UCL Université catholique de Louvain

École polytechnique de Louvain (EPL)

epl

Comparison of diffusion MRI models of the brain microstructure using the generalization error

Dissertation presented by
Antoine DURVIAUX

for obtaining the Master's degree in
Mathematical Engineering

Supervisor
Benoît MACQ

Readers
Pierre-Antoine ABSIL, Gaëtan RENSONNET

Academic year 2016-2017

Acknowledgements

Tout au long de ce mémoire j'ai été amené à rencontrer différentes personnes qui, de près ou de loin, m'ont permis de mener à bien ce travail. Je tiens à leur faire part de toute ma gratitude.

Tout d'abord, je voudrais remercier particulièrement mon promoteur Benoît Macq pour m'avoir permis de réaliser ce mémoire ainsi que de découvrir ce beau domaine.

Ensuite, je tiens à exprimer mes plus sincères remerciements à Gaëtan Rensonnet qui a passé de nombreuses heures à répondre à mes questions ainsi qu'à m'aider tout au long de ce mémoire.

Enfin, je tiens à remercier tous mes proches et particulièrement ma famille pour leur soutien tout au long de la réalisation de ce travail.

Contents

Abbreviations

AD	Axial Diffusivity
CSF	CerebroSpinal Fluid
DIAMOND	DIstribution of Anisotropic MicrOstructural eNvironments with Diffusion-weighted imaging
DTI	Diffusion Tensor Imaging
DWI	Diffusion-Weighted Imaging
DW-MRI	Diffusion-Weighted Magnetic Resonance Imaging
FA	Fractional Anisotropy
FID	Free Induction Decay
GE	Generalization Error
GM	Grey Matter
HARDI	High Angular Resolution Diffusion Imaging
MD	Mean Diffusivity
MC	Monte Carlo
MFM	Multi-Fascicle Model
MRI	Magnetic Resonance Imaging
NMR	Nuclear Magnetic Resonance
NODDI	Neurite Orientation Dispersion and Density Imaging
PGSE	Pulsed-Gradient Spin Echo
RD	Radial Diffusivity
RF	Radio Frequency
SNR	Signal-to-Noise Ratio
SPD	Symmetric Positive-Definite
SE	Spin Echo
TE	Echo Time
WM	White Matter

Symbols

b	b-value	s m^{-2}
δ	Duration of gradient pulse	s
Δ	Diffusion time	s
τ	Time between RF pulses in spin echo sequence	s
f_0	Larmor frequency	Hz
T1	Spin-lattice relaxation time	s
T2	Spin-spin relaxation time	s
$T2^*$	Effective spin-spin relaxation time	s
TE	Echo time	s
$\mathbf{D}$	Diffusion tensor	m^2 s^{-1}
E	DW-MRI signal attenuation	/
$\mathrm{Tr}(\boldsymbol{A})$	Trace of matrix $\boldsymbol{A}$	/
$\lvert\boldsymbol{A}\rvert$	Determinant of matrix $\boldsymbol{A}$	/
$\mathbb{E}[x]$	Statistical expectation of random variable x	/
$\Re(z)$	Real part of complex number z	/

k	Boltzmann's constant	1.38×10^{-23} J K^{-1}
h	Planck's constant	6.6×10^{-34} J s
D	Coefficient of diffusion for free water at 37 °C	3×10^{-9} m^2 s^{-1}
γ	Gyromagnetic ratio of hydrogen	$2\pi \times 42.577 \times 10^6$ rad s^{-1} T^{-1}

List of Figures

List of Tables

Introduction

Diffusion-weighted magnetic resonance imaging (DW-MRI) is a non-invasive medical imaging tool which is sensitive to the movement of water molecules present in the human body. When this technique is applied to provide informations about the brain microstructure, it is called *microstructure imaging*. Inside the human body, there is more than 50% of water. As the body temperature is non-zero, the water molecules experience a movement of diffusion due to thermal agitation. Thanks to magnetic gradients, DW-MRI allows to probe this movement. Therefore, it is possible to guess how neurons are oriented inside the brain. This technique is thus useful on patients suffering from brain microstructure alterations such as axon injury, demyelination, cerebral edema or neuroinflamation [1].

The output of DW-MRI is the signal emitted by water molecules. From this output, it is possible to construct 3D images with a gray intensity proportional to the signal in each voxel (volume element) of the image. Figure 1 shows an example of a brain image obtained by DW-MRI.

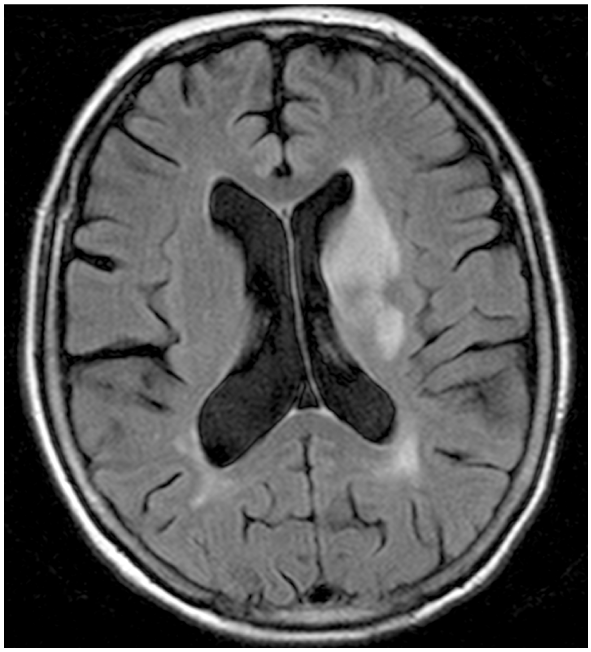

Figure 1: **Example of a 2D brain image obtained by DW-MRI** [2].

In order to have more details about the brain microstructure than a simple image, there exists several mathematical models that try to adjust parameters to fit the DW-MRI signal. For example, the simplest model, called *diffusion tensor imaging*, fits a second-order tensor $\boldsymbol{D}$ for each voxel to the diffusion signal. This tensor can be view as an ellipsoid and gives informations about the average diffusion direction and strength in each voxel. Other models divide the diffusion signal in several compartments, each of them associated with a part of the neurons structure. In this document, we compare four different models : the diffusion tensor imaging model, its amelioration and two more complex models.

The inputs of DW-MRI are parameters such as the norm of the magnetic gradients and the duration of their application. In order to increase the accuracy of the information about the brain microstructure, several DW-MRI are often realized on the same patient with different protocols, *i.e.* different input parameters. Unfortunately, it takes time and this is not always possible. Furthermore, the longer a patient spends time in a scanner, the more inaccurate is the signal due to artifacts coming from the patient movements.

This document proposes to compare four microstructure models by using the generalization error. This error is defined as the capacity of a model to predict the signal from a new protocol. A small error means that the model accurately predicts a new signal. Accurately predicting new signals is of great interest to reduce the acquisition time and to provide informations on the brain microstructure. Figure 2 illustrates the process by which it is possible to predict new signals using diffusion signals coming from DW-MRI acquisitions.

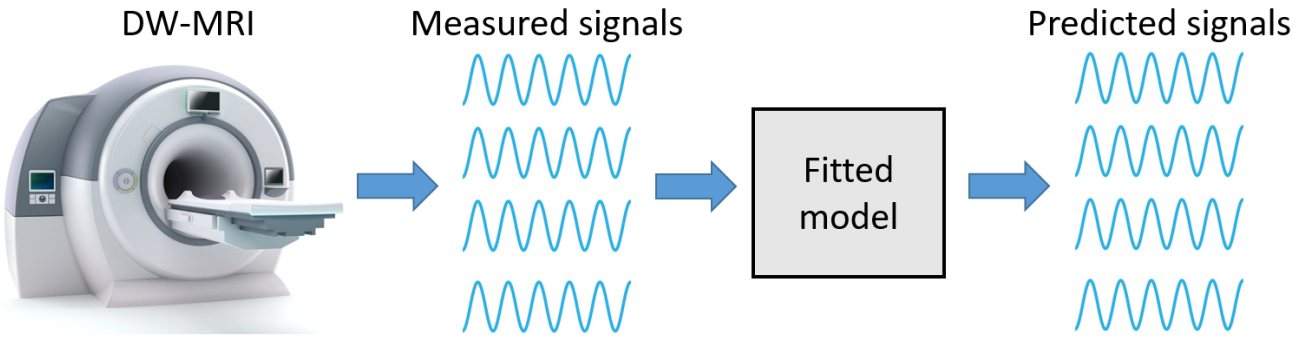

Figure 2: **Interest of accurate microstructural models**. First, several DW-MRI are made with different protocols. This gives rise to multiple signals which are used to fit a mathematical model. Once the model fitted, it can predict new signals associated with different new protocols.

The first chapter presents all the theoretical background needed to understand DW-MRI as well as the different kinds of microstructure models. The second chapter details the four chosen models and explains the limitations and advantages of each of them. In the third chapter, the tool used to compare the models, known as the generalization error, is developed. Finally, the last chapter exposes the data used to fit the different models and compares the errors made by each of them.

Chapter 1

Theoretical background

This first chapter presents the basics of a medical imaging process called diffusion-weighted magnetic resonance imaging (DW-MRI). This technique enables the diffusion of water molecules in the body of a patient to be characterized. DW-MRI applied to the study of the brain is referred to as brain microstructure imaging which is the main topic of this document.

DW-MRI is a complex topic that require strong knowledge in various domains. In the first section, the magnetic resonance imaging (MRI) is introduced. The second section presents the physical process of diffusion with both the microscopic and macroscopic approaches. The third section combines the MRI and the diffusion to introduce the DW-MRI. Afterward, a section explains a method called Monte-Carlo simulation to acquire DW-MRI synthetic data. Finally, this chapter ends with a section presenting the different types of DW-MRI models.

1.1 Physical basis of magnetic resonance imaging

Magnetic resonance imaging is a medical imaging technique used in radiology. It allows the formation of images of the anatomy as well as physiological processes in a patient. This is a non-invasive medical imaging technique because MRI scanners only use magnetic fields and radio waves which do not expose patients to harmful ionizing radiation.

1.1.1 Nuclear magnetic resonance

MRI is based upon the nuclear magnetic resonance (NMR) properties of certain atomic nuclei which are able to absorb and emit electromagnetic radiation in the presence of a magnetic field. These nuclei have an odd number of protons and/or neutrons and are thus quantum mechanically characterized by a nonzero spin quantum number. These particles are also called *spins*. Hydrogen is a good example of spin especially since it is the most abundant element in the human body. A spin can be visualized as a particle rotating around its own axis. Since a spin is charged, it creates an electric current which induces a magnetic field characterized by a magnetic moment $\boldsymbol{\mu}$. The orientation of this magnetic moment is random due to thermal random motion. When a constant uniform and external magnetic gradient is applied along the z-axis ($\boldsymbol{B_0} = B_0\boldsymbol{e_z}$), $\boldsymbol{\mu}$ will try to align with $\boldsymbol{B_0}$. For the hydrogen nuclei, spin is either parallel/up (low energy state) or antiparallel/down (high energy state) to $\boldsymbol{e_z}$. The magnetic moment experiences a torque that causes it to precess around $\boldsymbol{B_0}$ with the following frequency :

$$f_0 = \frac{\gamma}{2\pi} B_0 \quad [Hz] \,,$$

called the Larmor frequency. The constant γ is the gyromagnetic ratio [rad $\mathrm{s}^{-1}\mathrm{T}^{-1}$] and is characteristic for each nucleus, *e.g.* for hydrogen nucleus :

$$\frac{\gamma}{2\pi} = 42.58 \ \ [\text{MHz T}^{-1}] \,.$$

Figure 1.1 illustrates the magnetic moment of a spin precessing around an externally-applied magnetic field.

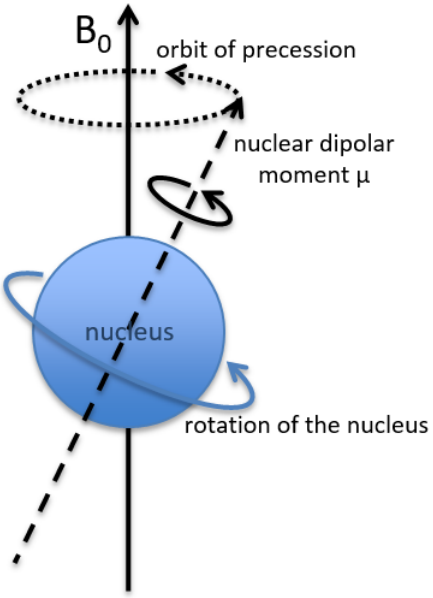

Figure 1.1: **Representation of a spin in an external magnetic field $\boldsymbol{B_0}$** [3].

In practice, an MRI scanner deals with a collection of hydrogen spins, each possessing a magnetic moment $\boldsymbol{\mu_i}$. The net macroscopic magnetization $\boldsymbol{M}$ is defined by the sum of all individual magnetic moments :

$$\boldsymbol{M} = \sum_i \boldsymbol{\mu_i} \ .$$

In the absence of an external magnetic field, $\boldsymbol{M} = \boldsymbol{0}$ since the spins are randomly oriented. If an external magnetic field $\boldsymbol{B_0}$ is applied, spins are either up or down as previously seen. It turns out that in thermal equilibrium, the ratio of the occupations follows a Boltzmann distribution [4] :

$$\frac{N_\uparrow}{N_\downarrow} = \exp(\frac{\Delta E}{kT}) \ ,$$

with $\Delta E = E_\downarrow - E_\uparrow = hf_0$ the energy difference between spin up and spin down. h is Planck's constant, k Boltzmann's constant and T the absolute temperature. This distribution means that more spins are up (lower energy) than down (higher energy). Therefore, $\boldsymbol{M} = M_0\boldsymbol{e_z} \neq \boldsymbol{0}$, *i.e.* the net macroscopic magnetization is longitudinal. Indeed, the transverse magnetization in the xy-plane is zero because of the random distribution of all the spins phases.

It is difficult to measure the longitudinal magnetization due to spins because it is much smaller (several orders of magnitude) than the external field B_0. Indeed, at 310 kelvins (*i.e.* the body temperature) one cubic millimeter (*i.e.* the characteristic size of a voxel[1]) of water contains 3.3456×10^{19} water molecules [5]. The number of hydrogen atoms is 6.6913×10^{19}, leading to :

$$\frac{M}{B_0} = 3.0882 \times 10^{-12} \ .$$

That is why it is better to measure the magnetization in the xy-plane where there is no other field. To create a transverse magnetization, the scanner sends a radio frequency (RF) pulse at the Larmor frequency to the spins population which will enter in resonance. This resonance occurs in two phases.

[1] A voxel (contraction of *volume element*) is a pixel in 3 dimensions.

In the first one, called excitation phase, the RF pulse is absorbed by the system. This modifies the energy state of the spins. Spins in the low energy state will move to the higher energy state. The longitudinal magnetization decreases due to a difference in the number of spins up and down. A transverse magnetization appears because the RF pulse causes spins to precess in phase. The magnetization $\boldsymbol{M}$ is tilted away from the z-axis and can be decomposed in two components :

$$\boldsymbol{M} = M_z \boldsymbol{e}_z + M_{xy} \boldsymbol{e}_{xy} \ .$$

A RF pulse is characterized by its flip angle α and is then called an α-pulse. For example, a 90°-pulse will make the longitudinal magnetization disappear, *i.e.* the proportion of parallel and anti-parallel spins is equal, and a transverse magnetization appears, *i.e.* all spins are in phase.

The second step called relaxation phase occurs when the scanner stops emitting RF pulses. The high-energy state being less stable, the system tends to come back to its initial thermal equilibrium (*i.e.* $\boldsymbol{M} = M_0 \boldsymbol{e_z}$). This is done by two separate processes called longitudinal relaxation and transverse relaxation. Longitudinal relaxation is the process by which the spin population returns to the equilibrium Boltzmann distribution. It is characterized by a time constant $T1$ which describes how fast the recovery occurs. During transverse relaxation, the spins come to thermal equilibrium and dephase themselves. It is characterized by a time constant $T2$. However, besides transverse relaxation, spins tend to get out of phase with one another when there are inhomogeneities in the external magnetic field $\boldsymbol{B_0}$. This is taken into account with the effective time constant $T2^*$.

1.1.2 Spin echo

The oscillating behavior of the transverse magnetization will induce a signal known as the free induction decay (FID). This signal is not directly measurable because of the severe signal loss due to $T2^*$-relaxation. To measure the transverse magnetization, the spin echo (SE) sequence is often used. This method is a pulse sequence consisting in two RF-pulses. At time $t = 0$, a 90°-pulse converts the longitudinal magnetization into transverse magnetization (Figure 1.2 B). The spins will dephase due to inhomogeneities in the external magnetic field (Figure 1.2 C and D). Then, at $t = \tau$, a 180°-pulse flips all the magnetic moments into the opposite direction and reverse the direction of rotation (Figure 1.2 E). As a result, at $t = 2\tau$, called the echo time (TE), they refocus (Figure 1.2 F) and create a spin echo (Figure 1.2 G). The transverse magnetization can be measured with an amplitude only affected by $T2$-relaxation. After $t = TE$, the spins will dephase again. Figure 1.2 illustrates a spin echo sequence.

1.1.3 Magnetic resonance imaging

Magnetic resonance imaging is a technique that uses NMR to produce 2D or 3D images. For that purpose, it is necessary to excite only the area of interest of the patient and localize the MR-signals. This is done by using magnetic field gradients. A gradient field is a magnetic field oriented (*e.g.* along z) that varies with position along some spatial direction (*e.g.* x). Therefore, the linear gradient field is $G_x x$ and the slope $G_x = \partial B_z / \partial x$ is called the magnetic gradient. G_y and G_z are defined similarly. If gradients are applied in the three spatial directions in the presence of the same external magnetic field $\boldsymbol{B_0} = B_0 \boldsymbol{e_z}$, the total field becomes :

$$\boldsymbol{B}(\boldsymbol{r}) = (B_0 + \boldsymbol{G} \cdot \boldsymbol{r}) \boldsymbol{e_z} = B(\boldsymbol{r}) \boldsymbol{e_z} \ ,$$

where $\boldsymbol{G} = (G_x, G_y, G_z)$ is the magnetic field gradient and $\boldsymbol{r} = (x, y, z)$ is the position. In practice, the external magnetic field is much bigger than the gradient field. Now that the field is position dependent, so is the Larmor frequency :

$$f_0(\boldsymbol{r}) = \frac{\gamma}{2\pi} B(\boldsymbol{r}) \ .$$

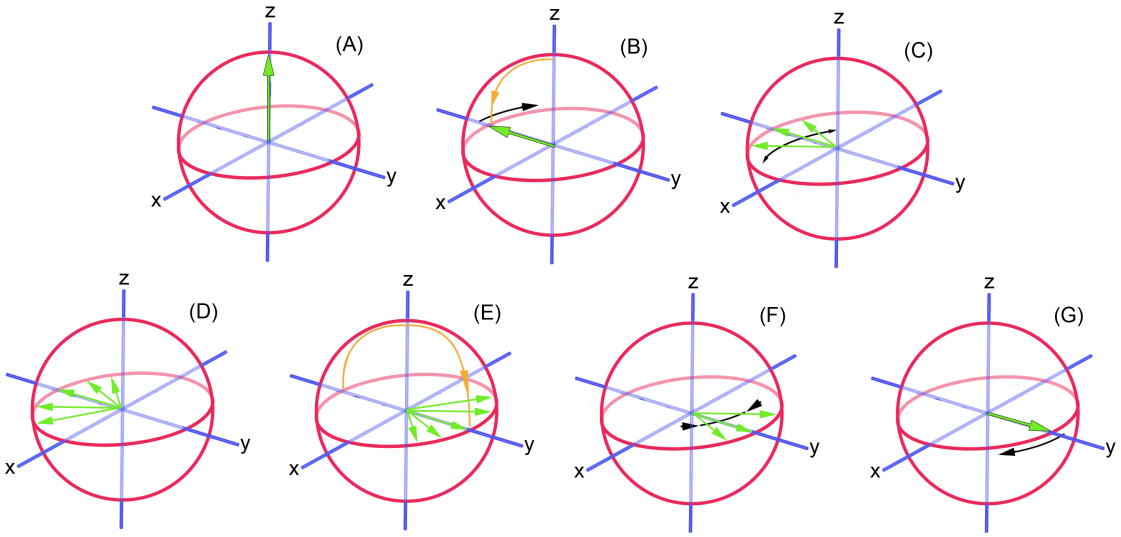

Figure 1.2: **Illustration of the spin echo sequence** [6].
(A) : Spins are aligned with the external magnetic field $\boldsymbol{B}_0$.
(B) : Longitudinal magnetization is converted into transverse magnetization due to a 90°-pulse emitted at $t = 0$.
(C) and (D) : Due to inhomogeneities in $\boldsymbol{B}_0$, spins are dephased.
(E) : Magnetic moments are flipped into opposite directions due to a 180°-pulse emitted at $t = \tau$.
(F) : Refocusing of the spins.
(G) : At $t = 2\tau$, a spin echo is created.

Therefore, to excite only a slice of spins of thickness Δz perpendicular to z, we have to use a RF-pulse of bandwidth :

$$\Delta f = \frac{\gamma}{2\pi} G_z \Delta z \ .$$

In order to get an image, spatial information has to be encoded. This is done by doing first a phase encoding and then a frequency encoding [4].

1.2 Diffusion

Random molecular collisions occur billions of times per second in fluids. The energy transfer associated with the collisions modifies the motion of the molecules. All those microscopic motions result, in a macroscopic level, in a net mass transfer of the considered chemical species, *e.g.* water molecules in human brain. This process is called diffusion. Thus, diffusion is a macroscopic manifestation of the Brownian motion that is taking place at the microscopic level. Therefore there are two main approaches by which the diffusion has been studied : a microscopic and a macroscopic way [7].

1.2.1 Microscopic approach : random walk and brownian motion

From a microscopic point of view, diffusion can be seen as the motion of molecules colliding with each other because of the thermal agitation. This motion can be considered as a random walk, *i.e.* a stochastic process that describes a path consisting on a succession of random steps. Each step in the random walk is independent of the preceding one. This lack of memory of the process is called *Markov property*. Therefore, the stochastic process $X(t)$, representing the position of a brownian particle, follows a *Markov process*. Figure 1.3 illustrates a random walk of a brownian particle.

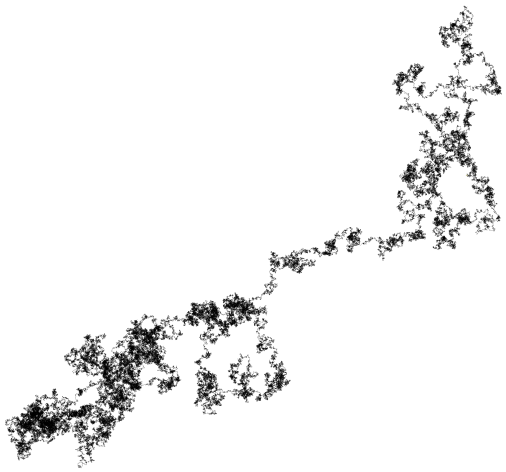

Figure 1.3: **Illustration of a brownian particle random walk** [8].

It can be proved [9] that after a sufficiently long time, if $\boldsymbol{r}(t)$ is the position at time t and $\boldsymbol{r}_0$ is the initial position, the probability distribution of $\boldsymbol{r}(t) - \boldsymbol{r}_0$ converges to a Gaussian distribution of zero mean and variance proportional to the time t :

$$\sigma^2 = 2Dt \ ,$$

where D [m^2 s^{-1}] is the diffusion coefficient.

