3390/s21041043.

- A. C. Stickland and I. Murray. Diverse Ensembles Improve Calibration. *arXiv:2007.04206 [cs, stat]*, July 2020.
- E. G. Tabak and C. V. Turner. A Family of Nonparametric Density Estimation Algorithms. *Communications on Pure and Applied Mathematics*, 66(2):145–164, 2013. ISSN 1097-0312. doi: 10.1002/cpa.21423.
- E. G. Tabak and E. Vanden-Eijnden. Density estimation by dual ascent of the log-likelihood. *Communications in Mathematical Sciences*, 8(1):217–233, Mar. 2010. ISSN 1539-6746, 1945-0796.
- D. Tamada. Review: Noise and artifact reduction for MRI using deep learning. *arXiv:2002.12889 [physics]*, Feb. 2020.
- R. Tanno, D. E. Worrall, E. Kaden, A. Ghosh, F. Grussu, A. Bizzi, S. N. Sotiropoulos, A. Criminisi, and D. C. Alexander. Uncertainty modelling in deep learning for safer neuroimage enhancement: Demonstration in diffusion MRI. *NeuroImage*, 225:117366, Jan. 2021. ISSN 10538119. doi: 10.1016/j.neuroimage.2020.117366.
- S. Tavaré, D. J. Balding, R. C. Griffiths, and P. Donnelly. Inferring coalescence times from DNA sequence data. *Genetics*, 145(2):505–518, Feb. 1997. ISSN 0016-6731. doi: 10.1093/genetics/145.2.505.
- C. M. Tax, M. Bastiani, J. Veraart, E. Garyfallidis, and M. Okan Irfanoglu. What’s New and What’s Next in Diffusion MRI Preprocessing. *NeuroImage*, page 118830, Dec. 2021. ISSN 10538119. doi: 10.1016/j.neuroimage.2021.118830.
- A. Tejero-Cantero, J. Boelts, M. Deistler, J.-M. Lueckmann, C. Durkan, P. Gonçalves, D. Greenberg, and J. Macke. Sbi: A toolkit for simulation-based inference. *JOSS*, 5(52):2505, Aug. 2020. ISSN 2475-9066. doi: 10.21105/joss.02505.
- C. Tian, L. Fei, W. Zheng, Y. Xu, W. Zuo, and C.-W. Lin. Deep learning on image denoising: An overview. *Neural Networks*, 131:251–275, Nov. 2020. ISSN 08936080. doi: 10.1016/j.neunet.2020.07.025.
- L. Tierney and J. B. Kadane. Accurate Approximations for Posterior Moments and Marginal Densities. *Journal of the American Statistical Association*, 81(393):82–86, Mar. 1986. ISSN 0162-1459. doi: 10.1080/01621459.1986.10478240.
- J.-D. Tournier, F. Calamante, D. G. Gadian, and A. Connelly. Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. *NeuroImage*, 23(3):1176–1185, Nov. 2004. ISSN 10538119. doi: 10.1016/j.neuroimage.2004.07.037.
- 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. *NeuroImage*, 35(4):1459–1472, May 2007. ISSN 10538119. doi: 10.1016/j.neuroimage.2007.02.016.
- D. S. Tuch. Q-ball imaging. *Magnetic Resonance in Medicine*, 52(6):1358–1372, Dec. 2004. ISSN 0740-3194, 1522-2594. doi: 10.1002/mrm.20279.
- B. Uria, M.-A. Côté, K. Gregor, I. Murray, and H. Larochelle. Neural Autoregressive Distribution Estimation, May 2016.

- J. Veraart, J. Sijbers, S. Sunaert, A. Leemans, and B. Jeurissen. Weighted linear least squares estimation of diffusion MRI parameters: Strengths, limitations, and pitfalls. *Neuroimage*, 81:335–346, Nov. 2013. ISSN 1095-9572. doi: 10.1016/j.neuroimage.2013.05.028.
- J. Veraart, E. Fieremans, and D. S. Novikov. Diffusion MRI noise mapping using random matrix theory: Diffusion MRI Noise Mapping. *Magnetic Resonance in Medicine*, 76(5): 1582–1593, Nov. 2016a. ISSN 07403194. doi: 10.1002/mrm.26059.
- J. Veraart, D. S. Novikov, D. Christiaens, B. Ades-aron, J. Sijbers, and E. Fieremans. Denoising of diffusion MRI using random matrix theory. *NeuroImage*, 142:394–406, Nov. 2016b. ISSN 10538119. doi: 10.1016/j.neuroimage.2016.08.016.
- J. Veraart, E. Fieremans, and D. S. Novikov. On the scaling behavior of water diffusion in human brain white matter. *NeuroImage*, 185:379–387, Jan. 2019. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2018.09.075.
- Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, Apr. 2004. ISSN 1057-7149. doi: 10.1109/TIP.2003.819861.
- S. Warrington, K. L. Bryant, A. A. Khrapitchev, J. Sallet, M. Charquero-Ballester, G. Douaud, S. Jbabdi, R. B. Mars, and S. N. Sotiropoulos. XTRACT - Standardised protocols for automated tractography in the human and macaque brain. *NeuroImage*, 217:116923, Aug. 2020. ISSN 10538119. doi: 10.1016/j.neuroimage.2020.116923.
- J. Wasserthal, P. Neher, and K. H. Maier-Hein. TractSeg - Fast and accurate white matter tract segmentation. *NeuroImage*, 183:239–253, Dec. 2018. ISSN 10538119. doi: 10.1016/j.neuroimage.2018.07.070.
- M.-A. Weber, A. M. Nagel, H. E. Kan, and M. P. Wattjes. Quantitative Imaging in Muscle Diseases with Focus on Non-proton MRI and Other Advanced MRI Techniques. *Seminars in Musculoskeletal Radiology*, 24(4):402–412, Aug. 2020. ISSN 1098-898X. doi: 10.1055/s-0040-1712955.
- V. J. Wedeen, P. Hagmann, W.-Y. I. Tseng, T. G. Reese, and R. M. Weisskoff. Mapping complex tissue architecture with diffusion spectrum magnetic resonance imaging. *Magnetic Resonance in Medicine*, 54(6):1377–1386, Dec. 2005. ISSN 0740-3194, 1522-2594. doi: 10.1002/mrm.20642.
- B. Whitcher, D. S. Tuch, J. J. Wisco, A. G. Sorensen, and L. Wang. Using the wild bootstrap to quantify uncertainty in diffusion tensor imaging. *Human brain mapping*, 29(3):346–362, Mar. 2008. ISSN 1065-9471. doi: 10.1002/hbm.20395.
- N. Wiest-Daesslé, S. Prima, P. Coupé, S. P. Morrissey, and C. Barillot. Non-Local Means Variants for Denoising of Diffusion-Weighted and Diffusion Tensor MRI. In D. Hutchinson, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, N. Ayache, S. Ourselin, and A. Maeder, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2007*, volume 4792, pages 344–351. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007. ISBN 978-3-540-75758-0 978-3-540-75759-7. doi: 10.1007/978-3-540-75759-7_42.
- N. Wiest-Daesslé, S. Prima, P. Coupé, S. P. Morrissey, and C. Barillot. Rician noise removal by non-Local Means filtering for low signal-to-noise ratio MRI: Applications to DT-MRI. *Med Image Comput Comput Assist Interv*, 11(Pt 2):171–179, 2008. doi: 10.1007/978-3-540-85990-1_21.

- D. P. Wipf and S. S. Nagarajan. A New View of Automatic Relevance Determination. page 8, 2007.
- D. Wolpert and W. Macready. No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, 1(1):67–82, Apr. 1997. ISSN 1941-0026. doi: 10.1109/4235.585893.
- M. W. Woolrich and K. E. Stephan. Biophysical network models and the human connectome. *NeuroImage*, 80:330–338, Oct. 2013. ISSN 10538119. doi: 10.1016/j.neuroimage.2013.03.059.
- J. C. Wu, A. F. Martin, G. A. Sanders, and R. N. Kacker. Bootstrap method versus analytical approach for estimating uncertainty of measure in ROC analysis on large datasets. July 2018.
- Z. Wu, T. Potter, D. Wu, and Y. Zhang. Denoising high angular resolution diffusion imaging data by combining singular value decomposition and non-local means filter. *Journal of Neuroscience Methods*, 312:105–113, Jan. 2019. ISSN 01650270. doi: 10.1016/j.jneumeth.2018.11.020.
- C. Ye. Tissue microstructure estimation using a deep network inspired by a dictionary-based framework. *Medical image analysis*, 42:288–299, Dec. 2017. ISSN 1361-8423. doi: 10.1016/j.media.2017.09.001.
- C. Ye, X. Li, and J. Chen. A deep network for tissue microstructure estimation using modified LSTM units. *Medical image analysis*, 55:49–64, July 2019. ISSN 1361-8423. doi: 10.1016/j.media.2019.04.006.
- C. Ye, Y. Li, and X. Zeng. An improved deep network for tissue microstructure estimation with uncertainty quantification. *Medical Image Analysis*, 61:101650, Apr. 2020. ISSN 13618415. doi: 10.1016/j.media.2020.101650.
- C.-H. Yeh, R. E. Smith, T. Dhollander, F. Calamante, and A. Connelly. Connectomes from streamlines tractography: Assigning streamlines to brain parcellations is not trivial but highly consequential. *NeuroImage*, 199:160–171, Oct. 2019. ISSN 10538119. doi: 10.1016/j.neuroimage.2019.05.005.
- A. Zare, A. Ozdemir, M. A. Iwen, and S. Aviyente. Extension of PCA to Higher Order Data Structures: An Introduction to Tensors, Tensor Decompositions, and Tensor PCA. *Proceedings of the IEEE*, 106(8):1341–1358, Aug. 2018. ISSN 1558-2256. doi: 10.1109/JPROC.2018.2848209.
- G. Zeng, Y. Guo, J. Zhan, Z. Wang, Z. Lai, X. Du, X. Qu, and D. Guo. A review on deep learning MRI reconstruction without fully sampled k-space. *BMC Medical Imaging*, 21(1):195, Dec. 2021. ISSN 1471-2342. doi: 10.1186/s12880-021-00727-9.
- 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. *NeuroImage*, 61(4):1000–1016, July 2012. ISSN 10538119. doi: 10.1016/j.neuroimage.2012.03.072.
- X. Zhang, J. Peng, M. Xu, W. Yang, Z. Zhang, H. Guo, W. Chen, Q. Feng, E. X. Wu, and Y. Feng. Denoise diffusion-weighted images using higher-order singular value decomposition. *NeuroImage*, 156:128–145, Aug. 2017. ISSN 10538119. doi: 10.1016/j.neuroimage.2017.04.017.

