

Bayesian Diffusion Tensor Estimation with Spatial Priors

Xuan Gu, Per Sidén, Bertil Wegmann, Anders Eklund, Mattias Villani and Hans Knutsson

The self-archived version of this journal article is available at Linköping University Electronic Press:

<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-139844>

N.B.: When citing this work, cite the original publication.

Gu, X., Sidén, P., Wegmann, B., Eklund, A., Villani, M., Knutsson, H., (2017), Bayesian Diffusion Tensor Estimation with Spatial Priors, *CAIP 2017*. https://dx.doi.org/10.1007/978-3-319-64689-3_30

Original publication available at:

https://dx.doi.org/10.1007/978-3-319-64689-3_30

Copyright: [Springer](#)



Bayesian diffusion tensor estimation with spatial priors

Xuan Gu^{1,3}, Per Sidén², Bertil Wegmann², Anders Eklund^{1,2,3},
Mattias Villani², and Hans Knutsson^{1,3}

¹ Division of Medical Informatics, Department of Biomedical Engineering, Linköping University, Linköping, Sweden

² Division of Statistics and Machine Learning, Department of Computer and Information Science, Linköping University, Linköping, Sweden

³ Center for Medical Image Science and Visualization, Linköping University, Linköping, Sweden

Abstract. Spatial regularization is a technique that exploits the dependence between nearby regions to locally pool data, with the effect of reducing noise and implicitly smoothing the data. Most of the currently proposed methods are focused on minimizing a cost function, during which the regularization parameter must be tuned in order to find the optimal solution. We propose a fast Markov chain Monte Carlo (MCMC) method for diffusion tensor estimation, for both 2D and 3D priors data. The regularization parameter is jointly with the tensor using MCMC. We compare FA (fractional anisotropy) maps for various b-values using three diffusion tensor estimation methods: least-squares and MCMC with and without spatial priors. Coefficient of variation (CV) is calculated to measure the uncertainty of the FA maps calculated from the MCMC samples, and our results show that the MCMC algorithm with spatial priors provides a denoising effect and reduces the uncertainty of the MCMC samples.

Keywords: spatial regularization, diffusion tensor, spatial priors, Markov chain Monte Carlo, fractional anisotropy

1 Introduction

Diffusion MRI is a technique used for studying brain connectivity. It has been reported that neighboring voxels show similar diffusion signals, due to their high degree of spatial coherence in terms of underlying fiber microstructure (Aboitiz et al., 1992). The spatial information is ignored by widely used voxel-by-voxel diffusion tensor estimation algorithms, e.g. ordinary least squares (OLS) and weighted least squares (WLS) (Chung et al., 2006), which assume that the data in each voxel is independent. Algorithms with this assumption make the parameter estimation more vulnerable to the image noise. The robustness and reliability of the estimated diffusion parameters can be improved by incorporating spatial information in the parameter estimation.

Spatial regularization is a technique that encourages similarity between parameters over a neighborhood of voxels. This is typically done by minimizing a cost function which is often the sum of a data error term and a regularization term (Raj et al., 2011; Wang et al., 2003). One of the drawbacks of this method is that the weight of the regularization term must be tuned in order to find the optimal solution. However, determining the appropriate regularization weight is a difficult problem, and there is no general solution. An alternative but similar concept is the use of probabilistic spatial priors in a Bayesian framework to spatially regularise the model fitting procedure (Demiralp and Laidlaw, 2011; King et al., 2009; Poupon et al., 2001; Sidén et al., 2017; Walker-Samuel et al., 2011). The regularization strength is then governed by a hyperparameter that scales the prior precision matrix which can be estimated directly during the fitting procedure, rather than being estimated separately. Poupon et al. (2001) used a Markov random field (MRF) framework to obtain a regularized fiber orientation map, in order to improve the diffusion tractography. Martín-Fernández et al. (2003, 2004) proposed both Gaussian and non-Gaussian MRF approaches for the 2D spatial regularization of diffusion tensor fields. Martín-Fernández et al. (2004) extended the 2D Gaussian MRF method to 3D using multivariate Gaussian Markov random field (GMRF) ideas. The maximum a posteriori (MAP) estimator was found by the simulated annealing algorithm in all the three approaches. King et al. (2009) tackled the crossing-fiber problem using an MRF model and Markov chain Monte Carlo (MCMC) sampling for the multi-fiber ball and stick model outlined in (Behrens et al., 2007), which offered a useful solution for the crossing-fiber problem in diffusion tractography. Walker-Samuel et al. (2011) proposed a MCMC method using GMRF to incorporate spatial information when estimating the apparent diffusion coefficient (ADC).

