Inverting brain grey matter models with likelihood-free inference: a tool for trustable cytoarchitecture measurements

Maëliss Jallais, Pedro L C Rodrigues, Alexandre Gramfort, Demian Wassermann

Inverting brain grey matter models with likelihood-free inference: a tool for trustable cytoarchitecture measurements

Maëliss Jallais maeliss.jallais@inria.fr
Université Paris-Saclay, Inria, CEA, Palaiseau, France
Pedro L. C. Rodrigues pedro.rodrigues@inria.fr
Université Paris-Saclay, Inria, CEA, Palaiseau, France
Alexandre Gramfort alexandre.gramfort@inria.fr
Université Paris-Saclay, Inria, CEA, Palaiseau, France
Demian Wassermann demian.wassermann@inria.fr
Université Paris-Saclay, Inria, CEA, Palaiseau, France

Abstract

Effective characterisation of the brain grey matter cytoarchitecture with quantitative sensitivity to soma density and volume remains an unsolved challenge in diffusion MRI (dMRI). Solving the problem of relating the dMRI signal with cytoarchitectural characteristics calls for the definition of a mathematical model that describes brain tissue via a handful of physiologically-relevant parameters and an algorithm for inverting the model. To address this issue, we propose a new forward model, specifically a new system of equations, requiring a few relatively sparse b-shells. We then apply modern tools from Bayesian analysis known as likelihood-free inference (LFI) to invert our proposed model. As opposed to other approaches from the literature, our algorithm yields not only an estimation of the parameter vector $\boldsymbol{\theta}$ that best describes a given observed data point $\boldsymbol{x}_0$, but also a full posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{x}_0)$ over the parameter space. This enables a richer description of the model inversion, providing indicators such as credible intervals for the estimated parameters and a complete characterization of the parameter regions where the model may present indeterminacies. We approximate the posterior distribution using deep neural density estimators, known as normalizing flows, and fit them using a set of repeated simulations from the forward model. We validate our approach on simulations using `dmipy` and then apply the whole pipeline on two publicly available datasets.

Keywords: Diffusion MRI, Brain Microstructure, Likelihood-Free Inference.

1. Introduction

Obtaining quantitative measurements of brain grey matter microstructure with a dedicated soma representation is a growing field of interest in the diffusion MRI (dMRI) community (Palombo et al., 2020; Jelescu et al., 2021). Unlike histology, dMRI permits to quantify brain tissue characteristics non-invasively and could, for example, help understanding dementia and cognitive deficits (Douaud et al., 2013). However, current methods require demanding acquisitions with several q-shells (equivalently b-shells) and rely on non-linear models for which several parameter values may yield the same observation, also known as parameter indeterminacy (Jelescu et al., 2016; Novikov et al., 2018b). This leads to numerically unstable solutions which are hard to interpret. Another major challenge is the fact

that quantifying brain tissue microstructure directly from the dMRI signal is an inherently difficult task, because of the large dimensionality of the collected data and its stochastic nature (Fick et al., 2016). For instance, each one of the $\approx 10^4$ voxels in a dMRI acquisition is sampled along nine dimensions, three directions in the q-space for each direction in the $\mathbb{R}^3$ spatial domain, which leads to very large data structures to process.

Many works have tackled the non-linear inverse problem of inferring brain tissue parameters from dMRI signal. A popular solution is NODDI (e.g. Zhang et al., 2012), which stabilises the solution by imposing constraints on model parameters that are not biologically plausible (Novikov et al., 2018b). However, this has the effect of biasing the parameter estimation and the inverse problem remains largely degenerate (Jelescu et al., 2016). There has also been some interest in applying methods from the machine learning literature to solve the inverse problem. This is the case of the SANDI model (Palombo et al., 2020), in which the authors employ random forest regressors. Although the method provides rather acceptable accuracy in real-case scenarios, it can only output one set of tissue parameters for a given observed dMRI signal, masking, therefore, other biologically plausible solutions that could generate the same observed signal.

To overcome such limitations, we present three contributions. First, we use a three-compartment model for brain tissue and introduce a new parameter that jointly encodes soma radius and inner diffusivity without imposing constraints on these values. This new parameter reduces indeterminacies in the model and has relevant physiological interpretations. Second, we present a method to fit the model through summary features of the dMRI signal based on a large and small q-value analysis using boundary approximations. These rotationally-invariant features relate directly to the tissue parameters and enable us to invert the model without manipulating the raw dMRI signals. Such summary statistics ensure a stable solution of the parameter estimations, as opposed to the indeterminate models used in Zhang et al. (2012); Palombo et al. (2020). Third, we employ modern tools from Bayesian analysis known as likelihood-free inference (LFI, (Cranmer et al., 2020)) to solve our non-linear inverse problem under a probabilistic framework and determine the posterior distribution of the fitted parameters. Such approach offers a full description of the solution landscape and can point out degeneracies, as opposed to the usual deterministic least-squares based solution (Jelescu et al., 2016; Novikov et al., 2018b).

2. Related Works

Current brain tissue models are predominantly based on the two compartment Standard Model (SM) (Zhang et al., 2012; Novikov et al., 2018a). Recent evidence shows that the SM, mainly used in white matter, does not hold for grey matter microstructure analysis (Veraart et al., 2020). Several assumptions aim at explaining this issue such as increased permeability in neurite membranes (Veraart et al., 2020), or curvy projections along with longer pulse duration (Novikov et al., 2018a). We follow the hypothesis that the SM doesn't hold due to an abundance of cell bodies in grey matter (Palombo et al., 2020). Our proposed biophysical model is then based on three compartments (Palombo et al., 2020): neurites, somas, and extra-cellular space (ECS). Despite its increased complexity, the main advantage of such model is the possibility to jointly estimate the characteristic features of each compartment.

Inferring parameters of the brain tissue model directly from dMRI signals has proven to be a very challenging task and has motivated the development of new approaches to reduce the dimensionality of the data to be processed. For instance, Novikov et al. (2018b) proposes the LEMONADE system of equations, based on a Taylor expansion of the diffusion signal and a set of rotationally invariant moments. In this work, we extend the LEMONADE equations to a setting with a three compartment model and further develop the method so to extract more features from the observed signal. We call the resulting vector of features defined by these quantities the 'summary statistics' of the dMRI signal.

The usual way of applying a Bayesian approach to solve non-linear inverse problems (Stuart, 2010) is to define two quantities: a prior distribution encoding initial knowledge of the parameter values (e.g. intervals which are physiologically relevant) and the likelihood function of the forward model being studied. One can then either obtain an analytic expression of the posterior distribution via Bayes' formula or use a Markov-Chain Monte Carlo (MCMC) procedure to numerically sample the posterior distribution (Gelman et al., 2013). However, the likelihood function of complex models such as the one that we consider here is often very hard to obtain and makes the Bayesian approach rather challenging to use. Likelihood-free inference (LFI) bypasses this bottleneck by recurring to several simulations of the forward model using different parameters and learning an approximation to the posterior distribution from these examples (Cranmer et al., 2020).

The first contributions on LFI are also known as approximate Bayesian computation (ABC) and have been applied to invert models from ecology, population genetics, and epidemiology (Sisson, 2018). Some of the limitations of these techniques include the large number of simulations required for the posterior estimations and the need of defining a distance function to compare the results of two simulations. Recently, there has been a growing interest in the machine learning community in improving the limitations of ABC methods through deep generative modeling, i.e. neural network architectures specially tailored to approximate probability density functions from a set of examples (Goodfellow et al., 2016). Normalizing flows (Papamakarios et al., 2019) are a particular class of such neural networks that have demonstrated promising results for likelihood-free inference in different research fields (Cranmer et al., 2020; Gonçalves et al., 2020; Greenberg et al., 2019).

3. Theory

In this section, we present the theoretical background underlying our three main contributions. **Section 3.1** describes the three-compartment model for brain grey matter. We also introduce a new parameter that captures both soma radius and inner diffusivity and avoids the usual indeterminacy in estimating them separately. **Section 3.2** presents the summary statistics used to reduce the dimensionality of the dMRI signal into a 7-dimensional feature vector that can be used to determine the physiological parameters of interest. **Section 3.3** describes the Bayesian approach used to solve our non-linear inverse problem and obtain a posterior distribution for the physiological parameters that best describe a given dMRI signal.

3.1 Modeling the Brain grey Matter with a 3-compartment Model

To characterize cortical cytoarchitecture, we propose a method that relates the diffusion MRI signal to specific tissue parameters. To that aim, we first define a forward model based on a biophysical modeling of brain grey matter.

Research in histology has demonstrated that grey matter is composed of neurons embedded in a fluid environment. Each neuron is composed of a soma, corresponding to the cellular body, surrounded by neurites connecting neurons together. Following this tissue biophysical composition, we model the grey matter tissue as three-compartmental (Palombo et al., 2020), moving away from the usual standard model (SM) designed for white matter. We further assume that our acquisition protocol is insensitive to exchanges between the compartments, i.e. molecules moving from one compartment to another have a negligible influence on the signal (Palombo et al., 2020). Many works include also a dot compartment into the SM with zero apparent diffusivity and no exchange. However, we have not considered such assumption, because its presence has been considered very unlikely in grey matter by previous works (Veraart et al., 2019; Tax et al., 2020). The observed diffusion signal is considered as a convex mixture of signals arising from somas, neurites, and extra-cellular space (ECS). Unlike white matter-centric methods (Jelescu and Budde, 2017), we are not interested in the fiber orientation and only estimate orientation-independent parameters. This enables us to work on the direction-averaged dMRI signal, denoted $\bar{S}(q)$, known as the power averaged signal. This consideration mainly matters for neurites, as their signal is not isotropic, as opposed to the proposed model for somas and ECS. Our direction-averaged grey matter signal model is then:

$$\frac{\bar{S}(q)}{S(0)} = f_n \bar{S}_{\text{neurites}}(q, D_n) + f_s \bar{S}_{\text{somas}}(q, D_s, r_s) + f_{\text{ECS}} \bar{S}_{\text{ECS}}(q, D_e) \ . \tag{1}$$

In this equation, f_n, f_s, and f_{ECS} represent signal fractions for neurites, somas, and ECS respectively ($f_n + f_s + f_{\text{ECS}} = 1$). Note that the relative signal fractions do not correspond to the relative volume fractions of the tissue compartments as they are also modulated by different T2 values (Novikov et al., 2018a). D_n corresponds to axial diffusivity inside neurites, while D_s and D_e correspond to somas and extra-cellular diffusivities. r_s is the average soma radius within a voxel. This model is the same as the model SANDI proposed by Palombo et al. (2020), with $f_n = f_{ic} f_{in}$, $f_s = f_{ic} f_{is}$ and $f_{\text{ECS}} = f_{ec}$. We use q-values for more readability and harmonization throughout the paper, but a direct conversion to b-values is also possible, using $b = (2\pi q)^2 \tau$ with $\tau = \Delta - \delta/3$.

