

# A Fully Bayesian Inference Framework for Population Studies of the Brain Microstructure

Maxime Taquet, Benoît Scherrer,  
Jurriaan M. Peters, Sanjay P. Prabhu, and Simon K. Warfield

Computational Radiology Laboratory, Harvard Medical School, Boston, USA

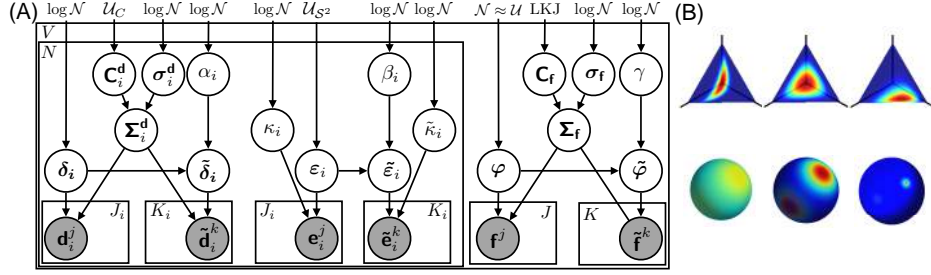
**Abstract.** Models of the diffusion-weighted signal are of strong interest for population studies of the brain microstructure. These studies are typically conducted by extracting a scalar property from the model and subjecting it to null hypothesis statistical testing. This process has two major limitations: the reported p-value is a weak predictor of the reproducibility of findings and evidence for the absence of microstructural alterations cannot be gained. To overcome these limitations, this paper proposes a Bayesian framework for population studies of the brain microstructure represented by multi-fascicle models. A hierarchical model is built over the biophysical parameters of the microstructure. Bayesian inference is performed by Hamiltonian Monte Carlo sampling and results in a joint posterior distribution over the latent microstructure parameters for each group. Inference from this posterior enables richer analyses of the brain microstructure beyond the dichotomy of statistical testing. Using synthetic and in-vivo data, we show that our Bayesian approach increases reproducibility of findings from population studies and opens new opportunities in the analysis of the brain microstructure.

**Keywords:** Microstructure, Diffusion Imaging, Bayesian Inference

## 1 Introduction

Novel models of the microstructure from diffusion-weighted imaging (DWI) provide insights into the cellular architecture of the healthy and diseased brain [4]. Population studies of the brain microstructure are commonly conducted by extracting a property of interest (e.g., diffusivities of a fascicle [8], principal direction of diffusion [7] or fraction of the isotropic compartment [5]) and subjecting it to null hypothesis statistical testing (NHST). Because the p-value is uniformly distributed under the null hypothesis, no evidence in favor of the null can be gained in this process [3]. The null hypothesis will thus eventually be rejected if a sufficient number of attempts are made, which significantly hampers the building of scientific knowledge and reduces the reproducibility of results [3]. These limitations are critical in microstructure imaging, wherein reproducibility is a major concern and researchers want evidence for the absence of abnormalities in certain brain regions alongside the detection of abnormalities in other regions.

Bayesian approaches to population studies address these limitations. They proceed by defining a generative model of the data in both populations and



**Fig. 1.** (A) Generative model of the microstructure at the population level. Circles are random variables and shaded circles have observations. Indices relate to fascicles and superscripts to subject (omitted in the text for clarity). The left box is repeated for the  $N$  fascicles in the voxel and the whole is repeated for all  $V$  voxels. (B) Instances of logistic-normal distribution over the simplex and Watson distribution over the sphere.

inferring a joint posterior probability over the model parameters. The posterior probability can then be interrogated to answer research questions [1]. Defining a generative model of the microstructure at the population level, however, is challenging due to the geometry of the space of parameters. This paper tackles this challenge and presents a novel Bayesian inference framework for population studies the brain microstructure. Section 2 presents our assumptions. Section 3 and 4 describe our model and its inference. Section 5 shows experimental results.