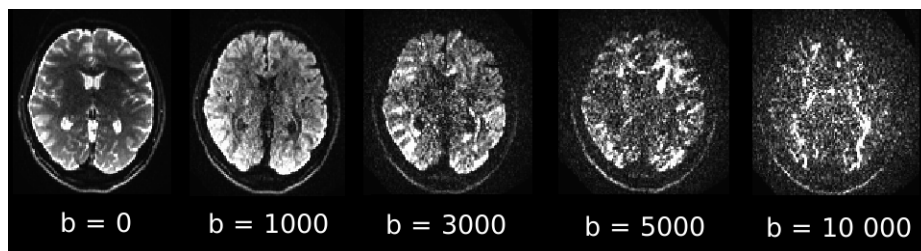


Fig. 1. Diffusion images using b-values of 0, 1000, 3000, 5000, and 10,000 s/mm^2 show progressively more diffusion weighting, but also reduced SNR.

Figure 1 shows a sequence of diffusion images for 5 different b-values. Increasing the b-value from 0 to 10,000 s/mm^2 creates greater diffusion weighting of the image, at the expense of a significant reduction of the signal-to-noise-ratio (SNR). The SNR issue becomes important at higher b-value diffusion imaging due to the larger exponential attenuation of the signal, longer echo time and in-

creased susceptibility (Graessner, 2011). To use the spatial dependency between the signals in neighbouring voxels for accurate inference, we extend a MCMC algorithm with spatial priors, previously used for fMRI analysis (Sidén et al., 2017), to diffusion MRI. A preconditioned conjugate gradient (PCG) approach is applied to make the MCMC algorithm a practical option for the whole brain analysis. The main contribution of this paper is that we use the PCG approach and MCMC to compute the posterior distribution with spatial priors for the whole brain, which was previously infeasible due to the high computational complexity. We calculate fractional anisotropy (FA) maps from three diffusion tensor estimation methods: least-squares, and MCMC with and without spatial priors for various b-values. The coefficient of variation (CV) is calculated to measure the dispersion of the FA maps from the MCMC samples.

2 Theory

2.1 Diffusion tensor estimation

In a diffusion experiment, the diffusion-weighted signal S_i of the i th measurement for one voxel is modeled by

$$S_i = S_0 \exp(-b\mathbf{g}_i^T \mathbf{D}\mathbf{g}_i), \quad \text{for } i = 1, 2, \dots, T, \quad (1)$$

where S_0 is the signal without diffusion weighting, b is the diffusion weighting factor, $\mathbf{D} = \begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{xy} & D_{yy} & D_{yz} \\ D_{xz} & D_{yz} & D_{zz} \end{bmatrix}$ is the diffusion tensor in the form of a 3×3 positive definite matrix, \mathbf{g}_i is a 3×1 unit vector of the gradient direction, and T is the total number of measurements. Using the log transform, the equation above becomes

$$\ln(S_i) = \ln(S_0) - b\mathbf{g}_i^T \mathbf{D}\mathbf{g}_i, \quad (2)$$

which, assuming additive noise on the log scale, can be structured into the well-known multiple linear regression form

$$\mathbf{y} = \mathbf{X}\mathbf{w} + \boldsymbol{\varepsilon}, \quad (3)$$

where $\mathbf{y} = [\ln(S_1), \ln(S_2), \dots, \ln(S_T)]^T$ represents the logarithm of the measured signal, $\mathbf{w} = [D_{xx}, D_{yy}, D_{zz}, D_{xy}, D_{xz}, D_{yz}, \ln S_0]^T$ are the unknown regression coefficients, \mathbf{X} is a $T \times 7$ design matrix containing the different diffusion gradient directions,