1.2.2 Macroscopic approach : diffusion equation

The macroscopic approach uses continuous physics to derive the so-called diffusion equation. This equation is the combination of two well-known equations :

- The mass conservation equation :

$$\frac{\partial C}{\partial t}(\boldsymbol{r}, t) = -\boldsymbol{\nabla} \cdot \boldsymbol{J}(\boldsymbol{r}, t) \ ,$$

  where C [mol m^{-3}] or [kg m^{-3}] is the concentration of the studied species, $\boldsymbol{J}$ [mol m^{-2} s^{-1}] or [kg m^{-2} s^{-1}] is the diffusive flux, $\boldsymbol{r}$ is the position and t the time. This equation states that since the system is closed to all transfers of matter and energy, the mass of the system must remain constant over time.

- Fick's law which is a constitutive equation for the flux $\boldsymbol{J}$:

$$\boldsymbol{J}(\boldsymbol{r}, t) = -\boldsymbol{D} \cdot \boldsymbol{\nabla} C(\boldsymbol{r}, t) \ ,$$

  where $\boldsymbol{D}$ [m^2 s^{-1}] is a 3×3 symmetric, positive-definite tensor known as the diffusion tensor. This equation states that the studied substance diffuses from regions of high concentration to regions of low concentration.

Combining these two equation gives the *diffusion equation*

$$\frac{\partial C}{\partial t}(\boldsymbol{r}, t) = \boldsymbol{\nabla} \cdot (\boldsymbol{D} \cdot \boldsymbol{\nabla} C(\boldsymbol{r}, t)) \ .$$

In the case of isotropic diffusion (see next subsection), the diffusion equation simplifies to

$$\frac{\partial C}{\partial t}(\boldsymbol{r}, t) = D \Delta C(\boldsymbol{r}, t))$$

since the diffusion is the same in all directions. This is a form of *heat equation*.

1.2.3 Types of diffusion

We conclude this section with a few words about the different types of diffusion that we can encounter.

- **Isotropic diffusion** : When the diffusion is the same in every direction, the diffusion is said to be *isotropic*. This diffusion is characterized by a single diffusion coefficient D.
- **Anisotropic diffusion** : When the medium is highly structured like in most biological tissues, the diffusion coefficients are different depending on the direction. The diffusion is *anisotropic*. As a consequence, the diffusion is not characterized by a single number anymore but by a 3×3 diffusion tensor $\boldsymbol{D}$. Practically, this tensor is always symmetric :

$$\mathbf{D} = \begin{pmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{xy} & D_{yy} & D_{yz} \\ D_{xz} & D_{yz} & D_{zz} \end{pmatrix},$$

  where the diagonal elements represent diffusion coefficients measured along each of the principal directions. The six off-diagonal elements reflect the correlation between random motions corresponding to each pair of principal directions.
- **Restricted diffusion** : Restricted diffusion refers to the diffusion in restricted geometries where diffusing species are physically constrained to remain within a bounded domain.
- **Hindered diffusion** : The diffusion is said to be hindered when obstacles hinder the diffusion of substance under study. It is not because a diffusion is hindered that it is restricted.

Figure 1.4 illustrates the different types of diffusion that can occur in the brain.

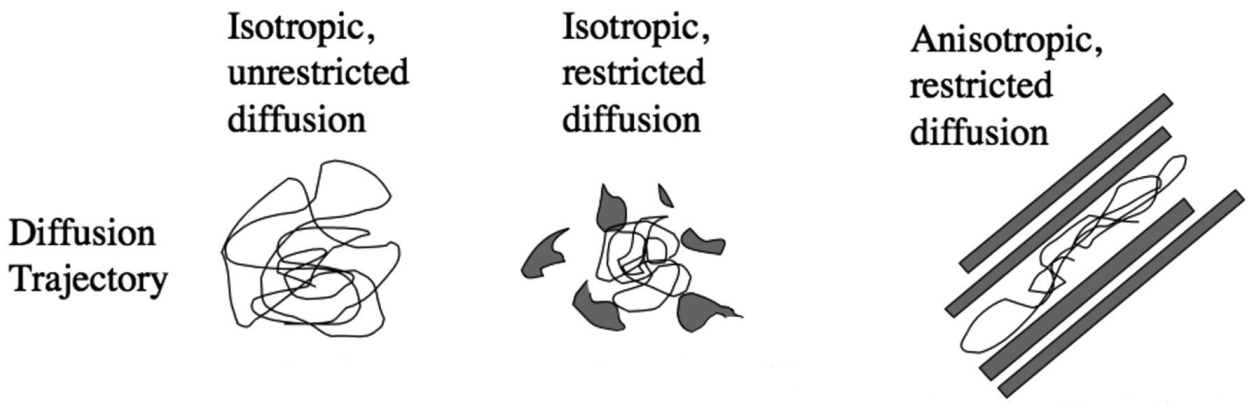

Figure 1.4: **The different types of diffusion** [10]. Examples of diffusion trajectory for isotropic unrestricted, isotropic restricted and anisotropic restricted diffusions.

In this document, we mainly study the diffusion within the brain. In that case, the degree of anisotropy and restriction depends on the microstructure. In highly organized structures like white matter fascicles, the diffusion is highly anisotropic and highly restricted in the direction orthogonal to fascicles because molecules diffuse better in directions where the resistance is the least. On the contrary, in less coherent structures, the diffusion is almost isotropic and scarcely restricted. Finally, in the cerebrospinal fluid (CSF), the diffusion is isotropic and unrestricted.

1.3 Diffusion-weighted magnetic resonance imaging

Diffusion-weighted magnetic resonance imaging (DW-MRI), sometimes abbreviated diffusion-weighted imaging (DWI), is a particular MRI modality that is sensitive to the diffusion process of water molecules. Its goal is to probe this by combining classical MRI sequences like the spin echo with successive applications of magnetic gradients. By doing this, the signal becomes sensitive to the diffusion of molecules during the acquisition. It is therefore possible to infer information about the underlying tissue microstructure.

This section first presents the most known DWI sequence : the pulsed-gradient spin echo (PGSE). Then it focuses on the two approaches developed in the previous section : the microscopic and macroscopic approaches.

1.3.1 Pulsed-gradient spin echo

A PGSE sequence is a spin echo sequence where two linear magnetic fields of magnitude $B_1(\boldsymbol{r}) = \boldsymbol{G} \cdot \boldsymbol{r}$ are successively superposed to the uniform and constant magnetic field $\boldsymbol{B_0} = B_0 \boldsymbol{e_z}$. A PGSE sequence is characterized by three parameters illustrated in Figure 1.5 :

- $\boldsymbol{G}$: the 3-dimensional magnetic gradient of magnitude G and unit direction $\hat{\boldsymbol{g}}$,
- δ : the time during which each gradient is applied,
- Δ : the time between the onset of the first and second gradient.

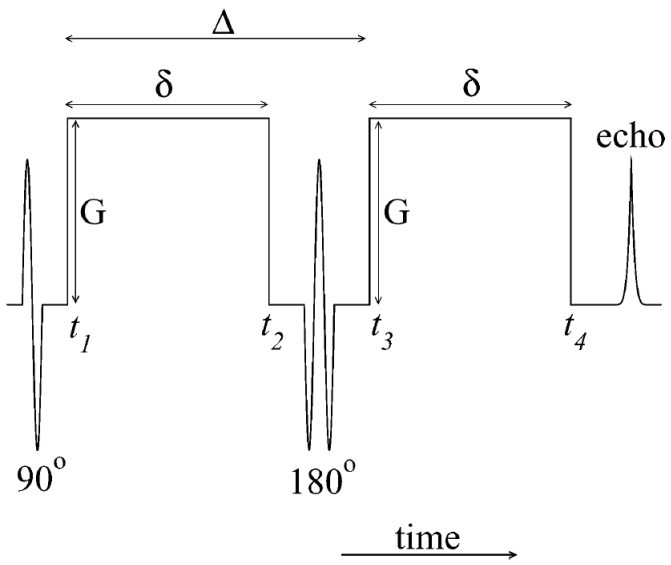

Figure 1.5: **Schematic representation of a pulsed-gradient spin echo sequence** [11]. At $t = 0$, a 90°-pulse is emitted. Immediately after the RF pulse at $t = t_1$, a gradient G is applied for a time δ until $t_2 = t_1 + \delta$. Then a 180°-pulse is emitted at $t = \tau$. A second gradient of same magnitude and same duration begins at $t_3 = t_1 + \Delta$ and ends at $t_4 = t_1 + \Delta + \delta$. Finally, the echo occurs at $t = 2\tau = TE$.

As for the spin echo sequence, the PGSE sequence first starts at $t = 0$ with a 90°-pulse which converts the longitudinal magnetization of the spins into a transverse magnetization. Then, at $t = t_1$, a first gradient is applied during a time δ. The effect of this gradient is to dephase the spins that were precessing together, therefore making them experience distinct Larmor frequencies. At $t = \tau$, a 180°-pulse flips all the magnetic moments into the opposite direction and reverse the direction of

rotation. At $t_3 = t_1 + \Delta$, the second gradient, which has the same duration, direction and intensity as the first one, is applied. Its effect is to refocus all the spins. If the spins did not move, their phase are perfectly refocused after the application of the second gradient and the measured signal is the same as for a SE sequence. However, due to random diffusion through the tissues, the refocusing is imperfect and the signal attenuation can't be completely compensated for. The amount of remaining signal attenuation is related to the amount of motion that occurs along the gradient direction.

1.3.2 Microscopic approach

It is possible to derive a mathematical expression for the attenuation of the transverse magnetization associated to the PGSE sequence by using the microscopic approach presented in the previous section.

Assume a Brownian motion composed of N Brownian particles (*i.e.* spins) in a diffusion environment Ω (*i.e.* a voxel of human brain). The initial 90°-pulse aligns spin phases perpendicular to the applied static field $\boldsymbol{B}_0 = B_0 \boldsymbol{e_z}$. All spins have a Larmor frequency $f = \gamma B_0 / 2\pi$. Then the first gradient is applied and dephases the spins. The phase shift of spin k after the application of the first gradient is the following :

$$\phi_k^{(1)} = \gamma \boldsymbol{B}_0 \tau + \gamma \int_{t_1}^{t_2} \boldsymbol{G} \cdot \boldsymbol{r}_k(t) dt \qquad k = 1, 2, ..., N \ ,$$

where $\boldsymbol{r}_k(t)$ is the position of spin k at time t. The times τ, t_1 and t_2 are the same as in Figure 1.5. The first term is the phase shift due to the static field $\boldsymbol{B}_0$ and the second is due to the gradient pulse.

After the application of the 180°-pulse, a second gradient is applied and each spin experiences a second phase shift :

$$\phi_k^{(2)} = \gamma \boldsymbol{B}_0 \tau + \gamma \int_{t_3}^{t_4} \boldsymbol{G} \cdot \boldsymbol{r}_k(t) dt \qquad k = 1, 2, ..., N \ .$$

Finally, the net phase shift ϕ_k experienced by a spin k at the echo time $t = TE$ is :

$$\begin{aligned} \phi_k &= \phi_k^{(1)} - \phi_k^{(2)} \\ &= \gamma \textstyle\int_{t_1}^{t_2} \boldsymbol{G} \cdot \boldsymbol{r}_k(t) dt - \gamma \int_{t_3}^{t_4} \boldsymbol{G} \cdot \boldsymbol{r}_k(t) dt \end{aligned} \qquad k = 1, 2, ..., N \tag{1.1}$$

where the minus sign reflects the effect of the inverting 180°-pulse applied at $t = \tau$.

Imagine that the spins are static during the application of the gradient pulses, *i.e.* the *narrow gradient pulse* condition $\delta << \Delta$ is assumed [12]. Then, a particular spin k located at position $\boldsymbol{r}_1$ after the first gradient pulse and at position $\boldsymbol{r}_2$ after the second gradient pulse, has the following net phase shift :

$$\phi_k = \gamma \delta \boldsymbol{G} \cdot (\boldsymbol{r}_2 - \boldsymbol{r}_1) \ .$$

This net phase shift is null if the spin remained static, *i.e.* $\boldsymbol{r}_1 = \boldsymbol{r}_2$. If the spin has moved due to diffusion, it will acquire a phase shift and the signal will be attenuated.

When the phase ϕ_k of each spin is known, it is possible to calculate the attenuation E of the transverse magnetization associated to the PGSE sequence of parameters $p_{pgse} = (\boldsymbol{G}, \Delta, \delta)$ in a diffusion environment Ω :

$$E = \frac{S(TE)}{\tilde{S}_0} = \frac{e^{-\frac{TE}{T2}} \sum_{k=1}^{N} \frac{M_0}{N} e^{i\phi_k}}{M_0 e^{-\frac{TE}{T2}}} = \frac{1}{N} \sum_{k=1}^{N} e^{i\phi_k} \ ,$$

where $\tilde{S}_0$ is the unweighted signal, $S(TE)$ is the measured signal at echo time, $T2$ is the spin-spin relaxation time and M_0 is the net magnetization in the xy-plane after the application of the 90°-pulse and is written in complex notation :

$$M_0 = M_{0x} + iM_{0y} \ .$$

With the microscopic approach, it is therefore possible to calculate the attenuation of the transverse magnetization by simply providing the Brownian trajectories $\boldsymbol{r}_k(t)$ which depends on the tissue microstructure Ω. The purpose of methods that generate synthetic data such as Monte Carlo simulation is to generate these N random trajectories in Ω. This is the subject of the next section.

1.3.3 Macroscopic approach

The Bloch-Torrey equations are a system of coupled linear PDE's which describe the evolution of the transverse magnetization $M_{xy} = M_x + iM_y$ in a diffusion domain Ω. This system is challenging to solve for arbitrary domains.

However, it is possible to obtain a simple expression for the signal S if isotropic Gaussian diffusion is assumed. This gives the following Stejskal-Tanner equation [13] :

$$S = \tilde{S}_0 e^{-\frac{TE}{T2}} e^{-\gamma^2\delta^2G^2(\Delta-\delta/3)D} \ , \tag{1.2}$$

where D is the diffusion coefficient, $T2$ is the time constant for spin-spin relaxation and $\tilde{S}_0$ is the signal obtained if no signal attenuation occurred.

In the field of diffusion MRI, a useful quantity named *b-value* is used and contains the parameters of a PGSE sequence. It is defined as follows :

$$b \triangleq \gamma^2\delta^2G^2(\Delta - \delta/3) \ .$$

Similarly, a notation commonly used in the literature is to group together the unweighted signal with the decreased exponential depending on $T2$, giving a new signal S_0 :

$$S_0 = \tilde{S}_0 e^{-\frac{TE}{T2}} \ .$$

With these new notations, equation 1.2 becomes :

$$S = S_0 e^{-bD} \ . \tag{1.3}$$

The signal attenuation depends only on the b-value and has an exponential profile as illustrated on Figure 1.6 :

$$E(b) = \frac{S(b)}{S_0} = e^{-bD} \ . \tag{1.4}$$

Figure 1.6 shows a schematic representation where a diffusion process occurs between two parallel barriers represented by two black lines. Perpendicular to these barriers, the diffusion is restricted whereas parallel to them it is unrestricted. The figure also illustrates the signal attenuation for both cases. The unrestricted attenuation is obtained with equation 1.4. The restricted one is simulated as the diffusion signal attenuation generated within a cylinder along the direction perpendicular to the axis of the cylinder [14]. Grey dots correspond to the attenuation with $b = 1000$ [s/mm^2]. The orange curves represent a diffusion time τ_2. For a given b-value, the attenuation is lower for unrestricted diffusion than for restricted diffusion. Indeed, when diffusion is restricted by the presence of the barrier, the spins are on average subject to a smaller net displacement and the signal is less attenuated.

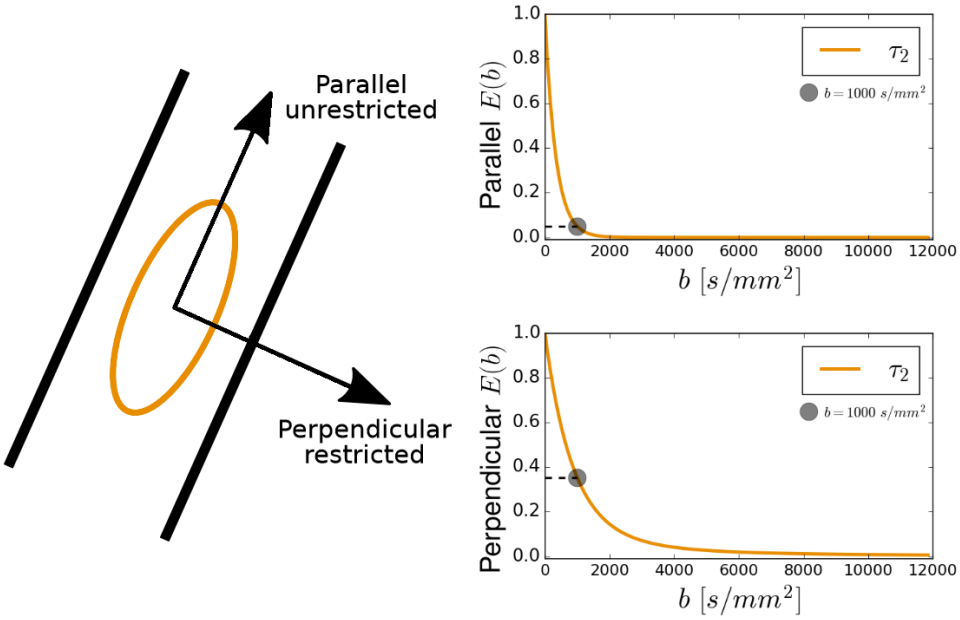

Figure 1.6: **Effect of the b-value on the signal attenuation** [12].
On the left : representation of a diffusion process between two black parallel barriers after a diffusion time τ_2.
On the right : graphics representing the signal attenuation in both unrestricted and restricted diffusion. The two grey dots highlight a b-value of 1000 s/mm^2.

Equation 1.3 is largely used in diffusion models such as MFM and NODDI[2] to represent compartments where the diffusion is free.

It is possible to generalize the Stejskal-Tanner equation for the Gaussian anisotropic case. The result is the following :

$$S = S_0 e^{-b\hat{\boldsymbol{g}}^T \boldsymbol{D} \hat{\boldsymbol{g}}} \ , \tag{1.5}$$

where $\boldsymbol{D}$ is the diffusion tensor. Equation 1.5 is the basis of the DTI model which is explained in the next chapter.

The three parameters TE, δ and Δ are chosen for the acquisition and must be such that

$$TE \geq \Delta + \delta \ ,$$

which is clear when looking at Figure 1.5. In practice, one wants a TE as short as possible to avoid a loss of signal, due to $e^{-TE/T2}$, and thus a loss of the signal-to-noise ratio (SNR) since the noise is approximately constant.

DWI are scalar images which contain the value of the measured signal S at each voxel for a given b-value b and gradient $\boldsymbol{G}$. For microstructure containing heterogeneity, multiple DWI are generally acquired using various gradient pulse directions and b-values.

The gradient field is generated by electromagnetic coils in the three cartesian directions. One hardware constraint is that each coil can create a magnetic field of a maximum amplitude. Therefore, to increase the b-value, Δ and/or δ must be increased. This results in an increased echo time which has two negative effects : first it increases the acquisition time of the images and second it decreases the SNR

[2]These two models will be detailed in Chapter 2.

of the images because of the factor $e^{-TE/T2}$. It is possible to increase the SNR by averaging multiple DWI acquired with the same parameters. Averaging N DWI increases the SNR by a factor $\sqrt{N}$ but the acquisition takes N times longer. However, there exists an acquisition scheme named cube and sphere (CUSP) which enable to increase the b-value without increasing the echo time [15].

1.4 Monte-Carlo simulation

In this document, the DW-MRI signals come from real data obtained on real subjects[3]. However, such real data are not always available because it takes time to acquire them on patients. Therefore, scientists developed methods to acquire synthetic data. These synthetic data are useful to compare estimated parameters to the known parameters of the simulation and investigate the effect of noise. The goal of these methods is to generate an accurate DW-MRI signal attenuation $E(p_{\text{pgse}}, \Omega)$ from given PGSE parameters $p_{pgse} = (\boldsymbol{g}, \Delta, \delta)$ and a given diffusion environment Ω. In this section, the Monte Carlo (MC) simulation [11], which is the reference for the generation of synthetic DW-MRI data, is presented followed by its limitations and advantages.

1.4.1 Description

A Monte Carlo simulation is based on the microscopic approach of diffusion where N Brownian particles (*i.e.* spins) follow a discrete Brownian motion as presented in Section 1.3.2. The simulation consists in generating the trajectories of these N particles which are uniformly distributed across the domain Ω.

Let $\{t_0, t_1, t_2, ..., t_{T-1}, t_T\}$ be a partition of the time interval between the initial 90°-pulse ($t_0 = 0$) and the echo time ($t_T = TE$) and $\boldsymbol{r}_k(t)$ be the position of a particular spin k. Then, from a random initial position $\boldsymbol{r}_k(t_0)$ in the domain, the position of spin k is updated as follows :

$$\boldsymbol{r}_k(t_{i+1}) = \boldsymbol{r}_k(t_i) + \Delta\boldsymbol{r}_i \qquad i = 0, ..., T \ ,$$

where the step $\Delta\boldsymbol{r}_i$ has a random orientation and a fixed length

$$L = \|\Delta\boldsymbol{r}_i\| = \sqrt{2nD\delta t_i} \ ,$$

with n the spatial dimension of the diffusion environment and

$$\delta t_i = t_i - t_{i-1} \quad i = 1, ..., T \ .$$

By the central limit theorem, the distribution of all n components of the position of spin k $\boldsymbol{r}_k(t)$ at time t will tend to a Gaussian distribution with a mean equal to zero by symmetry and a variance equal to $2Dt$ which is exactly what was obtained in Section 1.2.1.

It is possible to discretise the time integrals in equation 1.1 with a rectangle-like method to obtain the accumulated phase ϕ_k of spin k. The formula for the update at each iteration is

$$\phi_k(t_{i+1}) = \phi_k(t_i) + \gamma f(t_i)\boldsymbol{g} \cdot \boldsymbol{r}_k(t_i)\delta t_i \ ,$$

where $f(t)$ is the temporal profile of the applied magnetic gradient. For a PGSE sequence, $f(t)$ takes the values 0,1 or -1 as illustrated on Figure 1.7.

The final signal attenuation is computed as the sample mean of $e^{i\phi}$:

$$E(p_{\text{pgse}}, \Omega) = \frac{1}{N}\sum_{k=1}^{N} e^{i\phi_k} \ .$$

[3]More details about the data can be found in Appendix C.

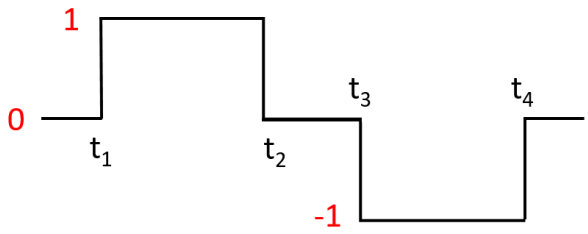

Figure 1.7: **Values taken by $f(t)$ for a PGSE sequence**. The -1 comes from the fact that between t_2 and t_3, a 180°-pulse is emitted. The times t_1, t_2, t_3 and t_4 are the same as on Figure 1.5.

Note that it suffices to compute the real component of the signal since it has been shown that DW-MRI signals are real and positive under assumptions that are usually met in our experiments [16].

Finally, spins that encounter a barrier are reflected elastically since we consider perfectly impermeable membranes. This is illustrated in Figure 1.8.

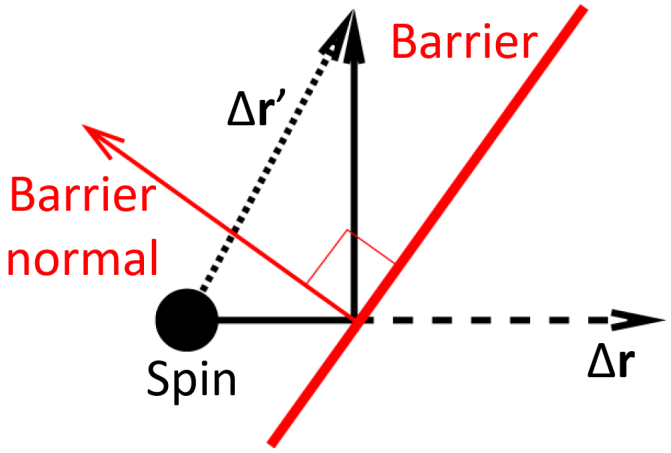

Figure 1.8: **Illustration of a membrane allowing perfect reflection**. Figure inspired from [11].