## 2 Parameterization and Assumptions

Multi-fascicle models represent the signal from each compartment in each voxel as a separate Gaussian encoded by a tensor  $\mathbf{D}_i$ , weighed by  $f_i$  [6]:

$$S(b, \mathbf{g}) = S_0 \sum_{i=1}^N f_i e^{-b\mathbf{g}^T \mathbf{D}_i \mathbf{g}}, \text{ where } b \text{ is the b-value and } \mathbf{g} \text{ is the gradient.}$$

One tensor is isotropic [4,5,6] and the two smallest eigenvalues of each tensor are equated and represent the radial diffusivity [4]. The biophysical properties of this model are: the vector  $\mathbf{d}_i = (\mathbf{d}_{\parallel,i}, \mathbf{d}_{\perp,i})$  of axial and radial diffusivities and the principal direction  $\mathbf{e}_i$  of Fascicle  $i$ , and the vector of fractions  $\mathbf{f}$ . These variables have observations  $\mathbf{d}_i^j$ ,  $\mathbf{e}_i^j$  and  $\mathbf{f}^j$  for Subject  $j$ . Inter-subject variability in these properties (encoded in the likelihood) is assumed much larger than DWI noise. An extra level in the hierarchical model could otherwise link parameters to DWI [9]. Fascicles are assumed independent from one another. By symmetry, directions are independent from diffusivities and fractions. Signal fractions mostly relate to partial voluming and are thus independent from fascicle properties. Diseases may affect several properties of the microstructure, so that these assumptions are only valid conditional on the diagnoses. In summary, we have:

$$\mathbf{d}_i \hat{=} (\mathbf{d}_{\parallel,i}, \mathbf{d}_{\perp,i}), \mathbf{d}_i \perp \mathbf{d}_{j \neq i}, \quad \mathbf{e}_i \perp \mathbf{e}_{j \neq i}, \quad \mathbf{e}_i \perp \mathbf{d}_i, \quad \mathbf{e}_i \perp \mathbf{f}, \quad \text{and } \mathbf{f} \perp \mathbf{d}_i.$$

### 3 Model

Fig. 1 shows our hierarchical model for the latent diffusivities ( $\boldsymbol{\delta}$ ,  $\tilde{\boldsymbol{\delta}}$ ), principal directions ( $\boldsymbol{\varepsilon}_i$ ,  $\tilde{\boldsymbol{\varepsilon}}_i$ ) and signal fractions ( $\boldsymbol{\varphi}$ ,  $\tilde{\boldsymbol{\varphi}}$ ) for the control and patient (tilded) groups respectively. Owing to the biophysical parameterization of the model, the posterior for these variables is of the form (conditioning on diagnoses is implicit):

$$p(\boldsymbol{\delta}, \tilde{\boldsymbol{\delta}}, \boldsymbol{\varepsilon}, \tilde{\boldsymbol{\varepsilon}}, \boldsymbol{\varphi}, \tilde{\boldsymbol{\varphi}} | \mathbf{f}, \mathbf{D}) = p(\boldsymbol{\varphi}, \tilde{\boldsymbol{\varphi}} | \mathbf{f}) \prod_{i=1}^N p(\boldsymbol{\delta}_i, \tilde{\boldsymbol{\delta}}_i | \mathbf{d}_i) p(\boldsymbol{\varepsilon}_i, \tilde{\boldsymbol{\varepsilon}}_i | \mathbf{e}_i).$$

A fully Bayesian hierarchical model is obtained by representing unknowns as random variables and assigning priors over their value [1]. We assign uniform priors for bounded variables, and log-normal priors for unbounded positive-definite variables. The log-normal prior only requires a vague idea about the order of magnitude of the variable, encoded as its mean. Its standard deviation is set to 5 to allow for mistakes in our estimate of the order of magnitude.

**Modeling diffusivities.** Diffusivities are in  $\mathbb{R}^+ \times \mathbb{R}^+$ . We represent distributions over their logarithm to assign zero probability to negative values [8]. The inter-subject variability within a group is represented as a normal law:

$$\log \mathbf{d}_i | \boldsymbol{\delta}_i, \boldsymbol{\Sigma}_i^d \sim \mathcal{N}(\log \boldsymbol{\delta}_i, \boldsymbol{\Sigma}_i^d) \quad \text{and} \quad \log \tilde{\mathbf{d}}_i | \tilde{\boldsymbol{\delta}}_i, \boldsymbol{\Sigma}_i^d \sim \mathcal{N}(\log \tilde{\boldsymbol{\delta}}_i, \boldsymbol{\Sigma}_i^d).$$