$$\mathbf{X} = -b \begin{bmatrix} g_{1x}^2 & g_{1y}^2 & g_{1z}^2 & 2g_{1x}g_{1y} & 2g_{1x}g_{1z} & 2g_{1y}g_{1z} & -\frac{1}{b} \\ g_{2x}^2 & g_{2y}^2 & g_{2z}^2 & 2g_{2x}g_{2y} & 2g_{2x}g_{2z} & 2g_{2y}g_{2z} & -\frac{1}{b} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ g_{Tx}^2 & g_{Ty}^2 & g_{Tz}^2 & 2g_{Tx}g_{Ty} & 2g_{Tx}g_{Tz} & 2g_{Ty}g_{Tz} & -\frac{1}{b} \end{bmatrix}, \quad (4)$$

and $\boldsymbol{\varepsilon} = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_T]^T$ are the error terms. We will consider diffusion data containing T volumes with N voxels ordered in a $T \times N$ matrix \mathbf{Y} . The linear regression model given in Equation 3 can be rewritten for simultaneous estimation of all the parameters, according to

$$\mathbf{Y} = \mathbf{X}\mathbf{W} + \mathbf{E}, \quad (5)$$

where \mathbf{W} is a $7 \times N$ matrix of regression coefficients and \mathbf{E} is a $T \times N$ matrix error terms.

2.2 Bayesian inference

In the Bayesian inference framework, the posterior distribution quantifies the uncertainty of the parameters \mathbf{W} given the data \mathbf{Y} , according to Bayes theorem:

$$p(\mathbf{W}|\mathbf{Y}) \propto p(\mathbf{Y}|\mathbf{W})p(\mathbf{W}), \quad (6)$$

where $p(\mathbf{W})$ is the prior information about the parameters, and $p(\mathbf{Y}|\mathbf{W})$ is the likelihood of \mathbf{Y} given \mathbf{W} . A closed form expression for the posterior distribution is normally not available, but MCMC can be used to produce samples from the posterior leading to inference that is asymptotically exact.

2.3 MCMC algorithm

Here we assume that the noise in each voxel is modeled as independent and identically distributed (i.i.d.) Gaussian noise. The likelihood then becomes

$$p(\mathbf{Y}|\mathbf{W}, \boldsymbol{\lambda}) = \prod_{n=1}^N \mathcal{N}(\mathbf{Y}_{:,n}; \mathbf{X}\mathbf{W}_{:,n}, \lambda_n^{-1}\mathbf{I}_T), \quad (7)$$

with $\mathbf{Y}_{:,n}$ and $\mathbf{W}_{:,n}$ denoting the n th column of \mathbf{Y} and \mathbf{W} , λ_n is the noise precision of voxel n and \mathbf{I}_T is the $T \times T$ identity matrix. The spatial part of the model is incorporated via the following prior on the regression coefficients \mathbf{W}

$$p(\mathbf{W}|\boldsymbol{\alpha}) = \prod_{k=1}^7 p(\mathbf{W}_{k,\cdot}^T|\alpha_k), \quad \mathbf{W}_{k,\cdot}^T|\alpha_k \sim \mathcal{N}(0, \alpha_k^{-1}\mathbf{D}_w^{-1}), \quad (8)$$

where $\mathbf{W}_{k,\cdot}^T$ denotes the transposed k th row of \mathbf{W} , \mathbf{D}_w is a $N \times N$ spatial precision matrix and $\boldsymbol{\alpha} = [\alpha_1, \alpha_1, \dots, \alpha_7]^T$ are hyperparameters that controls the smoothness, which are to be estimated from the data for each regressor in \mathbf{X} . We choose the unweighted graph-Laplacian (UGL) (Penny et al., 2005) which has 6's on the diagonal when modeling the 3D brain and 4's when modeling a 2D slice, and $\mathbf{D}_w(i, j) = -1$ if i and j are adjacent. The log likelihood can be expressed as