Appendix A

List of Abbreviations

ABC	Approximate Bayesian Computation
AC	Anterior Commissure
ACF	Autocorrelation Function
ADC	Apparent Diffusion Coefficient
ANN	Artificial Neural Network
APT	Automatic Posterior Transform
AR	Auto-regressive
ARD	Automatic Relevance Determination
BedpostX	Bayesian Estimation of Diffusion Parameters Obtained using Sampling Techniques
BNN	Bayesian Neural Network
BRISQUE	Blind/Reference-less Image Spatial Quality Evaluator
CC	Corpus Callosum
CNN	Convolutional Neural Network
CNR	Contrast-to-Noise Ratio
CPU	Central Processing Unit
CR	Corona Radiate
CRLB	Cramer-Raw Lower Boundary
CS	Compressed Sensing
CSD	Constrained Spherical Deconvolution
CSF	Cerebro-Spinal Fluid
DCNN	Deep Convolutional Neural Network
DCT	Discrete Cosine Transform
DIPY	Diffusion Imaging in Python
DL	Deep Learning
dMRI	diffusion MRI
dODF	Distribution ODF
DOF	Degrees of Freedom
DTI	Diffusion Tensor Imaging
EAP	Ensemble Average Propagator
EDDEN	Evaluating DMRI DENoising
ELBO	Evidence Lower Bound
EPI	Echo-Planar Imaging
ESS	Effective-Sample Size
FA	Fractional Anisotropy
fMRI	functional MRI
fODF	Fibre ODF
FRT	Funk-Radon Transform
FSL	FMRIB Software Library
FWHM	Full-Width Half-Maximum
FX	Fornix

GAN	Generative Adversarial Network
GFlowNets	Generative Flow Networks
GM	Gray Matter
GMM	Gaussian Mixture Model
GP	Gaussian Process
GPU	Graphical Processing Unit
GRAPPA	Generalized Autocalibrating Partial Parallel Acquisition
HCP	Human Connectome Project
IID	Independent and Identically Distributed
IQA	Image Quality Assessment
LA	Laplace Approximation
LFI	Likelihood-Free Inference
LM	Levenberg-Marquadt
LRA	Low-Rank Approximation
LSTM	Long-Short Term Memory
MAF	Masked Auto-regressive Flow
MADE	Masked Autoencoder Density Estimation
MAP	Maximum A Posteriori
MAP-MRI	Mean Apparent Propagator MRI
MD	Meand Diffusivity
MDN	Mixture Density Network
MCMC	Markov-Chain Monte Carlo
MH	Metropolis-Hastings
MIP	Maximum Intensity Projection
ML	Machine Learning
MLE	Maximum Likelihood Estimation
MLP	Multilayer Perceptron
MNI	Montreal Neurological Institute
MPPCA	Marchenko-Pastur PCA
MRI	Magnetic Resonance Imaging
MSE	Mean Squared Error
NDE	Neural Density Estimator
NDI	Neurite Density Model
NF	Normalizing Flow
NLE	Neural Likelihood Estimation
NLM	Non-Local Means
NODDI	Neurite Orientation Dispersion and Density Imaging
NORDIC	NOise Reduction with DIstribution Corrected
NPE	Neural Posterior Estimation
NRE	Neural Ration Estimation
NSF	Neural Spline Flow
NUQ	Noise Uncertainty Quantification
ODF	Orientation Distribution Function
OLS	Ordinary Least Squares
P2S	Patch2Self
PC	Principal Component
PCA	Principal Component Analysis
PGSE	Pulsed Gradient Spin Echo
PI	Parallel Imaging
PSNR	Peak SNR
PVM	Partial Volume Fraction
QBI	Q-Ball Imaging
QC	Quality Control
q-DL	q-space Deep Learning

RF	Random Forest
RMT	Random Matrix Theory
RMS	Root Mean Square
RNN	Recurrent Neural Network
RSoS	Root-Sum of Squares
RWM	Random-Walk Metropolis
SBI	Simulation-Based Inference
SENSE	Sensitivity Encoding
SH	Spherical Harmonics
SLF	Superior Longitudinal Fasciculus
SMC	Sequential Monte Carlo
SMT	Spherical Mean Technique
SNPE	Sequential Neural Posterior Estimation
SNLE	Sequential Neural Likelihood Estimation
SNR	Signal-to-Noise Ratio
SNRE	Sequential Neural Ratio Estimation
SSIM	Structural Similarity Index Measure
SVD	Singular Value Decomposition
TE	Echo Time
TR	Repetition Time
uODF	Uncertainty ODF
VAE	Variational Autoencoder
VB	Variational Bayes
VI	Variational Inference
VST	Variance Stabilization Transform
WM	White Matter

Appendix B

Initialization and other MCMC considerations

B.1 Spherical to Cartesian coordinate conversion

A number of parameters in the considered model in this thesis (e.g. fibre orientations v_i or gradient directions vectors g_i) are defined on the unit sphere. This is done for computational convenience as using spherical coordinates representation in the models we have one parameter less to estimate; for visualisation we convert these to Cartesian coordinates. To transform from Cartesian coordinates to spherical coordinates, we use the following identities:

$$\begin{aligned} r &= \sqrt{x^2 + y^2 + z^2} \\ \theta &= \arccos\left(\frac{z}{r}\right) \\ \phi &= \operatorname{atan2}(y, x) \end{aligned}$$

and from spherical coordinates to Cartesian coordinates:

$$\begin{aligned} x &= r \cdot \cos(\phi) \cdot \sin(\theta) \\ y &= r \cdot \sin(\phi) \cdot \sin(\theta) \\ z &= r \cdot \cos(\theta) \end{aligned}$$

For unit vectors, $r = 1$, so v becomes:

$$\begin{aligned} v &= [x, y, z]_{\text{Cartesians}} \\ v &= [\sin(\phi) \cdot \cos(\theta), \sin(\theta) \cdot \sin(\phi), \cos(\theta)]_{\text{spherical}} \end{aligned}$$

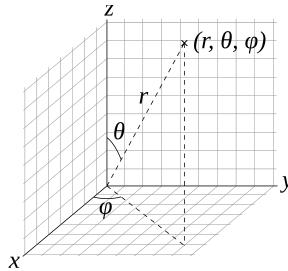


Figure B.1: Angle convention followed

B.2 MCMC Initialization

As explained in the main text, non-linear fitting of the Ball&Sticks model is performed to initialize the model parameters in the MCMC, ω_0 , in this case by the Levenberg-Marquardt (LM) algorithm. However, the LM itself requires an initial vector x_0 of values from where start to do the optimization process; this x_0 can be obtained by fitting a DTI model (see eq.2.22) to the data in each voxel.

B.2.1 Tensor model fitting

The first step is to estimate the diffusion tensor D , which is a generalization of the diffusion propagator (see eq.2.12) without assuming homogeneity in the medium of propagation, so different diffusion values are allowed along different directions.

$$\mathbf{D} = \begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{yx} & D_{yy} & D_{yz} \\ D_{zx} & D_{zy} & D_{zz} \end{bmatrix} \quad (\text{B.1})$$

As D is assumed to be symmetric and constant per voxel, there are 6 unknown parameters that need to be estimated from D . Let b_i indicate the b-value of gradient i (it is typically assumed constant and with values around 1000 s/mm^2 [Jones et al., 1999]), g_i a gradient directions vector with (x,y,z) coordinates (ideally evenly distributed in the unit sphere to obtain rotationally invariant properties of the parameters [Jones, 2004]), and S_i the signal measurements, with $i = 1, \dots, N$. Using the definition given by the DTI model

$$S_i(g_i, b_i) = S_0 \cdot \exp(-b_i \cdot g_i^T \cdot D \cdot g_i), \quad (\text{B.2})$$

and taking logs on both sides of eq.B.2, we obtain

$$\begin{aligned} -\ln\left(\frac{S_i}{S_0}\right) &= b_i \cdot g_i^T \cdot D \cdot g_i \\ &= b_i g_i(x)^2 D_{xx} + b_i g_i(y)^2 D_{yy} + b_i g_i(z)^2 D_{zz} \\ &\quad + 2b_i g_i(x) g_i(y) D_{xy} + 2b_i g_i(x) g_i(z) D_{xz} + 2b_i g_i(y) g_i(z) D_{yz} \end{aligned}$$

or in matrix form:

$$S = C \cdot \hat{D}_{est} \quad (\text{B.3})$$

where S is the vector of the $-\ln(\frac{S_i}{S_0})$ measurements, C is the $N \times N$ matrix of known coefficients, and \hat{D}_{est} is the estimated tensor D in vector form, i.e.