A prior over  $(\boldsymbol{\delta}_i, \tilde{\boldsymbol{\delta}}_i)$  is best expressed conditionally:  $p(\boldsymbol{\delta}_i, \tilde{\boldsymbol{\delta}}_i) = p(\tilde{\boldsymbol{\delta}}_i | \boldsymbol{\delta}_i) p(\boldsymbol{\delta}_i)$ . The order of magnitude of diffusivities is around  $10^{-4}$ ,  $10^{-3}$  mm<sup>2</sup>/s, so  $\boldsymbol{\delta}_i \sim \log \mathcal{N}(-3.5, 5)$ . Given  $\boldsymbol{\delta}_i$  and without data from patients, our best guess for  $\tilde{\boldsymbol{\delta}}_i$  is  $\boldsymbol{\delta}_i$ , which we encode as  $\tilde{\boldsymbol{\delta}}_i | \boldsymbol{\delta}_i, \alpha_i \sim \log \mathcal{N}(\log \boldsymbol{\delta}_i, \alpha_i^2 \mathbf{I}_2)$  for an unknown variance  $\alpha_i^2$ . We assign a log-normal hyperprior to  $\alpha_i$  with zero mean, to express that  $\tilde{\boldsymbol{\delta}}_i$  and  $\boldsymbol{\delta}_i$  likely have, a priori, a similar order of magnitude. Regarding  $\boldsymbol{\Sigma}_i^d$ , the correlation between the axial and radial diffusivities is unknown a priori whereas we have a vague idea of the order of magnitude of the standard deviation of each diffusivity. This prior information is effectively encoded by factorizing  $\boldsymbol{\Sigma}_i^d = \text{diag}(\boldsymbol{\sigma}_i^d) \mathbf{C}_i^d \text{diag}(\boldsymbol{\sigma}_i^d)$  [1], where  $\boldsymbol{\sigma}_i^d$  is the vector of standard deviations and  $\mathbf{C}_i^d$  is a  $2 \times 2$  correlation matrix. We assign a log-normal prior to  $\boldsymbol{\sigma}_i^d$  with zero mean to express that diffusivities likely have the same order of magnitude, and a uniform prior on  $[-1, 1]$  for the correlation coefficient in  $\mathbf{C}_i^d$ .

**Modeling directions.** Directions are dipoles on the sphere. We model their likelihood as a Watson distribution [7], with unknown concentrations  $\kappa_i$ ,  $\tilde{\kappa}_i$ :

$$p(\mathbf{e}_i | \boldsymbol{\varepsilon}_i, \kappa_i) = A(\kappa_i) \exp(\kappa_i (\boldsymbol{\varepsilon}_i \cdot \mathbf{e}_i)), \quad p(\tilde{\mathbf{e}}_i | \tilde{\boldsymbol{\varepsilon}}_i, \tilde{\kappa}_i) = A(\tilde{\kappa}_i) \exp(\tilde{\kappa}_i (\tilde{\boldsymbol{\varepsilon}}_i \cdot \tilde{\mathbf{e}}_i)).$$

We express the priors over latent group variables conditionally:  $p(\boldsymbol{\varepsilon}_i, \tilde{\boldsymbol{\varepsilon}}_i) = p(\tilde{\boldsymbol{\varepsilon}}_i | \boldsymbol{\varepsilon}_i) p(\boldsymbol{\varepsilon}_i)$ . The prior  $p(\boldsymbol{\varepsilon}_i)$  is uniform over the sphere and, conditionally on  $\boldsymbol{\varepsilon}_i$ , we model our prior knowledge about  $\tilde{\boldsymbol{\varepsilon}}_i$  as another Watson distribution with mean  $\boldsymbol{\varepsilon}_i$  and unknown concentration  $\beta_i$ . At concentrations of  $10^{-5}$ , the Watson is essentially uniform and at  $10^5$ , it is extremely peaky. We therefore assign log-normal hyperpriors to all concentration parameters  $\beta_i, \kappa_i, \tilde{\kappa}_i \stackrel{\text{iid}}{\sim} \log \mathcal{N}(0, 5)$ .