$$\begin{aligned} \log p(\mathbf{Y}|\mathbf{W}, \boldsymbol{\lambda}) &= \frac{T}{2} \sum_{n=1}^N \log(\lambda_n) \\ &\quad - \frac{1}{2} \sum_{n=1}^N \lambda_n (\mathbf{Y}_{:,n}^T \mathbf{Y}_{:,n} - 2\mathbf{Y}_{:,n}^T \mathbf{X}\mathbf{W}_{:,n} + \mathbf{W}_{:,n}^T \mathbf{X}^T \mathbf{X}\mathbf{W}_{:,n}), \end{aligned} \quad (9)$$

where we have ignored everything that is constant with respect to the parameters. Since \mathbf{Y} and \mathbf{X} will not change during the MCMC algorithm, quantities such as $\mathbf{X}^T\mathbf{X}$ can be effectively pre-computed, removing the time dimension from the likelihood which leads to significant speed up. We obtain closed form expressions for all full conditional posteriors, and can therefore perform Gibbs sampling. The full conditional posterior of \mathbf{W} is given by

$$\log p(\mathbf{W}|\mathbf{Y}, \boldsymbol{\lambda}, \boldsymbol{\alpha}) = -\frac{1}{2}\mathbf{w}_r^T \tilde{\mathbf{B}} \mathbf{w}_r + \mathbf{b}_w^T \mathbf{w}_r, \quad (10)$$

where $\mathbf{b}_w = \text{vec}(\text{diag}(\boldsymbol{\lambda})\mathbf{Y}^T\mathbf{X})$, $\tilde{\mathbf{B}} = \mathbf{X}^T\mathbf{X} \otimes \text{diag}(\boldsymbol{\lambda}) + \text{diag}(\boldsymbol{\alpha}) \otimes \mathbf{D}_w$, $\mathbf{w}_r = \text{vec}(\mathbf{W}^T)$. The full conditional posterior of $\boldsymbol{\lambda}$ can be written as

$$\log p(\boldsymbol{\lambda}|\mathbf{Y}, \mathbf{W}, \boldsymbol{\alpha}) = (\tilde{u}_2 - 1) \sum_{n=1}^N \log(\lambda_n) - \sum_{n=1}^N \frac{\lambda_n}{\tilde{u}_{1n}}, \quad (11)$$

where $\frac{1}{\tilde{u}_{1n}} = \frac{1}{2} (\mathbf{Y}_{:,n}^T \mathbf{Y}_{:,n} - 2\mathbf{Y}_{:,n}^T \mathbf{X} \mathbf{W}_{:,n}^T + \mathbf{W}_{:,n}^T \mathbf{X}^T \mathbf{X} \mathbf{W}_{:,n}) + \frac{1}{u_1}$, $\tilde{u}_2 = \frac{T}{2} + u_2$. The full conditional posterior of $\boldsymbol{\alpha}$ is given by

$$\log p(\boldsymbol{\alpha}|\mathbf{Y}, \mathbf{W}, \boldsymbol{\lambda}) = (\tilde{q}_2 - 1) \sum_{k=1}^7 \log(\alpha_k) - \sum_{k=1}^7 \frac{\alpha_k}{\tilde{q}_{1k}}, \quad (12)$$

where $\frac{1}{\tilde{q}_{1k}} = \frac{1}{2} (\mathbf{Y}_{:,n}^T \mathbf{X} \mathbf{W}_{:,n}^T) + \frac{1}{q_1}$, $\tilde{q}_2 = \frac{N}{2} + q_2$. We refer to Sidén et al. (2017) for further details.