1.4.2 Limitations

A first limitations of the MC simulations is the computation time and the memory requirements. Indeed, they are both proportional to the number of spins N, the number of time steps T and the number of PGSE sequences generated. Since these parameters are often high, the simulation time is therefore very big. To give an idea, for the simplest geometry, it can last for more than an hour [17].

Another limitation is that the parameters N and T have to be precisely determined for every diffusion environment and every PGSE sequence. This determination is done through convergence analysis and needs additional simulation time.

Finally, only few simple geometries are implemented in current software.

1.4.3 Advantages

The main advantages of MC simulations are its accuracy and its flexibility. Indeed, the equations presented above have theoretical foundations and their results are thus correct.

Furthermore, any magnetic gradients and any geometries can be implemented, even if the simulation time can be a limitation as just discussed.

Finally, since each of the N spins evolves independently, it is possible to benefit from parallel computing.

1.5 DW-MRI models

DW-MRI returns a signal altered by the diffusion of water. A DW-MRI model is a model that fits this measured signal to a parametrized analytical expression that incorporates information about the brain microstructure. Thanks to these models, we can provide valuable informations about the brain structure. There are two main categories of DW-MRI models : the phenomenological ones and the microstructural ones. This section first introduces the superposition principle which is a useful property used by microstructural models. Then, the two categories are explained.

1.5.1 Superposition principle

The DW-MRI signal benefits fom a useful property called superposition principle. The statement of this principle is the following [18] :

If the diffusion environment Ω can be expressed as the union of K mutually-disjoint compartments $\Omega_1, ..., \Omega_K$ with volumes $\|\Omega_1\|, ..., \|\Omega_K\|$ and with perfectly-reflecting boundaries, then the total signal $S_{pgse}(\Omega)$ in the diffusion environment Ω is obtained as :

$$S_{pgse}(\Omega) = \sum_{i=1}^{K} f_i S_{pgse}(\Omega_i) \ ,$$

where $f_i = \frac{\|\Omega_i\|}{\|\Omega\|}$ and $S_{pgse}(\Omega_i)$, for $i = 1, ..., K$ are respectively the volume fractions and the individual signal contributions of each compartment.

This result holds if the $T2$ characteristic time and the proton density is uniform across all subdomains Ω_i.

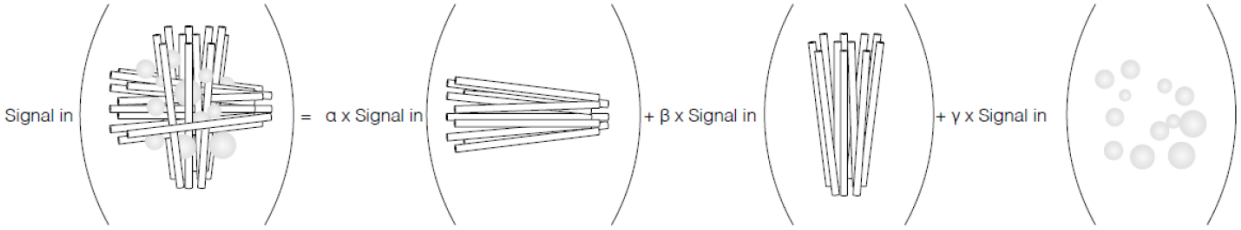

Figure 1.9: **Illustration of the superposition principle** [17]. A complex signal can be divided in different parts where the diffusion is similar. The total signal is equal to these different parts weighted by the fraction occupied by each of them.

1.5.2 Phenomenological models

The first category of models is the so-called phenomenological models. In those models, the DW-MRI signal is fitted with a mathematical expression containing parameters that do not directly represent the biological microstructure of the brain. They have simple mathematical formulations and their estimated parameters represent diffusion features. This means that they do not provide informations about the microstructural features like axon radii. The estimated parameters of phenomenological models are quantities like the principal direction of diffusion, the degree of anisotropy, etc. It is sometimes unclear what they biologically represent. The most known model of this category is the diffusion tensor imaging (DTI) model which fits to each voxel a diffusion tensor. This model is detailed in the next chapter.

1.5.3 Microstructural models

The second category of models comprises microstructural models which provide a more direct link between the brain microstructure and the measured signal. They usually are multi-compartment models and have a more complex mathematical formulation. A multi-compartment model is a model using the superposition principle by dividing the measured signal into several compartments, each of them containing a fraction of the total signal. Therefore, the measured DW-MRI signal is the sum of the signals S_i contributed by each population of water molecules multiplied by the fraction f_i of each population :

$$S = \sum_{i=1}^{N} f_i S_i \text{ such that } \sum_{i=1}^{N} f_i = 1 \ .$$

Another formulation is :

$$S = S_0 \sum_{i=1}^{N} f_i A_i \ ,$$

since the non diffusion-attenuated signal S_0 is the same for all compartment. A_i is the attenuation signal of the compartment i.

For all the A_i, a microstructural environment, where the water diffusion has a specific characteristic such as isotropic or hindered anisotropic diffusion, is assigned. The compartments can represent various microstructural configurations but generally have one compartment dedicated to the cerebrospinal fluid (CSF) where the diffusion of water is free and isotropic. Another compartment represents the axon where the diffusion is highly restricted in the direction perpendicular to the principal direction of the axon. There is also a compartment representing the extra-axonal space where the diffusion can be hindered.

In reality, the compartments are not completely impermeable which means that it is possible that water molecules move from one compartment to another. If the rate of exchange is slow compared to the diffusion time, the fraction of water molecules in each compartments can be assumed constant during the diffusion process. This assumption is called the *non-exchangeability of water molecules* and is generally made for multi-compartment models.

The principal advantage of microstructural models is that they provide a more intuitive insight of the brain microstructure. Unfortunately, they have a much more complicated mathematical formulation and strong assumptions have to be made in order to get an acceptable complexity. A good example of that kind of model is the NODDI model which is described in the next chapter.

Chapter 2

Presentation of the different models

This chapter presents the four diffusion MRI models that will be compared using the generalization error in the fourth chapter. They are developed from the simplest one to the most complex one, namely

- Diffusion Tensor Imaging (DTI),
- Multi-Fascicle Model (MFM),
- Neurite Orientation Dispersion and Density Imaging (NODDI),
- DIstribution of Anisotropic MicrOstructural eNvironments with Diffusion-weighted imaging (DIAMOND).

For each model, its limitations and advantages are exposed. The chapter ends with a summary table of the four models.

In this chapter, some biological terms concerning the neurons structure are used. The reader may refer to Appendix A for more details on these terms.

2.1 Diffusion Tensor Imaging (DTI)

The diffusion tensor imaging model [19] is the simplest and most widely used model of diffusion signal. Assuming homogeneous Gaussian diffusion, it models the average diffusion direction and strength at each voxel with a second-order tensor $\boldsymbol{D}$. This tensor is estimated from diffusion measurements in several directions. The Stejskal-Tanner equation links the measured diffusion-weighted signal S_k along a gradient direction $\boldsymbol{g}_k$ to the non-attenuated signal S_0 :

$$S_k(\boldsymbol{D}) = S_0 e^{-b_k \boldsymbol{g}_k^T \boldsymbol{D} \boldsymbol{g}_k} ,$$

where b_k is the b-value.

The tensor $\boldsymbol{D}$ can be interpreted as an ellipsoid the principal direction of which indicates where the diffusion is the strongest. The two orthogonal directions are where the diffusion is more restricted.

Mathematically speaking, an ellipsoid is represented by :

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

a 3×3 positive definite matrix. It turns out that for uncharged molecules such as water, $\boldsymbol{D}$ is symmetric[1]. Therefore, there are $p_{\mathrm{DTI}} = 6$ parameters[2] required to fully define the ellipsoid. The eigenvalues of $\boldsymbol{D}$, conventionally denoted $\lambda_1 \geq \lambda_2 \geq \lambda_3$, give the magnitudes of diffusion in the direction of the eigenvector associated, *i.e.* λ_i is the diffusion coefficient in the direction $\boldsymbol{e_i}$, the i-th eigenvector of $\boldsymbol{D}$.

It is therefore possible to infer the diffusion occurring in a voxel by looking at the shape of the ellipsoid. Free water, where the diffusion is isotropic, will give rise to a spherical ellipsoid while environments where the diffusion is highly anisotropic will give thin ellipsoids. Indeed, the diffusion coefficient is the same in all directions for isotropic diffusion whereas diffusion is favored along a principal direction for anisotropic diffusion.

To represent the parameters estimated by DTI into a scalar image, we commonly calculate the axial diffusivity (AD), the radial diffusivity (RD), the mean diffusivity (MD) and the fractional anisotropy (FA). The diffusion along the principal axis, λ_1, is called the axial diffusivity. The radial diffusivity is the average diffusion along the two others axis :

$$\mathrm{RD} = \frac{\lambda_2 + \lambda_3}{2} \ .$$

The mean diffusivity is the diffusion averaged in the three directions. It is calculated as followed :

$$\mathrm{MD} = \frac{1}{3}\mathrm{Tr}(\boldsymbol{D}) = \frac{\lambda_1 + \lambda_2 + \lambda_3}{3} \ .$$

The fractional anisotropy gives an idea of the degree of asymmetry between the diffusivity in the three directions. It is defined as :

$$\mathrm{FA} = \sqrt{\frac{3}{2}} \frac{\sqrt{(\lambda_1 - \mathrm{MD})^2 + (\lambda_2 - \mathrm{MD})^2 + (\lambda_3 - \mathrm{MD})^2}}{\sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}} \ .$$

FA is between 0 if the diffusion is completely isotropic and 1 if the diffusion is highly anisotropic, *i.e.* the water molecules can only move in one direction.

These DTI-based measures are computed in each voxel yielding to scalar images as shown on Figure 2.1.

These measures can be altered by a change of diffusion due to some pathologies. For example, when myelin sheaths are damaged, MD increases since myelin sheaths act like a barrier to diffusion. On the other hand, FA decreases because the main direction of diffusion is smaller. As a consequence, RA increases and DA decreases.

[1] $D_{ij} = D_{ji} \ \forall \ i, j \in \{x, y, z\}, \ i \neq j$.
[2] $D_{xx}, D_{xy}, D_{xz}, D_{yy}, D_{yz}$ and D_{zz}.

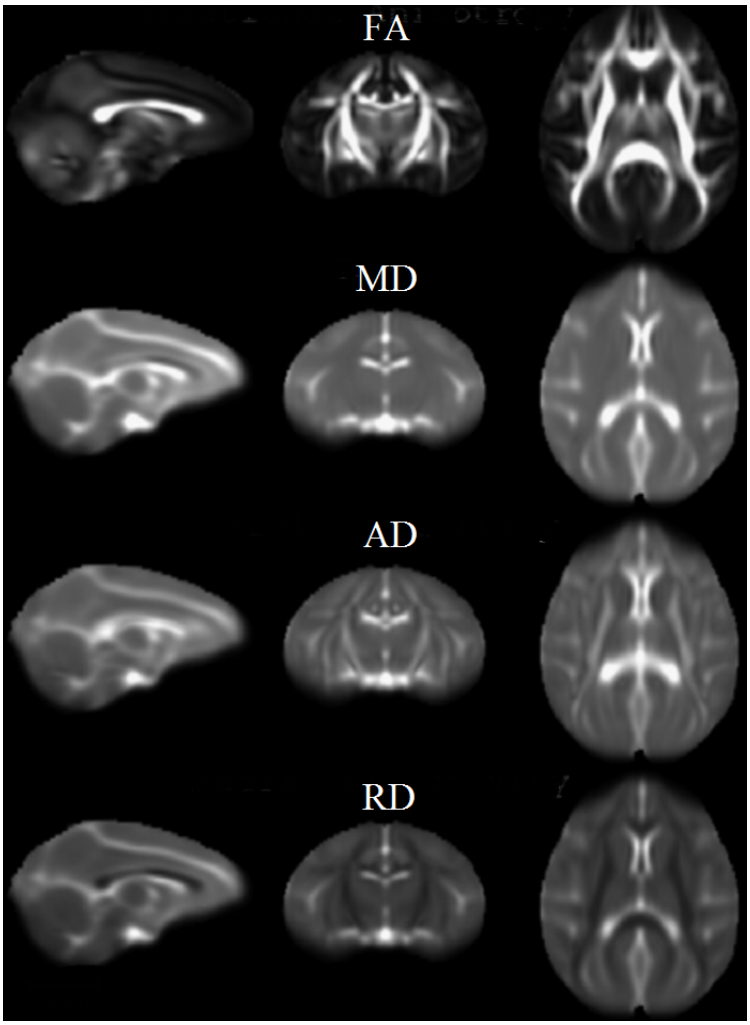

Figure 2.1: **DTI-based measures** [20]. Representation of images respectively plotted accorded to the Fractional Anisotropy, the Mean Diffusivity, the Axial Diffusivity and the Radial Diffusivity.

2.1.1 Limitations

The DTI model is simple but suffers from severe disadvantages. Indeed, it assumes that the diffusion at each voxel is Gaussian with either no preferential direction in the case of isotropic diffusion or one preferential direction for anisotropic diffusion. This is correct only if all axons in the voxel are into one fascicle with a specific orientation. If the voxel contains fascicles with heterogeneous orientations, the model is not able to predict the correct microstructure within that voxel. It can also lead to misinterpretations. For example, when two fascicles are crossing within the same voxel, DTI will fit a single wide tensor instead of two thinner as shown on Figure 2.2. This will decrease the fractional anisotropy and might be interpreted as altered myelin for the fascicle. In reality, a study shows that the orientation of the fascicles is heterogeneous in 60 to 90% of the voxels present in the white matter [21]. Therefore, the DTI model must be improved in order to detect several fascicles.

Another limitation of DTI, due to voxels that are at the interface between different tissues (like gray and white matter or between the cerebrospinal fluid and a tissue), is the partial volume effect. The diffusion signal arising from these different compartments will give a single average value. Since DTI assumes that the diffusion is purely restricted or purely unrestricted, the different compartments will conflate. This can give rise to a lower FA and thus a biological misinterpretation.

Finally, DTI being a phenomenological model, it does not directly represent the biological microstructure

of the brain.

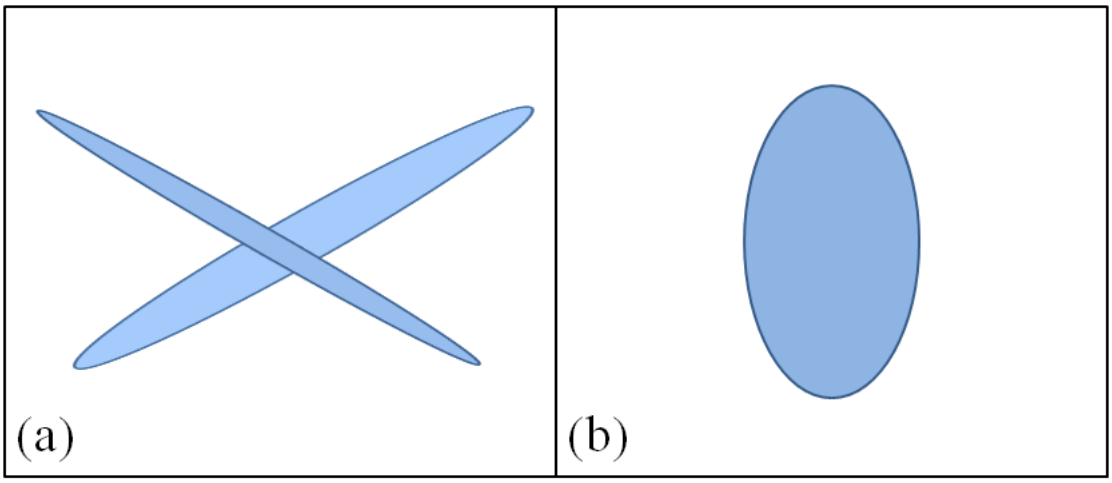

Figure 2.2: **Example of mistake made by the DTI model.** A voxel in which two fascicles are crossing (a) will be fitted by DTI with only one ellipsoid (b) representing the averaged diffusion within this voxel.

2.1.2 Advantages

As previously said, DTI is the simplest and most widely used model of diffusion signal. This is due to the fact that it only needs to estimate a single symmetric tensor, *i.e.* 6 parameters. So even for large protocol, the estimation does not require much calculation time.

2.2 Multi-Fascicle Model (MFM)

The multi-fascicle model [15] is a multi-compartment model which considers a mixture of independent fascicles with heterogeneous orientations at each voxel. To this, it adds an isotropic component to model the diffusion of unrestricted water as in the CSF. Making the assumption of non-exchangeability of water molecules, the diffusion signal in each voxel is modeled as a mixture of the diffusion signal arising from each individual fascicle plus the isotropic component. The diffusion-weighted signal $S_k(\boldsymbol{D}, \boldsymbol{f})$ along a gradient direction $\boldsymbol{g}_k$ can be described as followed :

$$S_k(\boldsymbol{D}, \boldsymbol{f}) = S_0(f_0 S_k^{\text{free\_water}} + \sum_{j=1}^{N_f} f_j S_{k,j}^{\text{fascicle}}) \ ,$$

where $\boldsymbol{D} = [\boldsymbol{D}_1, ..., \boldsymbol{D}_{N_f}]$ represents the diffusion tensors of the N_f fascicles, $S_0(f_0 S_k^{\text{free\_water}})$ is the diffusion signal arising from the unrestricted water and $S_{k,j}^{\text{fascicle}}$ is the diffusion signal arising from a single fascicle. Finally, $\boldsymbol{f} = [f_0, ..., f_{N_f}]$ describes the fraction of occupancy of each compartment and is such that :

$$\sum_{j=1}^{N_f} f_j = 1, \quad f_j \geq 0 \ \ \forall \ j \ .$$

The diffusion of free water is modeled assuming isotropic Gaussian displacements :

$$S_k^{\text{free\_water}} = e^{-b_k D_{iso}} \ ,$$

where D_{iso} is the diffusion coefficient for free water at 37 °C. Each of the N_f fascicle is modeled by a DTI model leading to :

$$S_{k,j}^{\text{fascicle}} = e^{-b_k \boldsymbol{g}_k^T \boldsymbol{D}_j \boldsymbol{g}_k} .$$

Diffusion parameters such as AD, RD, MD and FA can be computed for each fascicle independently. The number of parameters that needs to be estimated is[3] $p_{\text{MFM}} = 1 + 7N_f$. The estimation of the number of fascicles N_f is challenging. In [22], the authors determine the number of fascicles at each voxel by assessing the generalization error.

2.2.1 Limitations

Like the DTI model, the MFM model assumes a Gaussian diffusion within the voxels. The number of fascicles within a voxel is not known in advance, so the model has to be fitted with different values of N_f. To know which of these models best fits the signal, the generalization error is computed for each model. The lower the generalization error, the better the model [22]. The number of parameters that needs to be estimated is higher than for the DTI model, making the fitting of the MFM model slower. Furthermore, the model needs to be fitted several times and the generalization error computed.

2.2.2 Advantages

The MFM model is an improvement of the DTI model because it takes into account the signal coming from several fascicles in the same voxel. Since [21] suggests that at typical clinical spatial resolutions, voxels containing crossing fascicles could account for 60 to 90% of all voxels, the fact that MFM takes into account multiple fascicles is of great interest.

2.3 Neurite Orientation Dispersion and Density Imaging (NODDI)

NODDI, proposed in [23], is a two-level model. The first level consists of decomposing the contents of each voxel into two parts : the tissues and the cerebrospinal fluid. The second level decomposes the tissues into an intra-cellular part and an extra-cellular part. Therefore, there are three types of microstructural environment :

- The intra-cellular space which is the space bounded by the membrane of dendrites and axons, known collectively as neurites.
- The extra-cellular space which is the space around the neurites. It is occupied by various types of glial cells and somas in the grey matter.
- The cerebrospinal fluid (CSF) which is the main fluid of the brain.

Those compartments define different water diffusions and give rise to a separate signal. The full normalized signal S is written as :

$$S = (1 - f_{iso})(f_{ic} S_{ic} + f_{ec} S_{ec}) + f_{iso} S_{iso} ,$$

where

- S_{ic}, S_{ec} and S_{iso} are respectively the normalized signals coming from intra-cellular space, extra-cellular space and CSF.
- f_{iso}, f_{ic} and f_{ec} are respectively the fraction occupied by the CSF, the intra-cellular space and the extra-cellular space.

[3] f_0 for the unrestricted water and $f_i, D_{xx}^i, D_{xy}^i, D_{xz}^i, D_{yy}^i, D_{yz}^i, D_{zz}^i$ for each fascicle $i \in \mathbb{N}_0,\ i \leq N_f$.

Due to the two-level nature of NODDI,

$$f_{ec} + f_{ic} \triangleq 1 \ .$$

Therefore, only two fraction parameters are needed.

Intra-cellular model

In this compartment, the diffusion of water is highly restricted perpendicular to neurites and unhindered along them. It is therefore modeled as a set of sticks, *i.e.* cylinders of zeros radius.

$$S_{ic} = \int \rho(\boldsymbol{n}) e^{-bd_{\parallel}(\boldsymbol{g}\cdot\boldsymbol{n})^2} d\boldsymbol{n} \ ,$$

where $\boldsymbol{g}$ is the gradient orientation, b is the b-value and ρ is the orientation distribution function and is such that $\rho(\boldsymbol{n}) = \rho(-\boldsymbol{n})$ and $\int \rho(\boldsymbol{n}) = 1$. Therefore, $\rho(\boldsymbol{n})d\boldsymbol{n}$ gives the probability of finding sticks along orientation $\boldsymbol{n}$. The term $e^{-bd_{\parallel}(\boldsymbol{g}\cdot\boldsymbol{n})^2}$ gives the attenuation due to unhindered diffusion along a stick with intrinsic diffusivity $d_{\parallel}$ and orientation $\boldsymbol{n}$. The orientation distribution function is modeled with a Watson distribution [24] because it can capture the dispersion in orientation easily and provides a good representation either for high orientation dispersion as in gray matter or for low orientation dispersion as in white matter :

$$\rho(\boldsymbol{n}) = M(\frac{1}{2}, \frac{3}{2}\kappa)^{-1} e^{\kappa(\boldsymbol{\mu}\cdot\boldsymbol{n})^2} \ .$$

M is a confluent hypergeometric function, $\boldsymbol{\mu}$ is the mean orientation, *i.e.* the unit vector about which the distribution is cylindrically symmetric, and κ is the concentration parameter which controls the extent of orientation dispersion about $\boldsymbol{\mu}$.

Extra-cellular model

Due to neurites, the diffusion of water in the extra-cellular space is hindered but not restricted. The diffusion can therefore be modeled with a Gaussian anisotropic diffusion. The normalized signal for this compartment S_{ec} is such that :

$$S_{ec} = e^{-b\boldsymbol{g}^T \mathbb{D} \boldsymbol{g}} \ ,$$

where b is the b-value and $\boldsymbol{g}$ the gradient orientation. The apparent diffusion tensor $\mathbb{D}$ of the compartment is defined as :

$$\mathbb{D} = \int \rho(\boldsymbol{n}) \boldsymbol{D}(\boldsymbol{n}) d\boldsymbol{n} \ ,$$

where ρ is the distribution of axon orientation and $\boldsymbol{D}(\boldsymbol{n})$ is the diffusion tensor for axons with principal direction of diffusion $\boldsymbol{n}$. The diffusion coefficient parallel to $\boldsymbol{n}$ is $d_{\parallel}$ and $d_{\perp}$ is the one perpendicular to $\boldsymbol{n}$. The coefficient $d_{\parallel}$ is the same as in the intra-cellular compartment and $d_{\perp}$ is set to $d_{\parallel}(1 - f_{ic})$. This so-called *tortuosity model* is an important assumption of the model.