**Modeling fractions.** Fractions belong to an  $N$ -simplex:  $\{\mathbf{f}, f_i \geq 0, \sum_i f_i = 1\}$ . The conventional Dirichlet density, with a single parameter for all variances and covariances, is too limited to represent their likelihood. Instead, we apply a change of variables to map  $\mathbf{f}$  to an  $(N - 1)$ -vector of unconstrained variables  $\mathbf{y}$ :

$$\mathbf{y} \triangleq \mathbf{F}(\mathbf{f}) : \quad y_i = F_i(\mathbf{f}) = \text{logit} \left( \frac{f_i}{1 - \sum_{i'=1}^{i-1} f_{i'}} \right) + \log(N - i).$$

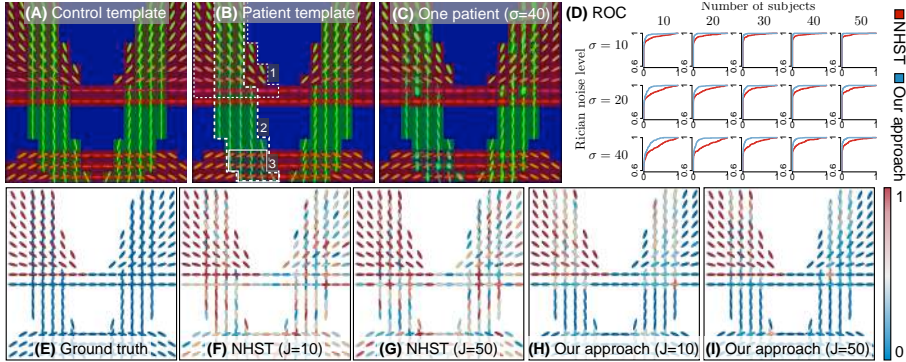
The logit maps its argument from  $(0, 1)$  to  $(-\infty, +\infty)$ . The second term is added so that  $\mathbf{y} = \mathbf{0} \Leftrightarrow f_i = 1/N, \forall i$ . We model the likelihood of  $\mathbf{y}$  as a multivariate normal distribution with mean  $\mathbf{F}(\boldsymbol{\varphi})$  for controls,  $\mathbf{F}(\boldsymbol{\zeta})$  for patients and unknown covariance  $\boldsymbol{\Sigma}_{\mathbf{f}}$ . This leads to a normal logistic distribution (Fig. 1B) whose expression is derived from the Gaussian density and the Jacobian of  $\mathbf{F}$ :

$$p(\mathbf{f}|\boldsymbol{\varphi}, \boldsymbol{\Sigma}_{\mathbf{f}}) = \frac{\prod_{i=1}^N \frac{1}{f_i}}{|2\pi\boldsymbol{\Sigma}_{\mathbf{f}}|^{1/2}} \exp \left[ -\frac{1}{2} (\mathbf{F}(\mathbf{f}) - \mathbf{F}(\boldsymbol{\varphi}))^T \boldsymbol{\Sigma}_{\mathbf{f}}^{-1} (\mathbf{F}(\mathbf{f}) - \mathbf{F}(\boldsymbol{\varphi})) \right].$$

A non-informative prior on  $\boldsymbol{\varphi}$  is obtained by a normal with large variance:  $\mathbf{F}(\boldsymbol{\varphi}) \sim \mathcal{N}(\mathbf{0}, 1000\mathbf{I}_{N-1})$ .  $\mathbf{F}(\boldsymbol{\zeta})|\mathbf{F}(\boldsymbol{\varphi})$  is assigned a multivariate normal prior with mean  $\mathbf{F}(\boldsymbol{\varphi})$  and covariance  $\gamma\mathbf{I}_{N-1}$  with unknown  $\gamma$ . As for  $\alpha_i$  and  $\beta_i$ ,  $\gamma$  is assigned a log-normal hyperprior with zero mean. The unknown covariance is decomposed as  $\boldsymbol{\Sigma}_{\mathbf{f}} = \text{diag}(\boldsymbol{\sigma}_{\mathbf{f}})\mathbf{C}_{\mathbf{f}}\text{diag}(\boldsymbol{\sigma}_{\mathbf{f}})$ . The correlation  $\mathbf{C}_{\mathbf{f}}$  is an  $(N-1) \times (N-1)$  matrix to which we assign a LKJ( $\eta=1$ ) prior which is uniform over all correlation matrices [1]. As before,  $\boldsymbol{\sigma}_{\mathbf{f}}$  is assigned a log-normal prior with zero mean.