2.4 PCG based sampling

Cholesky decomposition is usually the bottleneck for algorithms involving high dimensional GMRFs (Rue and Held, 2005). In practice it takes too long time to finish the whole 3D brain inference (if $N = 100,000$). Papandreou and Yuille (2010) proposed a sampling method from the posterior for \mathbf{W} that avoids the Cholesky decomposition. The main idea is to minimize $\mathbf{Y} - \mathbf{X}\mathbf{B}$ instead of solving $\mathbf{Y} = \mathbf{X}\mathbf{B}$. First we construct $\mathbf{B}_{data} = \text{diag}(\boldsymbol{\lambda}) \otimes \mathbf{X}^T\mathbf{X}$ and then we calculate $\mathbf{b} = (\text{blkdiag}[\sqrt{\alpha_k}\mathbf{G}_w])^T \mathbf{z}_1 + \mathbf{H}_w^T \mathbf{L}_{data} \mathbf{H}_w \mathbf{z}_2 + \mathbf{b}_w$. \mathbf{z}_1 and \mathbf{z}_2 are random draws from a Gaussian distribution with zero mean and identity covariance. \mathbf{G}_w needs to satisfy $\mathbf{D}_w = \mathbf{G}_w^T \mathbf{G}_w$. \mathbf{H}_w is defined as the permutation matrix such that $\text{vec}(\mathbf{W}) = \mathbf{H}_w \text{vec}(\mathbf{W}^T)$ (Penny et al., 2007). \mathbf{L}_{data} is the Cholesky factor of \mathbf{B}_{data} . Lastly compute \mathbf{M} as the incomplete Cholesky factor of $\tilde{\mathbf{B}}$ and solve $\tilde{\mathbf{B}}\mathbf{w}_r = \mathbf{b}$ approximately using preconditioned conjugate gradient (PCG) with preconditioner \mathbf{M} . PCG is a conjugate gradient method with some preconditions set to ensure fast convergence. Please note that $\tilde{\mathbf{B}}$ has to be reordered using a reordering method (Amestoy et al., 1996), and the same reordering has to be used for \mathbf{b} . Using the inverse reordering, posterior samples \mathbf{w}_r can be obtained.

3 Data

We use the MGH adult diffusion dataset from the Human Connectome Project (HCP) (Van Essen et al., 2013). Data were collected from 35 healthy adults scanned on a customized Siemens 3T Connectom scanner with 4 different b -values (1000, 3000, 5000 and 10,000 s/mm²). The data has already been pre-processed for gradient nonlinearity correction, motion correction and eddy current correction (Glasser et al., 2013). The data consists of 40 non-diffusion weighted volumes ($b = 0$), 64 volumes for $b = 1000$ and 3000 s/mm², 128 volumes for $b = 5000$ s/mm² and 256 volumes for $b = 10,000$ s/mm², which yields 552 volumes of $140 \times 140 \times 96$ voxels with an 1.5 mm isotropic voxel size.

Data used in the preparation of this work were obtained from the Human Connectome Project (HCP) database (<https://ida.loni.usc.edu/login.jsp>). The HCP project (Principal Investigators: Bruce Rosen, M.D., Ph.D., Martinos Center at Massachusetts General Hospital; Arthur W. Toga, Ph.D., University of Southern California, Van J. Weeden, MD, Martinos Center at Massachusetts General Hospital) is supported by the National Institute of Dental and Craniofacial Research (NIDCR), the National Institute of Mental Health (NIMH) and the National Institute of Neurological Disorders and Stroke (NINDS). HCP is the result of efforts of co-investigators from the University of Southern California, Martinos Center for Biomedical Imaging at Massachusetts General Hospital (MGH), Washington University, and the University of Minnesota.

4 Results

In Figure 2 and 3 we present FA maps obtained from three diffusion tensor estimation methods: a standard least-squares method, and MCMC with and without 2D/3D spatial prior. The formula for FA is

$$\text{FA} = \sqrt{\frac{(\lambda_1 - \lambda_2)^2 + (\lambda_2 - \lambda_3)^2 + (\lambda_3 - \lambda_1)^2}{2(\lambda_1^2 + \lambda_2^2 + \lambda_3^2)}}. \quad (13)$$