CSF model

As for the free water compartment in MFM, the CSF is modeled as an isotropic Gaussian diffusion with diffusivity D_{iso}. Therefore, the signal coming from this part of the brain is written as :

$$S_{iso} = e^{-bD_{iso}} \ ,$$

where D_{iso} is the diffusion coefficient for free water at 37 $^\circ$C and b is the b-value.

Orientation dispersion index

From the concentration parameter κ, we can define an orientation dispersion index (OD) as :

$$OD = \frac{2}{\pi} arctan\left(\frac{1}{\kappa}\right) .$$

This index lies between 0 and 1 and is therefore easy to visualize. In high orientation dispersion tissues like the gray matter, OD will be close to 1. On the contrary, in low orientation dispersion tissues like the white matter, it will be almost zero.

Number of parameters estimated

The number of parameters that needs to be estimated for the NODDI model is $p_{\text{NODDI}} = 5$:

- f_{iso} : the fraction occupied by the CSF in each voxel,
- f_{ic} : the intra cellular volume fraction in each voxel,
- κ : the concentration parameter,
- θ : angle of the fiber orientation (see Figure 2.3),
- φ : angle of the fiber orientation (see Figure 2.3).

The two angles parameters θ and φ are necessary to calculate the fiber direction. Figure 2.3 shows how these two angles are defined.

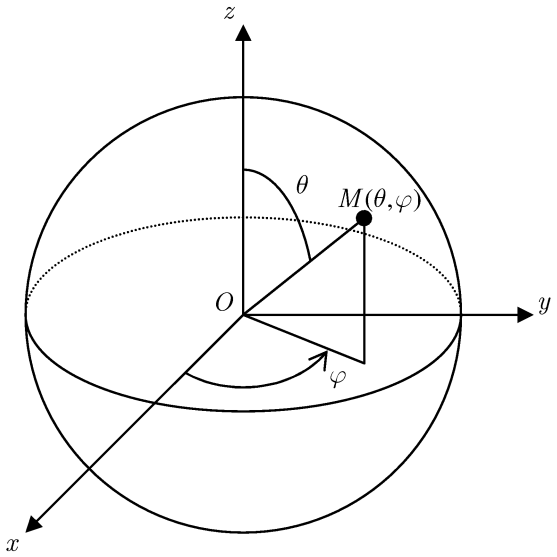

Figure 2.3: **Representation of the angles defining the fiber orientation** [25]. The figure illustrates the way the angles θ and φ are defined. The fiber is the segment $\vec{\text{OM}}$ and its length is $\|\vec{\text{OM}}\|$.

From this definition, the fiber direction can be calculated as follows :

$$\begin{cases} x &= \sin(\theta)\cos(\varphi) \\ y &= \sin(\theta)\sin(\varphi) \\ z &= \cos(\theta) \end{cases} .$$

2.3.1 Limitations

The NODDI model relies on a representation with a single white matter fascicle while fascicles with various microstructures have been observed such as in the body of the corpus callosum [26]. NODDI ignores the intra-axonal radial diffusivity and, similarly to the DTI model, considers only a single fascicle compartment per voxel while fascicles crossings occur a lot in the human brain. NODDI may capture crossing fascicles as increased dispersion but can't characterize each of them separately [26].

Furthermore, due to its much more complex mathematical formulation compared to DTI or MFM, the computational time is higher even if it has a relatively low number of parameters to estimate.

2.3.2 Advantages

The NODDI model is a multi-compartment microstructural model which provide a more direct link with the brain microstructure than a phenomenological model such as DTI. It focuses on the modeling of neurites which are key markers of brain development and aging [23].

2.4 DIstribution of Anisotropic MicrOstructural eNvironments with Diffusion-weighted imaging (DIAMOND)

Description

The DIAMOND model, proposed in [26], is another type of diffusion compartment model, *i.e.* a model which parameters reflect the tissue compartments present in each voxel. In this model, we consider measurements arising from a large number of 3D spin packets such that each homogeneous spin packet undergoes local anisotropic 3D Gaussian diffusion represented by a diffusion tensor $\boldsymbol{D}$. It proposes to model the measured diffusion signal S_k, for the b-value b_k and gradient direction $\boldsymbol{g}_k$, by integration of the contribution of all spin packets.

$$S_k = S_0 \int_{\boldsymbol{D}\in \mathrm{Sym}^+(3)} P(\boldsymbol{D}) e^{-b_k \boldsymbol{g}_k^T \boldsymbol{D} \boldsymbol{g}_k} d\boldsymbol{D} \ , \tag{2.1}$$

where S_0 is the signal without diffusion, $\mathrm{Sym}^+(3)$ is the set of 3×3 symmetric positive-definite (SPD) matrices and $P(\boldsymbol{D})$ is the matrix-variate distribution which gives the fraction of spin packets described by a 3D diffusivity $\boldsymbol{D}$ in the voxel.

If a voxel contains exactly one homogeneous microstructural environment whose diffusion tensor is $\boldsymbol{D}^0$, then the matrix-variate distribution is a delta function :

$$P(\boldsymbol{D}) = \delta(\boldsymbol{D} - \boldsymbol{D}^0) \ .$$

In that case, the DIAMOND model is equivalent to the DTI model.

If a voxel contains several homogeneous microstructural environments, then the matrix-variate distribution is a mixture of delta functions such as represented on Figure 2.4 (b).

In reality, each microstructural environment contains heterogeneities and is then best described by a population of spin packets which is modeled with a peak-shaped matrix-variate distribution centered around its expectation $\boldsymbol{D}^0$ and defined over the space of SPD matrices as illustrated on Figure 2.4 (c).

The DIAMOND model uses the matrix-variate Gamma (mv-Γ) distribution which is the generalization of the Wishart distribution[4] for non integer numbers of degrees of freedom. The formal definition of the mv-Γ distribution is the following :

[4] Actually, Wishart distribution is a generalization to multiple dimensions of the chi-squared distribution. Similarly, mv-Γ distribution is a generalization to multiple dimensions of the gamma distribution.

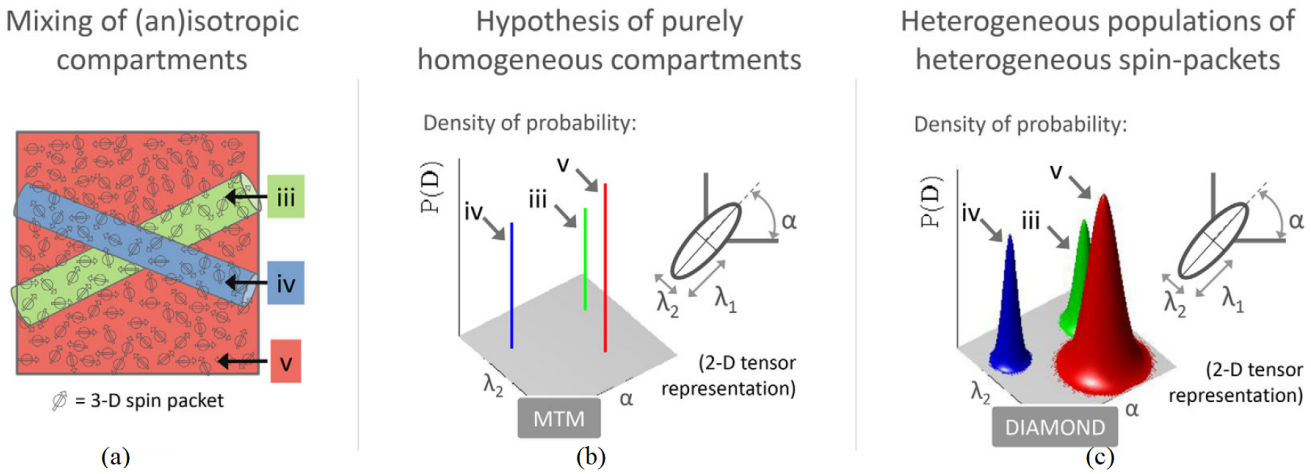

Figure 2.4: **Illustration of microstructural environments in the DIAMOND model** [26].
(a) Mixing of (an)isotropic compartments : Example of multi-compartment model with a mix of a isotropic (red) and two anisotropic (green and blue) compartments.
(b) Hypothesis of purely homogeneous compartments : If the compartments are homogeneous with no exchange, the signal arising from (a) is describe by a multi-tensor model.
(c) Heterogeneous populations of heterogeneous spin-packets : In reality, compartments have some degree of heterogeneity. Therefore, the DIAMOND model uses peak-shaped distributions of multi-dimensional diffusivities to model the heterogeneity.

A $p \times p$ SPD random matrix $\boldsymbol{D} \in \mathrm{Sym}^+(p)$ follows a mv-Γ distribution with shape parameter $\kappa > (p-1)/2$ and scale parameter $\Sigma \in \mathrm{Sym}^+(p)$ if it has a density

$$P_{\kappa,\Sigma}(\boldsymbol{D}) = \frac{|\boldsymbol{D}|^{\kappa-(p+1)/2}}{|\Sigma|^{\kappa}\Gamma_p(\kappa)} e^{\mathrm{Tr}(\Gamma^{-1}\boldsymbol{D})} , \tag{2.2}$$

where Γ_p is the multi-variate gamma function :

$$\Gamma_p(\kappa) = \pi^{p(p-1)/4} \prod_{j=1}^{p} \Gamma(\kappa - \frac{j-1}{2}) ,$$

and Γ is the usual gamma function :

$$\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} dx \quad \text{with } \Re(z) > 0 .$$

The expectation and mode of the density 2.2 are respectively $\boldsymbol{D}^0 = \kappa\Sigma$ and $\boldsymbol{M}^0 = \boldsymbol{D}^0(\kappa-1)/\kappa$.

The concentration is determined by the shape parameter κ, *i.e.* for $\boldsymbol{D}^0$ constant, the density 2.2 is more concentrated around $\boldsymbol{D}^0$ if κ increases. Therefore, a distribution with a sharp peak indicates a population with homogeneous microstructure whereas a distribution with a broad peak indicates a population with heterogeneous microstructure.

Considering N_p populations of spin packets in slow exchange in a voxel and representing the composition of each population with a mv-Γ of spin packets $P_{\kappa_j,\Sigma_j}(\boldsymbol{D})$ gives the following expression for the matrix-variate distribution :

$$P(\boldsymbol{D}) = \sum_{j=1}^{N_p} f_j P_{\kappa_j,\Sigma_j}(\boldsymbol{D}) \quad \text{with} \quad \sum_{j=1}^{N_p} f_j = 1 , \tag{2.3}$$

where f_j, κ_j and Σ_j are respectively the MR-visible fraction of occupancy, the shape parameter and the scale parameter of spin packet j.

The combination of Equations 2.1 and 2.3 gives :

$$S_k = S_0 \sum_{j=1}^{N_p} f_j \int_{\boldsymbol{D}\in \mathrm{Sym}^+(3)} P_{\kappa_j,\Sigma_j}(\boldsymbol{D}) e^{-b_k \boldsymbol{g}_k^T \boldsymbol{D} \boldsymbol{g}_k} d\boldsymbol{D} \ .$$

It turns out that these integrals are Laplace transforms of $P_{\kappa_j,\Sigma_j}(\boldsymbol{D})$ and have known analytical expressions [26]. The final expression for the DW signal with b-value b_k and gradient direction $\boldsymbol{g}_k$ is the following:

$$\begin{aligned} S_k &= S_0 \sum_{j=1}^{N_p} f_j \left(1 + \frac{b_k \boldsymbol{g}_k^T \boldsymbol{D}_j^0 \boldsymbol{g}_k}{\kappa_j}\right)^{-\kappa_j} \\ &= S_0 \sum_{j=1}^{N_p} f_j \exp\left(-\kappa_j \log\left(1 + \frac{b_k \boldsymbol{g}_k^T \boldsymbol{D}_j^0 \boldsymbol{g}_k}{\kappa_j}\right)\right) \\ &= \sum_{j=1}^{N_p} f_j \mathcal{D}\left(\boldsymbol{D}_j^0, \kappa_j\right) \ , \end{aligned}$$

where $\boldsymbol{D}_j^0$ is the expectation of the j^{th} mv-Γ distribution and $\mathcal{D}\left(\boldsymbol{D}_j^0, \kappa_j\right)$ is the non-monoexponential decaying signal arising from spin packets described by $P_{\kappa,\Sigma}$.

Incorporation of the different types of diffusion

The last equation is a generic expression of the DW signal arising from heterogeneous populations of spin packets in each voxel. It is possible to incorporate informations about the type of diffusion into the model.

- Free diffusion can be modeled with a mv-Γ distribution $P_{\kappa_{\text{free}},\Sigma_{\text{free}}}$ with isotropic mode

$$\boldsymbol{M}_{\text{free}}^0 = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{pmatrix} \times 10^{-9} \quad \left[\text{m}^2 \ \text{s}^{-1}\right] \ ,$$

  and two free parameters : the shape parameters κ_{free} and the volume fraction f_{free}.

- Isotropic restricted diffusion can be modeled with a mv-Γ distribution $P_{\kappa_{\text{iso,r}},\Sigma_{\text{iso,r}}}$ with three free parameters : the isotropic mode $M_{\text{iso,r}}^0$, the shape parameter $\kappa_{\text{iso,r}}$ and the volume fraction $f_{\text{iso,r}}$.

- Finally, anisotropic restricted and hindered diffusion can be both represented by a single mv-Γ distribution with anisotropic cylindrical $\boldsymbol{M}_j^0$, shape parameter κ_j and volume fraction f_j. Index j represents the fascicle that hindered the diffusion.

Putting all together gives the following signal generation model :

$$S_k = f_{\text{free}} \mathcal{D}\left(\boldsymbol{D}_{\text{free}}^0, \kappa_{\text{free}}\right) + f_{\text{iso,r}} \mathcal{D}\left(\boldsymbol{D}_{\text{iso,r}}^0, \kappa_{\text{iso,r}}\right) + \sum_{j=1}^{N_f} f_j \mathcal{D}\left(\boldsymbol{D}_j^0, \kappa_j\right) \ ,$$

where $N_f = N_p - 2$ is the number of fascicle in the voxel. Therefore, the DIAMOND model has $p_{\text{DIAMOND}} = 6N_f + 4$ free parameters[5].

The parameters at each voxel of the DIAMOND model are estimated using a maximum a posteriori approach [26].

[5] One for $\boldsymbol{M}_{\text{iso,r}}^0$, $N_f + 2$ concentration parameters κ, $N_f + 1$ volume fraction f and four for each $\boldsymbol{D}_j^0$ with $j = 1, 2, ..., N_f$.

2.4.1 Limitations

The DIAMOND model is the more complex of the four models. Similarly to MFM, the number of fascicles within a voxel is not known in advance : the model needs to be fitted for different values of N_f and the generalization error computed for these different models. Another limitation is that the echo time TE is assumed constant.

2.4.2 Advantages

The DIAMOND model extends the tensor model by replacing the matrix $\boldsymbol{D}$ by a matrix-variate gamma distribution over matrices $\boldsymbol{D}$ with a mean $\boldsymbol{D}^0$. This mean remains a tensor but the variability around represents the heterogeneity in microstructural properties. Doing so enables DIAMOND to simultaneously account for multiple sources of heterogeneity.

2.5 Summary of the four models

To conclude this chapter on the different models that will be used in the remainder of this document, Table 2.1 presents a summary of the models explained.

Table 2.1: **Summary of the four models.**

	DTI	**MFM**	**NODDI**	**DIAMOND**
Parameters	6	$1 + 7N_f$	5	$4 + 6N_f$
Type of model	Phenomenological	Phenomenological	Microstructural	Phenomenological
Compartments	- One fascicle : either *Gaussian isotropic diffusion* or *Gaussian anisotropic diffusion*.	- Independant fascicles with heterogeneous orientations : *Gaussian anisotropic diffusion*. - CSF : *Gaussian isotropic diffusion*.	- Intra-celluar : cylinders of zeros radius whose orientation distribution function is modeled by a *Watson distribution*. - Extra-cellular : *Gaussian anisotropic diffusion*. - CSF : *Gaussian isotropic diffusion*.	- 3D spin packets modeled by a *matrix-variate Gamma distribution*. - *Free diffusion*. - *Isotropic restricted diffusion*. - *Anisotropic restricted* and *hindered diffusion*.
Advantages	- Very simple. - Fast.	- Several fascicles.	- Orientation distribution function. - Direct link with the brain microstructure. - More compartments.	- Accounts for heterogeneity within each compartment. - Enables compartment specific properties to be estimated.
Limitations	- Gaussian diffusion. - Only one fascicle. - Partial volume effect.	- Gaussian diffusion. - N_f unknown.	- Only one fascicle. - Ignores intra-axonal radial diffusivity.	- N_f unknown. - More complex. - *TE* constant.

Chapter 3

The generalization error

This chapter explains the metric used in this work to compare the different models that have been explained in the previous chapter. This metric is called the *generalization error* and in practice, must be estimated.

The first section defines the generalization error and explains why it needs to be estimated. The second section presents the different estimators and discuss their bias.

3.1 The generalization error

Each diffusion model has inputs such as the b-values and the gradient directions and a unique output : the signal attenuated due to diffusion. In practice, different b-values are combined with different gradient directions and other MRI parameters to obtain a lot of signals in order to infer the brain microstructures. Let n be the number of observed data points and $z_i = (x_i, y_i)$, $i = 1, ..., n$ a particular data point where x_i contains the inputs and y_i the output. The set of the observed data points, called *training set*, is $\boldsymbol{z} = \{z_1, ..., z_n\}$. This set is used to build a generative model $r_{\boldsymbol{z}}(x)$ that tries to predict the output y from an input x. Ideally, the best model would minimize the error made on a new data point $z_0 = (x_0, y_0) \notin \boldsymbol{z}$. This error is called the *generalization error* and is defined as followed :

$$E_g|\boldsymbol{z} = \mathbb{E}_{z_0 \sim F}\left[|y_0 - r_{\boldsymbol{z}}(x_0)|^2|\boldsymbol{z}\right] \ , \tag{3.1}$$

where

- $E_g|\boldsymbol{z}$ is the generalization error conditional on the observed data $\boldsymbol{z}$,
- $\mathbb{E}[\ .\ |\ .\ ]$ is the conditional statistical expectation,
- $z_0 \sim F$ indicates that the statistical expectation is taken over a new data point z_0 that follows the distribution F.

Here the training set $\boldsymbol{z}$ is fixed. To account for the variability of the training set, the unconditional generalization error can be defined as the statistical expectation of Equation 3.1 over all $\boldsymbol{z}$:

$$\begin{aligned} E_{g,n} &= \mathbb{E}_{z_i \overset{iid}{\sim} F}[E_g|\boldsymbol{z}] \\ &= \mathbb{E}_{z_i \overset{iid}{\sim} F}\Big[\mathbb{E}_{z_0 \sim F}\left[|y_0 - r_{\boldsymbol{z}}(x_0)|^2|\boldsymbol{z}\right]\Big] \ , \end{aligned} \tag{3.2}$$

where z_i's are independent and identically distributed (iid) random variables and n indicates the number of these z_i's used to optimize the model $r_{\boldsymbol{z}}$.

It turns out that Equations 3.1 and 3.2 can't be directly computed because the distribution F is unknown. Therefore, they have to be estimated. Below we describe 5 estimators of the generalization error : the fitting error, the K-fold cross-validation error, the leave-one-out cross-validation error, the bootstrap error and finally the .632 boostrap error.

3.2 Estimators of the generalization error

3.2.1 The fitting error

A first estimator of the generalization error can be computed if the distribution $F(z)$ is estimated with the following empirical distribution :

$$\hat{F}(z_i) = \frac{1}{n} \ \forall \ z_i \in \boldsymbol{z} \ .$$

Therefore, Equation 3.1 becomes :

$$\begin{aligned} \hat{\mathrm{E}}_g^{fit} &= \mathbb{E}_{z_0 \sim \hat{F}} \left[|y_0 - r_{\boldsymbol{z}}(x_0)|^2 | \boldsymbol{z} \right] \\ &= \frac{1}{n} \sum_{i=1}^{n} |y_i - r_{\boldsymbol{z}}(x_i)|^2 \ . \end{aligned}$$

This estimator is known as the *fitting error*. It is a biased estimator of E_g. In fact, the data $\boldsymbol{z}$ are used both to optimize the parameters of the model $r_{\boldsymbol{z}}$ and to estimate its error $\hat{\mathrm{E}}_g^{fit}$. Furthermore, the fitting error consistently decreases with model complexity, dropping to zero if we increase the model complexity enough [27]. This estimator is thus downward biased. Other estimators of E_g need to be investigated.

3.2.2 *K*-fold cross-validation error

The *cross-validation* (CV) is the simplest and most widely used method for estimating the prediction error. Ideally, we would like to have a validation set from which we don't fit the model and used it to assess the performance of our prediction model. Unfortunately, the data are often scarce so we can't have that kind of validation set. The K-fold cross-validation uses one part of the data to fit the model and another different part to test it. This is done by splitting the data into K roughly equal-sized parts. For each group k, it fits the model on the remaining $K-1$ groups and test it on the k^{th} group.

Let $\kappa(i)$ be an indexing function that indicates the partition to which observation i is allocated.

$$\kappa : \{1, ..., n\} \mapsto \{1, ..., K\} \ .$$

Let $r_{\boldsymbol{z}_{(-k)}}$ be the fitted function computed with the k^{th} part of the data removed. The cross-validation estimate of the prediction error is therefore the following average :

$$\hat{\mathrm{E}}_g^{CV} = \frac{1}{n} \sum_{i=1}^{n} |y_i - r_{\boldsymbol{z}_{(-\kappa(i))}}(x_i)|^2 \ .$$

The result is an unbiased estimate of $\mathrm{E}_{g,n-\frac{n}{K}}$. For large n, $\hat{\mathrm{E}}_g^{CV}$ as an estimator of $\mathrm{E}_{g,n}$ has a low positive bias.

3.2.3 The leave-one-out cross validation error

The *leave-one-out cross validation* (LOOCV) is a special case of the K-fold cross-validation when $K = n$. In that case, $\kappa(i) = i$ meaning that for the i^{th} observation, the model is fitted with the full data except the i^{th}. Therefore, the leave-one-out cross-validation estimate is simplified in :

$$\hat{\mathrm{E}}_g^{LOOCV} = \frac{1}{n} \sum_{i=1}^{n} |y_i - r_{\boldsymbol{z}_{(-i)}}(x_i)|^2 \ .$$

The LOOCV estimator is an unbiased estimate of $\mathrm{E}_{g,n-1}$ since it uses $n-1$ points. For large n, $\hat{\mathrm{E}}_g^{LOOCV}$ as an estimator of $\mathrm{E}_{g,n}$ has a low positive bias. However, its variance is high because the

n subsets are very similar to one another. From a computational point of view, this method is not optimal since it requires to fit the model n times.

3.2.4 The bootstrap error

Another estimator of E_g as in Equation 2.3 is the *bootstrap* (BS). This technique lowers the variance of the cross-validation estimator by using bootstrap resampling. In bootstrap resampling, B bootstrap samples are created. In each of these samples, n data points are randomly picked with replacement among $\boldsymbol{z}$. This leaves B bootstrap datasets $\boldsymbol{Z}_1, \boldsymbol{Z}_2, ..., \boldsymbol{Z}_B$, each of them of size n. The bootstrap then fits the model to each of the bootstrap datasets. An illustration of the bootstrap method is shown on Figure 3.1.