## 4 Inference

The model needs to be estimated in each voxel. Efficient estimation of its posterior is therefore critical to its use in practice. The posterior has no closed-form expression due to interactions between variables and the introduction of non-conjugate priors. Variational Bayes approximations, often used in this case, may introduce biases that would jeopardize the reliability of group differences [2]. We therefore use Markov Chain Monte Carlo (MCMC) sampling. Because our log-posterior is differentiable, we exploit Hamiltonian Monte Carlo (HMC) sampling which converges much faster (in  $O(D^{5/4})$  for  $D$  dimensions) than conventional Gibbs and Metropolis sampling ( $O(D^2)$ ) [2]. Specifically, we use the No-U-Turn Sampler (NUTS) which automatically optimizes the parameters of the HMC sampling [2]. We draw four Markov chains of 2000 samples and discard the first 1000 burn-in samples. Convergence is monitored by the scale reduction factor  $\hat{R}$  [1]. From samples of the joint posterior over the model parameters  $\boldsymbol{\theta}$  (all variables in Fig. 1), marginal posteriors of any statistics of interest  $\mathbf{T}(\boldsymbol{\theta})$  are simply obtained by computing their values from the posterior samples. Importantly, because the joint posterior is not altered by the computation of  $\mathbf{T}(\boldsymbol{\theta})$ , many statistics can be simultaneously analyzed without the need to correct for multiple comparisons as in NHST. False discoveries are mitigated by the shrinkage towards the prior. For the validation, we focus on the posterior probabilities



**Fig. 2.** (A) Template for the control population (fractions encoded as RGB). (B) Template for the patient population with three overlapping affected areas (see text). (C) A simulated patient with noise at level  $\sigma = 40$ . (D) ROC analysis demonstrates that our approach outperforms NHST for all noise levels and all number of subjects. (E) Ground truth difference in radial diffusivity. (F-G) NHST provides evidence for an alteration in Area 1 but fails to provide evidence that no alteration is present elsewhere. Color encodes  $1 - p$ -value. (H-I) Bayesian posterior maps provide both evidence that fascicles in Area 1 are altered (high posterior) and that tensors elsewhere are unaltered (low posterior). This evidence increases as the number of subjects increases.

that the differences in latent group variables fall outside a region of practical equivalence (ROPE):

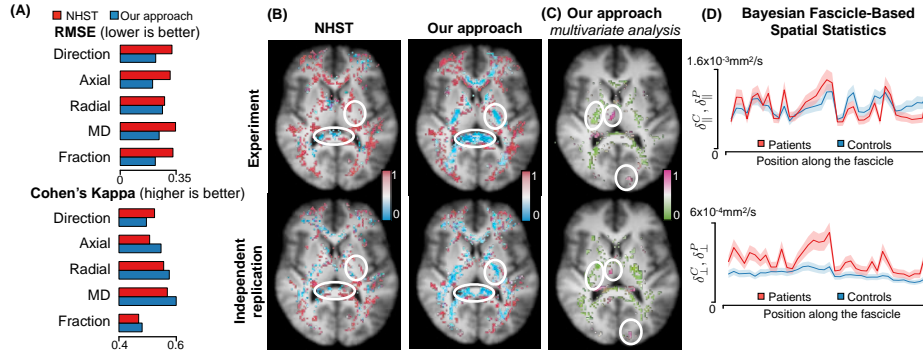
$$P\left(|\delta_{.,i} - \tilde{\delta}_{.,i}| > \epsilon_d | \text{Data}\right), P\left(|\varphi_i - \tilde{\varphi}_i| > \epsilon_f | \text{Data}\right), P\left(|\text{acos}(\boldsymbol{\varepsilon}_i \cdot \tilde{\boldsymbol{\varepsilon}}_i)| > \epsilon_e | \text{Data}\right).$$

The bounds of ROPE vary with the empirical context. In our experiments, we set  $\epsilon_d = 10^{-6} \text{mm}^2/\text{s}$ ,  $\epsilon_f = 0.01$ ,  $\epsilon_e = 1^\circ$ . Posterior values close to 1 are strong evidence for the presence of a microstructure abnormality, akin to small p-values in NHST. Posterior values close to zero are strong evidence that the difference between the groups is within the ROPE. This property of the posterior has no equivalent in NHST and has, as will be seen, far-reaching consequences in population studies.

## 5 Experiments and Results

**Implementation.** HMC sampling was implemented in Stan 2.2 ([mc-stan.org](https://mc-stan.org)), took an average 0.03 sec per chain and always reached full convergence ( $\hat{R} = 1$ ).