$$\begin{bmatrix} -\ln(\frac{S_1}{S_0}) \\ -\ln(\frac{S_2}{S_0}) \\ \vdots \\ -\ln(\frac{S_N}{S_0}) \end{bmatrix} = \begin{bmatrix} b_1 g_1(x)^2 & b_1 g_1(y)^2 & b_1 g_1(z)^2 & 2b_1 g_1(x) g_1(y) & 2b_1 g_1(x) g_1(z) & 2b_1 g_1(y) g_1(z) \\ b_2 g_2(x)^2 & b_2 g_2(y)^2 & b_2 g_2(z)^2 & 2b_2 g_2(x) g_2(y) & 2b_2 g_2(x) g_2(z) & 2b_2 g_2(y) g_2(z) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ b_N g_N(x)^2 & b_N g_N(y)^2 & b_N g_N(z)^2 & 2b_N g_N(x) g_N(y) & 2b_N g_N(x) g_N(z) & 2b_N g_N(y) g_N(z) \end{bmatrix} \cdot \begin{bmatrix} D_{xx} \\ D_{yy} \\ D_{zz} \\ D_{xy} \\ D_{xz} \\ D_{yz} \end{bmatrix} \quad (\text{B.4})$$

This is a linear system where, at least, 6 measurements of S_i are needed to estimate D (usually many more than 6 non co-linear directions are used [Jones, 2004] to increase immunity against noise). By assuming zero-mean Gaussian noise and constant variance, ordinary least squares (OLS) are usually applied to solve linear systems [Goldberger, 1964]. These assumptions are reasonable for common scenarios in MRI, where $\text{SNR} > 3$ [Pajevic and Bassar, 2003]. The estimated tensor will be:

$$\hat{D}_{est} = (C^T \cdot C)^{-1} \cdot C^T \cdot S \quad (\text{B.5})$$

B.2.2 Non-linear fitting

Once the tensor is estimated, its eigenvectors Λ and eigenvalues λ are obtained (and sorted from high to low λ) and subsequently the fractional anisotropy (FA) and mean diffusivity (MD), such as:

$$MD = \frac{D_{xx} + D_{yy} + D_{zz}}{3} = \frac{\lambda_1 + \lambda_2 + \lambda_3}{3} \quad (\text{B.6})$$

$$FA = \sqrt{\frac{3 \cdot \sum_{i=1}^3 (\lambda_i - \lambda_{avg})^2}{2 \cdot \sum_{i=1}^3 \lambda_i^2}} \quad (\text{B.7})$$

These are then used to initialize x_0 such as:

$$\begin{aligned} d &= 2MD \\ f_1 &= FA \\ f_2 &= FA/2 \\ f_3 &= FA/2 \\ (\theta_i, \phi_i) &= (\Lambda_i)_{sph} \\ \rightarrow x_0 &= [S_0, d, \theta_1, \phi_1, f_1, \theta_2, \phi_2, f_2, \theta_3, \phi_3, f_3] \end{aligned}$$

These equivalencies have been found experimentally to provide good results and are used in softwares like BedpostX (FSL). Having set x_0 , non-linear fitting is used to obtain more accurate estimation of the initial values of parameters for the MCMC, ω_0 , where the cost function optimised is:

$$\omega_0 = \arg \min_{\omega_i} \frac{N}{2} \sum_{i=1}^N \log \left(\frac{(Y_i - S_i)^2}{2} \right) \quad (\text{B.8})$$

This is a minimisation problem between the observed data S_i and the forward predictions Y_i , i.e. obtaining the most likely parameter values ω given the observed data (maximum likelihood estimation). See appendix B.4 for more details. To solve this minimization problem, the Powell conjugate-directions or the Levenberg-Marquadt algorithms have shown reliable results in minimizing the cost function B.8 and are extensively implemented in statistical packages [Harms et al., 2017]. The LM is used in this thesis.

Algorithm 6 Model parameter initialization**Require:** S_i , Cost function, b_i , g_i

-
- ```

1: procedure MODEL PARAMETER INITIALIZATION FOR MCMC
2: $\hat{D}_{est} = DTIFit(S_i, bvals, bvecs)$ ▷
3: Eigendecomposition of tensor D_{est}
4:
5: $MD = \frac{sum(\lambda_i)}{len(\lambda)}$
6: $FA = \frac{\sqrt{3 \cdot \sum(\lambda_i - MD)^2}}{2 \cdot \sum(\lambda_i^2)}$
7:
8: $d = 2MD$ ▷ Experimental equivalencies
9: $f_1 = FA$
10: $f_2 = FA/2$
11: $f_3 = FA/2$
12: $(\theta_i, \phi_i) = (\Lambda_i)_{sph}$
13: $x_0 = [S_0, d, \theta_1, \phi_1, f_1, \theta_2, \phi_2, f_2, \theta_3, \phi_3, f_3]$
14:
15: $\omega_0 = \arg \min_{\omega_i} \frac{N}{2} \sum_{i=1}^N \log \left(\frac{(Y_i - S_i)^2}{2} \right)$ ▷ Run LM
16:
17: return ω_0 ▷ Parameters initialized for the MCMC

```
- 

**B.3 Defining priors for the MCMC**

The priors used in this work are defined as following:

- $\pi(d)$  is given by a Gamma prior  $\sim \Gamma(\alpha, \beta)$ , where the covariance is  $\alpha/\beta^2$ . The hyperparameters  $\alpha, \beta$  are then chosen to be  $>0$  to give very high variance (i.e. an uninformative prior) while ensuring positive values of  $d$ . Therefore, we can set as possible values  $\alpha_d = 1$  and  $\beta_d$  very small (e.g.  $< 10^{-7}$ ).
- The priors used for the orientations  $v_i$  ensure uniform distribution on the sphere. This cannot be achieved if spherical angles follow a uniform distribution, a correction term is needed:

$$\begin{aligned} \pi(v) &\sim \mathcal{U}(0, 1) \\ \pi(\theta, \phi) &\sim \pi(v) * |J(v \Rightarrow (\theta, \phi))| \end{aligned}$$

where  $|J(v \Rightarrow (\theta, \phi))| = |\sin \theta|$ . This distribution is used to be non-informative; using the geometrical definition of *solid angle*

$$\delta A = \sin \theta d\theta d\phi$$

we can check that:

$$\int_0^{2\pi} \int_0^\pi \frac{1}{4\pi} \sin \theta d\theta d\phi = \int_0^{2\pi} d\phi \int_0^\pi \frac{1}{4\pi} \sin \theta d\theta = \int_0^\pi \frac{2\pi}{4\pi} \sin \theta d\theta = \int_0^\pi \frac{\sin \theta}{2} d\theta = 1 \quad (\text{B.9})$$

Therefore, the prior distribution of the parameters are:

$$\pi(d) \sim \Gamma(\alpha_d, \beta_d) \sim \mathcal{U}(0, 1)$$

$$\begin{aligned}\pi(\theta, \phi) &\sim \frac{1}{2} |\sin \theta| \\ \pi(f) &\sim \mathcal{U}(0, 1) \\ \pi(S_0) &\sim \mathcal{U}(0, \infty)\end{aligned}$$

Note that to make the priors as non-informative as possible, the unique boundary we are introducing is the positivity of the parameters.

## B.4 Integrating out precision $\tau$ from likelihood

Our likelihood functions contain inherently a noise variance parameter, or equivalently a precision  $\tau = 1/\sigma^2$ . We can integrate this parameter out of the likelihood if we are not interested in directly inferring on it. By using a Gamma distribution as a prior for precision  $\tau$ , and having  $\omega_1 = S_0, d, \theta, \phi, f$  and  $\omega = \omega_1, \tau$ :

$$\begin{aligned}\pi(Y|\omega) &= \int_0^\infty \pi(Y|\omega_1, \tau) \cdot \pi(\tau) d\tau = \int_0^\infty \frac{\sqrt{\tau}}{\sqrt{2\pi}} \cdot \exp\left(\frac{-\tau}{2} \sum_{i=1}^n (Y_i - S_i)^2\right) \cdot \frac{1}{\tau} d\tau \propto \\ &\propto \frac{1}{\sqrt{2\pi}} \int_0^\infty \tau^{1/2-1} \cdot \exp\left(\frac{-\tau}{2} \sum_{i=1}^n (Y_i - S_i)^2\right) d\tau\end{aligned}\quad (\text{B.10})$$

By comparing the integral to the probability density function of a Gamma distribution, we can use the identity  $\int_0^\infty t^{\alpha-1} e^{-\beta t} dt = \frac{\Gamma(\alpha)}{\beta^\alpha}$ , being  $\alpha, \beta > 0$ :

$$\pi(Y|\omega_1) \propto \frac{\Gamma(1/2)}{-\frac{1}{2} \cdot \sum_{i=1}^n (Y_i - S_i)^2} \quad (\text{B.11})$$

$$\Rightarrow \text{Log}(\pi(Y|\omega_1)) \propto \frac{1}{2} \cdot \sum_{i=1}^n (Y_i - S_i)^2 \quad (\text{B.12})$$

A similar result can be achieved when integrating out  $\sigma$  instead of the precision  $\tau$ , as  $\pi(\tau) = \pi(\frac{1}{\sigma^2}) = \sigma^2$ :

$$\begin{aligned}\pi(Y|\omega_1) &= \int_0^\infty \pi(Y|\omega_1, \sigma^2) \cdot \pi(\sigma^2) d\sigma^2 \propto \frac{1}{\sqrt{2\pi}} \int_0^\infty \left(\frac{1}{\sigma^2}\right)^{\frac{1}{2}} \cdot \exp\left(-\frac{\sum_{i=1}^n (Y_i - S_i)^2}{2\sigma^2}\right) \cdot \sigma^2 d\sigma \propto \\ &\propto \int_0^\infty \left(\frac{1}{\sigma^2}\right)^{\frac{1}{2}-1} \cdot \exp\left(-\frac{\sum_{i=1}^n (Y_i - S_i)^2}{2\sigma^2}\right) d\sigma\end{aligned}\quad (\text{B.13})$$

Setting  $t = \frac{1}{\sigma^2}$ , we obtain the identities  $\alpha = 1/2$  and  $\beta = -\sum_{i=1}^n (Y_i - S_i)^2/2$  in the inverse-gamma above, resulting in eqB.11 again.