We now review the model for each compartment, to make explicit the impact of each parameter on the diffusion MRI signal.

Neurite compartment. Neurites, as in the SM, are modeled as 0-radius impermeable cylinders ("sticks"), with effective diffusion along the parallel axis, and a negligible radial intra-neurite diffusivity. In our acquisition setting, this model has been shown to hold (Veraart et al., 2020). Its direction averaged signal is (Veraart et al., 2020):

$$\bar{S}_{\text{neurites}}(q) \simeq \frac{1}{4\sqrt{\pi \tau D_n}} \cdot q^{-1} \ . \tag{2}$$

Soma compartment. Somas are modeled as spheres, whose signal can be computed using the GPD approximation (Balinov et al., 1993):

$$-\log \bar{S}_{\text{somas}}(q) = C(r_s, D_s) \cdot q^2 \ . \tag{3}$$

We exploit this relation here to extract a parameter $C_s = C(r_s, D_s)[m^2]$ which, at fixed diffusivity D_s, is modulated by the radius of the soma r_s:

$$C(r_s, D_s) = \frac{2}{D_s \delta^2} \sum_{m=1}^{\infty} \frac{\alpha_m^{-4}}{\alpha_m^2 {r_s}^2 - 2} \cdot \left(2\delta - \frac{2 + e^{-\alpha_m^2 D_s(\Delta-\delta)} - e^{-\alpha_m^2 D_s \delta} - e^{-\alpha_m^2 D_s \Delta} + e^{-\alpha_m^2 D_s(\Delta+\delta)}}{\alpha_m^2 D_s} \right) ,$$

where α_m is the mth root of $(\alpha r_s)^{-1} J_{\frac{3}{2}}(\alpha r_s) = J_{\frac{5}{2}}(\alpha r_s)$, with $J_n(x)$ the Bessel functions of the first kind. In certain specific cases, C_s has a simpler and more interpretable expression. For instance, when we consider a narrow pulse regime, with small τ or large r_s, C_s loses its dependence on r_s (Balinov et al., 1993) and we obtain:

$$C_s = D_s \tau \ . \tag{4}$$

In the Neuman (wide pulse) regime, i.e. when $D_s \Delta \gg r_s^2$ and $D_s \delta \ll 1$, C_s becomes only dependent on the soma radius (Murday and Cotts, 1968):

$$C_s = \frac{1}{5} r_s^2 \ . \tag{5}$$

These two approximations permits us to better interpret the parameter C_s.

Extra-cellular space compartment. The extra-cellular space is approximated as isotropic Gaussian diffusion, i.e. a mono-exponential diffusion signal with a scalar diffusion constant D_e, which reflects the molecular viscosity of the fluid. This approximation assumes that the ECS is fully connected. The approximation is therefore:

$$-\log(\bar{S}_{\text{ECS}}(q)) = (2\pi q)^2 \tau D_e \ . \tag{6}$$

Because of the geometry of the problem, we estimate D_e as equal to one third of the diffusivity in the ventricles (considered as free diffusivity), given the same metabolic composition of the extracellular fluid and ventricles (Vincent et al., 2021).

3.2 An Invertible 3-compartment Model: dMRI Summary Statistics

The tissue model presented in Section 3.1 enables us to relate the dMRI signal with parameters representing grey matter tissue microstructure. However, solving the inverse problem directly from Eq. (1) is a difficult task, leading to indeterminacies and poorly estimated parameters with large variability. Current methods addressing this issue have not studied its stability (Palombo et al., 2020) but simpler models with only two compartments have been shown to be indeterminate (Novikov et al., 2018a).

To produce a method which addresses this indeterminacy, we introduce rotationally invariant summary statistics to describe the dMRI signal. The goal being to reduce the dimensionality of the data at hand and representing all the relevant information for statistical

inference with a few features. We then solve the inverse problem efficiently using Bayesian inference as described in Section 3.3. These dMRI-based summary statistics are extracted from our proposed model presented in Section 3.1 via the following analysis of the dMRI signal on the boundaries of large and small q-value cases.

Large q-value approximation: RTOP. We compute a q-bounded return-to-the-origin probability (RTOP), which measures the restrictions of the diffusing fluid molecule motion and gives us information about the structure of the media (Mitra et al., 1995):

$$\mathrm{RTOP}(q) = 4\pi \int_0^q \frac{\bar{S}(\eta)}{S(0)} \eta^2 d\eta \ . \tag{7}$$

For q large enough, the RTOP in our 3-compartment model in Eq. (1) yields a soma and extra-cellular signal which converges towards a constant value in q, while the neurites' contribution becomes quadratic in q. In this case, RTOP becomes:

$$\mathrm{RTOP}(q) = \underbrace{f_s \left(\frac{\pi}{C_s}\right)^{3/2} + \frac{f_{\mathrm{ECS}}}{8(\pi\tau D_e)^{3/2}}}_{a_{\mathrm{fit}}} + \underbrace{\frac{f_n}{2} \cdot \sqrt{\frac{\pi}{\tau D_n}}}_{b_{\mathrm{fit}}} \cdot q^2 + \gamma q^3 \ , \tag{8}$$

in which the last term of the equation, γq^3, is a nuisance parameter that describes a constant noise in the direction averaged signal. By accurately estimating the second derivative of $\mathrm{RTOP}(q)$ at q large enough, we can solve the coefficients of interest of the polynomial in Eq. (8): a_{fit} and b_{fit}. We do this efficiently by casting it as an overdetermined ordinary least squares problem which has a unique solution.

Small q-value approximation: Spiked LEMONADE. We propose a second approximation, based on a moment decomposition for small q-values (Novikov et al., 2018b):

$$\frac{S_{\hat{\mathbf{g}}}(q)}{S(0)} = 1 - b(q) M^{(2)}_{i_1 i_2} g_1 g_2 + \frac{b(q)^2}{2!} M^{(4)}_{i_1 \ldots i_4} g_1 \ldots g_4 - \ldots, \quad b(q) = (2\pi q)^2 \tau \ , \tag{9}$$

where i_k are the directional basis of the tensors M, $g_k = i_k \cdot \hat{\mathbf{g}} \in \mathbb{R}^3$, and $\hat{\mathbf{g}}$ the unit direction of the dMRI acquisition. From the moment tensors of this decomposition, LEMONADE (Novikov et al., 2018b) extracts rotational invariant scalar indices $M^{(i),j}$, $i, j \in \{0, 2, 4, \ldots\}$. These quantify white matter microstructure by plugging the 2-compartment SM into Eq. (9) (see Novikov et al., 2018b, app. C).

In this work, we extended LEMONADE to our 3-compartment model presented in Section 3.1 by plugging Eq. (1) into Eq. (9) and performing tedious arithmetic. This results in the following equation system, which now includes the soma parameter C_s, relating the dMRI signal to grey matter microstructure:

$$\begin{cases} M^{(2),0} = f_n D_n + 3 f_s \frac{C_s}{(2\pi)^2 \tau} + 3 f_{\mathrm{ECS}} D_e \\ M^{(2),2} = f_n D_n p_2 \\ M^{(4),0} = f_n {D_n}^2 + 5 f_s \left(\frac{C_s}{(2\pi)^2 \tau}\right)^2 + 5 f_{\mathrm{ECS}} D_e^2 \\ M^{(4),2} = f_n {D_n}^2 p_2 \end{cases} \tag{10}$$

where p_2 is a scalar measure of neurite orientation dispersion (Novikov et al., 2018b).

Note that only the shells with $b(q) \leq 2.5\,\mathrm{ms}\,\mu\mathrm{m}^{-2}$ are used, to get an unbiased estimation of the rotational invariant moments $M^{(2),0}$, $M^{(2),2}$, $M^{(4),0}$ and $M^{(4),2}$.

Complete system. Combining equations (8) and (10) and adding the constraint that the fractions for the three compartments sum to one, we obtain a non-linear system of 7 equations and 7 unknowns. Following Menon et al. (2020), we assume D_e nearly-constant per subject acquisition and estimated it as one third of the mean diffusivity in the subject's ventricles (Vincent et al., 2021). This assumption allows us to drop an unknown from the system, use D_e as a reference diffusivity and render our variables unitless as $D_n^u = \frac{D_n}{D_e}$ and $C_s^u = \frac{C_s}{(2\pi)^2 \tau D_e}$, obtaining:

Spiked LEMONADE Small q-values	**RTOP** Large q-values
$\begin{cases} \frac{M^{(2),0}}{D_e} = f_n D_n^u + 3 f_s C_s^u + 3 f_{\mathrm{ECS}} \\ \frac{M^{(2),2}}{D_e} = f_n D_n^u \cdot p_2 \\ \frac{M^{(4),0}}{D_e^2} = f_n D_n^{u2} + 5 f_s C_s^{u2} + 5 f_{\mathrm{ECS}} \\ \frac{M^{(4),2}}{D_e^2} = f_n D_n^{u2} \cdot p_2 \end{cases}$	$\begin{cases} a_{\mathrm{fit}} (\tau D_e)^{3/2} = \frac{f_s}{8(\pi C_s^u)^{3/2}} + \frac{f_{\mathrm{ECS}}}{8\pi^{3/2}} \\ b_{\mathrm{fit}} (\tau D_e)^{1/2} = \frac{f_n}{2}\sqrt{\frac{\pi}{D_n^u}} \end{cases}$
$f_n + f_s + f_{\mathrm{ECS}} = 1$	