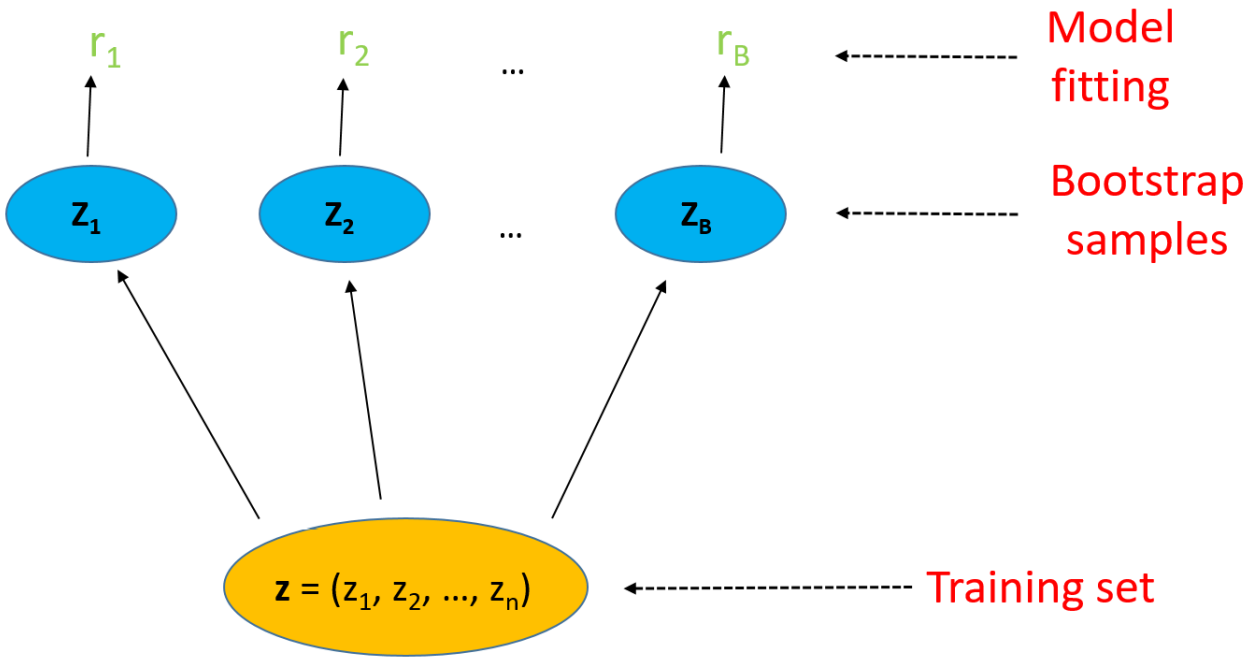

Figure 3.1: **Illustration of the bootstrap method**. From the initial training set, B bootstrap samples are created by randomly picking n z_i's with replacement from the training set. Then the model is fitted to these B new samples. Figure inspired from [27] §7.11.

Finally, the bootstrap estimator is calculated as followed :

$$\hat{\mathrm{E}}_g^{BS} = \frac{1}{n}\frac{1}{B}\sum_{i=1}^{n}\sum_{b=1}^{B}|y_i - r_b(x_i)|^2 \,,$$

where $r_b(x_i)$ is the model fitted with the b^{th} bootstrap sample and evaluated in x_i.

Unfortunately, this is not a particularly good estimator because the evaluated point x_i can be contained in the b^{th} bootstrap sample $\boldsymbol{Z}_b$ used to fit the model r_b, leaving $r_b(x_i)$ biased. This can be improved by the *leave-one-out booststrap* which uses the same trick as the LOOCV :

$$\hat{\mathrm{E}}_g^{LOOBS} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{|C^{-i}|}\sum_{b\in C^{-i}}|y_i - r_b(x_i)|^2 \,,$$

where C^{-i} is the set of indices of the bootstrap samples that do not contain observation i, and $|C^{-i}|$ is the number of such samples. In this way, if the data x_i is in the bootstrap sample $\boldsymbol{Z}_b$, the value of $r_b(x_i)$ won't be used in the computation of $\hat{\mathrm{E}}_g^{LOOBS}$.

Like the LOOCV estimator, the bootstrap estimator is upward biased because is relies on fewer points than the number of available samples n. This results from sampling with replacement. The average number of distinct observations in each bootstrap sample is about $0.632n$ (see Appendix B).

3.2.5 The .632 bootstrap error

Another estimator called the *.632 bootstrap* has been designed to lower the bias of the leave-one-out bootstrap. It is defined as :

$$\hat{\mathrm{E}}_g^{.632} = 0.368\ \hat{\mathrm{E}}_g^{fit} + 0.632\ \hat{\mathrm{E}}_g^{LOOBS}\ .$$

It combines the downward biased from the fitting error with the upper biased from the bootstrap error.

More details about the .632 coefficient

The coefficient .632 comes from a theoretical argument showing that the bootstrap samples used in computing $\hat{\mathrm{E}}_g^{LOOBS}$ are farther away on average than $\hat{\mathrm{E}}_g^{fit}$ by roughly a factor of 1/0.632. The adjustment of $\hat{\mathrm{E}}_g^{.632}$ corrects for this and makes it roughly unbiased for the true generalization error [28]. In his article [29], Efron himself says that the .632 coefficient has a weak motivation. However, he tries to explain it with the distance of the point to be predicted from the training set.

The fitting error is an error rate for points zero distance from the training set. But in the generalization error, the new point z_0 may lie some distance away from $\boldsymbol{z}$. Therefore, if the generalization error increases as the point being predicted moves away from $\boldsymbol{z}$, the fitting error underestimates the generalization error. As proved in Appendix B, the probability that the a point is zero distance from the training set for the bootstrap error is 0.632. Then, he proves that the points z_0 contributing to $\hat{\mathrm{E}}_g^{LOOBS}$ are on average 1/.632 too far from $\boldsymbol{z}$. To correct this, we have to multiply $\hat{\mathrm{E}}_g^{LOOBS}$ by .632 and hence multiply $\hat{\mathrm{E}}_g^{fit}$ by $1 - .632 = .368$.

Chapter 4

Presentation of the results and comparison between the different models

During all this document, we talk about four models : DTI, MFM, NODDI and DIAMOND. Since the complex code of DIAMOND is not yet published online, its author Benoît Scherrer kindly agreed to run his code on our data. It turns out that when running the DIAMOND model, he is also able to get the result for the MFM model, from which he is the co-author. Unfortunately, for reasons beyond our control, the MFM and DIAMOND codes did not work. The results presented in this chapter are therefore those of the DTI and NODDI models.

In this document, we choose to analyze the effect of two PGSE parameters on the generalization error : the effect of the diffusion time Δ and the one of the gradient norm $\|\boldsymbol{G}\|$. For each of these parameters, the results for each model, a comparison between them and a summary are presented. We also study the effect of mixing several shells belonging to the same multi-shell HARDI together. As a reminder, a multi-shell HARDI is a set of shells wherein only $\|\boldsymbol{G}\|$ varies, Δ and δ being fixed. Finally, we end with the analysis of the effect of the subprotocols size on the generalization error.

4.1 Presentation of the data

The generalization error has been calculated on two different data sets coming from two different parts of the brain :

- The *genu* is the anterior end of the corpus callosum which is the largest white matter structure in the brain connecting the left and right cerebral hemispheres. The genu is bent downward and backward and the fibers within it are approximately straight and parallel.

- The *fornix* is a C-shaped bundle of nerve fibers in the brain. Its name in latin means *arch*. In this region of the brain, the configuration of the fibers is more complex.

Each of these two regions contains 6 voxels. The study of these regions is interesting because the configurations of the fibers are different. We expect that the signals coming from the voxels of the genu are homogeneous due to the straight and parallel fibers contained in it. Conversely, we expect that the signals coming from the voxels of the fornix are more heterogeneous because of the complexity of the fibers in this region.

The whole data set comes from the *White Matter Modelling Challenge* and all the details are presented in Appendix C. From this large protocol, 40 subprotocols were invented in order to compute the generalization error for each of them. These subprotocols are presented in Appendix D.

The data contain the signal from in-vivo human brain white matter acquired on the *Connectom scanner* [30] at the Massachusetts General Hospital. This scanner uses gradients strength up to 300 mT/m which is a major stepping stone towards DW-MRI data that is more sensitive to microstructural anatomy and organization of brain tissues[31]. Indeed, studies like [32] illustrate the benefits of stronger gradients on animals to map the axon diameters. This was confirmed on live human subjects for the first time by [33] on the Connectom scanner. This rich data set offers a great opportunity for the comparison of DW-MRI models, over a very wide range of the measurement space.

For the results presented in this chapter, the generalization error has been calculated with a number of random bootstrap samples equal to 150. This is a good compromise between accuracy and simulation time since the NODDI model is very low. In order to fairly compare the models, we fixed the randomness, *i.e.* we generated 150 random bootstrap samples once and we used them for all the models. Furthermore, to interpret the results, we normalized the generalization error by dividing it by the average S_0 (signal without diffusion) for each voxel. Indeed, since the signal tends to decrease when TE increases (*i.e.* when δ or Δ increases), the errors may be smaller simply because the measured scale has changed. This normalization allows us to fairly compare the subprotocols between them. Finally, since the errors are of the same order of magnitude for the 6 voxels, we just average to allow better visibility on graphs.

4.2 Effect of the diffusion time on the generalization error

We first examine the impact of the diffusion time Δ on the generalization error. For that, the two parameters δ and $\|\boldsymbol{G}\|$ are fixed while Δ varies. The subprotocols used to examine the effect of the diffusion time are presented in Appendix D.1. We start by presenting the results when $\delta = 3$ ms and $\|\boldsymbol{G}\| = 63$ mT/m for the 6 subprotocols containing only one shell, the 3 subprotocols containing 2 shells and the 2 subprotocols containing 3 shells. Then the results when $\delta = 8$ ms and $\|\boldsymbol{G}\| \simeq 60$ mT/m[1] for the 5 subprotocols containing only one shell and the 2 subprotocols containing 2 shells are presented. This is reproduced for both genu and fornix data.

4.2.1 Genu

Figure 4.1 presents the evolution of the normalized generalization error made by the DTI and NODDI models in the region of the genu as the diffusion time varies. The two other parameters are fixed to $\delta = 3$ ms and $\|\boldsymbol{G}\| = 63$ mT/m. Linear regressions are made to highlight the trends.

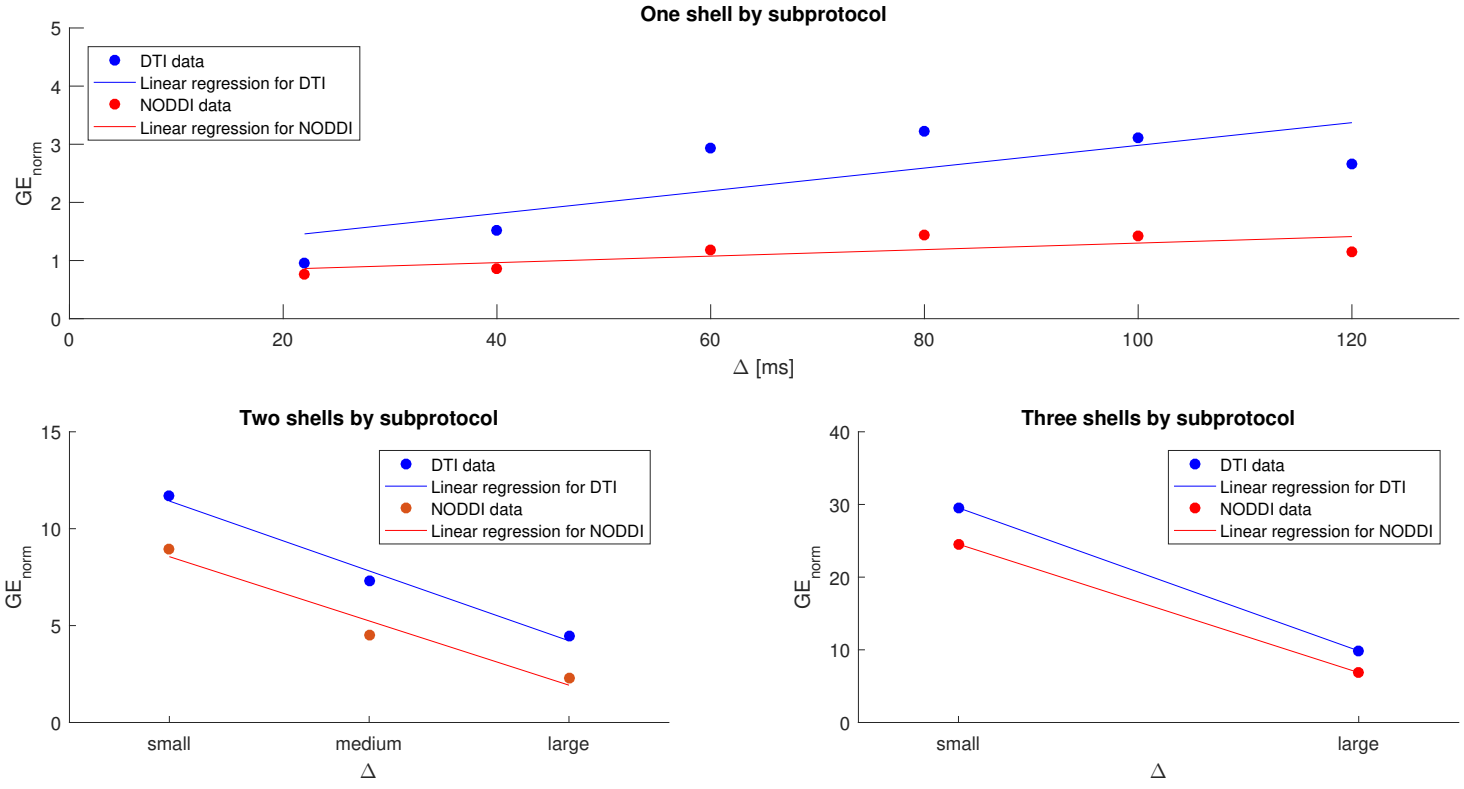

Figure 4.1: **Effect of the diffusion time for the DTI and NODDI models in the genu with $\delta = 3$ ms.**
Top : Generalization error of the six first subprotocols containing one shell.
Bottom left : Generalization error of subprotocols N° 12 (2 shells with small Δ), 13 (2 shells with medium Δ) and 14 (2 shells with large Δ).
Bottom right : Generalization error of subprotocols N° 17 (3 shells with small Δ) and 18 (3 shells with large Δ).

We first observe that the GE made by the DTI model are always greater than the ones made by the NODDI model. For subprotocols containing one shell, the GE seems to increase with the diffusion

[1]Indeed, the five gradient norms don't have the same value but have a mean of $(58 + 59 + 61 + 60 + 60)/5 = 59.6$ mT/m.

time, whereas it decreases for subprotocols containing two and three shells. The last observation is that the GE increases with the number of shell by subprotocol. All of this is valid for both the DTI and NODDI models.

Figure 4.2 presents the normalized generalization error made by the DTI and NODDI models in the region of the genu with $\delta = 8$ ms and $\|\boldsymbol{G}\| \simeq 60$ mT/m.

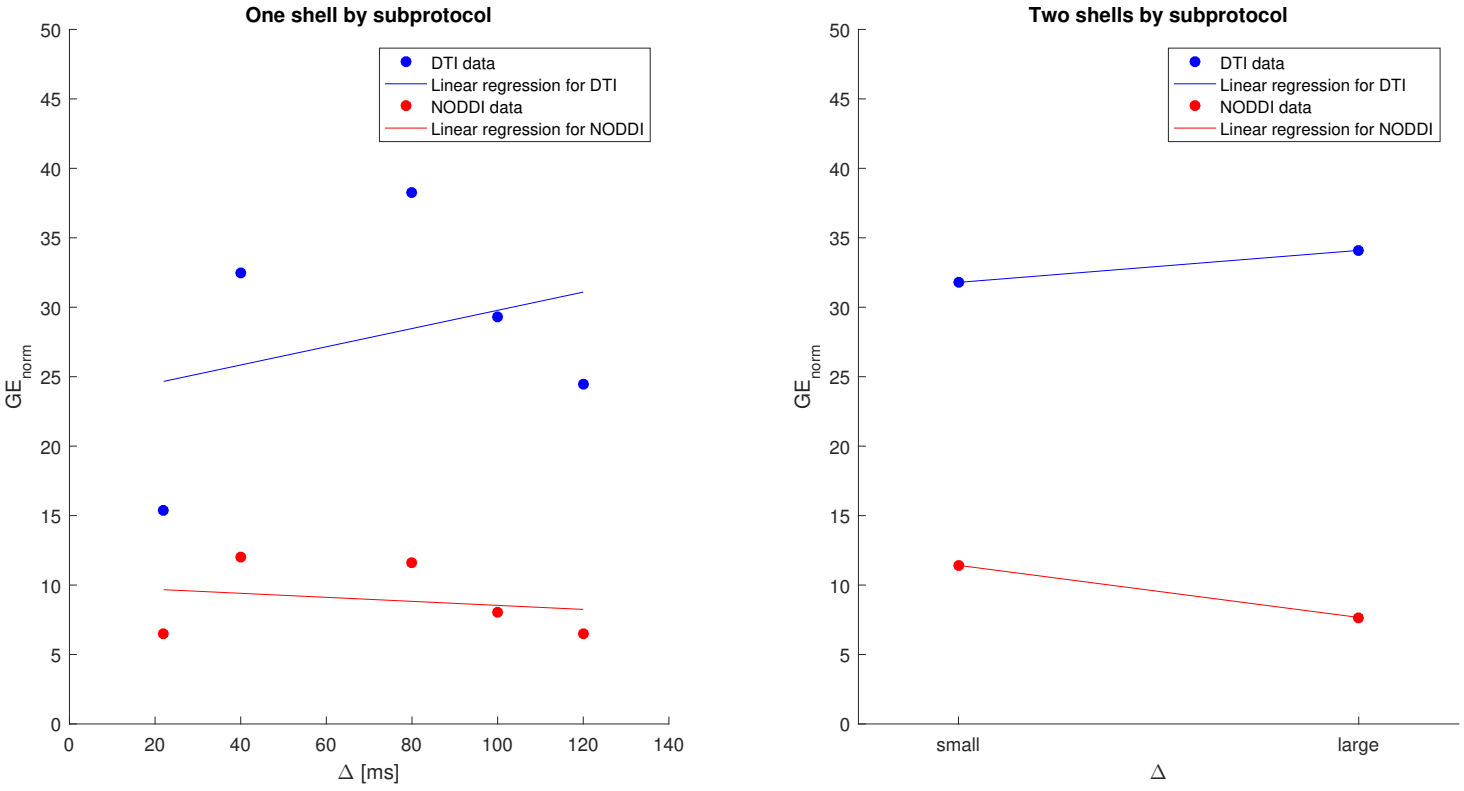

Figure 4.2: **Effect of the diffusion time for the DTI and NODDI models in the genu with $\delta = 8$ ms.**
Left : Generalization error of the five subprotocols containing one shell (subprotocol N° 7-8-9-10-11). Right : Generalization error of subprotocols N° 15 (2 shells with small Δ) and 16 (2 shells with large Δ).

The observation that the GE is greater for the DTI model than for the NODDI model is confirmed when $\delta = 8$ ms. The errors when $\delta = 8$ ms are bigger than when $\delta = 3$ ms. As in the previous figure, the GE of the DTI model increases with the diffusion time for subprotocols containing one shell. This is not the case for the NODDI model anymore. Its error decreases with the diffusion time. For subprotocols containing two shells, the error increases for the DTI model whereas it decreases for the NODDI model.

Finally, Table 4.1 shows the normalized generalization error for the last two subprotocols (N° 19 and 20) containing respectively 6 and 5 shells.

Table 4.1: **Normalized generalization error made by the DTI and NODDI models in the region of the genu for subprotocols N° 19 and 20.**

Subprotocol N°	**δ [ms]**	GE_{norm}	
		DTI	*NODDI*
19	3	79.07	57.18
20	8	169.2	44.79

Again, the normalized GE is bigger for subprotocols containing more shells. The only exception is for the NODDI model for subprotocols N° 19 and 20. There are more shells in subprotocol N° 19 than in N° 20 and the GE is bigger for N° 20. However, these two are not really comparable since they don't have the same δ. As before, the GE for the DTI model is greater than for the NODDI model.

4.2.2 Fornix

Figure 4.3 presents the evolution of the normalized generalization error made by the DTI and NODDI models in the region of the fornix as the diffusion time varies and for $\delta = 3$ ms.

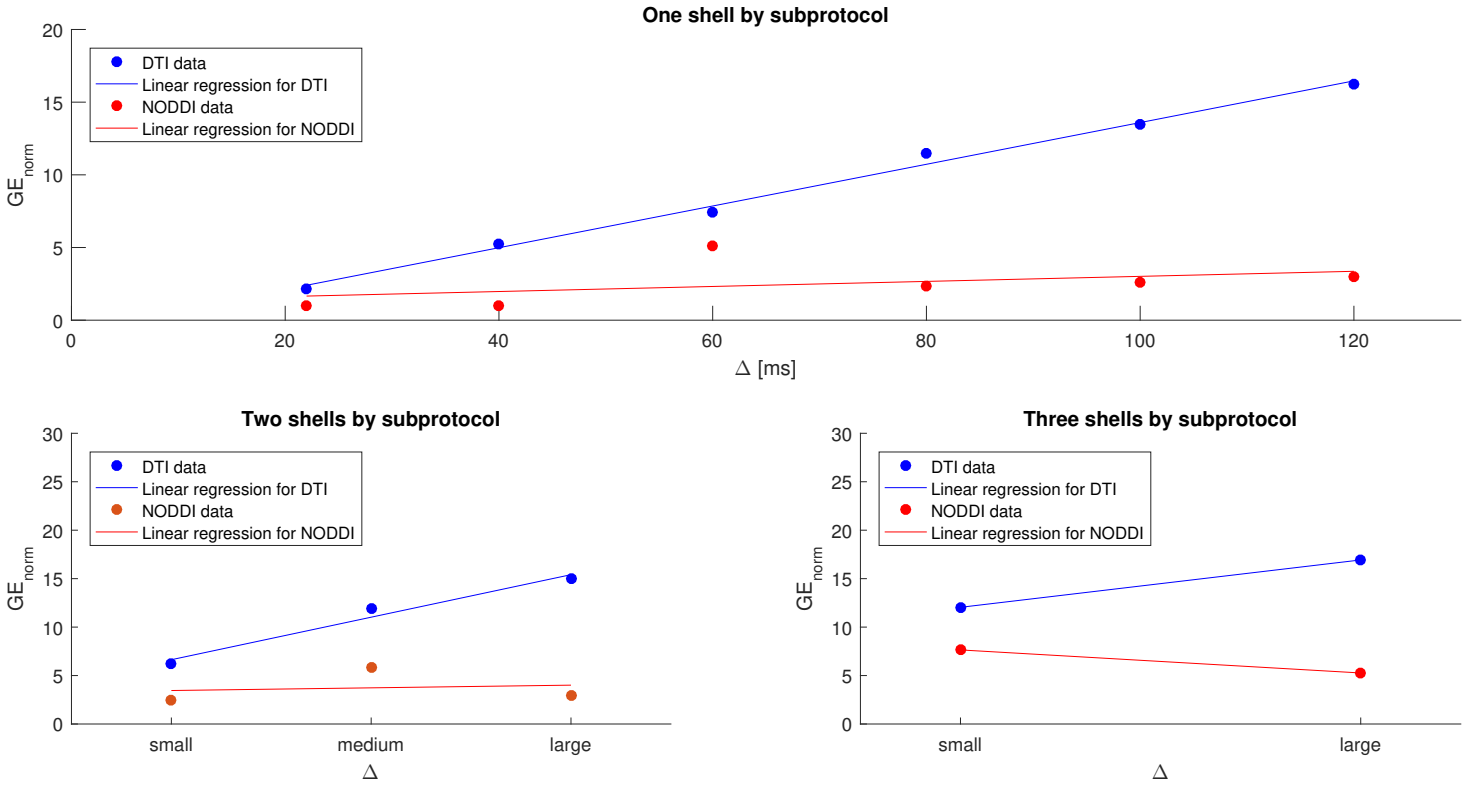

Figure 4.3: **Effect of the diffusion time for the DTI and NODDI models in the fornix with $\delta = 3$ ms.**
Top : Generalization error of the six first subprotocols containing one shell.
Bottom left : Generalization error of subprotocols N° 12 (2 shells with small Δ), 13 (2 shells with medium Δ) and 14 (2 shells with large Δ).
Bottom right : Generalization error of subprotocols N° 17 (3 shells with small Δ) and 18 (3 shells with large Δ).