**Data.** Directly simulating multi-fascicle models is biased because the impact of the noise on model parameters is unclear. We therefore simulated the process of estimating models from noisy DWI. We created a control and a patient template with various crossing angles (Fig. 2A-B). The patient template has overlapping areas of microstructural differences: Area 1 has radial diffusivity of oblique fascicles inflated by 10%, Area 2 has in-plane orientation alterations of  $10^\circ$  and Area



**Fig. 3.** (A) RMSE and Cohen’s  $\kappa$  coefficient both indicate that inference results from the Bayesian approach are more reproducible than those obtained with NHST. (B) The improved reproducibility for the axial diffusivity is explained by the presence of many voxels with low values of the posterior (blue voxels) for which NHST results in uniformly distributed p-values. (C) Our framework can be applied to more complex analyses such as the coupled increase in fraction of isotropic diffusion and radial diffusivity. The resulting posterior map can be interpreted exactly as the others. (D) Augmenting the model with spatial priors enables the analysis of fascicle-wise properties, such as the diffusivity profile along the arcuate fasciculus.

3 has increased fraction of isotropic diffusion by 0.1. For each simulated subject, 90 DWI on 3 shells at  $b=1,2,3000$ , were corrupted by Rician noise with scale parameters  $\sigma = 10, 20, 40$ . Multi-fascicle models were then estimated as in [6] (Fig. 2C). For each  $\sigma$ , groups of  $K$  patients and  $K$  controls were generated for  $K = 10, 20, 30, 40, 50$ . The process was repeated 10 times for each pair  $(\sigma, K)$  leading to 150 synthetic populations. In-vivo DWI ( $1.7 \times 1.7 \times 2 \text{mm}^3$ , Siemens 3T Trio, 32-channel head coil) were acquired using the CUSP-45 sequence [6] with b-values up to 3000 in 36 patients with Tuberous Sclerosis Complex (TSC), a condition with high prevalence of autism, and 36 age-matched controls. Two-fascicle models with one isotropic compartment were estimated.

**Preprocessing.** All multi-fascicle models were non-linearly registered to an atlas as in [8]. At each voxel, the tensors of all  $J$  controls and  $K$  patients were grouped in  $N$  clusters using the iterative scheme of [8], where  $N$  is the number of fascicles in the atlas. Due to model selection, the number of observations  $(J_i + K_i)$  varies between fascicles. The presence or absence of a fascicle may itself be an interesting group contrast and is reflected in the vector of fractions  $\mathbf{f}$  that may contain null values (replaced by an epsilon machine to avoid singularity).

**Validation on synthetic phantoms.** Population studies were conducted with all 150 simulated populations and for five properties: the radial, axial and mean diffusivities, the orientation of fascicles and the fractions of isotropic diffusion. NHST for log-diffusivities and logit-fractions were performed by independent  $t$ -tests as in [8] and NHST for directions used Schwartzman’s test for diffusion

tensors [7]. For all properties, true positive rates and true negative rates were computed at different thresholds leading to ROC curves (Fig. 2D). Consistently for all noise levels and all number of subjects, our Bayesian approach outperforms NHST. This is also the case independently for each property that presents a group contrast. The improvement of our Bayesian approach over NHST is akin to tripling the number of subjects. This can be attributed to two features of the Bayesian framework. First, the introduction of ROPE avoids the detection of false positives with negligible yet statistically significant magnitude. Second, low values of the Bayesian posteriors is evidence for the absence of a group difference whereas high p-values occur randomly (uniformly) if the null hypothesis is true. This situation is illustrated in Fig. 2E-I for the radial diffusivity. Bayesian posteriors outside of Area 1 takes on small values (Fig. 2H-I) whereas p-values are uniformly distributed in those voxels.

**Validation on in-vivo data.** In-vivo data were used to assess the reproducibility of results. Patient and control groups were each split in two subgroups, leading to two cohorts each with 18 controls and 18 patients. Population studies for directions, isotropic fraction, radial, axial and mean diffusivities were conducted in each cohort. For each property, the resulting two p-value maps were compared to one another in terms of root mean squared error (RMSE) and so were the two posterior maps. For all five variables of interests, the RMSE was significantly smaller between posterior maps (mean RMSE: 0.23) than between p-value maps (mean RMSE: 0.32; two-sample t-test:  $p < 0.0005$ ) with a mean improvement of 27.5% (Fig. 3A). The largest improvement was observed for the axial diffusivity (reduction of 35%). RMSE directly compares the statistical maps regardless of the decision made from these maps (after thresholding). We therefore also assess reproducibility of inference results based on Cohen’s  $\kappa$  coefficient of agreement after thresholding p-values at 0.05 and posteriors at 0.95:  $\kappa = \frac{P_a - P_e}{1 - P_e}$  where  $P_a$  is the observed agreement probability and  $P_e$  is the probability of agreeing by chance.  $\kappa$  was larger with the Bayesian approach for all properties except for the directions (Fig. 3A). The improvement was larger for axial diffusivity (+6.5%). The higher improvement in reproducibility observed for the study of axial diffusivity both in terms of RMSE and  $\kappa$  is elucidated by Fig. 3(B). Strong evidence for the absence of any substantive difference in axial diffusivity can be inferred and reproduced with the Bayesian approach (blue areas in Fig. 3B). In these regions, the p-value is uniformly distributed leading to poor RMSE and  $\kappa$ .