3.3 Inverting the model with Bayesian inference

Our main goal is to determine the values of the parameter vector

$$\boldsymbol{\theta} = (D_n^u, C_s^u, p_2, f_s, f_n, f_{\mathrm{ECS}}) \in \mathbb{R}^6$$

that best explain a given observed dMRI signal d. Because of the high-dimensionality of this kind of signal and the difficulties in obtaining stable estimates of $\boldsymbol{\theta}$ directly from it, we recur to the set of summary features defined in Section 3.2,

$$\boldsymbol{x} = \left(\frac{M^{(2),0}}{D_e}, \frac{M^{(2),2}}{D_e}, \frac{M^{(4),0}}{D_e^2}, \frac{M^{(4),2}}{D_e^2}, a_{\mathrm{fit}} (\tau D_e)^{3/2}, b_{\mathrm{fit}} \sqrt{\tau D_e} \right) \in \mathbb{R}^7 \ ,$$

and make the assumption that it carries all the information necessary for determining the $\boldsymbol{\theta}_0$ having generated a given dMRI signal $\boldsymbol{S}_0$. We denote the relation between these quantities as

$$\boldsymbol{x} = \mathcal{M}(\boldsymbol{\theta}) + \boldsymbol{n} \ , \tag{11}$$

where $\mathcal{M} : \mathbb{R}^6 \to \mathbb{R}^7$ is a multivariate function that implements the system of equations described in Section 3.2 and $\boldsymbol{n}$ is an additive noise capturing the imperfections of our modelling procedure, the limitations of the summary statistics, and the measurement noise.

The Bayesian formalism. We interpret the inverse problem of inferring the parameters that best describe a given observed summary feature vector $\boldsymbol{x}_0$ as that of determining the posterior distribution of $\boldsymbol{\theta}$ given an observation $\boldsymbol{x}_0$. By first choosing a prior distribution $p(\boldsymbol{\theta})$ describing our initial knowledge of the parameter values, we may use Bayes' theorem to write

$$p(\boldsymbol{\theta}|\boldsymbol{x}_0) = \frac{p(\boldsymbol{x}_0|\boldsymbol{\theta}) p(\boldsymbol{\theta})}{p(\boldsymbol{x}_0)} \ , \tag{12}$$

where $p(\boldsymbol{x}_0|\boldsymbol{\theta})$ is the likelihood of the observed data point and $p(\boldsymbol{x}_0)$ is a normalizing constant, commonly referred to as the evidence of the data. Note that such a probabilistic approach returns not only which $\boldsymbol{\theta}$ best fits the observed data (i.e. the parameter that maximizes the posterior distribution), but the full posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{x}_0)$. The latter can be possibly multimodal or flat, which would indicate the difficulty of summarizing it with a unique maximum.

Bypassing the likelihood function. Despite its apparent simplicity, it is usually difficult to use Eq. (12) to determine the posterior distribution, since the likelihood function for data points generated by complex non-linear models is often hard to write. To avoid such difficulty, we directly approximate the posterior distribution using a conditional density estimator, i.e. a family of conditional p.d.f. approximators $q_\phi(\boldsymbol{\theta}|\boldsymbol{x})$ parametrized by ϕ and that takes $\boldsymbol{\theta}$ (the parameter) and $\boldsymbol{x}$ (the observation) as input arguments. Our posterior approximation is then obtained by minimizing its average Kullback-Leibler divergence with respect to the conditional density estimator for different choices of $\boldsymbol{x}$, as per (Papamakarios and Murray, 2016)

$$\min_{\boldsymbol{\phi}} \quad \mathcal{L}(\phi) \quad \text{with} \quad \mathcal{L}(\phi) = \mathbb{E}_{\boldsymbol{x}\sim p(\boldsymbol{x})}\left[D_{\mathrm{KL}}(p(\boldsymbol{\theta}|\boldsymbol{x})\|q_\phi(\boldsymbol{\theta}|\boldsymbol{x}))\right] \ , \tag{13}$$

which can be rewritten as

$$\begin{aligned}
\mathcal{L}(\phi) &= \int D_{\mathrm{KL}}(p(\boldsymbol{\theta}|\boldsymbol{x})\|q_\phi(\boldsymbol{\theta}|\boldsymbol{x}))p(\boldsymbol{x})\mathrm{d}\boldsymbol{x} \ , \\
&= -\iint \log\big(q_\phi(\boldsymbol{\theta}|\boldsymbol{x})\big)p(\boldsymbol{\theta}|\boldsymbol{x})p(\boldsymbol{x})\mathrm{d}\boldsymbol{\theta}\mathrm{d}\boldsymbol{x} + C \ , \\
&= -\iint \log\big(q_\phi(\boldsymbol{\theta}|\boldsymbol{x})\big)p(\boldsymbol{x},\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}\mathrm{d}\boldsymbol{x} + C \ , \\
&= -\mathbb{E}_{(\boldsymbol{x},\boldsymbol{\theta})\sim p(\boldsymbol{x},\boldsymbol{\theta})}\left[\log\big(q_\phi(\boldsymbol{\theta}|\boldsymbol{x})\big)\right] + C \ ,
\end{aligned} \tag{14}$$

where C is a constant that does not depend on ϕ. Note, however, that in practice we actually consider a N-sample Monte-Carlo approximation of the loss function given by

$$\mathcal{L}(\phi) \approx \mathcal{L}^N(\phi) = -\frac{1}{N}\sum_{i=1}^{N} \log\Big(q_\phi(\boldsymbol{\theta}_i|\boldsymbol{x}_i)\Big) \ , \tag{15}$$

where the N data points $(\boldsymbol{\theta}_i, \boldsymbol{x}_i)$ are sampled from the joint distribution with $\boldsymbol{\theta}_i \sim p(\boldsymbol{\theta})$ and $\boldsymbol{x}_i \sim p(\boldsymbol{x}|\boldsymbol{\theta}_i)$. We can then use stochastic gradient descent to obtain a set of parameters ϕ which minimizes $\mathcal{L}^N$. If the class of conditional density estimators is sufficiently expressive, it is possible to show that the minimizer of Eq. (15) converges to $p(\boldsymbol{\theta}|\boldsymbol{y})$ when $N \to \infty$ (Greenberg et al., 2019). Note, also, that the parametrization ϕ that we obtain by the end of the optimization procedure yields a posterior which is amortized for different choices of $\boldsymbol{x}$. Thus, for a specific observation $\boldsymbol{x}_0$ we may simply write $q_\phi(\boldsymbol{\theta}|\boldsymbol{x}_0)$ to get an approximation of $p(\boldsymbol{\theta}|\boldsymbol{x}_0)$.

Neural density estimators. In this work, our conditional p.d.f. approximators belong to a class of neural networks called normalizing flows (Papamakarios et al., 2019). These flows are invertible functions capable of transforming vectors generated by a simple base distribution (e.g. the standard multivariate Gaussian distribution) into an approximation

of the true posterior distribution. An important advantage of normalizing flow versus other p.d.f. approximators such as generative adversarial network (GAN (Goodfellow et al., 2014)) and variational auto-encoders (VAE (Kingma and Welling, 2014)) is that it provides both the likelihood of any sample point of interest and it is also straightforward to sample new data points from it. Furthermore, certain classes of normalizing flows can be shown to be universal approximators of probability density functions. We refer the reader to Papamakarios et al. (2019) for more information on the different types of normalizing flows.

4. Materials and methods

This section presents the technical details on how we have implemented our theoretical contributions and describes the simulated and real datasets used in the numerical illustrations.

4.1 The likelihood-free inference setup

Using a likelihood-free inference approach for inverting the 3-compartment model relating tissue parameters $\boldsymbol{\theta}$ and dMRI summary statistics $\boldsymbol{x}$ relies on four important aspects:

(1) **The forward model.** As explained in Section 3.3, we obtain an approximation of the amortized posterior distribution using a dataset containing several paired examples of a parameter $\boldsymbol{\theta}_i$ and its corresponding summary statistics $\boldsymbol{x}_i$. In what follows, we adopt the usual assumptions from the inverse problems literature and consider the additive noise $\boldsymbol{n}$ from Eq. (11) small enough to be ignored in the data generation, so that we have $\boldsymbol{x}_i \approx \mathcal{M}(\boldsymbol{\theta}_i)$.

(2) **Prior distribution.** The simplest way of defining a prior distribution $p(\boldsymbol{\theta})$ in our setting is to use an uniform distribution with limits within physiologically relevant intervals for each parameter. From Section 3.1, we know that the fractions f_s, f_n, and f_{ECS} have values between 0 and 1 and all sum up to one. To encode this information in $p(\boldsymbol{\theta})$, we define new parameters k_1 and k_2 and relate them with the fractions by

$$f_n = k_2\sqrt{k_1}, \quad f_s = \sqrt{k_1}(1-k_2), \quad f_{\text{ECS}} = 1-\sqrt{k_1}\ . \tag{16}$$

In this way, whenever we want to generate a prior sample for $(f_n, f_s, f_{\text{ECS}})$ we first generate a sample of $k_1, k_2 \sim \mathcal{U}([0,1])$ and then transform them according to (16) to get a set of fractions which is uniformly sampled in the region $\{f_n, f_s, f_{\text{ECS}} \in [0,1]\ : f_n + f_s + f_{\text{ECS}} = 1\}$. We follow the usual assumption that the diffusivity of the compartments are inferior or equal to the self-diffusion coefficient of free water, which is $3\,\mu\text{m}\,\text{ms}^{-2}$ (Li et al., 2016). We fix, therefore, the interval for neurite diffusivity (D_n) as between 10^{-5} and $3\ \mu\text{m}\,\text{ms}^{-2}$. The newly introduced parameter C_s is parametrized by soma radius and diffusivity. To account for a soma radius comprised between 2 and $30\,\mu\text{m}$ (Palombo et al.) and a diffusivity range as previously defined, we used the C_s interval [50, 2500] μm^2. Parameter p_2, which measures the dispersion of neurites orientation, is comprised in the interval [0,1], with 1 indicating an anisotropic orientation distribution function.