For the DTI model, whatever the number of shell by subprotocol, the normalized GE increases with the diffusion time in this region of the brain. We also note that globally, the GE for both models are higher in the fornix than in the genu.

Figure 4.4 presents the normalized generalization error made by the DTI and NODDI models in the region of the fornix with $\delta = 8$ ms and $\|\boldsymbol{G}\| \simeq 60$ mT/m.

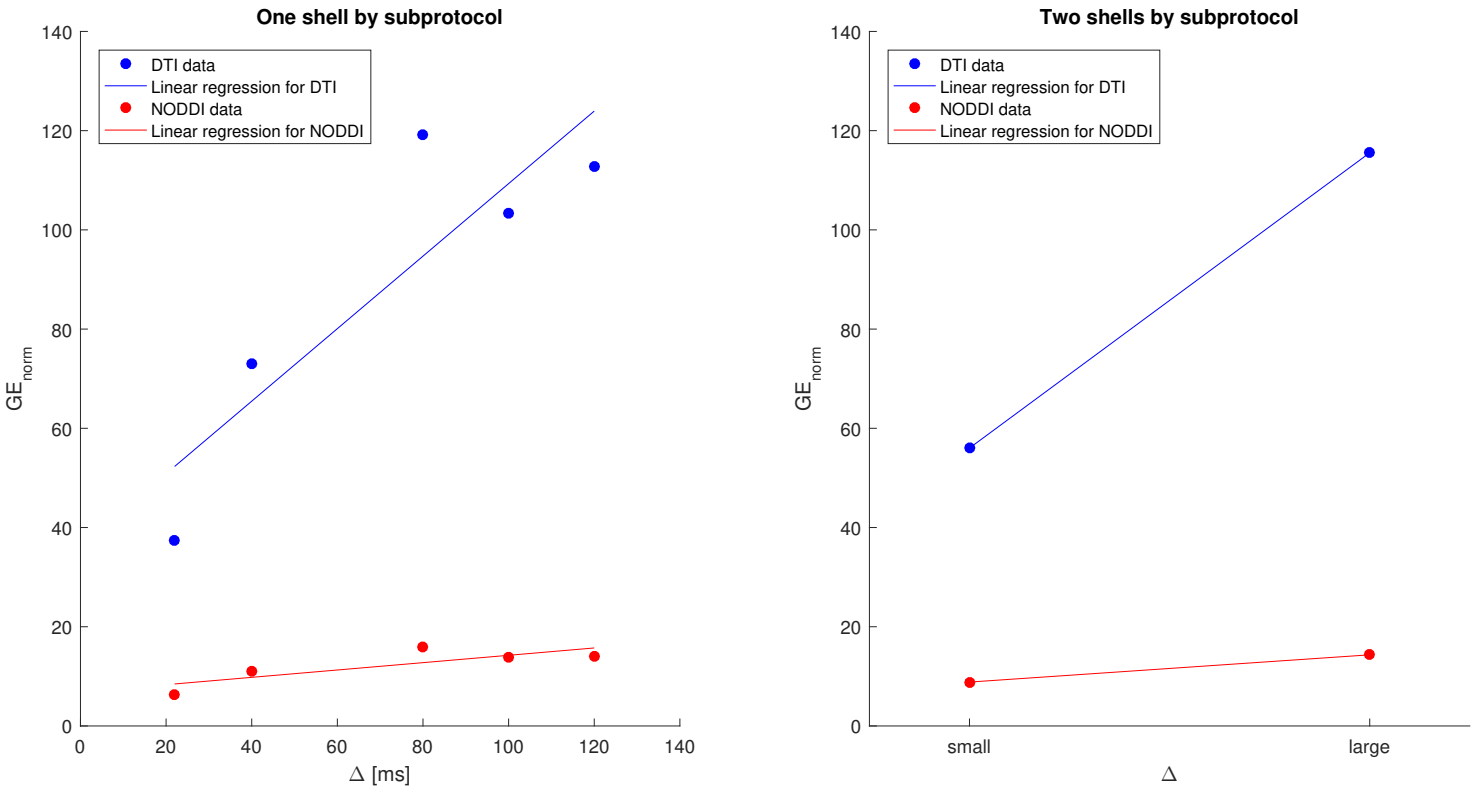

Figure 4.4: **Effect of the diffusion time for the DTI and NODDI models in the fornix with $\delta = 8$ ms.**
Left : Generalization error of the five subprotocols containing only one shell (subprotocol N° 7-8-9-10-11).
Right : Generalization error of subprotocols N° 15 (2 shells with small Δ) and 16 (2 shells with large Δ).

For both models, the GE increases with the diffusion time. The errors made by the DTI model are much greater than the ones made by the NODDI model. GE in Figure 4.4 are much higher than in Figure 4.3, confirming that the generalization error increases with the duration of the gradient pulse.

Finally, the normalized generalization error in the fornix for the last two subprotocols (N° 19 and 20), containing respectively 6 and 5 shells, are presented in Table 4.2.

Table 4.2: **Normalized generalization error made by the DTI and NODDI models in the region of the fornix for subprotocols N° 19 and 20.**

Subprotocol N°	**δ [ms]**	GE_{norm}	
		DTI	*NODDI*
19	3	37.25	21.53
20	8	143.34	23.45

As previously observed in the genu, to a fixed δ, the GE increases for subprotocols containing more shells. It also increases with the duration of the gradient pulse.

4.2.3 Summary

To conclude this section about the effect of the diffusion time on the generalization error, we summarize the results presented and try to explain them.

Table 4.3 summarizes the trends of the results of the two models.

Table 4.3: **Comparison of the DTI and NODDI models : effect of the diffusion time on the generalization error.**

Variation of GE with Δ		**DTI**		**NODDI**	
		Genu	*Fornix*	*Genu*	*Fornix*
1 shell	$\delta = 3\ ms$	↗	↗	↗	↗
	$\delta = 8\ ms$	↗	↗	↘	↗
2 shells	$\delta = 3\ ms$	↘	↗	↘	↗
	$\delta = 8\ ms$	↗	↗	↘	↗
3 shells	$\delta = 3\ ms$	↘	↗	↘	↘

Globally, we can say that the GE increases with the diffusion time. This is strangely not the case for the NODDI model in the genu. We can't explain why. Since the trends are clear for the DTI model as well as for the NODDI model in the fornix, there should be no reason for it to be different in a more homogeneous region.

Concerning the two models, NODDI is clearly better than DTI. For $\delta = 3$ ms, the difference between DTI and NODDI is not very big, even if the greater the diffusion time, the greater the difference. However, for $\delta = 8$ ms, the DTI error increases much faster than the NODDI error, leading to fairly considerable errors.

The GE in the fornix are globally greater than in the genu. This is due to the fact that since the fibers orientation is more complex in the fornix, models have more difficulties to predict a new heterogeneous signal. In the genu, the fibers are approximately straight and parallel making it easier for models to predict a new signal.

Finally, the GE increases with the duration of the gradient pulse and more generally with the echo time which is the sum of the diffusion time and the duration of gradient pulse ($\Delta + \delta = TE$). Indeed, the errors are greater when $\delta = 3$ ms than when $\delta = 8$ ms. Intuitively, it makes sense that the GE increases with the echo time because it means that we take more time to measure the signal, allowing more time for the water molecules to diffuse. It is then more complicated to predict a signal emitted by water molecules having moved (*i.e.* diffused) a lot. On the contrary, for subprotocols with a small echo time, the measurement of the signal is quickly made, leaving not much time for the spins to diffuse. The models can more easily predict a new signal in that range of echo time.

4.3 Effect of the gradient norm on the generalization error

In a second time, we examine the impact of the gradient norm $\|\boldsymbol{G}\|$ on the generalization error. For that, the parameter δ is fixed and we take three subprotocols containing 4 shells. Each of the three subprotocols has shells with the same diffusion time parameter Δ so that this parameter is somehow fixed too[2]. The subprotocols used to examine the effect of the gradient norm are presented in Appendix D.2. There are three subprotocols with $\delta = 3$ ms and three others with $\delta = 8$ ms. This is reproduced for both genu and fornix data.

Table 4.4 shows the average value of the gradient norm for each subprotocol.

Table 4.4: **Average value of the gradient norm for subprotocols used to analyze the effect of the gradient norm on the generalization error.**

Subprotocol N°	δ **[ms]**	$\|\boldsymbol{G}\|_{av}$ **[mT/m]**
21	3	62.5
22	3	122
23	3	243
24	8	53
25	8	147.5
26	8	287.75

[2]In each subprotocol, there are four shells with $\Delta = 22, 40, 60$ and 80 ms.

4.3.1 Genu

Figure 4.5 shows the evolution of the generalization error with subprotocols having small, medium and large gradient norms in the genu.

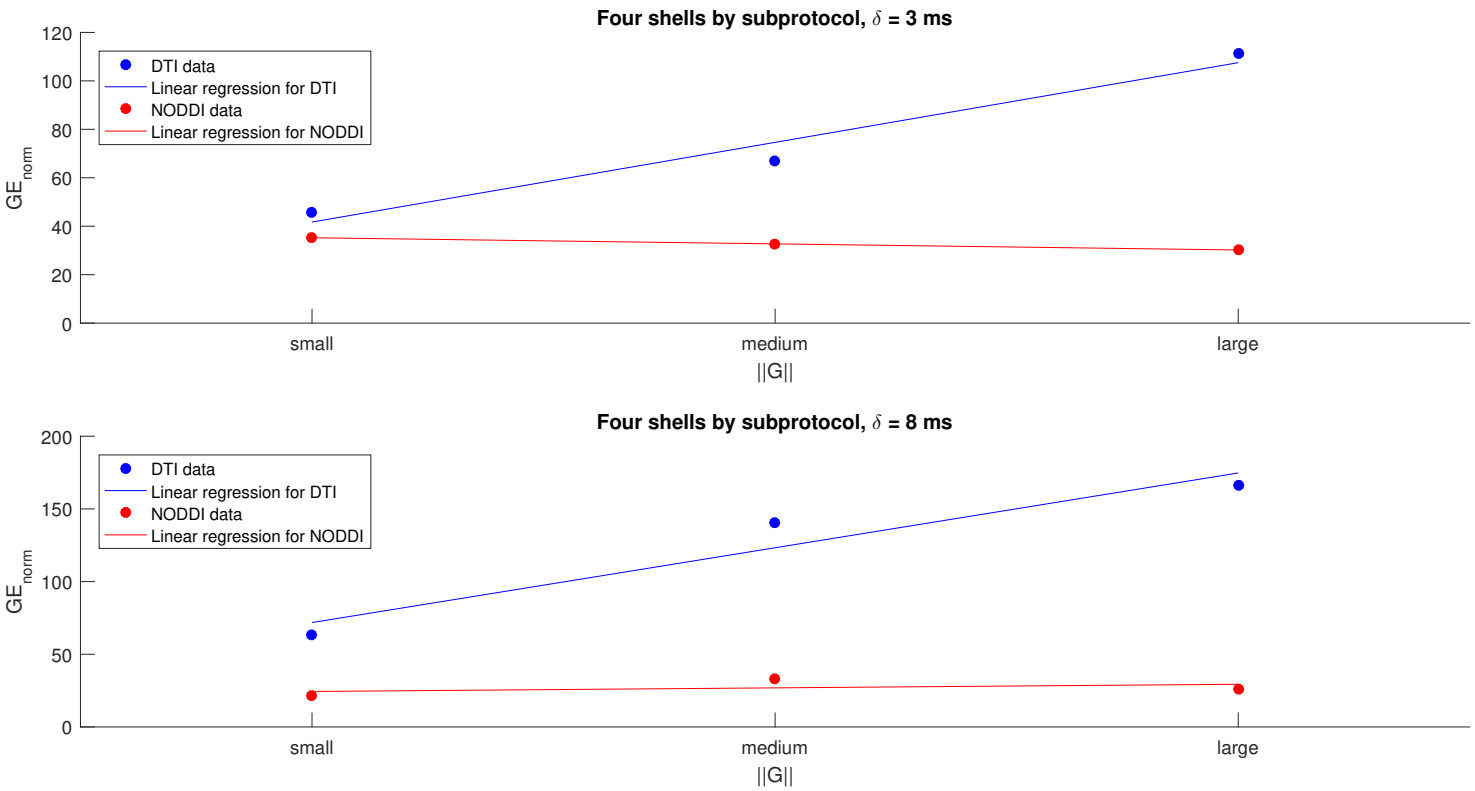

Figure 4.5: **Effect of the gradient norm for the DTI and NODDI models in the genu.**
Top : Generalization error of the three subprotocols containing four shells with $\delta = 3$ ms (subprotocol N° 21-22-23).
Bottom : Generalization error of the three subprotocols containing four shells with $\delta = 8$ ms (subprotocol N° 24-25-26).

The first observation is that the errors are greater than in the previous section. This can be explained by the fact that the subprotocols contained four shells. The effect of the subprotocols size is investigated in Section 4.5. Again, the DTI errors are greater than the NODDI errors. For the DTI model, the GE increases with the gradient norm. For the NODDI model, the effect of the gradient norm seems less important. Indeed, for $\delta = 3$ ms, the GE slightly decreases whereas for $\delta = 8$ ms, it slightly increases. When δ increases, the GE increases as previously observed for the DTI model. This is strangely not the case for the NODDI model.

4.3.2 Fornix

Figure 4.6 shows the evolution of the generalization error with subprotocols having small, medium and large gradient norms in the fornix.

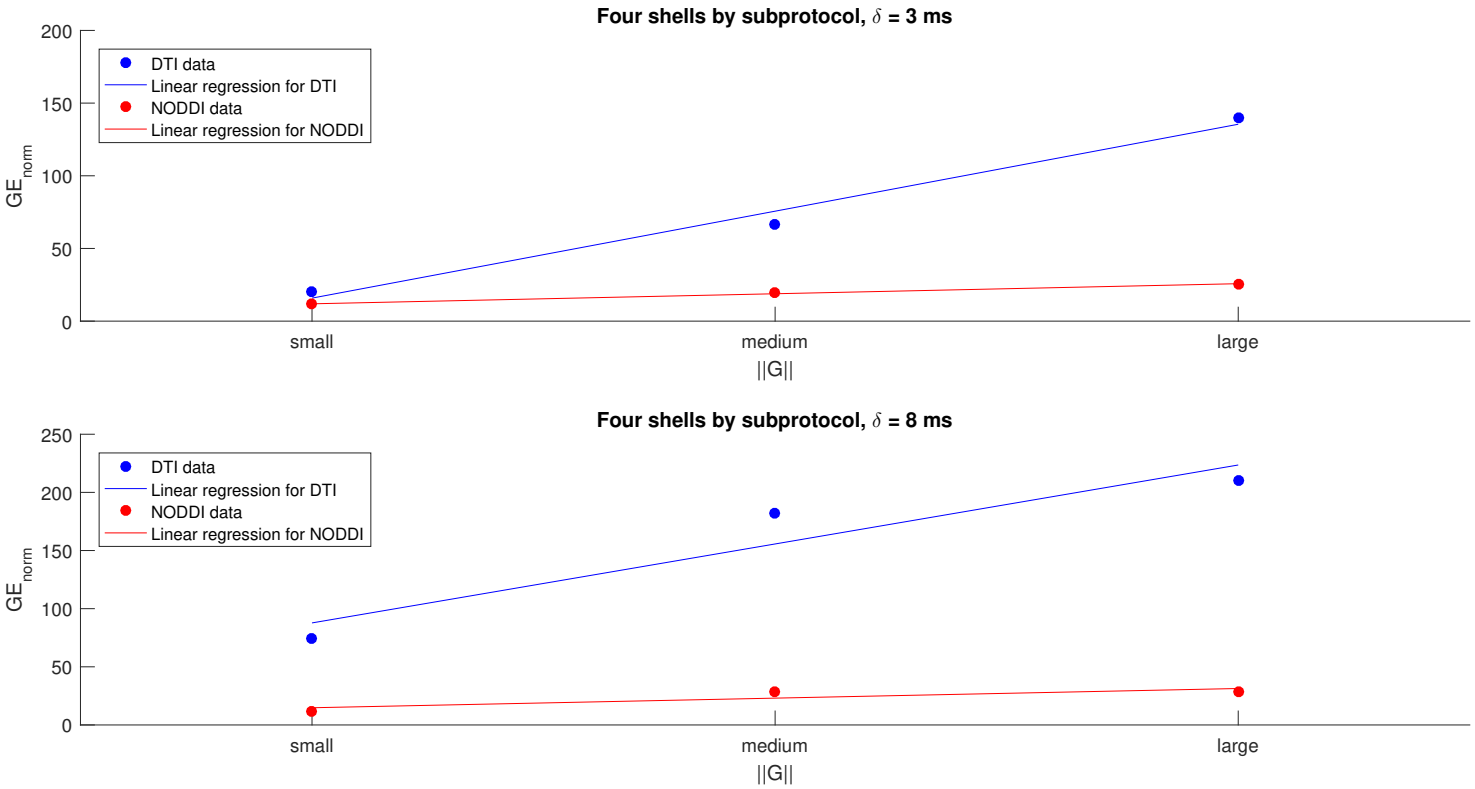

Figure 4.6: **Effect of the gradient norm for the DTI an NODDI models in the fornix.**
Top : Generalization error of the three subprotocols containing four shells with $\delta = 3$ ms (subprotocol N° 21-22-23).
Bottom : Generalization error of the three subprotocols containing four shells with $\delta = 8$ ms (subprotocol N° 24-25-26).

The results in the fornix are very similar to the ones in the genu. The GE clearly increases with the gradient norm for the DTI model. This is less pronounced but also the case for the NODDI model.

4.3.3 Summary

To conclude this section about the effect of the gradient norm on the generalization error, we summarize the results presented and try to explain them.

Table 4.5 summarizes the trends of the results of the two models.

Table 4.5: **Comparison of the DTI and NODDI models : effect of the gradient norm on the generalization error.**

Variation of GE with $\|\boldsymbol{G}\|$	**DTI**		**NODDI**	
	Genu	*Fornix*	*Genu*	*Fornix*
$\delta = 3$ *ms*	↗	↗	↘	↗
$\delta = 8$ *ms*	↗	↗	↗	↗

Globally, we can say that the generalization error increases with the gradient norm. For the DTI model, this is very clear. For the NODDI model, this increase is less strong but nonetheless present. We saw in Section 4.1 that strong gradients are more sensitive to small structures like thin axons in white matter. For the DTI model, there is thus a compromise between the quality of predicted signals and the sensitivity to small structures. This compromise is less important for the NODDI model, since the GE don't increase a lot with the gradient strength. Therefore, the NODDI model is also better than the DTI model from this point of view.

Intuitively, it makes sense that the GE increases with the gradient norm. The gradient norm can be viewed as the strength with which we "shake" the water molecules during a PGSE sequence. The more they are "shaken", the more they diffuse. And the more they diffuse, the more difficult it is for models to predict a new signal as explained in Section 4.2.3.

The same conclusion as in Section 4.2.3 can be drawn with regard to the fact that errors in the fornix are globally greater than the ones in the genu for the DTI model. Finally, we have the confirmation that the GE increases with the echo time since the errors are greater for $\delta = 3$ ms than for $\delta = 8$ ms.

4.4 Effect of mixing on the generalization error

In a third time, we examine the impact of mixing on the generalization error. We refer to the word *mixing* to talk about all possible combinations of shells contained in a multi-shell HARDI. In the full scanning protocol of Appendix C used in this document, each multi-shell HARDI contains three shells. We chose to analyze the effect of mixing on two multi-shell HARDI with medium echo time, one with $\delta = 3$ ms and the other with $\delta = 8$ ms. Section D.3 of Appendix D shows the 12 subprotocols built from the two multi-shell HARDI. Subprotocols from the multi-shell HARDI with $TE = 87$ ms are called *Mixing (1)* whereas those from the multi-shell HARDI with $TE = 92$ ms are under the name *Mixing (2)*.

4.4.1 Genu

Figure 4.7 shows the results obtained for Mixing (1) by the DTI and NODDI models in the region of the genu.

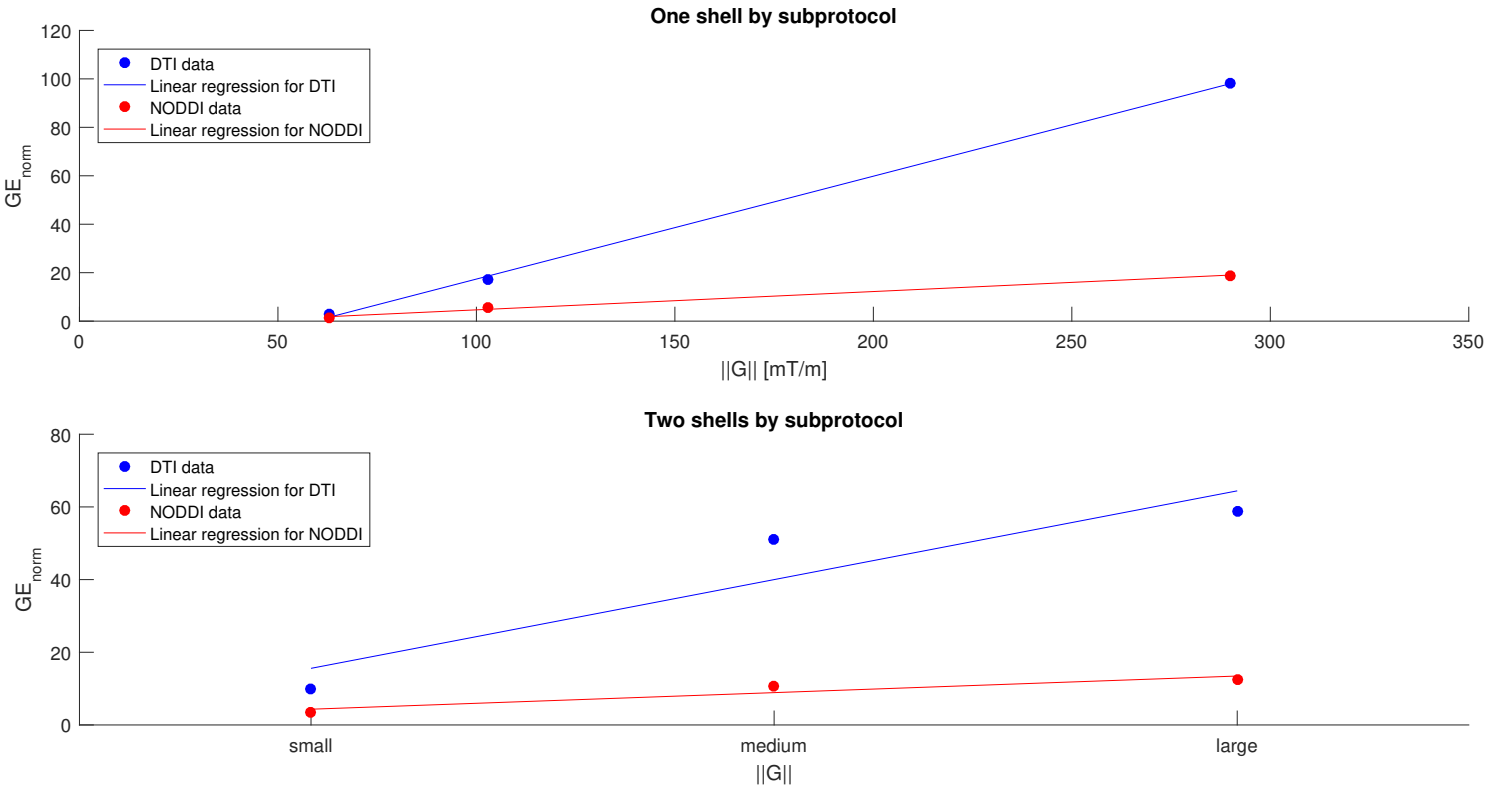

Figure 4.7: **Effect of mixing for the DTI and NODDI models in the genu (1).**
Top : Evolution of the normalized generalization error with the gradient norm for subprotocols containing one shell.
Bottom : Evolution of the normalized generalization error with the gradient norm for subprotocols containing two shells.