**Prospectives.** So far, we focused on simple statistics to fairly compare our approach to NHST. But more complex statistics may too be of interest. For instance, the presence of myelin injury with neuroinflammation may result in a coupled increase in fraction of isotropic diffusion due to cell swelling and in radial diffusivity due to demyelination. This association remains hypothetical, but our capability to detect it may increase specificity of microstructure imaging. Detection of these complex responses is not easily framed as an NHST, but is readily computable with our Bayesian approach:  $p(\tilde{\varphi}_{\text{iso}} > \varphi_{\text{iso}}, \tilde{\delta}_{\perp,i} > \delta_{\perp,i} | \text{Data})$ , depicted in Fig. 3C. Our Bayesian framework also enables incorporation of prior

information, such as spatial priors to increase coherence of the posterior. To illustrate this possibility, we augment our model with a Gaussian Markov random field prior over latent diffusivities in adjacent voxels along a fascicle of interest: the median tract of the left arcuate fasciculus (AF). The results are joint posteriors of all diffusivities along the tract, leading to fascicle-wise (rather than voxel-wise) analysis. The profiles in Fig. 3D are maximum a posteriori (lines) and 95% posterior intervals (shades). They indicate that TSC patients have substantially higher radial but unaltered axial diffusivities along the AF.

## 6 Conclusion

This paper introduced a Bayesian framework to conduct population studies of the brain microstructure. A key property of this framework is its ability to build evidence for the absence of alterations alongside the detection of abnormalities. This key property makes it less prone to false discoveries than NHST and improves reproducibility of findings. By estimating the full posterior distribution over latent variables, our Bayesian framework therefore enables richer and more reliable analyses of the brain microstructure and its alterations in diseases.

***Acknowledgments*** This work was supported in part by F.R.S.-FNRS, WBI, NIH grants R01 NS079788, R01 LM010033, R01 EB013248, P30 HD018655, and by a research grant from the Boston Children’s Hospital Translational Research Program.

## References

1. Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B.: Bayesian data analysis. CRC press (2013)
2. Hoffman, M.D., Gelman, A.: The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research* (2013)
3. Nuzzo, R.: Scientific method: statistical errors. *Nature* 506(7487), 150–152 (2014)
4. Panagiotaki, E., Schneider, T., Siow, B., Hall, M.G., Lythgoe, M.F., Alexander, D.C.: Compartment models of the diffusion MR signal in brain white matter: a taxonomy and comparison. *Neuroimage* 59(3), 2241–2254 (2012)
5. Pasternak, O., Westin, C.F., Bouix, S., et al.: Excessive extracellular volume reveals a neurodegenerative pattern in schizophrenia onset. *The Journal of Neuroscience* 32(48), 17365–17372 (2012)
6. Scherrer, B., Warfield, S.K.: Parametric representation of multiple white matter fascicles from cube and sphere diffusion MRI. *PLoS one* 7(11), e48232 (2012)
7. Schwartzman, A., Dougherty, R., Taylor, J.: Cross-subject comparison of principal diffusion direction maps. *Magnetic Resonance in Medicine* 53(6), 1423–1431 (2005)
8. Taquet, M., Scherrer, B., Commowick, O., Peters, J.M., Sahin, M., Macq, B., Warfield, S.K.: A mathematical framework for the registration and analysis of multi-fascicle models for population studies of the brain microstructure. *Medical Imaging, IEEE Transactions on* 33(2), 504–517 (2014)
9. Taquet, M., Scherrer, B., Boumal, N., Macq, B., Warfield, S.K.: Estimation of a multi-fascicle model from single b-value data with a population-informed prior. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) *Medical Image Computing and Computer-Assisted Intervention, LNCS*, vol. 8149, pp. 695–702 (2013)