(3) **Posterior approximator.** We use an autoregressive architecture for normalizing flows implemented via the masked autoencoder for distribution estimation (MADE) (Ger-

main et al., 2015). We follow the same setup from Greenberg et al. (2019) for LFI problems, stacking five MADEs, each with two hidden layers of 50 units, and a standard normal base distribution. This choice provides a sufficiently flexible function capable of approximating complex posterior distributions.

(4) **Training procedure.** The parametrization of our normalizing flow is obtained by minimizing the loss function (15) using the ADAM optimizer (Kingma and Ba, 2017) with default parameters, a learning rate of 5.10^{-4} and a batch size of 100. Except for a few validation experiments, we have used $N = 10^5$ simulated data points to approximate the posterior distribution.

4.2 Simulated dMRI data

We first validate our proposed method using simulated dMRI data. For this, we fix a parameter vector $\boldsymbol{\theta}_0$ based on a plausible biophysical configuration (Palombo et al.) and generate a simulated observation $\boldsymbol{x}_0$ associated to it. Our goal, then, is to check whether the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{x}_0)$ concentrates around the ground truth parameter, i.e. if it is peaked around the true values of the parameters in $\boldsymbol{\theta}_0$. If this is the case, we can assert that the LFI procedure is capable of inverting our non-linear model successfully.

The simplest way of generating an observation from the ground truth parameter would be to use the forward model defined in Section 4.1. However, such validation could yield unrealistically good results, since the posterior approximation is trained on data points generated in the same way. We have decided, therefore, to use a different approach in which the dMRI signals are simulated following a setup which is closer to what we would expect from real experimental experiments. This is based on two steps. Firstly, we use the `dmipy` simulator (Fick et al., 2019) to simulate the three-compartment model described in Section 3.1 and obtain a dMRI signal $\boldsymbol{S}_0$. Then, we calculate the summary statistics of this signal as defined in Section 3.2 to reduce the dimensionality of the observation and obtain a feature vector $\boldsymbol{x}_0$.

We have carried out our simulations on `dmipy` considering three different kinds of acquisition setup. They all have b-shells with 128 uniformly distributed directions, but they differ in their b-values and acquisition times:

- Setup `Ideal` corresponds to a rather "confortable" case with 10 b-values between 0 and $10\,\mathrm{ms}\,\mu\mathrm{m}^{-2}$. We use $\delta/\Delta = 12.9/21.8\,\mathrm{ms}$ as in the HCP MGH database.
- Setup `HCP MGH` reproduces the setup from the HCP MGH dataset, with 5 b-values: 0, 1, 3, 5, and $10\,\mathrm{ms}\,\mu\mathrm{m}^{-2}$ and $\delta/\Delta = 12.9/21.8\,\mathrm{ms}$. Since the Spiked LEMONADE approximation (3.2) requires at least three b-values inferior to $2.5\,\mathrm{ms}\,\mu\mathrm{m}^{-2}$, we extrapolated an extra b-shell at $0.1\,\mathrm{ms}\,\mu\mathrm{m}^{-2}$ using `MAPL` (Fick et al., 2016).
- Setup `HCP 1200` reproduces the setup from the HCP 1200 dataset, with only 3 b-values: 1, 2, and $3\,\mathrm{ms}\,\mu\mathrm{m}^{-2}$ and $\delta/\Delta = 10.6/43.1\,\mathrm{ms}$ An extra b-shell at $2.8\,\mathrm{ms}\,\mu\mathrm{m}^{-2}$ has been interpolated to be used in the RTOP approximation.

Note that in the simulations with all setups we have used the three b-shells with the lowest b-values for the small q-value approximation (Spiked LEMONADE), and the three largest q-values for the RTOP approximation.

4.3 HCP MGH dataset

After validating our proposal on different simulated settings, we carried out our analysis on two publicly available databases. Our goal was to estimate the brain tissue parameters for each voxel in the dMRI acquisitions and determine how these parameters vary accross the human brain grey matter. Because of the probabilistic framework that we use, these estimates are accompanied of credible intervals that can be used to inform our degree of confidence of these estimates.

Our first analysis was on the HCP MGH Adult Diffusion database (Setsompop et al., 2013). This database is composed of 35 subjects with $\delta/\Delta = 12.9/21.8\,\text{ms}$ and $b = 1$, 3, 5, $10\,\text{ms}\,\mu\text{m}^{-2}$. We used the 3, 5 and $10\,\text{ms}\,\mu\text{m}^{-2}$ b-values for the RTOP approximation (i.e. the large q-value analysis). For the Spiked LEMONADE approximation, we fitted `MAPL` on the 0, 1 and $3\,\text{ms}\,\mu\text{m}^{-2}$ b-values to reduce noise, and interpolated a b-value of $0.1\,\text{ms}\,\mu\text{m}^{-2}$ to improve the estimations. D_e was estimated as 1/3 of the mean diffusivity in the ventricles.

The spatial distribution of the estimated parameters were mapped to the MNI template, averaged over all subjects, and then projected onto an inflated cortical surface using `FreeSurfer`. We have then evaluated whether such distributions seemed physiologically plausible by using the Brodmann atlas (Brodmann, 1909; Zilles, 2018), which is a parcellation of the brain based on cytoarchitecture features. In addition, we compared qualitatively the results of soma estimations with nissl-stained histological images of cytoarchitecture (Allman et al., 2010; Amunts et al., 1999; Geyer et al., 1999).

4.4 HCP 1200

We proceeded with our analysis on real data using a more challenging database, in which the dMRI signals were acquired with only a few b-values. Our goal was to demonstrate that the credible regions obtained via the posterior approximation can be used to inform which parameters remain possible to estimate even in very challenging situations. Note that this unlocks the door to the analysis of any dMRI database, since the estimates always come with a "quality certificate".

We applied our pipeline to a subset of the HCP 1200 database. We randomly picked 30 subjects, to have an identical number of subjects to that in our analysis of the HCP MGH database. The data were acquired for b-values equal to 1, 2 and $3\,\text{ms}\,\mu\text{m}^{-2}$, with $\delta/\Delta = 10.6/43.1\,\text{ms}$. Using `MAPL`, we interpolated a b-shell at $2.8\,\text{ms}\,\mu\text{m}^{-2}$ to improve the computation of the summary statistics. We used all the three lowest b-values for the Spiked LEMONADE approximation, and $b = 2$, 2.8 and $3\,\text{ms}\,\mu\text{m}^{-2}$ for the high b-value approximation based on RTOP. Similarly to the HCP MGH dataset, we have averaged the parameter estimations in a common space (MNI) and then projected the resulting parametric maps onto an inflated cortical surface.

5. Results

In this section, we describe our results obtained on simulated data and two real datasets. All experiments have been implemented with Python (Python Software Fundation, 2017) using several scientific packages: dMRI signals were simulated with the package `dmipy` (Fick et al., 2019) and processed using `dipy` (Garyfallidis et al., 2014) or custom implementations

based on `numpy` (Harris et al., 2020). We used the `sbi` (Tejero-Cantero et al., 2020) and `nflows` (Durkan et al., 2020) packages for carrying out the LFI procedures and combined them with data structures and functions from `pyTorch` (Paszke et al., 2019). The figures of results on real experimental data were generated with `mayavi` (Ramachandran and Varoquaux, 2011).

5.1 Simulated data

Validating the LFI pipeline. In this first round of experiments, our aim was to check whether the LFI pipeline worked correctly on a setting where we knew the true values of the parameter $\boldsymbol{\theta}_0$ (ground truth) generating the observed data $\boldsymbol{x}_0$. Furthermore, we wanted to confirm whether the use of summary statistics for the dMRI signal actually conveyed any improvements in the parameter estimation. We have considered, therefore, three different cases:

- **Case 1.** Generate $\boldsymbol{x}_0$ directly from $\boldsymbol{\theta}_0$ using the forward model defined in Section 4.1. This is a rather favorable case for our posterior approximation, since it is applied on a data point generated in the same way as the dataset in which it was trained.
- **Case 2.** Generate a dMRI signal $\boldsymbol{S}_0$ from $\boldsymbol{\theta_0}$ using `dmipy` and then obtain $\boldsymbol{x}_0$ by calculating the summary statistics presented in Section 3.2. We use the same posterior approximator from **Case 1**, meaning that the data point in inference time is generated differently from those for the training procedure. This is the actual realistic case that interests us the most.
- **Case 3.** Generate a dMRI signal $\boldsymbol{S}_0$ from $\boldsymbol{\theta_0}$ using `dmipy` and do not use any summary statistics for the model inversion, i.e. consider $\boldsymbol{x}_0 = \boldsymbol{S}_0$. Note that the posterior approximator has to be trained on a dataset with observations consisting of dMRI signals, so it is different from the approximators in **Case 1** and **Case 2**. Depending on how the LFI pipeline behaves on this case, the use of summary statistics can be justified or not.