The observation already made in the previous section seems to be confirmed again. Indeed, the GE increases with the gradient norm for both DTI and NODDI models. The GE is also higher for the subprotocols containing two shells than for the ones containing one shell.

Figure 4.8 shows the results obtained for Mixing (2) by the DTI and NODDI models in the region of the genu.

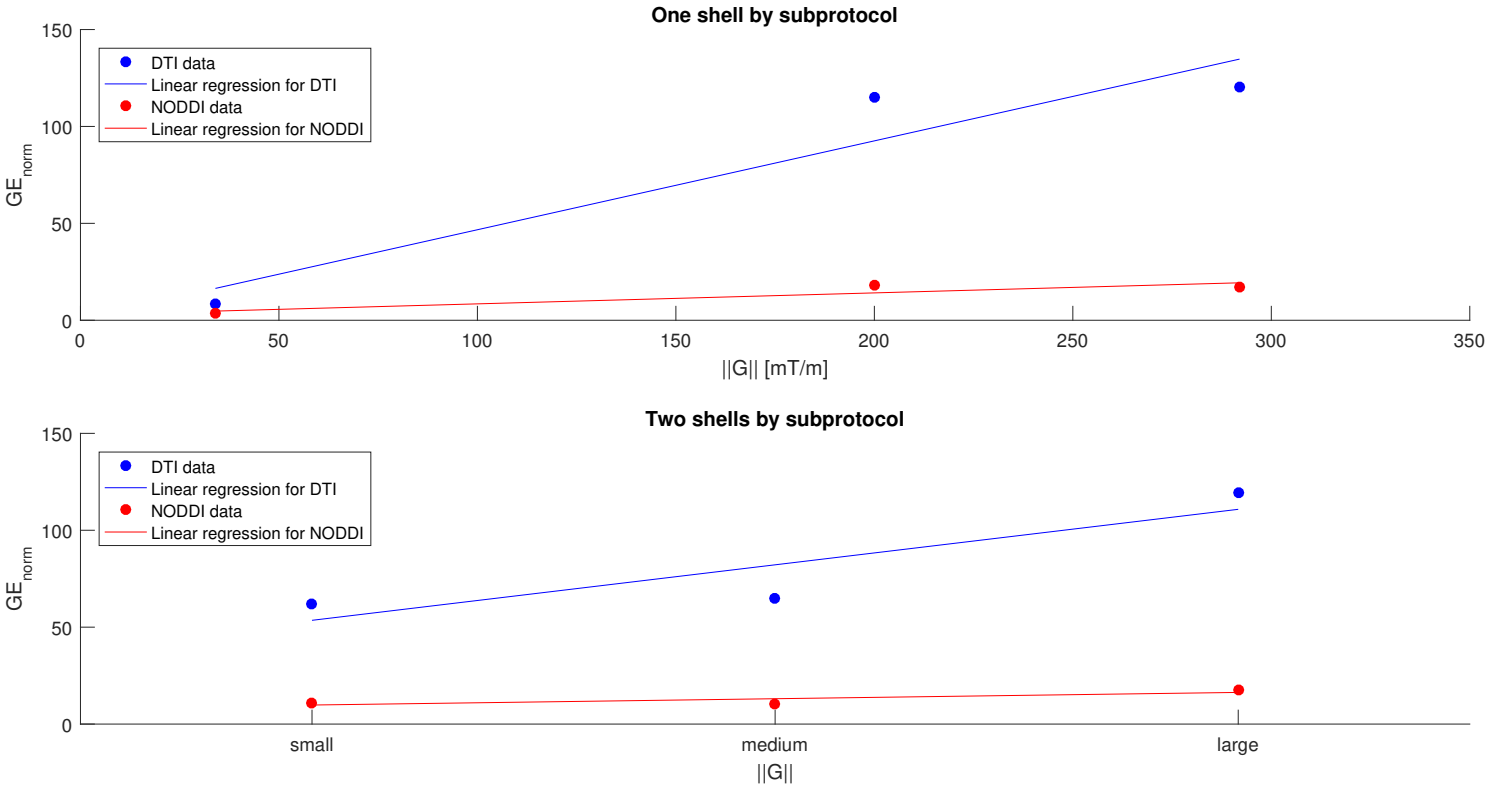

Figure 4.8: **Effect of mixing for the DTI and NODDI models in the genu (2).**
Top : Evolution of the normalized generalization error with the gradient norm for subprotocols containing one shell.
Bottom : Evolution of the normalized generalization error with the gradient norm for subprotocols containing two shells.

The conclusions are the same as in the genu. We add that the errors for Mixing (2) are higher than for Mixing (1) because the echo time of Mixing (2) is greater.

Finally, Table 4.6 shows the values of the normalized GE made by the DTI and NODDI models for subprotocols N° 32 and 39, *i.e.* the two subprotocols containing 3 shells. Again, these values are larger than those made on subprotocols containing 2 shells, confirming that the GE increases with the subprotocol size.

Table 4.6: **Normalized generalization error made by the DTI and NODDI models in the region of the genu for subprotocols N° 32 and 39.**

Subprotocol	$\boldsymbol{TE}$	$\boldsymbol{GE_{norm}}$	
N°	**[ms]**	*DTI*	*NODDI*
32	87	40.54	8.87
39	92	83.11	13.34

4.4.2 Fornix

Figure 4.9 shows the results obtained for Mixing (1) by the DTI and NODDI models in the region of the fornix.

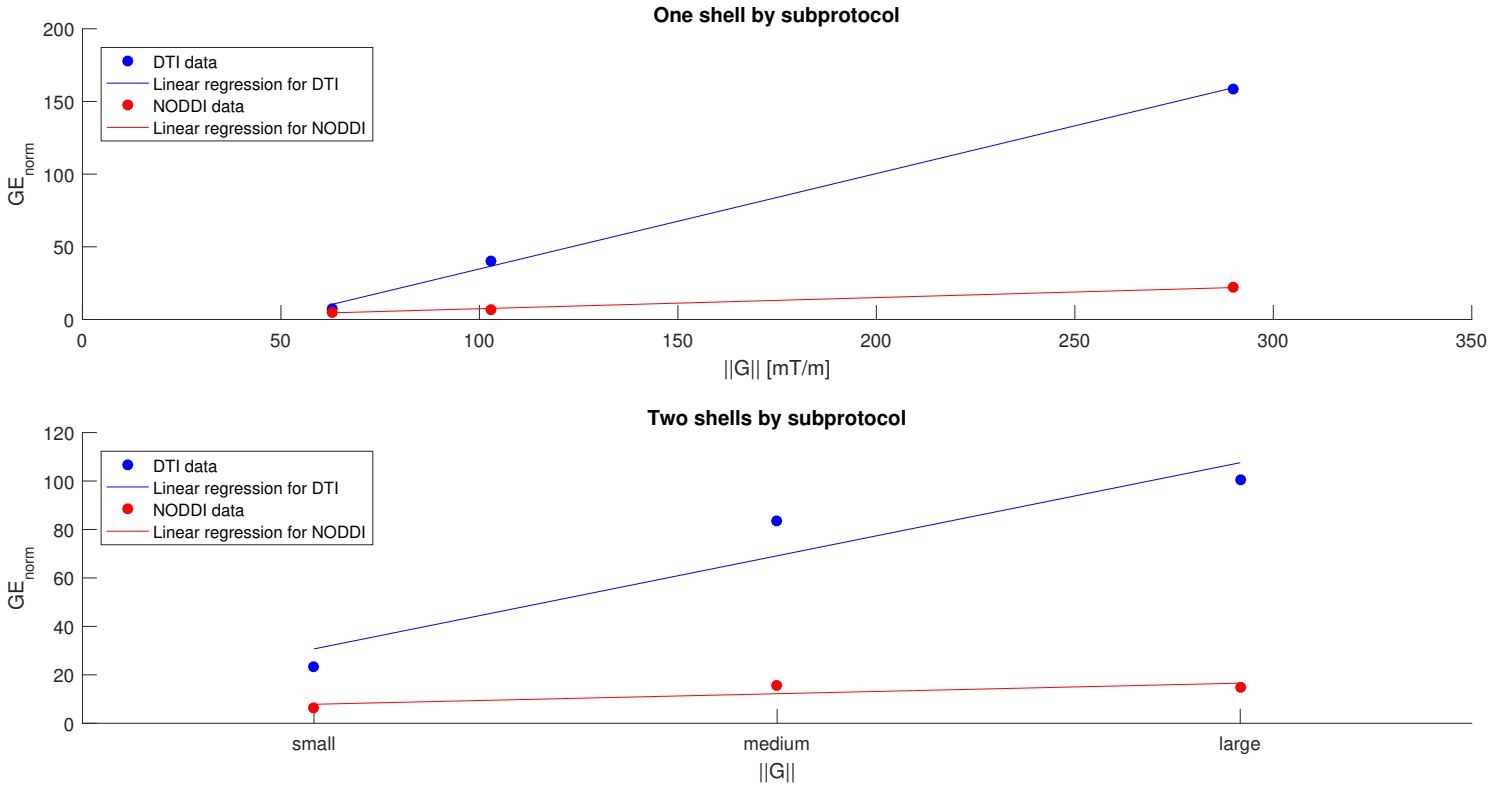

Figure 4.9: **Effect of mixing for the DTI and NODDI models in the fornix (1).**
Top : Evolution of the normalized generalization error with the gradient norm for subprotocols containing one shell.
Bottom : Evolution of the normalized generalization error with the gradient norm for subprotocols containing two shells.

Again, the GE strongly increases with the gradient norm for the DTI model whereas it slightly increases for the NODDI model.

Figure 4.10 shows the results obtained for Mixing (2) by the DTI and NODDI models in the region of the fornix.

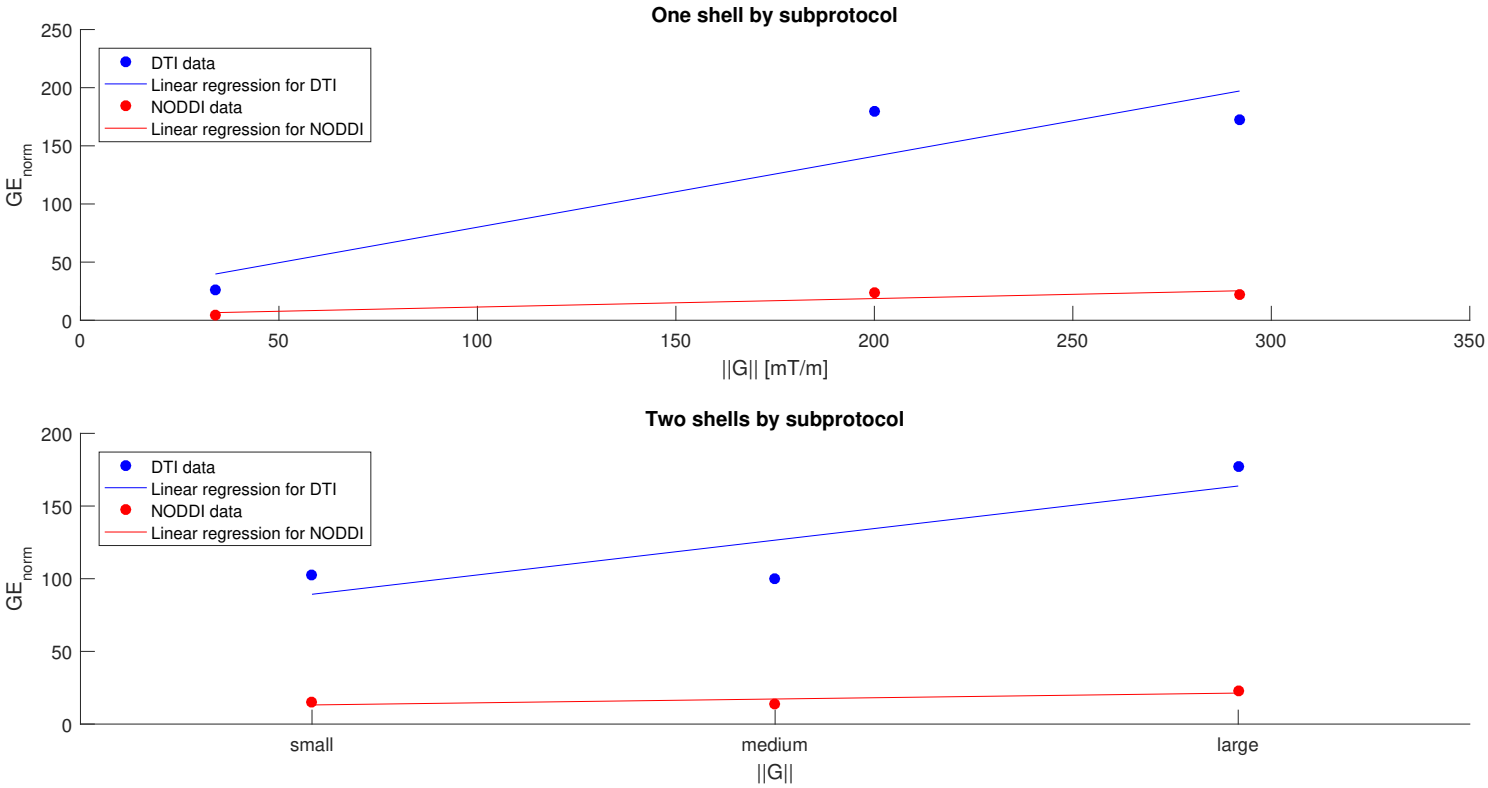

Figure 4.10: **Effect of mixing for the DTI and NODDI models in the fornix (2).**
Top : Evolution of the normalized generalization error with the gradient norm for subprotocols containing one shell.
Bottom : Evolution of the normalized generalization error with the gradient norm for subprotocols containing two shells.

This figure is similar to the previous one, confirming all the trends that have been presented in this section.

Finally, the normalized generalization errors of the two subprotocols containing 3 shells are presented in Table 4.7. We note that these values are higher than those in the genu, confirming that the models have more difficulties to predict a new signal in a region where the fibers orientation is more complex.

Table 4.7: **Normalized generalization error made by the DTI and NODDI models in the region of the fornix for subprotocols N° 32 and 39.**

Subprotocol	$\boldsymbol{TE}$	$\boldsymbol{GE_{norm}}$	
N°	**[ms]**	*DTI*	*NODDI*
32	87	69.29	12.35
39	92	127.14	17.82

4.4.3 Summary

Within a multi-shell HARDI, the diffusion time Δ and the duration of the gradient pulse δ are constant. Only the gradient norm varies. It is thus normal that the results are the same as in the previous section where the effect of the gradient norm on the GE was studied. This result is that the generalization error increases with the gradient norm.

For small values of $\|\boldsymbol{G}\|$, the DTI and NODDI errors are similar for subprotocols containing one shell. But as soon as the gradient norm increases, the DTI model makes larger errors than the NODDI model.

4.5 Effect of the subprotocols size

As seen in the previous sections, the impact of the subprotocols size on the generalization error is important. We have seen that the more shells by subprotocol, the more informations are available about the diffusion of water molecules.

However, we remark that globally, the more shells by subprotocol, the greater the generalization error. This is particularly the case for the DTI model. This can be explained by the fact that the only estimated parameter of the DTI model is the diffusion tensor $\boldsymbol{D}$. The more shells, the more informations on the diffusion and this is summarized in a single second order tensor. Therefore, it is complicated for the DTI model to estimate new signals for subprotocols containing several shells.

This is more complex for the NODDI model. Indeed, contrary to the DTI model, NODDI estimates 5 parameters as explained in Section 2.3. These parameters can contained more informations than a diffusion tensor. This is why the errors are similar for subprotocols having different sizes. Nevertheless, the GE error slightly increases with the subprotocols size.

A final remark for the NODDI model is that, in the figures of the previous section, it seems that the generalization errors for subprotocols containing two and three shells are the average of the errors made for subprotocols containing one shell. This observation is verified by making the calculations. This is done in Table 4.8 and the results are convincing. Indeed, the mean is very close to the real GE. This is therefore a good technique to predict new generalization errors within a multi-shell HARDI. The technique is the following : first we choose a multi-shell HARDI. Then we compute the GE for the subprotocols containing one shell. And with that, it is possible to predict the GE for subprotocols containing several of the initial shells.

Table 4.8: **Comparison between the generalization error and the mean for the NODDI model.** Mean(x;y) is the mean between subprotocol N° x and y.

Subprotocol	**Genu**		**Fornix**	
N°	$\boldsymbol{GE}_{norm}$	**Mean**	$\boldsymbol{GE}_{norm}$	**Mean**
3	1.17	/	5.11	/
27	5.72	/	6.98	/
28	18.87	/	22.07	/
29	3.41	Mean(3;27) = 3.44	6.18	Mean(3;27) = 6.05
30	10.73	Mean(3;28) = 10.02	15.52	Mean(3;28) = 13.59
31	12.54	Mean(27;28) = 12.3	14.9	Mean(27;28) = 14.53
32	8.87	Mean(3;27;28) = 8.58	12.35	Mean(3;27;28) = 11.39
33	3.38	/	4.8	/
34	17.88	/	23.58	/
35	16.91	/	22.18	/
36	11.07	Mean(33;34) = 10.63	14.86	Mean(33;34) = 14.19
37	10.06	Mean(33;35) = 10.15	13.84	Mean(33;35) = 13.49
38	17.57	Mean(34;35) = 17.4	23.02	Mean(34;35) = 22.88
39	13.34	Mean(33;34;35) = 12.72	17.82	Mean(33;34;35) = 16.85

Conclusion

In this work, we have compared DW-MRI models and investigated the effects of the PGSE parameters on them.

In order to do that, we chose four existing models and described them in detail. Each model having its own limitations and advantages, we presented them. This was the topic of Chapter 2.

To compare those different models, we needed a metric. As the acquisition time to obtain several DW-MRI can be long, we focused our comparison on the capacity of a model to predict new signals. The most appropriate metric to compare the models is therefore the generalization error. Chapter 3 develops this metric and presents the different ways to estimate it.

Unfortunately, only two models (DTI and NODDI) have been compared. However, the results allowed us to draw significant conclusions. The first and more important one is that the NODDI model is always better than the DTI model which is too simple. This means that from a purely predictive point of view, in any situation it is not preferable to use the DTI model since its predicted signals are worse than the NODDI model. However, in practice, the NODDI model is much more slower than the DTI model.

Concerning the effects of the PGSE parameters, we have clearly seen that the generalization error increases with the echo time and the gradient norm. Since [33] proved that stronger gradients allow better sensibility to small structures, there is a compromise to do between a better sensibility and a good ability to predict new signals.

Finally, as predicted intuitively, the generalization error made by a simple model like DTI increases with the subprotocols size. Concerning the NODDI model, this is more complex but it seems that it slightly increases too. We were also able to establish a technique to predict new generalization errors for the NODDI model.

In the future, it will be interesting to add to this comparison the two missing models : MFM and DIAMOND. Indeed, the MFM model, which is the amelioration of the DTI model, allows to take into account several fascicles in one voxel. This is probably an interesting model for regions containing crossing fascicles like the fornix. Finally, the DIAMOND model combines a statistical approach with a biophysical model and can be of great interest due to its complexity.

Appendix A

Structure of a neuron

This appendix details the structure of a typical neuron.

A neuron is composed of a *cell body* (also called *soma*) containing the *nucleus*. To this cell body are linked *dendrites* the role of which is to propagate the electrochemical stimulation received from other neural cells to the soma. The *axon* conducts electrical impulses away from the neuron's cell body. It transmits the information to different neurons, muscles, etc. This axon can be covered by several *myelin sheaths* whose role is to electrically isolated the nerve fibers. Between two successive myelin sheaths is a *node of Ranvier* which is conductive. Thereby the nerve impulse jumps from one node of Ranvier to another along the axon, increasing the propagation speed. Finally, *Schwann cells* are a type of glial cells which form the myelin sheaths.

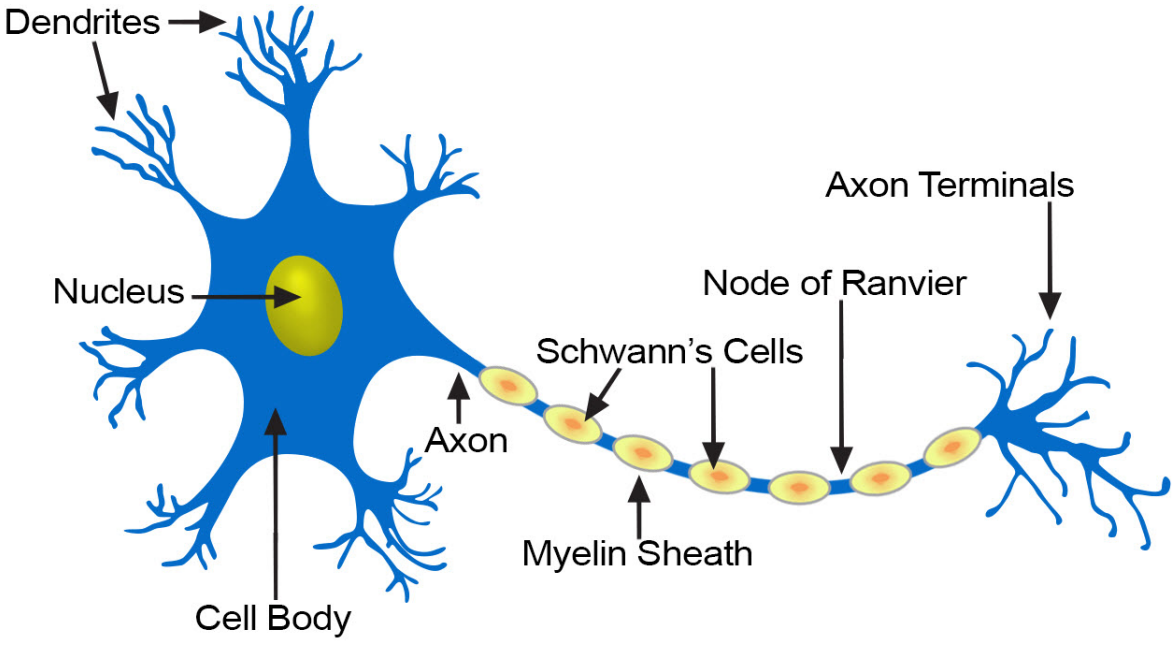

Figure A.1: **Structure of a typical neuron** [34].

Appendix B

Average number of distinct observations in each bootstrap sample

This appendix proves that the average number of distinct observations in each bootstrap sample is $0.632n$ with n the number of observed data points.