## B.5 MCMC Posterior Energy

Having noise precision integrated out, the MCMC can draw samples from a distribution independent of  $\tau$  and dependent only on the model parameters of interest:

$$\pi(\omega|Y) \propto \pi(Y|\omega) \cdot \pi(\omega) \quad (\text{B.14})$$

$$\pi(\omega|Y) \propto \pi(Y|\omega) \cdot \pi(\theta, \phi) \cdot \pi(f_i) \cdot \pi(d) \cdot \pi(S_0) \quad (\text{B.15})$$

Expressing the posterior distribution in terms of the energy  $\mathcal{E}$ :

$$\begin{aligned} \mathcal{E} &= -\log(\pi(\omega|Y)) = -\log(\pi(Y|\omega) \cdot \pi(\omega)) \propto \\ &\propto -\log \left( \frac{\Gamma(1/2)}{-\frac{1}{2} \cdot \sum_{i=1}^n (Y_i - S_i)^2} \cdot \frac{|\sin(\theta_i)|}{2} \cdot \left(-\frac{1}{f_i}\right) \cdot \pi(d) \cdot \pi(S_0) \right) \end{aligned} \quad (\text{B.16})$$

Non-informative priors can be discarded, as  $\pi(\omega) = 1 \rightarrow \log(\pi(\omega)) = 0$ . Therefore, the energy function results as:

$$\boxed{\mathcal{E} \propto \frac{1}{2} \cdot \sum_{i=1}^n (Y_i - S_i)^2 - \log\left(\frac{|\sin(\theta_i)|}{2}\right) + \log(f_i)} \quad (\text{B.17})$$

Note that the prior  $\pi(f_i)$  applied here is the ARD prior in its Gaussian form (see below section B.6), applied to the volume fraction of secondary compartments (i.e.,  $i \geq 2$ ). For  $f_1$  (or if not ARD is applied), the prior will be non-informative and can be discarded from B.17 as  $\pi(d)$  or  $\pi(S_0)$ .

## B.6 Implementation of the multi-fibre model with ARD

The Automatic Relevance Determination is a technique used in dMRI model to perform online model selection of the compartments supported by the data. The ARD prior can be implemented either with a Gaussian distribution with zero mean and unknown variance or a Beta distribution with mode at zero and unknown width:

$$ARD \sim \mathcal{N}(0, \sigma)$$

$$ARD \sim \beta(1, k)$$

The idea of the ARD is that if a second or third compartment is not supported, the width of the distribution ( $\sigma$  or  $k$ ) is very small, so that  $f_i$  is driven to the mode, which is zero. However, if  $f_i$  is supported by the data, then the variance has a very large value that allows  $f_i$  to be away from the zero mode and take any value in  $[0,1]$ . Both of the Beta- and Normal-based ARD prior exhibit similar behavior (penalize large  $f_i$  values).

### B.6.1 Beta based ARD prior for $f_i$

$$\pi\left(\frac{f_i}{k}\right) \sim \beta(1, k) \quad (\text{B.18})$$

To get an explicit formula for  $\pi(f_i)$ , we can integrate out  $k$ :

$$\pi(f_i) = \int_0^\infty \pi(f_i|k) \pi(k) dk = \int k(1-f_i)^{k-1} \frac{1}{k} dk = \int \frac{(1-f_i)^k}{1-f_i} dk \quad (\text{B.19})$$

Using the identity  $\int a^{cx} dx = \frac{1}{c \ln(a)} a^{cx}$  (for  $a > 1, a \neq 1$ ):

$$\begin{aligned} \pi(f_i) &= \frac{1}{1-f_i} \frac{1}{\ln(1-f_i)} [(1-f_i)^k]_0^\infty = \frac{1}{1-f_i} \frac{1}{\ln(1-f_i)} \left[ \lim_{k \rightarrow \infty} (1-f_i)^k - (1-f_i)^0 \right] = \\ &= \frac{1}{(1-f_i) \ln(1-f_i)} \end{aligned} \quad (\text{B.20})$$

### B.6.2 Gaussian based ARD prior for $f_i$

As noted in [Behrens et al., 2007], we can simplify the computation if assuming  $\pi(f_i) \sim \mathcal{N}(0, \sigma^2)$  and applying the Jeffrey's prior to  $\sigma$ , as  $\pi(\sigma) \sim \frac{1}{\sigma}$ :

$$\begin{aligned} \pi(f_i) &= \int_0^\infty \pi(f_i|\sigma^2)\pi(\sigma^2)d\sigma^2 = \int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp -\frac{f_i^2}{2\sigma^2} \frac{1}{\sigma^2} d\sigma^2 = \\ &= -\frac{1}{\sqrt{2\pi}} \int_0^\infty \left(\frac{1}{\sigma^2}\right)^{1/2-1} \cdot \exp -\frac{f_i^2}{2\sigma^2} d\sigma^2 \end{aligned} \quad (\text{B.21})$$

Comparing with the inverse-Gamma again, as in the previous section:

$$\pi(f_i) \propto \frac{\Gamma(1/2)}{\left(\frac{f_i^2}{2}\right)^{1/2}} \propto -\frac{1}{f_i} \quad (\text{B.22})$$

$$\cdot \rightarrow \text{Log}(\pi(f_i)) = \text{Log}\left(-\frac{1}{f_i}\right) = f_i \quad (\text{B.23})$$

This is the ARD prior applied in the main text B.23.

## Appendix C

# Laplace Approximation

### C.1 Demonstration 1: Laplace Approximation is located around the posterior mode

The Laplace approximation replaces the problem of integrating a function with the problem of maximizing it. In order to compute the Laplace approximation, we have to compute the location of the mode, which is an optimization problem. Often, this problem is faster to solve using conventional non-linear optimizers rather than integrating the same function.

Suppose we have a function  $g(x) \in \mathcal{L}^2$  with maximum at  $x_0$ . We want to compute

$$\int_a^b g(x)dx \quad (\text{C.1})$$

Let  $h(x) = \log(g(x))$

$$\int_a^b g(x)dx = \int_a^b \exp(h(x))dx \quad (\text{C.2})$$

From here we can take a Taylor series approximation of  $h(x)$  around the point  $x_0$ :

$$\int_a^b \exp(h(x))dx \approx \int_a^b \exp(h(x_0) + h'(x_0)(x - x_0) + \frac{1}{2}h''(x_0)(x - x_0)^2)dx \quad (\text{C.3})$$

Because we assumed  $h(x)$  achieves its maximum at  $x_0$ , we know  $h'(x_0) = 0$ . Therefore, we can simplify the above expression to be

$$= \int_a^b \exp(h(x_0) + \frac{1}{2}h''(x_0)(x - x_0)^2)dx \quad (\text{C.4})$$

Given that  $h(x_0)$  is a constant that doesn't depend on  $x$ , we can pull it outside the integral. Rearranging some terms, we have

$$= \exp(h(x_0)) \int_a^b \exp(-\frac{1}{2} \frac{(x - x_0)^2}{-h''(x_0)^{-1}})dx \quad (\text{C.5})$$

Now, we have a quantity proportional to a Normal distribution  $\mathcal{N}(x_0, -h''(x_0)^{-1})$ .

The next step is to compute the normalizing constant. Let  $\Phi(x|\mu, \sigma^2)$  be the cumulative distribution function for the  $\mathcal{N}(\mu, \sigma^2)$  and  $\varphi$  its density function. The expressions above can be rewritten as:



$$\begin{aligned}
&= \exp(h(x_0)) \sqrt{\frac{2\pi}{-h''(x_0)}} \int_a^b \varphi(x|x_0, -h''(x_0)^{-1}) dx \\
&= \exp(h(x_0)) \sqrt{\frac{2\pi}{-h''(x_0)}} [\Phi(b|x_0, -h''(x_0)^{-1}) - \Phi(a|x_0, -h''(x_0)^{-1})]
\end{aligned}$$

If  $b = \infty$  and  $a = -\infty$ ,  $[\Phi(b|x_0, -h''(x_0)^{-1}) - \Phi(a|x_0, -h''(x_0)^{-1})] \rightarrow 1$ , making the Laplace approximation equal to the value of the function  $g(x)$  at its mode multiplied by a constant that depends on the curvature of the function  $h$ :

$$\int_a^b g(x) dx = \int_a^b \exp(h(x)) dx = \exp(h(x_0)) \sqrt{\frac{2\pi}{-h''(x_0)}} \quad (\text{C.6})$$

### C.1.1 Computing the posterior mean

The posterior mean of a parameter  $\omega$  given the data  $Y$  is can be expressed in terms of the expected value  $\mathbb{E}[\omega]$ :

$$\mathbb{E}[\omega] = \int \pi(\omega|y) \cdot \omega \cdot d\omega = \frac{\int f(Y|\omega) \cdot \omega \cdot \pi(\omega) \cdot d\omega}{\int f(Y|\omega) \cdot \pi(\omega) \cdot d\omega} \quad (\text{C.7})$$

Introducing the exponential-log trick, we can obtain an expression composed by integrals in the same form that the ones solved in by using the Laplace Approximation:

$$= \frac{\int \exp(\log(f(Y|\omega)\pi(\omega))) \cdot \omega \cdot d\omega}{\int \exp(\log(f(Y|\omega)\pi(\omega))) \cdot d\omega} = \frac{\int \exp(h(\omega)) \cdot \omega \cdot d\omega}{\int \exp(h(\omega)) \cdot d\omega} \quad (\text{C.8})$$

Given that  $h(\omega)$  its a simply monotonic transformation (log) of a function proportional to the posterior density:

$$\max(h(\omega)) = \mathbb{E}[\omega] \quad (\text{C.9})$$

So, if we let be  $\hat{\omega}$  the posterior mode of  $(\omega|y)$ , then we have:

$$\int \omega \cdot p(\omega|y) dx \approx \dots = \hat{\omega} \quad (\text{C.10})$$

Hence, the Laplace approximation to the posterior mean is equal to the posterior mode. This approximation is likely to work well when the posterior is unimodal and relatively symmetric around the mode. Furthermore, the more concentrated the posterior around  $\hat{\omega}$ , the better.