All simulations were carried out with setup `Ideal`, which corresponds to an ideal dMRI acquisition scheme, and the posterior approximators were trained with $N = 10^5$ simulated data points. While we have validated the LFI pipeline on multiple choices of physiologically relevant ground truth parameters $\boldsymbol{\theta}_0$, Figure 1 portrays the results only for

$$\boldsymbol{\theta}_0 = (D_n, C_s, p_2, f_s, f_n, f_{\text{ECS}}) = (2.5\ \mu\text{m}\,\text{ms}^{-2}, 617\ \mu\text{m}^2, 0.50, 0.15, 0.45, 0.40)\ .$$

This choice of parameters represent a sample tissue containing pyramidal neurons of radius $12\,\mu$m (Palombo et al.) and diffusivity $3\,\mu\text{m}\,\text{ms}^{-2}$. The three compartment proportions were chosen in accordance to reported values from histology of human grey matter tissue (Shapson-Coe et al., 2021). Results are presented in Figure 1. We see that in **Case 1** the marginals of the posterior distribution are well concentrated around the values of the ground truth parameter $\boldsymbol{\theta}_0$. This confirms that the posterior approximator successfully inverts the non-linear model using the examples in the training set. We also see that the parameter estimation in **Case 2** captures very well most of the true values of the parameters, indicating both that our posterior approximator is robust to observed data generated

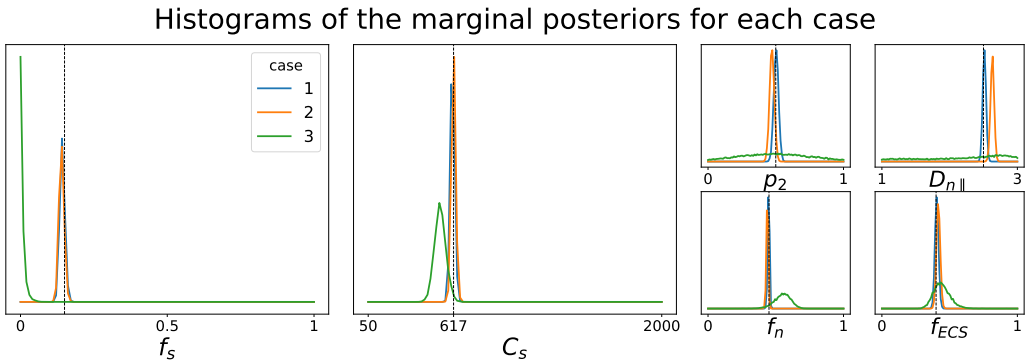

Figure 1: Histograms of 10^4 samples of the approximate posterior distribution in three different base cases (see text for details). Vertical black dashed lines represent ground truth values of $\boldsymbol{\theta}_0$ which generated the observed dMRI signals. We observe that the marginals tend to concentrate around the ground truth parameters when the observed summary statistics are obtained directly from the parameters (**Case 1**) and have a small bias when the signals are generated using `dmipy` (**Case 2**). The figure also shows that inverting the model directly from the dMRI signals leads to rather poor results (**Case 3**).

differently from its training set and that our summary statistics are descriptive enough to synthesize each tissue configuration. Finally, the poor results in the estimation for **Case 3** reflect the largely indeterminate system of equations that results from trying to infer the tissue parameters directly from the dMRI signals. This behavior was expected, as similar degeneracies were shown for a simpler model in Novikov et al. (2018a).

Influence of the number of b-values. We have also investigated how different choices of b-values in the acquisition scheme affect the quality of the parameter estimation using the posterior approximation. Note that these choices have no influence on how the posterior approximator is obtained, since it is trained on data generated directly from the relations between tissue parameters and the diffusion summary statistics, in which the b-values do not interact. In fact, the different acquisition setups that we consider have only an impact over the observed data point generated via `dmipy`. Figure 2 portrays the marginal posterior distributions for each tissue parameter in setups `Ideal`, `HCP MGH`, and `HCP 1200`. We see that estimations in the `Ideal` setup (equivalent to Case 2 in Figure 1) are very much concentrated around the true values of the parameters. Results with `HCP MGH` present a small bias for most parameters but capture well the soma parameters (f_s and C_s), whereas the results with setup `HCP 1200` present a larger bias for most parameters. They nevertheless remain reliable for the estimation of soma parameters. Note that the `Ideal` and `HCP MGH` ground truth value of C_s (vertical black dashed line) is different from the one of the `HCP 1200` setup (vertical green dashed line), because of their different diffusion times. These simulations allow us to have a fair confidence in the estimations on real data, presented in Sections 5.2 and 5.3.

Our new parameter avoids model indeterminacy. In Section 3.1, we introduce the parameter C_s, which serves as a proxy of the soma radius and provides key information on

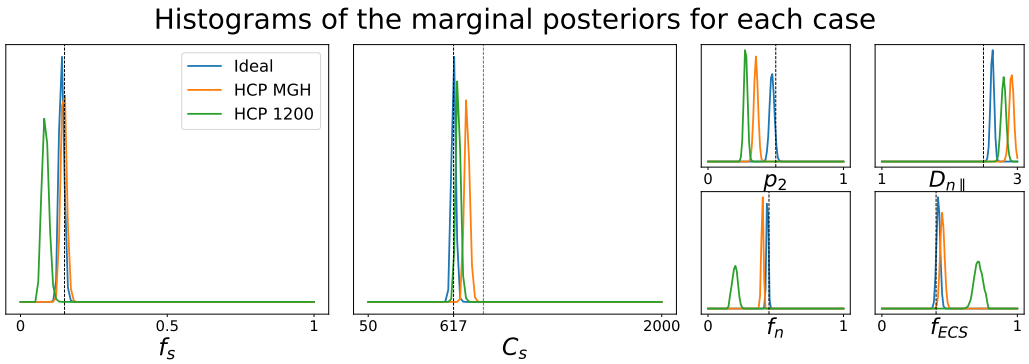

Figure 2: Histograms of 10^4 samples of the approximate posterior distribution with observed dMRI signals generated under three acquisition setups (see text for details). Vertical black dashed lines represent ground truth values of $\boldsymbol{\theta}_0$ which generated the observed signals (the green dashed line indicates the C_s value expected for the `HCP 1200` setup). We see that while the `Ideal` case delivers very good estimates, the results for the two other setups are only reliable for a subset of the tissue parameters.

the soma compartment. In this experiment, we illustrate the results of our model inversion if we had not defined parameter C_s.

Figure 3 presents the marginal posterior distributions of r_s and D_s as well as their joint distribution using the setup `Ideal`. To obtain these results, we have altered our LFI pipeline so to consider an extended parameter vector including r_s and D_s. The prior distribution reflects our assumption that $r_s \in [10^{-5}, 30]$ µm and $D_s \in [1, 3]$ µm ms^{-2} and we consider ground truth parameters $r_s = 17.5\,$µm and $D_s = 2.3\,$µm ms^{-2}. We note that in addition to larger marginal posterior distributions for each parameter, the joint posterior has a valley of large values for the (r_s, D_s) pair, including the ground truth parameters. This result is typical of non-injective models, i.e. models for which several input parameters may yield the same output observation, and is an important asset of a probabilistic framework such as ours.

Assessing the variances of estimated parameters. Deriving the posterior distributions of the parameter vectors allows us to report the values of the most likely tissue parameters for a given observation, along with our certitude regarding our inference. Figure 4 presents the logarithm of the standard deviation of the marginal posterior samples for different ground truth parameter choices (varying f_s and f_n) under setup `Ideal`. These values indicate how sharp a posterior distribution is and, therefore, quantify the quality of the fit. We observe larger standard deviations in the absence (or weak presence) of soma compartments in the mixture signal, e.g., the standard deviation of C_s is large when few or no somas are present ($f_s \approx 0$). This is to be expected, since the lack of contribution from the somas in the diffusion signal makes it difficult to estimate parameters related to them.

5.2 HCP MGH

Our results on simulated data show that, although we manage to invert very well the brain tissue parameter on settings for which the dMRI signal is obtained with several b-values,

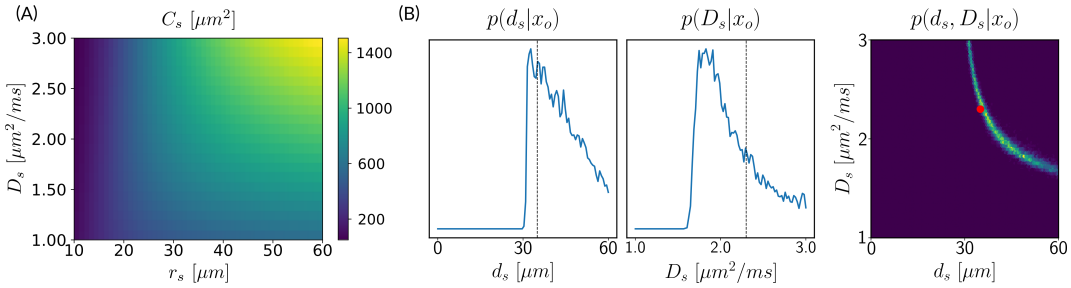

Figure 3: (A) C_s dependence on soma radius r_s and diffusivity D_s. We see that there are several values of (d_s, D_s) that yield the same C_s. (B) and (C) Histograms of 10^4 samples from the marginal and joint posterior distributions of $d_s = 2r_s$ and D_s. The ridge in the joint distribution indicates that there are several possible values for the pair (d_s, D_s) with high probability, which are those yielding the same C_s. Estimating C_s directly bypasses this indeterminacy.

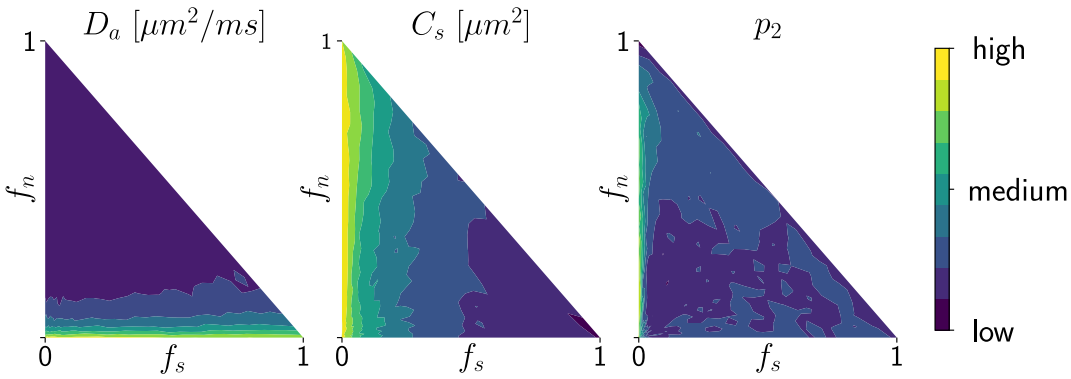

Figure 4: Logarithm of the standard deviations for the marginal posterior distribution of D_n, C_s, and p_2 with different choices of ground truth parameters (varying f_s and f_n). Since the ranges of values for each plot were quite different, we labeled the colorbar in terms of {'high', 'medium', 'low'} values of standard deviations so to provide mainly qualitative information to the reader. We see that when the signal fraction of somas decreases ($f_s \to 0$) the standard deviation of the C_s-estimation increases; and when less neurites are present ($f_n \to 0$) the standard deviation of p_2 and D_n increase.

the estimates for more realistic settings are less robust and demand a more subtle analysis. Indeed, we have observed that for both setups `HCP MGH` and `HCP 1200` the soma parameters seem to be rather well estimated without too much bias, which has lead us to consider mainly these parameters in our interpretations of the results.