The training set $\boldsymbol{z} = \{z_1, z_2, ..., z_n\}$ contains n items. Bootstrap resampling picks items with replacement from this original set until another set of size n is built. The probability of choosing any one item (*e.g.* z_1) on the first draw is :

$$P\left(\text{picking } z_i \text{ on the first draw}\right) = \frac{1}{n} \ , \ \ i \in [1, 2, ..., n] \ .$$

It follows that :

$$P\left(\text{not picking } z_i \text{ on the first draw}\right) = 1 - \frac{1}{n} \ , \ \ i \in [1, 2, ..., n] \ .$$

This is just for the first draw. Since there are n independent draws, the probability of never choosing this item on any draws is :

$$P\left(\text{not picking } z_i \text{ on all the draws}\right) = \left(1 - \frac{1}{n}\right)^n \ , \ \ i \in [1, 2, ..., n] \ . \tag{B.1}$$

Let us see which value takes this probability when n becomes large. In that case, the first order Taylor expansion of $e^{\frac{x}{n}}$ is :

$$e^{\frac{x}{n}} \simeq 1 + \frac{x}{n} + \mathcal{O}\left(x^2\right) \ .$$

Since $(e^{\frac{x}{n}})^n = e^x$, if we drop the higher order terms, we find :

$$e^x \simeq \left(1 + \frac{x}{n}\right)^n \ . \tag{B.2}$$

The right hand side of equation B.2 is the same as the right hand side of equation B.1 if $x = -1$. Therefore, we get the following result :

$$\lim_{n \to \infty} \left(1 - \frac{1}{n}\right)^n = e^{-1} \simeq 0.3679 \ .$$

The probability of an instance being sampled at least once is thus $1 - 0.368 = 0.632$. The average number of distinct observations in each bootstrap sample is therefore $0.632n$.

Figure B.1 shows the function $F(n) = \left(1 - \frac{1}{n}\right)^n$ as n increases. We clearly see an asymptote in $y = e^{-1}$. The approximation $\left(1 - \frac{1}{n}\right)^n \simeq e^{-1}$ makes sense for n larger than 80. In the computations of this document n will always be much larger than 80 making this approximation valid.

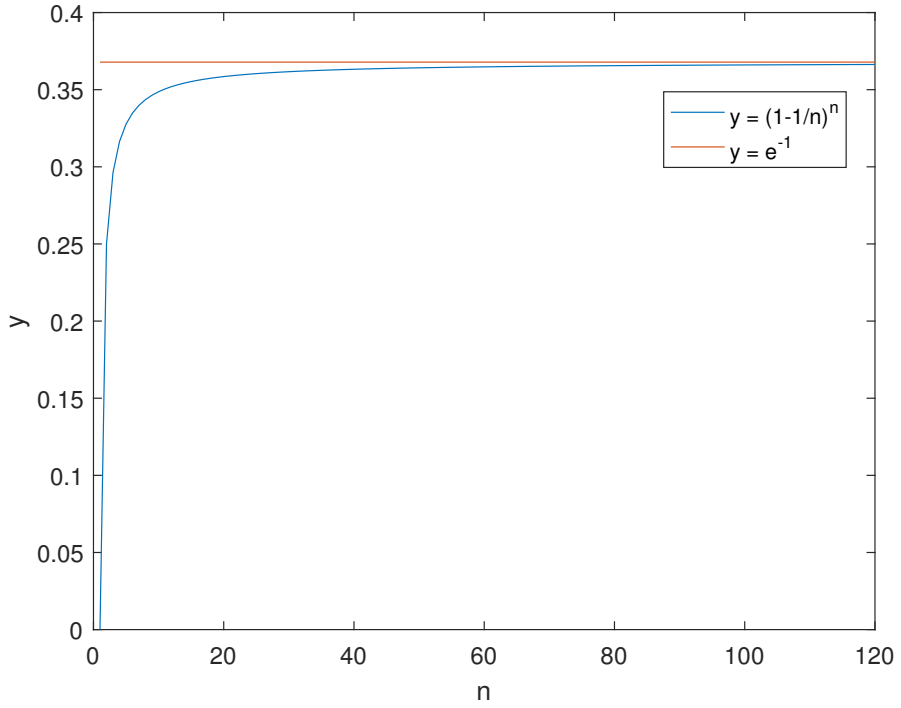

Figure B.1: **Approximation of the probability that an item won't be picked in a bootstrap sample**.

Appendix C

Acquisition protocol

This appendix details the acquisition protocol used to obtain the data with which the comparison between the different DW-MRI models was done. These data derived from the *White Matter Modelling Challenge*[1], one of the International Symposium on Biomedical Imaging (ISBI) challenges that took place in April 2015.

Regions of interest

The data come from two different regions, each containing 6 voxels :

- The *genu* in the corpus callosum where the fibers are approximately straight and parallel.
- The *fornix* where the configuration of fibers is more complex and includes fascicle crossings.

Figure C.1 illustrates the location of these regions in the brain.

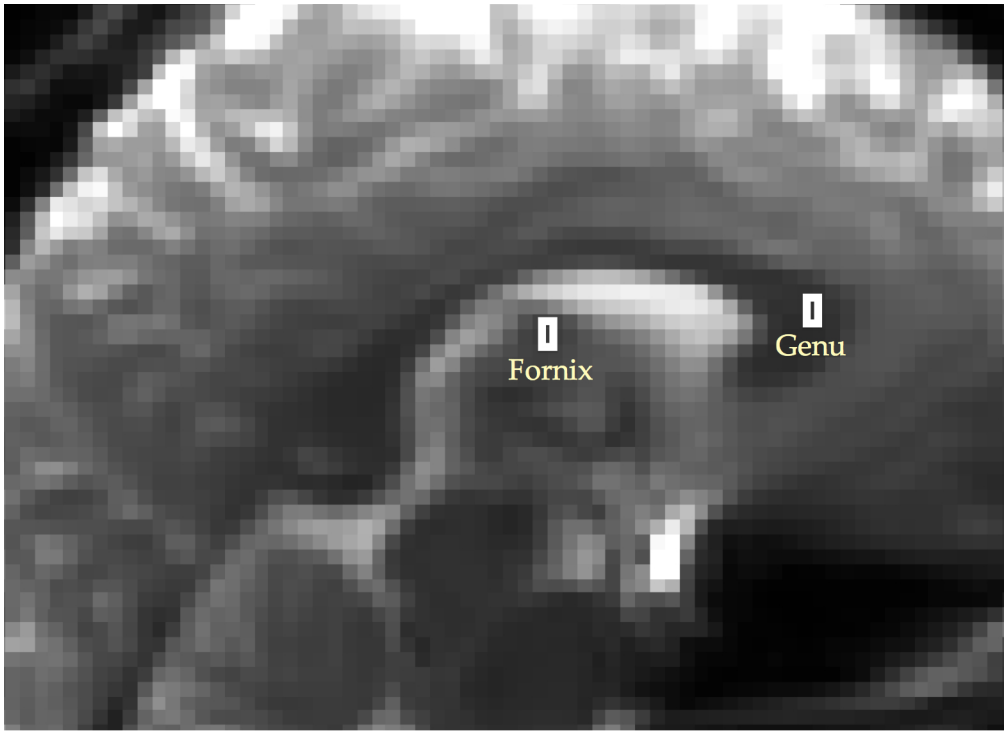

Figure C.1: **Regions of interest : *genu* and *fornix*** [31].

[1]See `http://cmic.cs.ucl.ac.uk/wmmchallenge/` for more informations.

The complete experimental protocol

Table C.1 shows the full acquisition protocol consisting of 36 high angular resolution diffusion imaging (HARDI) shells [31].

Table C.1: **The full scanning protocol.**

$\delta = 3$ [ms]					$\delta = 8$ [ms]				
N°	Δ [ms]	TE [ms]	$\|\boldsymbol{G}\|$ [mT/m]	b [s/mm^2]	N°	Δ [ms]	TE [ms]	$\|\boldsymbol{G}\|$ [mT/m]	b [s/mm^2]
1	22	49	61	50	10	22	58	58	300
2	22	49	86	100	11	22	58	190	3200
3	22	49	192	500	12	22	58	275	6700
4	60	87	63	150	13	60	92	34	300
5	60	87	103	400	14	60	92	200	10500
6	60	87	290	3200	15	60	92	292	22350
7	100	127	63	250	16	100	132	60	1600
8	100	127	101	650	17	100	132	100	4450
9	100	127	200	2550	18	100	132	292	38050
19	40	67	63	100	28	40	72	59	600
20	40	67	200	1000	29	40	72	100	1700
21	40	67	289	2100	30	40	72	292	14550
22	80	107	63	200	31	80	112	61	1300
23	80	107	99	500	32	80	112	100	3550
24	80	107	201	2050	33	80	112	292	30200
25	120	147	63	300	34	120	152	60	1950
26	120	147	199	3050	35	120	152	200	21500
27	120	147	291	6500	36	120	152	292	45900

A HARDI scheme consists in probing diffusion in tissues with a fixed b-value from many different angles. This means that the values δ, Δ and $\|\boldsymbol{G}\|$ are fixed while the gradient directions are uniformly distributed across the unit sphere.

In multi-shell HARDI, the values of δ and Δ are fixed and M shells of different b-values are created by varying the gradient intensity $\|\boldsymbol{G}\|$, each shell containing N gradient directions. Maintaining δ and Δ

constant throughout all DW-MRI acquisition shells has the advantage of leading to the same Eddy currents distortion patterns and making subsequent image alignment easier.

Here are the values used to obtain the different HARDI shells :

- $\|\boldsymbol{G}\| = \{60, 100, 200, 300\}$ [mT/m]
- $\delta = \{3, 8\}$ [ms]
- $\Delta = \{22, 40, 60, 80, 100, 120\}$ [ms].

From these values, we deduce that b-value $\in [50; 45900]$ [s/mm^2]. For δ and Δ fixed, $M = 4$ shells of different b-values. We thus have $2 \times 6 \times 4 = 48$ shells. But for δ and Δ fixed, one of the $M = 4$ shells is hidden[2], leaving $2 \times 6 \times 3 = 36$ shells. Each shell has $N = 90$ gradient directions (45 acquired in both positive and negative direction to allow for eddy current distortion correction). The protocol is designed to cover the measurement space as widely as possible.

[2]The original challenge was to predict unseen data. These hidden shells are the unseen data.

Appendix D

Choices of subprotocols

This appendix presents the subprotocols on which the generalization error was computed for each model. In this document, we chose to analyze the effects of the diffusion time and the gradient norm on the generalization error. We also tried to mix the shells together within the same multi-shell HARDI.

D.1 Effect of the diffusion time

To analyze the effect of the diffusion time on the generalization error, the two parameters δ and $\|\boldsymbol{G}\|$ are fixed and Δ varies. We first took one shell by subprotocol leading to 6 subprotocols with $\delta = 3$ ms and 5 subprotocols with $\delta = 8$ ms. Table D.1 shows these subprotocols.

Table D.1: **Effect of the diffusion time (1).**

Subprotocol N°	Shell N°	Δ [ms]	δ [ms]	TE [ms]	$\|\boldsymbol{G}\|$ [mT/m]	b [s/mm^2]
1	1	22	3	49	61	50
2	19	40	3	67	63	100
3	4	60	3	87	63	150
4	22	80	3	107	63	200
5	7	100	3	127	63	250
6	25	120	3	147	63	300
7	10	22	8	58	58	300
8	28	40	8	72	59	600
9	31	80	8	112	61	1300
10	16	100	8	132	60	1600
11	34	120	8	152	60	1950

Then, to a fixed δ, we grouped shells by two with approximately the same Δ leading to three subprotocols for $\delta = 3$ ms ("small Δ" - "medium Δ" - "large Δ") and two subprotocols for $\delta = 8$ ms ("small Δ" - "large Δ"). Table D.2 shows these subprotocols.

Table D.2: **Effect of the diffusion time (2).**

Subprotocol N°	Shell N°	Δ [ms]	δ [ms]	TE [ms]	$\|\boldsymbol{G}\|$ [mT/m]	b [s/mm^2]
12	1	22	3	49	61	50
	19	40	3	67	63	100
13	4	60	3	87	63	150
	22	80	3	107	63	200
14	7	100	3	127	63	250
	25	120	3	147	63	300
15	10	22	8	58	58	300
	28	40	8	72	59	600
16	16	100	8	132	60	1600
	34	120	8	152	60	1950

Finally, for $\delta = 3$ ms, we grouped shells by three with approximately the same Δ leading to two subprotocols ("small Δ" - "large Δ"). We also grouped all the shells with the same δ and $\|\boldsymbol{G}\|$ together (subprotocols 19 and 20 in table D.3).

These 20 first subprotocols allow the analysis of the effect of Δ on the generalization error. It is also possible to analyze how the generalization error behaves as a function of the number of shells per subprotocol.

Table D.3: **Effect of the diffusion time (3).**

Subprotocol N°	Shell N°	Δ [ms]	δ [ms]	TE [ms]	$\|\boldsymbol{G}\|$ [mT/m]	b [s/mm^2]
17	1	22	3	49	61	50
	19	40	3	67	63	100
	4	60	3	87	63	150
18	22	80	3	107	63	200
	7	100	3	127	63	250
	25	120	3	147	63	300
19	1	22	3	49	61	50
	19	40	3	67	63	100
	4	60	3	87	63	150
	22	80	3	107	63	200
	7	100	3	127	63	250
	25	120	3	147	63	300
20	10	22	8	58	58	300
	28	40	8	72	59	600
	31	80	8	112	61	1300
	16	100	8	132	60	1600
	34	120	8	152	60	1950

D.2 Effect of the gradient norm

To analyze the effect of the gradient norm on the generalization error, the two parameters Δ and δ are fixed and $\|\boldsymbol{G}\|$ varies. For both $\delta = 3$ ms and $\delta = 8$ ms, we chose 4 shells by subprotocol with increasing Δ. Each of the 6 subprotocols has the same diffusion times. The only things that change between subprotocols are the gradient norms. There are three subprotocols for $\delta = 3$ ms ("small $\|\boldsymbol{G}\|$" - "medium $\|\boldsymbol{G}\|$" - "large $\|\boldsymbol{G}\|$") and three subprotocols for $\delta = 8$ ms ("small $\|\boldsymbol{G}\|$" - "medium $\|\boldsymbol{G}\|$" - "large $\|\boldsymbol{G}\|$"). This is illustrated on table D.4.

These 6 subprotocols allow the analysis of the effect of the gradient norm on the generalization error.

Table D.4: **Effect of the gradient norm.**

Subprotocol N°	Shell N°	Δ [ms]	δ [ms]	TE [ms]	$\|\boldsymbol{G}\|$ [mT/m]	b [s/mm^2]
21	1	22	3	49	61	50
	19	40	3	67	63	100
	4	60	3	87	63	150
	22	80	3	107	63	200
22	2	22	3	49	86	100
	20	40	3	67	200	1000
	5	60	3	87	103	400
	23	80	3	107	99	500
23	3	22	3	49	192	500
	21	40	3	67	289	2100
	6	60	3	87	290	3200
	24	80	3	107	201	2050
24	10	22	8	58	58	300
	28	40	8	72	59	600
	13	60	8	92	34	300
	31	80	8	112	61	1300
25	11	22	8	58	190	3200
	29	40	8	72	100	1700
	14	60	8	92	200	10500
	32	80	8	112	100	3550
26	12	22	8	58	275	6700
	30	40	8	72	292	14550
	15	60	8	92	292	22350
	33	80	8	112	292	30200

D.3 Effect of mixing

Finally, to analyze the effect of mixing, we took two multi-shell HARDI (one with $\delta = 3$ ms and one with $\delta = 8$ ms) and made all possible combinations of the shell contained in each multi-shell HARDI. As each multi-shell HARDI contains 3 shells (a, b and c), the 6 subprotocols containing the 6 possible combinations are :

- *a-b-c* : subprotocols with one shell,
- *ab-ac-bc* : subprotocols with two shells,
- *abc* : subprotocol with three shells.

Table D.5: **Effect of mixing (1).**

Subprotocol N°	Shell N°	Δ [ms]	δ [ms]	TE [ms]	$\|\boldsymbol{G}\|$ [mT/m]	b [s/mm^2]
3 (already done)	4	60	3	87	63	150
27	5	60	3	87	103	400
28	6	60	3	87	290	3200
29	4	60	3	87	63	150
	5	60	3	87	103	400
30	4	60	3	87	63	150
	6	60	3	87	290	3200
31	5	60	3	87	103	400
	6	60	3	87	290	3200
32	4	60	3	87	63	150
	5	60	3	87	103	400
	6	60	3	87	290	3200

Table D.6: **Effect of mixing (2).**

Subprotocol N°	Shell N°	Δ [ms]	δ [ms]	TE [ms]	$\|\boldsymbol{G}\|$ [mT/m]	b [s/mm^2]
33	13	60	8	92	34	300
34	14	60	8	92	200	10500
35	15	60	8	92	292	22350
36	13	60	8	92	34	300
	14	60	8	92	200	10500
37	13	60	8	92	34	300
	15	60	8	92	292	22350
38	14	60	8	92	200	10500
	15	60	8	92	292	22350
39	13	60	8	92	34	300
	14	60	8	92	200	10500
	15	60	8	92	292	22350

Bibliography

[1] Taquet Maxime. *Multi-Fascicle Models of the Brain Microstructure for Population Studies: Acquisition, Estimation, Registration and Statistical Analysis.* PhD thesis, Université Catholique de Louvain, 2013.

[2] Silke Walter, Panagiotis Kostpopoulos, Anton Haass, Stefan Helwig, Isabel Keller, Tamara Licina, Thomas Schlechtriemen, Christian Roth, Panagiotis Papanagiotou, Anna Zimmer, et al. Bringing the hospital to the patient: first treatment of stroke patients at the emergency site. *PLoS One*, 5(10):e13758, 2010.

[3] Akbar Abdolmohammadi. Chapitre 6 : Rmn du proton, principes, April 2017. `http://borzuya-university.org/index.php/fr/2016/10/24/chapter-6-nmr-proton-principles/`.

[4] Bol Anne, Lee John, Macq Benoît, and Peeters Frank. Course notes *GBIO2050* medical imaging. Technical report, Université Catholique de Louvain, 2016-2017.

[5] Zhi-Pei Liang and Paul C Lauterbur. *Principles of magnetic resonance imaging: a signal processing perspective*. SPIE Optical Engineering Press, 2000.

[6] Wikipedia. Eco de espín, April 2017. `https://es.wikipedia.org/wiki/Eco_de_esp%C3%ADn`.

[7] Roland Keunings. Course notes *LINMA2720* modélisation mathématique de problèmes physiques. Technical report, Université Catholique de Louvain, 2016-2017.

[8] Thompson Jeff. Random walk: Square root of two, June 2017. `https://www.jeffreythompson.org/blog/2012/01/04/random-walk-square-root-of-two/`.

[9] Mark H Holmes. *Introduction to the foundations of applied mathematics*, volume 56. Springer Science & Business Media, 2009.

[10] P Mukherjee, JI Berman, SW Chung, CP Hess, and RG Henry. Diffusion tensor mr imaging and fiber tractography: theoretic underpinnings. *American journal of neuroradiology*, 29(4):632–641, 2008.

[11] Matt G Hall and Daniel C Alexander. Convergence and parameter choice for monte-carlo simulations of diffusion mri. *IEEE transactions on medical imaging*, 28(9):1354–1364, 2009.

[12] Rutger Fick, Marco Pizzolato, Demian Wassermann, and Rachid Deriche. Diffusion mri anisotropy: Modeling, analysis and interpretation. 2017.

[13] Edward O Stejskal and John E Tanner. Spin diffusion measurements: spin echoes in the presence of a time-dependent field gradient. *The journal of chemical physics*, 42(1):288–292, 1965.

[14] Paul T Callaghan. Pulsed-gradient spin-echo nmr for planar, cylindrical, and spherical pores under conditions of wall relaxation. *Journal of magnetic resonance, Series A*, 113(1):53–59, 1995.

[15] Benoit Scherrer and Simon K Warfield. Parametric representation of multiple white matter fascicles from cube and sphere diffusion mri. *PLoS one*, 7(11):e48232, 2012.

[16] Van J Wedeen, Patric Hagmann, Wen-Yih Isaac Tseng, Timothy G Reese, and Robert M Weisskoff. Mapping complex tissue architecture with diffusion spectrum magnetic resonance imaging. *Magnetic resonance in medicine*, 54(6):1377–1386, 2005.

[17] Rensonnet Gaëtan. *Generation of a dictionary of DW-MRI signals from arbitrary brain microstructure configurations.* Université Catholique de Louvain, 2015.

[18] Gaëtan Rensonnet, Damien Jacobs, Benoît Macq, and Maxime Taquet. A hybrid method for efficient and accurate simulations of diffusion compartment imaging signals. In *11th International Symposium on Medical Information Processing and Analysis (SIPAIM 2015)*, pages 968107–968107. International Society for Optics and Photonics, 2015.

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

[20] Nagesh Adluru, Hui Zhang, Andrew S Fox, Steven E Shelton, Chad M Ennis, Anne M Bartosic, Jonathan A Oler, Do PM Tromp, Elizabeth Zakszewski, James C Gee, et al. A diffusion tensor brain template for rhesus macaques. *Neuroimage*, 59(1):306–318, 2012.

[21] Ben Jeurissen, Alexander Leemans, Jacques-Donald Tournier, Derek K Jones, and Jan Sijbers. Investigating the prevalence of complex fiber configurations in white matter tissue with diffusion magnetic resonance imaging. *Human brain mapping*, 34(11):2747–2766, 2013.

[22] Benoit Scherrer, Maxime Taquet, and Simon K Warfield. Reliable selection of the number of fascicles in diffusion images by estimation of the generalization error. In *Information processing in medical imaging: proceedings of the... conference*, volume 23, page 742. NIH Public Access, 2013.

[23] Hui Zhang, Torben Schneider, Claudia A Wheeler-Kingshott, and Daniel C Alexander. Noddi: practical in vivo neurite orientation dispersion and density imaging of the human brain. *Neuroimage*, 61(4):1000–1016, 2012.

[24] Hui Zhang and Daniel Alexander. Axon diameter mapping in the presence of orientation dispersion with diffusion mri. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2010*, pages 640–647, 2010.

[25] Delaunay David. Géométrie des surfaces, June 2017. `http://mp.cpgedupuydelome.fr/cours.php?id=11965&idPartie=53695`.

[26] Benoit Scherrer, Armin Schwartzman, Maxime Taquet, Mustafa Sahin, Sanjay P Prabhu, and Simon K Warfield. Characterizing brain tissue by assessment of the distribution of anisotropic microstructural environments in diffusion-compartment imaging (diamond). *Magnetic resonance in medicine*, 2015.

[27] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. *The elements of statistical learning*, volume 1. Springer series in statistics Springer, Berlin, 2001.

[28] Robert J Tibshirani and B Efron. An introduction to the bootstrap. *Monographs on Statistics and Applied Probability*, 57:1–436, 1993.

[29] Bradley Efron. Estimating the error rate of a prediction rule: improvement on cross-validation. *Journal of the American statistical association*, 78(382):316–331, 1983.

[30] Kawin Setsompop, R Kimmlingen, E Eberlein, Thomas Witzel, Julien Cohen-Adad, Jennifer A McNab, Boris Keil, M Dylan Tisdall, P Hoecht, P Dietz, et al. Pushing the limits of in vivo diffusion mri for the human connectome project. *Neuroimage*, 80:220–233, 2013.

[31] Uran Ferizi, Benoit Scherrer, Torben Schneider, Mohammad Alipoor, Odin Eufracio, Rutger HJ Fick, Rachid Deriche, Markus Nilsson, Ana K Loya-Olivas, Mariano Rivera, et al. Diffusion mri microstructure models with in vivo human brain connectom data: results from a multi-group comparison. *arXiv preprint arXiv:1604.07287*, 2016.

[32] Tim B Dyrby, Matt G Hall, Maurice Ptito, Daniel Alexander, et al. Contrast and stability of the axon diameter index from microstructure imaging with diffusion mri. *Magnetic resonance in medicine*, 70(3):711–721, 2013.

[33] Jennifer A McNab, Brian L Edlow, Thomas Witzel, Susie Y Huang, Himanshu Bhat, Keith Heberlein, Thorsten Feiweier, Kecheng Liu, Boris Keil, Julien Cohen-Adad, et al. The human connectome project and beyond: initial applications of 300mt/m gradients. *Neuroimage*, 80:234–245, 2013.

[34] National Cancer Institute U.S. National Institutes of Health. Seer training modules, anatomy & physiology, nervous system, April 2017. `https://training.seer.cancer.gov/anatomy/nervous/tissue.html`.

Rue Archimède, 1 bte L6.11.01, 1348 Louvain-la-Neuve www.uclouvain.be/epl