## C.2 Demonstration 2. Hessian matrix for the Ball&Sticks model with 3 compartments

The Laplace Approximation to the posterior distribution requires the calculation of the Hessian matrix  $H$  of such conditional density, i.e. the second partial derivatives of  $\pi(\omega|Y)$  respect to the parameters  $\omega_i$  given the data  $Y$ . We can express the posterior in terms of the Energy function,  $\mathcal{E}$ :

$$\mathcal{E} = -\log(\pi(\omega|Y)) = -\log(\pi(Y|\omega) \cdot \pi(\omega)) \quad (\text{C.11})$$

Assuming non-informative priors, i.e.  $\pi(\omega) = 1 \rightarrow \log(\pi(\omega)) = 0$ , the energy function has

the following final form:

$$\mathcal{E} = -\log(\pi(\omega|Y)) \propto -\log(\pi(Y|\omega)) = -\frac{N}{2} \log \left( \frac{\sum_{i=1}^N (Y_i - S_i)^2}{2} \right), \quad (\text{C.12})$$

where  $i$  refers to each volume of the image (gradients,  $i = 1, \dots, N$ ) and  $S_i$  is given by the Ball&Sticks model forward predictions (see eq.2.26). Note that, for the sake of simplicity, in this derivation we have obviated the prior terms of the fibre orientations and volume fractions where ARD is applied; the matrix  $H$  is calculated only to initialise the covariance in the Block-RWM and Hybrid approaches so we don't expect the lack of this term to harm the algorithm in any way. Further discussion can be done regarding the Independence Sampler.

With all the above, the Hessian matrix of  $E$  for a model with  $M = 3$  compartments is defined by:

$$H = \begin{pmatrix} \frac{\partial^2 \mathcal{E}}{\partial S_0^2} & \frac{\partial^2 \mathcal{E}}{\partial S_0 \partial d} & \frac{\partial^2 \mathcal{E}}{\partial S_0 \partial \theta_1} & \frac{\partial^2 \mathcal{E}}{\partial S_0 \partial \phi_1} & \frac{\partial^2 \mathcal{E}}{\partial S_0 \partial f_1} & \cdots & \frac{\partial^2 \mathcal{E}}{\partial S_0 \partial f_3} \\ \frac{\partial^2 \mathcal{E}}{\partial d \partial S_0} & \frac{\partial^2 \mathcal{E}}{\partial d^2} & \frac{\partial^2 \mathcal{E}}{\partial d \partial \theta_1} & \frac{\partial^2 \mathcal{E}}{\partial d \partial \phi_1} & \frac{\partial^2 \mathcal{E}}{\partial d \partial f_1} & \cdots & \frac{\partial^2 \mathcal{E}}{\partial d \partial f_3} \\ \frac{\partial^2 \mathcal{E}}{\partial \theta_1 \partial S_0} & \frac{\partial^2 \mathcal{E}}{\partial \theta_1 \partial d} & \frac{\partial^2 \mathcal{E}}{\partial \theta_1 \partial \theta_1} & \frac{\partial^2 \mathcal{E}}{\partial \theta_1 \partial \phi_1} & \frac{\partial^2 \mathcal{E}}{\partial \theta_1 \partial f_1} & \cdots & \frac{\partial^2 \mathcal{E}}{\partial \theta_1 \partial f_3} \\ \frac{\partial^2 \mathcal{E}}{\partial \phi_1 \partial S_0} & \frac{\partial^2 \mathcal{E}}{\partial \phi_1 \partial d} & \frac{\partial^2 \mathcal{E}}{\partial \phi_1 \partial \theta_1} & \frac{\partial^2 \mathcal{E}}{\partial \phi_1 \partial \phi_1} & \frac{\partial^2 \mathcal{E}}{\partial \phi_1 \partial f_1} & \cdots & \frac{\partial^2 \mathcal{E}}{\partial \phi_1 \partial f_3} \\ \frac{\partial^2 \mathcal{E}}{\partial f_1 \partial S_0} & \frac{\partial^2 \mathcal{E}}{\partial f_1 \partial d} & \frac{\partial^2 \mathcal{E}}{\partial f_1 \partial f_1} & \frac{\partial^2 \mathcal{E}}{\partial f_1 \partial \phi_1} & \frac{\partial^2 \mathcal{E}}{\partial f_1 \partial f_1} & \cdots & \frac{\partial^2 \mathcal{E}}{\partial f_1 \partial f_3} \\ \frac{\partial^2 \mathcal{E}}{\partial \theta_2 \partial S_0} & \frac{\partial^2 \mathcal{E}}{\partial \theta_2 \partial d} & \frac{\partial^2 \mathcal{E}}{\partial \theta_2 \partial \theta_1} & \frac{\partial^2 \mathcal{E}}{\partial \theta_2 \partial \phi_1} & \frac{\partial^2 \mathcal{E}}{\partial \theta_2 \partial f_1} & \cdots & \frac{\partial^2 \mathcal{E}}{\partial \theta_2 \partial f_3} \\ \frac{\partial^2 \mathcal{E}}{\partial \phi_2 \partial S_0} & \frac{\partial^2 \mathcal{E}}{\partial \phi_2 \partial d} & \frac{\partial^2 \mathcal{E}}{\partial \phi_2 \partial \theta_1} & \frac{\partial^2 \mathcal{E}}{\partial \phi_2 \partial \phi_1} & \frac{\partial^2 \mathcal{E}}{\partial \phi_2 \partial f_1} & \cdots & \frac{\partial^2 \mathcal{E}}{\partial \phi_2 \partial f_3} \\ \frac{\partial^2 \mathcal{E}}{\partial f_2 \partial S_0} & \frac{\partial^2 \mathcal{E}}{\partial f_2 \partial d} & \frac{\partial^2 \mathcal{E}}{\partial f_2 \partial f_1} & \frac{\partial^2 \mathcal{E}}{\partial f_2 \partial \phi_1} & \frac{\partial^2 \mathcal{E}}{\partial f_2 \partial f_1} & \cdots & \frac{\partial^2 \mathcal{E}}{\partial f_2 \partial f_3} \\ \frac{\partial^2 \mathcal{E}}{\partial \theta_3 \partial S_0} & \frac{\partial^2 \mathcal{E}}{\partial \theta_3 \partial d} & \frac{\partial^2 \mathcal{E}}{\partial \theta_3 \partial \theta_1} & \frac{\partial^2 \mathcal{E}}{\partial \theta_3 \partial \phi_1} & \frac{\partial^2 \mathcal{E}}{\partial \theta_3 \partial f_1} & \cdots & \frac{\partial^2 \mathcal{E}}{\partial \theta_3 \partial f_3} \\ \frac{\partial^2 \mathcal{E}}{\partial \phi_3 \partial S_0} & \frac{\partial^2 \mathcal{E}}{\partial \phi_3 \partial d} & \frac{\partial^2 \mathcal{E}}{\partial \phi_3 \partial \theta_1} & \frac{\partial^2 \mathcal{E}}{\partial \phi_3 \partial \phi_1} & \frac{\partial^2 \mathcal{E}}{\partial \phi_3 \partial f_1} & \cdots & \frac{\partial^2 \mathcal{E}}{\partial \phi_3 \partial f_3} \\ \frac{\partial^2 \mathcal{E}}{\partial f_3 \partial S_0} & \frac{\partial^2 \mathcal{E}}{\partial f_3 \partial d} & \frac{\partial^2 \mathcal{E}}{\partial f_3 \partial f_1} & \frac{\partial^2 \mathcal{E}}{\partial f_3 \partial \phi_1} & \frac{\partial^2 \mathcal{E}}{\partial f_3 \partial f_1} & \cdots & \frac{\partial^2 \mathcal{E}}{\partial f_3 \partial f_3} \end{pmatrix}$$

### C.2.1 First Derivatives

Making use of some properties of the Hessian matrix:

- Symmetry:  $\frac{\partial^2 \mathcal{E}}{\partial S_0 \partial d} = \frac{\partial^2 \mathcal{E}}{\partial d \partial S_0}$
- Semi-positive definite

and the following derivating rules:

$$\frac{\partial \sum_{i=1}^N f(x, y)}{\partial x} = \sum_{i=1}^N \frac{\partial f(x, y)}{\partial x} \quad (\text{C.13})$$

$$\frac{\partial(\log(f(x, y)))}{\partial x} = \frac{f'(x, y)}{f(x, y)} \quad (\text{C.14})$$

$$\frac{\partial f(x, y)^N}{\partial x} = N \cdot f(x, y)^{N-1} \cdot f'(x, y) \quad (\text{C.15})$$

$$\frac{\partial f(x, y) \cdot g(x, y)}{\partial x} = f'(x, y) \cdot g(x, y) - f(x, y) \cdot g'(x, y) \quad (\text{C.16})$$

we can calculate a general form for the first partial derivatives of E:

$$\begin{aligned} \boxed{\frac{\partial \mathcal{E}}{\partial \omega}} &= \frac{\partial(\frac{N}{2} \log(f(x, y)))}{\partial x} = \frac{N}{2} \frac{f'(x, y)}{f(x, y)} = \frac{N}{2} \frac{(\frac{1}{2} \sum_{i=1}^N (Y_i - S_i)^2)'}{\frac{1}{2} \sum_{i=1}^N (Y_i - S_i)^2} = \\ &= \frac{N}{2} \frac{\frac{1}{2} \sum_{i=1}^N (2(Y_i - S_i)(-S_i)')}{\frac{1}{2} \sum_{i=1}^N (Y_i - S_i)^2} = \boxed{-\mathbf{N} \frac{\sum_{i=1}^N ((\mathbf{Y}_i - \mathbf{S}_i) \cdot \mathbf{S}_i')}{\sum_{i=1}^N (\mathbf{Y}_i - \mathbf{S}_i)^2}} \quad (\text{C.17}) \end{aligned}$$