Figure 5 presents the results on the HCP MGH dataset. We have masked our results so to show only areas where parameters were deemed stable, i.e. when the values were larger than 2 times the LFI-obtained standard deviations of the fitted posterior. We observe, therefore, a lack of stability on small sections including the auditory cortex and the precentral gyrus fundus. Our figure assesses qualitatively the results on soma size by comparing with nissl-stained histological studies (Allman et al., 2010; Amunts et al., 1999; Geyer et al., 1999). Our comparison shows good agreement between different cortical areas and the parameter C_s, which, under nearly-constant intra-soma diffusion D_s, is modulated by soma size.

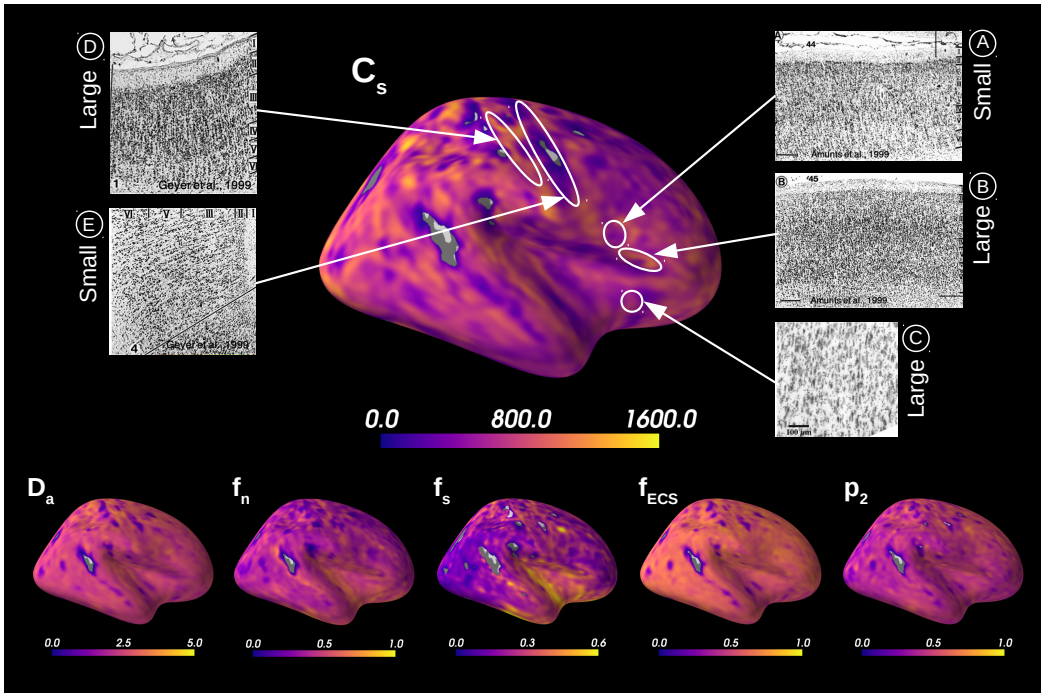

Figure 5: Microstructural measurements averaged over 31 HCP MGH subjects. We deemed stable measurements with a z-score larger than 2, where the standard deviation on the posterior estimates was estimated through our LFI fitting approach. In comparing with Nissl-stained cytoarchitectural studies we can qualitatively evaluate our parameter C_s: Broadmann area 44 (A) has smaller soma size in average than area 45 (B) (Amunts et al., 1999); large von Economo neurons predominate the superior anterior insula (C) (Allman et al., 2010); precentral gyrus (E) shows very small somas while post-central (D) larger ones (Geyer et al., 1999).

Interestingly, most regions of Figure 6 in which the parameter estimation has low confidence are located in the fundus of the sulci. Two main hypothesis could explain such behavior. Firstly, brain regions which are very curved may be more prone to mixing between tissue layers and CSF, which generates noisier signals. Thus, the estimation of summary

statistics becomes more difficult and the posterior distributions tend to be wider. Another possible explanation, based on cytoarchitecture considerations, points out the fact that the fundus of sulci is where sharp changes in cellular populations happen (Brodmann, 1909; Pandya et al., 2015). A mixing of several types of neurons in one voxel could generate multi-modal posterior distributions, and hence a region with large variance.

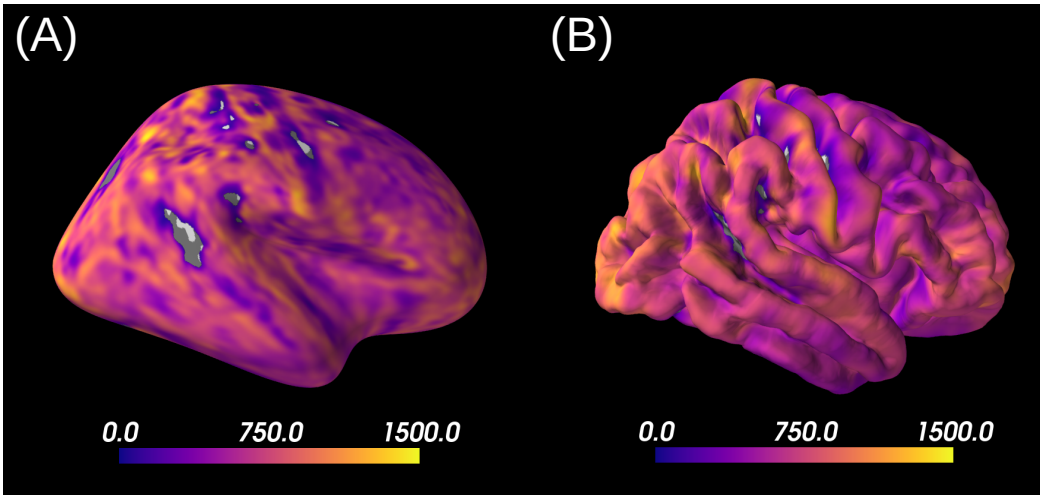

Figure 6: C_s estimations averaged over 31 HCP MGH subjects, with unstable results masked, projected onto inflated (A) and pial (B) surfaces. Interestingly, low confidence areas correspond to the fundus of the sulci.

Figure 7A reports the soma proportion (parameter f_s) averaged over 31 HCP MGH subjects, masking unstable results. Mean soma signal proportion in the cortex equals 0.32 ± 0.22 (mean computed in trusted estimations only). These results are coherent with the mean volume occupancy of $10 - 20\%$ observed in grey matter (Shapson-Coe et al., 2021). To interpret the results at a region-based level, we have superimposed the soma proportion estimations with the Brodmann atlas. We observe a general agreement between the estimations and the atlas, and more particularly in the somatosensory and Broca's areas. A clear difference in soma proportion can be observed in the 12 regions, as presented in the barplot.

Despite the C_s parameter being useful for avoiding indeterminacies in the model inversion, its biological interpretation remains difficult. With the goal of relating our C_s estimation with physiological insight, we estimated soma radius r_s by fixing soma diffusivity. Similarly to the SANDI method, we fixed soma diffusivity D_s to the value of the self-diffusion coefficient of free water ($3\,\mu m\,ms^{-2}$). Note that this value could be adjusted for each voxel, and is only used here in a matter of comparison and interpretation. Using a fixed-point method, we computed the soma radius r_s from the averaged C_s map obtained from our posterior distribution. The results are portrayed in Figure 7B. We see that the estimated soma radius vary between 8 and 14 µm, which is in accordance with histology (Palombo et al.). Mean C_s values are presented in the barplot beneath the soma radius estimations in the different Brodmann regions.

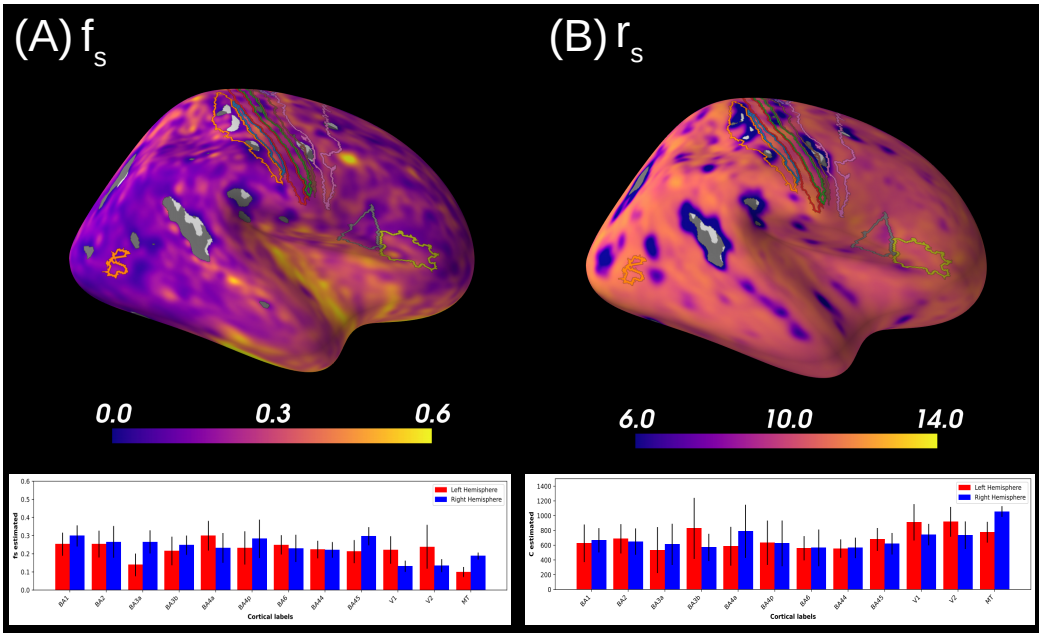

Figure 7: (Left) The average soma proportion (f_s) over the 31 HCP MGH subjects is projected onto an inflated cortical surface, with unstable results masked; see text for details. We also show the main Brodmann areas available on FreeSurfer. The mean values of f_s over these regions (using only reliable estimations) are reported on the bar plots below. (Right) Soma radius map computed from C_s with soma diffusivity fixed to $D_s = 3\,\mu\text{m}\,\text{ms}^{-2}$, averaged on 31 subjects. The bar plots below report the C_s mean values on main Brodmann areas. We observe a good agreement between our reported values and the Brodmann areas.