E And apply it to each parameter  $\omega$ :

$$\begin{aligned} \boxed{\frac{\partial \mathcal{E}}{\partial \mathbf{S}_0}} &= \frac{\partial \frac{N}{2} \cdot \log(\frac{1}{2} \sum_{i=1}^N (Y_i - S_i)^2)}{\partial S_0} = \\ &= \frac{\partial \frac{N}{2} \cdot \log(\frac{1}{2} \sum_{i=1}^N (Y_i - [S_0 \cdot [(1 - \sum_{j=1}^M f_j) e^{-b \cdot d} + \sum_{j=1}^M f_j \cdot e^{-b \cdot d \cdot (g \cdot v^T)^2}]]^2))}{\partial S_0} = \\ &= \boxed{-\mathbf{N} \frac{\sum_{i=1}^N [(\mathbf{Y}_i - \mathbf{S}_i) \cdot ((1 - \sum_{j=1}^M \mathbf{f}_j) \cdot \mathbf{e}^{-b \cdot d} + \sum_{j=1}^M \mathbf{f}_j \cdot \mathbf{e}^{-b \cdot d \cdot (g_i \cdot \mathbf{v}_i^T)^2})]}{\sum_{i=1}^N (\mathbf{Y}_i - \mathbf{S}_i)^2}} \quad (\text{C.18}) \end{aligned}$$

$$\begin{aligned}
\boxed{\frac{\partial \mathcal{E}}{\partial \mathbf{d}}} &= -N \frac{\sum_{i=1}^N [(Y_i - S_i) \cdot S_0 \cdot [-b \cdot (1 - \sum_{j=1}^M f_j) \cdot e^{-b_i d} + \sum_{j=1}^M f_j \cdot (-b_i) \cdot (g_i v_i^T)^2 \cdot e^{-b_i d (g_i v_i^T)^2}]]}{\sum_{i=1}^N (Y_i - S_i)^2} = \\
&= \boxed{N \frac{\sum_{i=1}^N [(\mathbf{Y}_i - \mathbf{S}_i) \cdot \mathbf{S}_0 \cdot \mathbf{b}_i \cdot [(1 - \sum_{j=1}^M \mathbf{f}_j) \cdot \mathbf{e}^{-\mathbf{b}_i \mathbf{d}} + \sum_{j=1}^M \mathbf{f}_j \cdot (\mathbf{g}_i \mathbf{v}_i^T)^2 \cdot \mathbf{e}^{\mathbf{b}_i \mathbf{d} (\mathbf{g}_i \mathbf{v}_i^T)^2}]]}{\sum_{i=1}^N (\mathbf{Y}_i - \mathbf{S}_i)^2}} \quad (\text{C.19})
\end{aligned}$$

$$\begin{aligned}
\boxed{\frac{\partial \mathcal{E}}{\partial \mathbf{f}_j}} &= -N \frac{\sum_{i=1}^N [(Y_i - S_i) \cdot (-S_0) \cdot (e^{-b_i d} + \sum_i e^{-b_i d (g_i v_i^T)^2})]}{\sum_{i=1}^N (Y_i - S_i)^2} = \\
&= \boxed{N \frac{\sum_{i=1}^N [(\mathbf{Y}_i - \mathbf{S}_i) \cdot \mathbf{S}_0 \cdot (\mathbf{e}^{-\mathbf{b}_i \mathbf{d}} - \sum_{i=1}^N \mathbf{e}^{-\mathbf{b}_i \mathbf{d} (\mathbf{g}_i \mathbf{v}_i^T)^2})]}{\sum_{i=1}^N (\mathbf{Y}_i - \mathbf{S}_i)^2}} \quad (\text{C.20})
\end{aligned}$$

Regarding the fibre orientation parameters  $(\theta, \phi)$ , it is worth to note that the gradient orientation vector,  $g_i$ , should be in the same reference system than  $v_i$ . So, being  $g_i = [\sin(\alpha) \cdot \cos(\beta), \sin(\alpha) \cdot \sin(\beta), \cos(\alpha)]$ ,  $v_i = [\sin(\theta) \cdot \cos(\phi), \sin(\theta) \cdot \sin(\phi), \cos(\theta)]$ , and  $dp = g_i v_i^T$ , their partial derivatives are given by:

$$\begin{aligned}
dp &= g_i v_i^T = \cos(\phi - \beta) \cdot \sin(\alpha) \cdot \sin(\theta) + \cos(\alpha) \cdot \cos(\theta) \\
dp1_\theta &= \frac{\partial dp}{\partial \theta} = \cos(\theta) \cdot (g(:, 1) \cdot \cos(\phi) + g(:, 2) \cdot \sin(\phi)) - g(:, 2) \cdot \sin(\theta) \\
dp1_\phi &= \frac{\partial dp}{\partial \phi} = \sin(\theta) \cdot (-g(:, 1) \cdot \sin(\phi) + g(:, 2) \cdot \cos(\phi)) \quad (\text{C.21})
\end{aligned}$$

Then, the partial derivatives of  $\mathbf{E}$  respect the fibre orientation parameters are:

$$\boxed{\frac{\partial \mathcal{E}}{\partial \theta_j} = N \frac{\sum_{i=1}^N [(\mathbf{Y}_i - \mathbf{S}_i) \cdot (\mathbf{S}_0 \cdot \mathbf{f} \cdot 2\mathbf{b}_i \cdot \mathbf{d} \cdot \mathbf{dp} \cdot \mathbf{dp}1_\theta) \cdot \mathbf{e}^{-\mathbf{b}_i \mathbf{d} (\mathbf{g}_i \mathbf{v}_i^T)^2}]}{\sum_{i=1}^N (\mathbf{Y}_i - \mathbf{S}_i)^2}} \quad (\text{C.22})$$

$$\boxed{\frac{\partial \mathcal{E}}{\partial \phi_j} = N \frac{\sum_{i=1}^N [(\mathbf{Y}_i - \mathbf{S}_i) \cdot (\mathbf{S}_0 \cdot \mathbf{f} \cdot 2\mathbf{b}_i \cdot \mathbf{d} \cdot \mathbf{dp} \cdot \mathbf{dp}1_\phi) \cdot \mathbf{e}^{-\mathbf{b}_i \mathbf{d} (\mathbf{g}_i \mathbf{v}_i^T)^2}]}{\sum_{i=1}^N (\mathbf{Y}_i - \mathbf{S}_i)^2}} \quad (\text{C.23})$$

For the sake of simplicity in the second derivatives, let's re-arrange some terms in the following form:

$$\begin{aligned}
SS &= \frac{1}{2} \sum_{i=1}^N (Y_i - S_i)^2 \\
A_i &= \exp(-b_i \cdot d) \\
B_i &= \exp(-b_i \cdot d \cdot (g \cdot v^T)^2) \\
W_i &= S_0 \left(1 - \sum_{j=1}^M f_j\right) \cdot \exp(-b_i \cdot d) = S_0 \left(1 - \sum_{j=1}^M f_j\right) \cdot A_i \\
X_i &= S_0 \cdot \sum_{j=1}^M f_j \cdot \exp(-b_i \cdot d \cdot (g \cdot v^T)^2) = S_0 \cdot \sum_{j=1}^M f_j \cdot B_i \\
K_i &= \left(1 - \sum_{j=1}^M f_j\right) \cdot \exp(-b_i \cdot d) + \sum_{j=1}^M f_j \cdot \exp(-b_i \cdot d \cdot (g_i \cdot v_i^T)^2) = \left(1 - \sum_{j=1}^M f_j\right) \cdot A_i + \sum_{j=1}^M f_j \cdot B_i \\
Z_i &= S_0 \left(1 - \sum_{j=1}^M f_j\right) \cdot \exp(-b_i \cdot d) + S_0 \cdot \sum_{j=1}^M f_j \cdot (g_i \cdot v_i^T)^2 \cdot \exp(-b_i \cdot d \cdot (g_i \cdot v_i^T)^2) = W_i + X_i (g_i v_i^T)^2
\end{aligned} \tag{C.24}$$

So introducing these in C.18, C.19, and C.20:

$$\begin{aligned}
\frac{\partial \mathcal{E}}{\partial S_0} &= N \cdot \frac{\sum_{i=1}^N (K_i \cdot (Y_i - S_i)^2)}{SS} \\
\frac{\partial \mathcal{E}}{\partial d} &= N \cdot \frac{\sum_{i=1}^N (K_i \cdot (Y_i - S_i)^2)}{SS} \\
\frac{\partial \mathcal{E}}{\partial f_j} &= N \cdot \frac{\sum_{i=1}^N (S_0 \cdot (Y_i - S_i)(A_i - B_i))}{SS}
\end{aligned} \tag{C.25}$$

## C.2.2 Second Derivatives

For the second partial derivatives the following derivative rules will be used:

$$\begin{aligned}
f(x) = \frac{u}{v} &\rightarrow f'(x) = \frac{u' \cdot v - u \cdot v'}{v^2} \\
g(x, y) = \frac{1}{f(x, y)^N} &\rightarrow \frac{\partial g(x, y)}{\partial x} = N \cdot f(x, y)^{n-1} \cdot f'(x, y)
\end{aligned}$$

Now, starting from the general form of the 1st derivative:

$$\frac{\partial \mathcal{E}}{\partial \omega_1} = -N \cdot \frac{\sum_{i=1}^N ((Y_i - S_i) \cdot S'_i)}{\sum_{i=1}^N (Y_i - S_i)^2} = -N \cdot \frac{\sum_{i=1}^N ((Y_i - S_i) \cdot \frac{\partial \mathcal{E}}{\partial \omega_1})}{\sum_{i=1}^N (Y_i - S_i)^2} \quad (\text{C.26})$$

We can obtain a general expression also for the second derivatives:

$$\begin{aligned} \boxed{\frac{\partial^2 \mathcal{E}}{\partial \omega_1 \partial \omega_2}} &= -N \frac{\frac{\partial(\sum_{i=1}^N ((Y_i - S_i) \cdot \frac{\partial S_i}{\partial \omega_1}))}{\partial \omega_2} \cdot \sum_{i=1}^N (Y_i - S_i)^2 - \sum_{i=1}^N ((Y_i - S_i) \cdot \frac{\partial S_i}{\partial \omega_1}) \cdot \frac{\partial(\sum_{i=1}^N (Y_i - S_i)^2)}{\partial \omega_2}}{(\sum_{i=1}^N (Y_i - S_i)^2)^2} = \\ &= -N \cdot \frac{\left( \sum_{i=1}^N \left( \frac{\partial S_i}{\partial \omega_2} \cdot \frac{\partial S_i}{\partial \omega_1} \right) - \sum_{i=1}^N ((Y_i - S_i) \cdot \frac{\partial S_i}{\partial \omega_1 \partial \omega_2}) \right) \cdot \sum_{i=1}^N (Y_i - S_i)^2}{(\sum_{i=1}^N (Y_i - S_i)^2)^2} \\ &\quad + \frac{\sum_{i=1}^N ((Y_i - S_i) \cdot \frac{\partial S_i}{\partial \omega_1}) \cdot 2 \cdot \sum_{i=1}^N ((Y_i - S_i) \cdot (-\frac{\partial S_i}{\partial \omega_2}))}{(\sum_{i=1}^N (Y_i - S_i)^2)^2} \end{aligned}$$

Most of the terms have been already calculated in C.2.1. Only  $\frac{\partial S_i}{\partial \omega_1 \partial \omega_2}$  needs to be calculated here. Therefore, the second partial derivatives for  $S_0$  result in:

$$\boxed{\frac{\partial^2 \mathcal{E}}{\partial S_0^2} = 0} \quad (\text{C.27})$$

$$\boxed{\frac{\partial^2 \mathcal{E}}{\partial S_0 \partial d} = -(1 - \sum_{j=1}^M f_j) \cdot b_i \cdot e^{-b_i \cdot d} + \sum_{j=1}^M f_j \cdot (-b_i) \cdot dp^2 \cdot e^{-b_i \cdot d \cdot dp^2}} \quad (\text{C.28})$$

$$\boxed{\frac{\partial^2 \mathcal{E}}{\partial S_0 \partial \theta_j} = f_j \cdot (-2 \cdot b_i) \cdot d \cdot dp \cdot dp 1_\theta \cdot e^{-b_i \cdot d \cdot dp^2}} \quad (\text{C.29})$$

$$\boxed{\frac{\partial^2 \mathcal{E}}{\partial S_0 \partial \phi_j} = f_j \cdot (-2 \cdot b_i) \cdot d \cdot dp \cdot dp 1_\phi \cdot e^{-b_i \cdot d \cdot dp^2}} \quad (\text{C.30})$$

$$\boxed{\frac{\partial^2 \mathcal{E}}{\partial S_0 \partial f_j} = -e^{-b_i d} + e^{-b_i \cdot d \cdot dp^2}} \quad (\text{C.31})$$

For  $d$ :

$$\boxed{\frac{\partial^2 \mathcal{E}}{\partial d^2} = S_0 \cdot \left( (1 - \sum_{j=1}^M f_j) \cdot b_i^2 \cdot e^{-b_i \cdot d} + \sum_{j=1}^M f_j \cdot b_i^2 \cdot dp^2 \cdot e^{-b_i \cdot (g_i \cdot v_i^T)^2} \right)} \quad (\text{C.32})$$

$$\frac{\partial^2 \mathcal{E}}{\partial d \cdot \partial \theta_j} = S_0 \cdot f_j \cdot (-2 \cdot b_i) \cdot dp \cdot dp1_{dth} \cdot (1 - b_i \cdot d \cdot dp^2) \cdot e^{-b_i \cdot d \cdot dp^2} \quad (C.33)$$

$$\frac{\partial^2 \mathcal{E}}{\partial d \cdot \partial \phi_j} = S_0 \cdot f_j \cdot (-2 \cdot b_i) \cdot dp \cdot dp1_{dph} \cdot (1 - b \cdot d \cdot dp^2) \cdot e^{-b_i \cdot d \cdot dp^2} \quad (C.34)$$

$$\frac{\partial^2 \mathcal{E}}{\partial d \cdot \partial f_j} = S_0 \left( f_j \cdot b_i \cdot e^{-b_i d} + (-b_i) \cdot dp^2 \cdot e^{-b \cdot d \cdot dp^2} \right) \quad (C.35)$$

For  $\theta_j$ :

$$\frac{\partial^2 \mathcal{E}}{\partial \theta_j^2} = S_0 \cdot f_j \cdot (-2b_i) \cdot d \cdot e^{-b_i \cdot d \cdot dp^2} \cdot (1 - 2b_i \cdot dp^2) - dp^2 \quad (C.36)$$

$$\frac{\partial^2 \mathcal{E}}{\partial \theta_j \cdot \partial \phi_j} = S_0 \cdot f_j \cdot (-2 \cdot b_i) \cdot d \cdot e^{-b_i \cdot d \cdot dp^2} \cdot dp \cdot dp2_{\theta\phi} \quad (C.37)$$

$$\frac{\partial^2 \mathcal{E}}{\partial \theta_j \cdot \partial f_j} = S_0 \cdot (-2 \cdot b_i) \cdot d \cdot dp \cdot dp1_{\theta} \cdot e^{-b_i \cdot d \cdot dp^2} \quad (C.38)$$

For  $\phi_j$ :

$$\frac{\partial^2 \mathcal{E}}{\partial \phi_j^2} = S_0 \cdot f_j \cdot (-2b_i) \cdot d \cdot e^{-b_i \cdot d \cdot dp^2} \cdot ((1 - 2b_i \cdot dp^2) \cdot dp1_{\phi^2} + dp \cdot dp2_{\phi^2}) \quad (C.39)$$

$$\frac{\partial^2 \mathcal{E}}{\partial \phi_j^2} = S_0 \cdot (-2 \cdot b_1) \cdot d \cdot dp \cdot dp1_{\phi} \cdot e^{-b_i \cdot d \cdot dp^2} \quad (C.40)$$

where

$$dp1_{\phi^2} = \frac{1 - \cos(2\theta_j)}{2}$$

$$dp2_{\phi^2} = -g_i[:, 1] \cdot v_j[1] - g_i[:, 2] \cdot v_j[2]$$

And for  $f_j$ :

$$\frac{\partial^2 \mathcal{E}}{\partial f_j^2} = 0 \quad (C.41)$$

## Appendix D

# Hybrid-MCMC results

### D.0.1 Dispersion maps in Dataset C - 3 fibres

The following maps show the standard deviation (or dispersion) provided by each MCMC algorithm. In agreement with the rest of results, the Block-RWM may suffer from a rigid structure given by the joint proposal and it is not able to correctly map the uncertainty of the estimates (i.e. the posterior distribution).

## D.1 Probabilistic Tractography and uncertainty propagation

In this section, we will provide results that complement the main text.

### D.1.1 Comparisons to the RWM

Following are comparisons of the spatial tract correlations obtained from different MCMC configurations with respect the default RWM used as gold-standard.