5.3 HCP 1200

Figure 8 shows the results obtained on a database with only three b-shells. We see that 55 % of the C_s estimations on brain grey matter is considered unstable and is, therefore, masked. Indeed, our q-bounded RTOP approximation relies on high b-values, where the signal is expected to have converged towards a value that depends on the radius of the soma. The larger the soma, the sooner the q-bounded RTOP converges. However, the largest b-value contained in this database equals $3\,\text{ms}\,\mu\text{m}^{-2}$, which is not enough for the signal to have converged. The poor quality of the summary statistics estimation leads to rather wide posterior distributions, resulting in unreliable results, as shown in simulations. Note, however, that estimations of the superior temporal gyrus for example are not masked, and both data sets seem to indicate large neurones in that region. The estimation of D_a is considered as unstable for 98.7% of the data set. This behavior was expected, given the results presented in Figure 2, for which a large bias is observed.

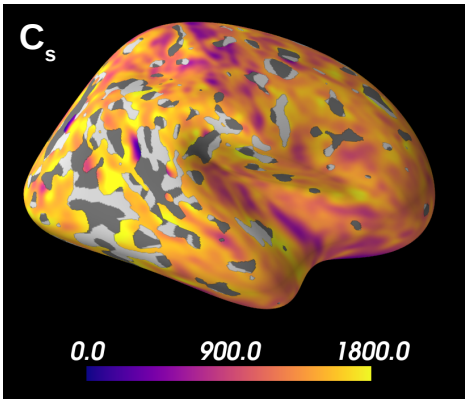

Figure 8: Spatial distribution of the C_s estimations averaged over 30 HCP 1200 subjects and projected onto an inflated cortical surface, with unstable estimations masked. Due to the scarce and rather low b-values ($\leq 3\,\mathrm{ms}\,\mu\mathrm{m}^{-2}$) used in the database, the results are very unstable and, therefore, many voxels are discarded.

6. Discussion

6.1 Validation simulated data

An important aspect of our work is the thorough validation that we have carried out on simulated data, using different acquisition setups and ways of generating diffusion signals. Part of this validation had the goal of ensuring that our method gave consistent results on simple standard cases, as confirmed by the results in Figure 1. Additionally, we have demonstrated the benefits of using summary statistics for describing dMRI signals, attaining better parameter estimates when using them instead of directly manipulating the diffusion signals. Such results are very encouraging and should push other researchers into using these summary statistics for processing their dMRI signals.

Another relevant byproduct of our validation was observing that the quality of the parameter estimations depends heavily on the distribution of b-values used to acquire and simulate the dMRI signals. Indeed, if only small b-values are available, the summary statistics of dMRI signals are poorly estimated and the parameter estimation too. Based on this observation, we were able to apply our LFI pipeline to real datasets HCP MGH and HCP 1200 knowing in advance the limitations of our methodology; for example, we knew from which parameters we could expect good estimates (mostly soma-related ones) and which ones should not be taken into account in our analysis.

6.2 C_s: A proxy to soma size

Estimating both soma radius (r_s) and diffusivity (D_s) with diffusion MRI is a challenging task. When trying to estimate them separately, we can expect a 'banana-shape' in their joint posterior distribution as shown in Figure 3. This indicates that several values of the pair (r_s, D_s) can explain the observed signal with high probability and, therefore, one is confronted with model indeterminacy.

The new parameter C_s that we introduce in this paper is modulated both by the soma radius and its diffusivity. Thus, estimating it directly avoids problems of indeterminacy, as shown in Figure 1, for example. Note, however, that avoiding such indeterminacy comes with the price of losing specificity and, therefore, physiological interpretations. Fortunately, acquisitions in the narrow or wide pulse regimes allow us to better interpret estimations of C_s, as it only depends on r_s or D_s (see Section 3.1).

6.3 Can I apply this approach to my data?

One of the main benefits of a probabilistic framework is that it can be applied to any data set or acquisition setup without too much hesitation, since the estimates always come with a "quality certificate" described by the credible regions derived from the posterior distribution. We have benefited from this feature when creating all figures related to databases HCP MGH and HCP 1200, since they allow us to mask regions for which the variance of the parameter estimation is too high. We can also use it as a proxy to identify regions for which the three-compartment model is adequate or not, or assessing whether the distribution of b-values used to acquire the observed diffusion signal is sufficiently informative.

6.4 Limitations and perspectives

There are many extensions that we could envision for the method that we propose.

For instance, the three-compartment model that we use approximates well the intracellular signal in brain grey matter tissue by adding a sphere compartment to account for soma presence (Palombo et al., 2020). However, this is a rather simplified model and could be improved; for instance, the geometry of ECS is very restricted and tortuous, and diffusion signals have been proven to deviate from a mono-exponential behavior (Vincent et al., 2021). A first improvement could be, therefore, to deviate from modelling the diffusion with a simple isotropic Gaussian and consider more complex geometric representations.

Also, soma-related parameters are better estimated in tissues where there are many somas. Consequently, the method that we present is more suitable to grey matter than for white matter. However, applying a LFI pipeline on a modified model in which the contributions of the somas are discarded would improve the estimates of tissue parameters for white matter. This new pipeline would then be seen as a new way to solve the LEMONADE system of equations proposed by Novikov et al. (2018b) within a probabilistic framework.

Finally, we could leverage the modularity of the probabilistic framework and apply our methodology to other kinds of models in medical imaging. This could be done by simply re-defining the forward model and the summary statistics defined in Section 4.1. Such flexibility confirms the generality of our approach and the potential impact that it might have in the medical imaging field.

7. Conclusion

Quantifying grey matter tissue composition is challenging. In this work, we have presented a methodology to estimate the parameters of a model that best fit an observed data point, and also their full posterior distribution. This rich description provides many useful tools, such as assessing the quality of the parameter estimation or characterizing regions in the

parameter space where it is harder to invert the model. The inclusion of such "quality certificate" accompanying our parameter estimation is very useful in practice and allows one to apply the pipeline on any kind of database and know to which degree one can trust the results. Moreover, our proposal alleviates limitations from current methods in the literature by not requiring physiologically unrealistic constraints on the parameters and avoiding indeterminacies when estimating them.

To conclude, we believe that our approach based on Bayesian inference with modern tools from neural networks is a promising one that can easily be applied to other applications in the medical imaging field: one only needs to define a sufficiently rich model describing a certain phenomenon of interest and the LFI pipeline will manage to invert it and provide a related posterior distribution. We expect, therefore, that other researchers will find this contribution valuable for their own applications and see such probabilistic approach more often used in the literature.

Acknowledgments

Experiments on diffusion MRI data were made possible thanks to DiPy (Garyfallidis et al., 2014), as well as the scientific Python ecosystem: Python (Python Software Fundation, 2017), Matplotlib (Hunter, 2007), Numpy (Harris et al., 2020), Scipy (Virtanen et al., 2020), Mayavi (Ramachandran and Varoquaux, 2011), SBI (Tejero-Cantero et al., 2020) and PyTorch (Paszke et al., 2019).

This work was supported by grants ERC-StG NeuroLang ID:757672 and ANR-NSF NeuroRef to DW and the ANR BrAIN (ANR-20-CHIA-0016) grant to AG.

Ethical Standards

The work follows appropriate ethical standards in conducting research and writing the manuscript, following all applicable laws and regulations regarding treatment of animals or human subjects.

Conflicts of Interest

We declare we don't have conflicts of interest.

References

John M. Allman, Nicole A. Tetreault, Atiya Y. Hakeem, Kebreten F. Manaye, Katerina Semendeferi, Joseph M. Erwin, Soyoung Park, Virginie Goubert, and Patrick R. Hof. The von Economo neurons in frontoinsular and anterior cingulate cortex in great apes and humans. Brain Structure and Function, 214:495–517, June 2010.

Katrin Amunts, Axel Schleicher, Uli Bürgel, Hartmut Mohlberg, Harry B.M. Uylings, and Karl Zilles. Broca's region revisited: Cytoarchitecture and intersubject variability. Journal of Comparative Neurology, 412(2):319–341, 1999.

Balin Balinov, Jönsson, Per Linse, and Olle Söderman. The NMR Self-Diffusion Method Applied to Restricted Diffusion. Simulation of Echo Attenuation form Molecules in Spheres and between Planes. Journal of Magnetic Resonance, Series A, 104(1):17–25, 1993.

Korbinian Brodmann. Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues. Barth, 1909.

Kyle Cranmer, Johann Brehmer, and Gilles Louppe. The frontier of simulation-based inference. Proceedings of the National Academy of Sciences, 117(48):30055–30062, 2020.

Gwenaëlle Douaud, Ricarda A. L. Menke, Achim Gass, Andreas U. Monsch, Anil Rao, Brandon Whitcher, Giovanna Zamboni, Paul M. Matthews, Marc Sollberger, and Stephen Smith. Brain microstructure reveals early abnormalities more than two years prior to clinical progression from mild cognitive impairment to alzheimer's disease. Journal of Neuroscience, 33(5):2147–2155, 2013. ISSN 0270-6474. doi: 10.1523/JNEUROSCI.4437-12.2013. URL https://www.jneurosci.org/content/33/5/2147.

Conor Durkan, Artur Bekasov, Iain Murray, and George Papamakarios. nflows: normalizing flows in PyTorch. November 2020. doi: 10.5281/zenodo.4296287.

Rutger H. J. Fick, Demian Wassermann, and Rachid Deriche. The Dmipy Toolbox: Diffusion MRI Multi-Compartment Modeling and Microstructure Recovery Made Easy. Frontiers in Neuroinformatics, 13:64, October 2019.