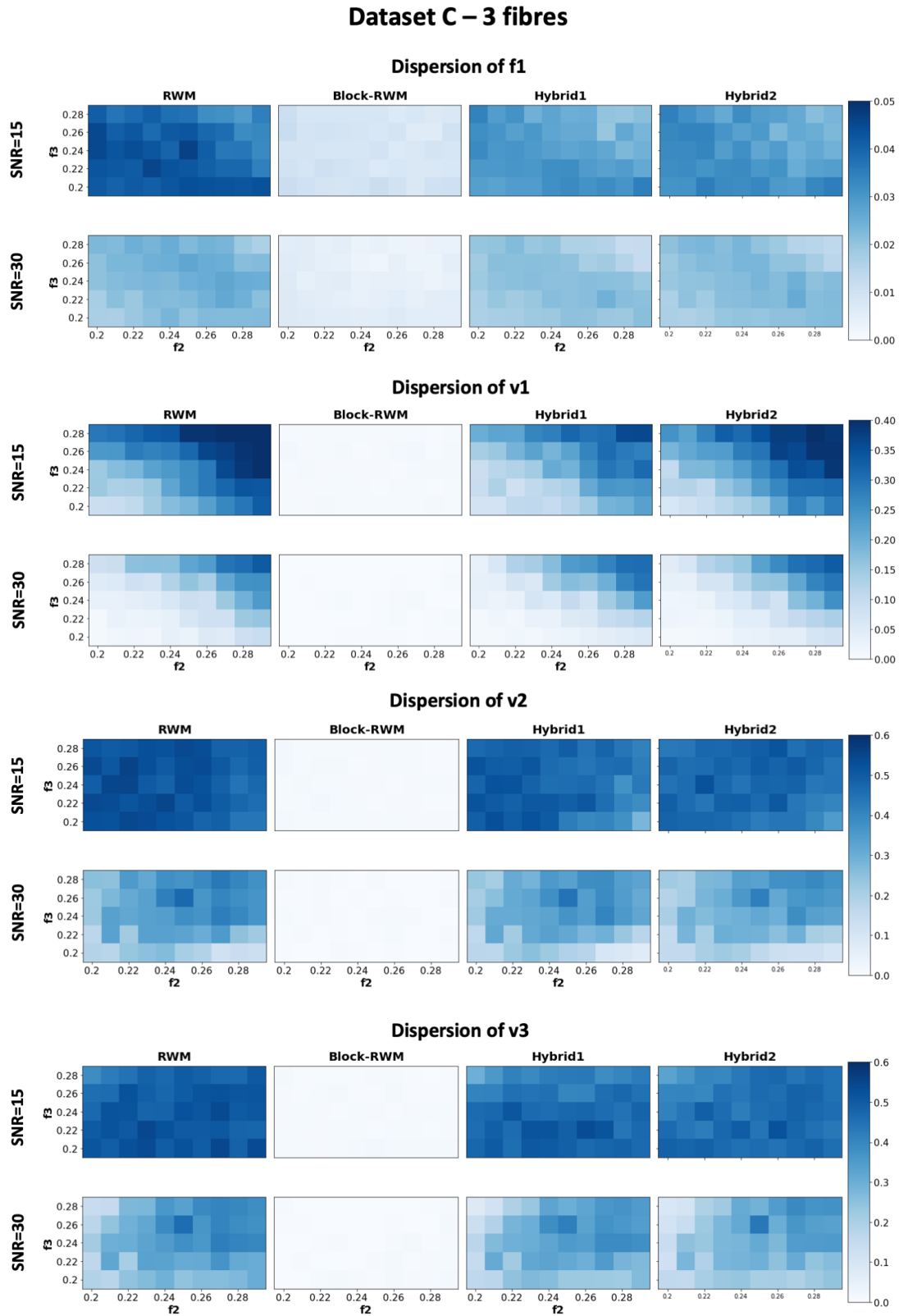
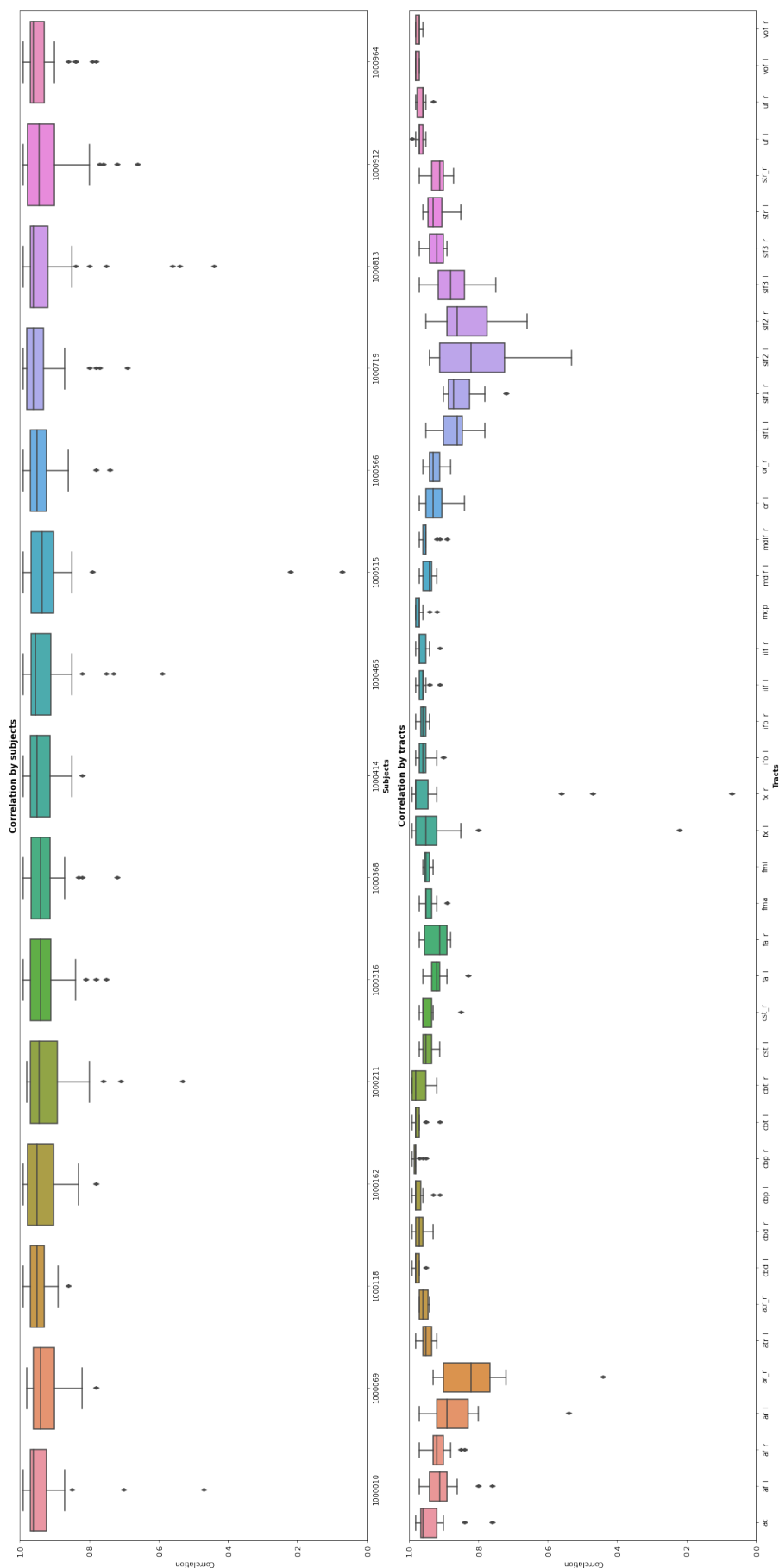


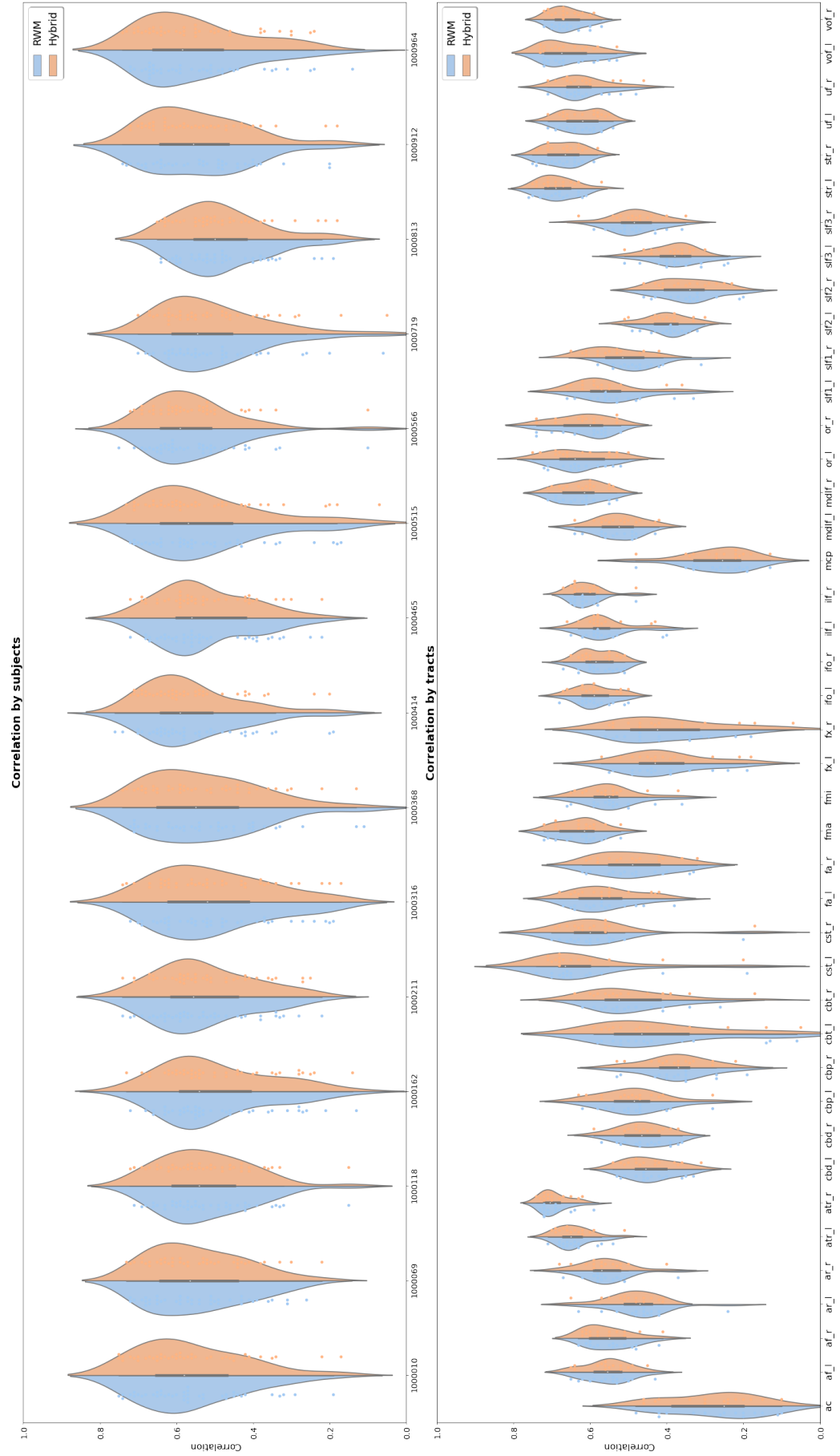
Figure D.1: *Standard deviation or dispersion* obtained by each method for  $f_1$  and the fibre orientations in Dataset C at different SNR levels.



**Figure D.2: Spatial Tract Correlations.** MCMC approach evaluated is the **Hybrid-MCMC** configuration 1 with increased number of jumps ( $\text{burnin}=3000$ ,  $\text{jumps}=5000$ ,  $\text{thinning}=100$ ). Correlations are grouped by: multiple subjects from the UK Biobank dataset (Top row); Defined tracts from the XTRACT protocol (Bottom row)

### D.1.2 Comparisons to the HCP White-Matter population average atlas

Following are comparisons of the spatial tract correlations obtained from different MCMC configurations.



**Figure D.3: Spatial Tract Correlations with the HCP population-averaged atlas.** MCMC approaches compared are the RWM (blue) and Hybrid-MCMC configuration 1 (orange) with MCMC parameters as follows: burnin=3000, jumps=5000, thinning=100. Correlations are grouped by: multiple subjects from the UKBiobank dataset (Top row); Defined tracts from the XTRACT protocol (Bottom row)