Rutger H.J. Fick, Demian Wassermann, Emmanuel Caruyer, and Rachid Deriche. MAPL: Tissue microstructure estimation using Laplacian-regularized MAP-MRI and its application to HCP data. NeuroImage, 134:365–385, July 2016.

Eleftherios Garyfallidis, Matthew Brett, Bagrat Amirbekian, Ariel Rokem, Stefan Van Der Walt, Maxime Descoteaux, and Ian Nimmo-Smith. Dipy, a library for the analysis of diffusion mri data. Frontiers in Neuroinformatics, 8:8, 2014. ISSN 1662-5196. doi: 10.3389/fninf.2014.00008. URL https://www.frontiersin.org/article/10.3389/fninf.2014.00008.

A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, and D.B. Rubin. Bayesian Data Analysis, Third Edition. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis, 2013. ISBN 9781439840955. URL https://books.google.fr/books?id=ZXL6AQAAQBAJ.

Mathieu Germain, Karol Gregor, Iain Murray, and Hugo Larochelle. Made: Masked autoencoder for distribution estimation. In Proceedings of the 32nd International Conference on Machine Learning, volume 37, pages 881–889. PMLR, 2015.

Stefan Geyer, Axel Schleicher, and Karl Zilles. Areas 3a, 3b, and 1 of Human Primary Somatosensory Cortex. NeuroImage, 10(1):63–83, July 1999. ISSN 10538119.

Pedro J Gonçalves, Jan-Matthis Lueckmann, Michael Deistler, Marcel Nonnenmacher, Kaan Öcal, Giacomo Bassetto, Chaitanya Chintaluri, William F Podlaski, Sara A Haddad, Tim P Vogels, David S Greenberg, and Jakob H Macke. Training deep neural density estimators to identify mechanistic models of neural dynamics. eLife, 9, September 2020.

Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep Learning. MIT Press, 2016. http://www.deeplearningbook.org.

Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks, 2014.

David Greenberg, Marcel Nonnenmacher, and Jakob Macke. Automatic posterior transformation for likelihood-free inference. In Proceedings of the 36th International Conference on Machine Learning, volume 97, pages 2404–2414. PMLR, 2019.

Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. Nature, 585(7825):357–362, September 2020. doi: 10.1038/s41586-020-2649-2. URL `https://doi.org/10.1038/s41586-020-2649-2`.

John D Hunter. Matplotlib: A 2d graphics environment. Computing in science & engineering, 9(3): 90–95, 2007.

Ileana O. Jelescu and Matthew D. Budde. Design and validation of diffusion MRI models of white matter. Frontiers in Physics, 5, November 2017. doi: 10.3389/fphy.2017.00061. URL `https://doi.org/10.3389/fphy.2017.00061`.

Ileana O. Jelescu, Jelle Veraart, Els Fieremans, and Dmitry S. Novikov. Degeneracy in model parameter estimation for multi-compartmental diffusion in neuronal tissue: Degeneracy in Model Parameter Estimation of Diffusion in Neural Tissue. NMR in Biomedicine, 29(1):33–47, January 2016.

Ileana O. Jelescu, Alexandre de Skowronski, Marco Palombo, and Dmitry S. Novikov. Neurite exchange imaging (nexi): A minimal model of diffusion in gray matter with inter-compartment water exchange, 2021.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017.

Diederik P Kingma and Max Welling. Auto-encoding variational bayes, 2014.

Hua Li, Xiaoyu Jiang, Jingping Xie, John C. Gore, and Junzhong Xu. Impact of transcytolemmal water exchange on estimates of tissue microstructural properties derived from diffusion MRI. Magnetic Resonance in Medicine, 77(6):2239–2249, June 2016. doi: 10.1002/mrm.26309. URL `https://doi.org/10.1002/mrm.26309`.

Vinod Menon, Guillermo Gallardo, Mark A Pinsk, Van-Dang Nguyen, Jing-Rebecca Li, Weidong Cai, and Demian Wassermann. Microstructural organization of human insula is linked to its macrofunctional circuitry and predicts cognitive control. elife, 9:e53470, 2020.

P. P. Mitra, L. L. Latour, R. L. Kleinberg, and C. H. Sotak. Pulsed-field-gradient NMR measurements of restricted diffusion and the return-to-origin probability. Journal of Magnetic Resonance, 114: 47–58, 1995.

J. S. Murday and R. M. Cotts. Self-diffusion coefficient of liquid lithium. The Journal of Chemical Physics, 48(11):4938–4945, 1968. doi: 10.1063/1.1668160.

Dmitry S. Novikov, Els Fieremans, Sune N. Jespersen, and Valerij G. Kiselev. Quantifying brain microstructure with diffusion MRI: Theory and parameter estimation: Brain microstructure with dMRI: Theory and parameter estimation. NMR in Biomedicine, page e3998, October 2018a. ISSN 09523480.

Dmitry S. Novikov, Jelle Veraart, Ileana O. Jelescu, and Els Fieremans. Rotationally-invariant mapping of scalar and orientational metrics of neuronal microstructure with diffusion MRI. NeuroImage, 174:518–538, July 2018b.

M Palombo, DC Alexander, and H Zhang. Large-scale analysis of brain cell morphometry informs microstructure modelling of gray matter.

Marco Palombo, Andrada Ianus, Michele Guerreri, Daniel Nunes, Daniel C. Alexander, Noam Shemesh, and Hui Zhang. SANDI: A compartment-based model for non-invasive apparent soma and neurite imaging by diffusion MRI. NeuroImage, 215:116835, 2020.

Deepak Pandya, Michael Petrides, and Patsy Benny Cipolloni. Cerebral cortex: architecture, connections, and the dual origin concept. Oxford University Press, 2015.

George Papamakarios and Iain Murray. Fast ϵ-free inference of simulation models with bayesian conditional density estimation. In Advances in Neural Information Processing Systems, volume 29, pages 1028–1036. Curran Associates, Inc., 2016.

George Papamakarios, Eric T. Nalisnick, Danilo Jimenez Rezende, Shakir Mohamed, and Balaji Lakshminarayanan. Normalizing flows for probabilistic modeling and inference. ArXiv, abs/1912.02762, 2019.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In Advances in Neural Information Processing Systems (NeurIPS), page 12, Vancouver, BC, Canada, 2019.

Python Software Fundation. Python Language Reference, version 3.6, 2017.

P. Ramachandran and G. Varoquaux. Mayavi: 3D Visualization of Scientific Data. Computing in Science & Engineering, 13(2):40–51, 2011. ISSN 1521-9615.

K. Setsompop, R. Kimmlingen, E. Eberlein, T. Witzel, J. Cohen-Adad, J.A. McNab, B. Keil, M.D. Tisdall, P. Hoecht, P. Dietz, S.F. Cauley, V. Tountcheva, V. Matschl, V.H. Lenz, K. Heberlein, A. Potthast, H. Thein, J. Van Horn, A. Toga, F. Schmitt, D. Lehne, B.R. Rosen, V. Wedeen, and L.L. Wald. Pushing the limits of in vivo diffusion MRI for the Human Connectome Project. NeuroImage, 80:220 – 233, 2013.

Alexander Shapson-Coe, Michał Januszewski, Daniel R. Berger, Art Pope, Yuelong Wu, Tim Blakely, Richard L. Schalek, Peter Li, Shuohong Wang, Jeremy Maitin-Shepard, Neha Karlupia, Sven Dorkenwald, Evelina Sjostedt, Laramie Leavitt, Dongil Lee, Luke Bailey, Angerica Fitzmaurice, Rohin Kar, Benjamin Field, Hank Wu, Julian Wagner-Carena, David Aley, Joanna Lau, Zudi Lin, Donglai Wei, Hanspeter Pfister, Adi Peleg, Viren Jain, and Jeff W. Lichtman. A connectomic study of a petascale fragment of human cerebral cortex. bioRxiv, 2021. doi: 10.1101/2021.05.29.446289. URL `https://www.biorxiv.org/content/early/2021/05/30/2021.05.29.446289`.

Scott A. Sisson. Handbook of Approximate Bayesian Computation. Chapman and Hall/CRC, September 2018.

A. M. Stuart. Inverse problems: A bayesian perspective. Acta Numerica, 19:451–559, 2010. doi: 10.1017/S0962492910000061.

Chantal M.W. Tax, Filip Szczepankiewicz, Markus Nilsson, and Derek K. Jones. The dot-compartment revealed? diffusion mri with ultra-strong gradients and spherical tensor encoding in the living human brain. NeuroImage, 210:116534, 2020.

Alvaro Tejero-Cantero, Jan Boelts, Michael Deistler, Jan-Matthis Lueckmann, Conor Durkan, Pedro J. Gonçalves, David S. Greenberg, and Jakob H. Macke. sbi: A toolkit for simulation-based inference. Journal of Open Source Software, 5(52):2505, 2020. doi: 10.21105/joss.02505.

Jelle Veraart, Els Fieremans, and Dmitry S. Novikov. On the scaling behavior of water diffusion in human brain white matter. NeuroImage, 185:379–387, January 2019.

Jelle Veraart, Daniel Nunes, Umesh Rudrapatna, Els Fieremans, Derek K Jones, Dmitry S Novikov, and Noam Shemesh. Noninvasive quantification of axon radii using diffusion MRI. eLife, 9, February 2020.

Mélissa Vincent, Mylène Gaudin, Covadonga Lucas-Torres, Alan Wong, Carole Escartin, and Julien Valette. Characterizing extracellular diffusion properties using diffusion-weighted mrs of sucrose injected in mouse brain. NMR in Biomedicine, 34(4):e4478, 2021. doi: https://doi.org/10.1002/nbm.4478. URL `https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/abs/10.1002/nbm.4478`.

Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature Methods, 17:261–272, 2020. doi: 10.1038/s41592-019-0686-2.

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.

Karl Zilles. Brodmann: a pioneer of human brain mapping—his impact on concepts of cortical organization. Brain, 141(11):3262–3278, October 2018. doi: 10.1093/brain/awy273. URL `https://doi.org/10.1093/brain/awy273`.