The number of MCMC samples we use is 1000, and the number of samples for burn-in is 500. For MCMC with and without spatial priors, we calculated the average FA maps over all MCMC draws. It is known that a higher b -value will result in diffusion data with lower SNR. To investigate the effect of the b -value, we decompose the diffusion data into 5 parts, each of which consists of volumes for a single b -value (0, 1000, 3000, 5000 and 10,000 s/mm²). In theory it is possible to estimate the diffusion tensor with at least two b -values. For each method, we calculate FA maps for 4 different combinations of b -values ($b = 0/1000$, $0/3000$, $0/5000$, $0/10,000$ s/mm²), and the entire dataset. Each FA map is normalized by its mean, to adjust the FA maps to a common scale. As showed in the first row of Figures 2 and 3, the FA maps from the least-squares method are quite noisy. Comparing the least-squares approach to the MCMC approach in Figures 2 and 3, we can see a denoising effect in MCMC

with spatial priors. Also, one can see a denoising effect for the FA maps from MCMC with spatial priors, compared with those from MCMC without spatial priors, as showed in the second row of Figures 2 and 3. The MCMC algorithm with a spatial prior provides a better denoising effect for data with higher b -values, since that data has lower SNR. However, the denoising effect of using a spatial prior is clearly weaker for the entire dataset, as we can see from the fifth column in Figures 2 and 3. The absolute differences between the FA maps from MCMC with and without spatial priors are rather small for the entire dataset. In the fourth row of Figures 2 and 3, one can see that the 3D spatial prior tends to provide a very close absolute difference to its 2D counterpart.

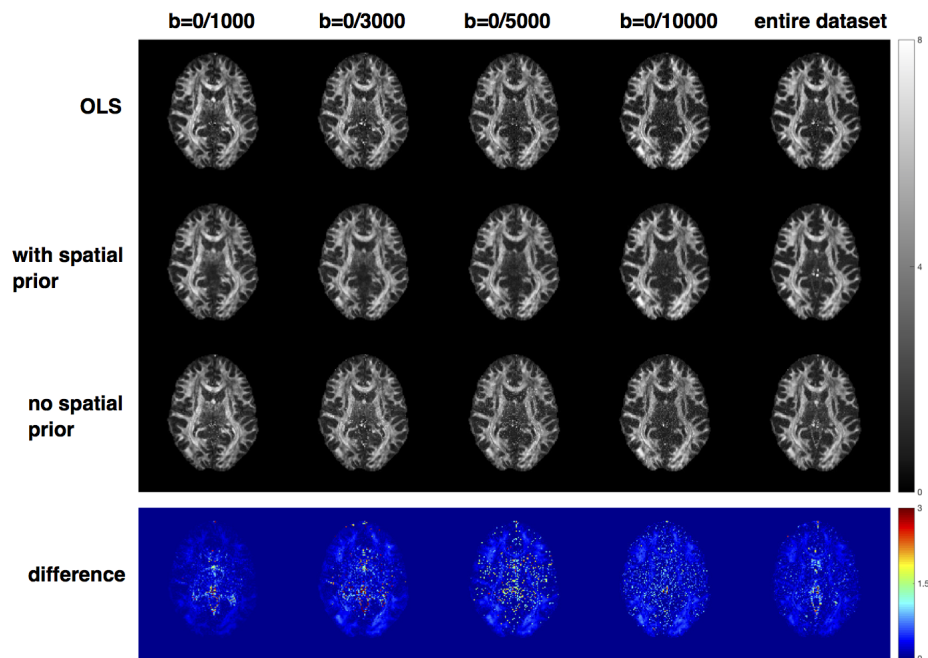


Fig. 2. FA maps from three diffusion tensor estimation methods. **First row:** least-squares method, **second row:** MCMC with 2D spatial priors, **third row:** MCMC without 2D spatial priors, **fourth row:** absolute difference of FA maps from MCMC with and without 2D spatial priors. **Columns from left to right:** $b = 0/1000, 0/3000, 0/5000, 0/10,000$ s/mm² and the entire dataset. Each FA map was normalized by its mean, to adjust the FA maps to a common scale. For the FA maps, Values no greater than 0 are mapped to the first color in the colormap, and values no less than 8 are mapped to the last color in the colormap. For the absolute difference of FA maps, the scaling interval is 0 to 3.

We use the CV (the ratio of the standard deviation to the mean) to measure the uncertainty of the FA maps calculated from the MCMC samples, as showed in

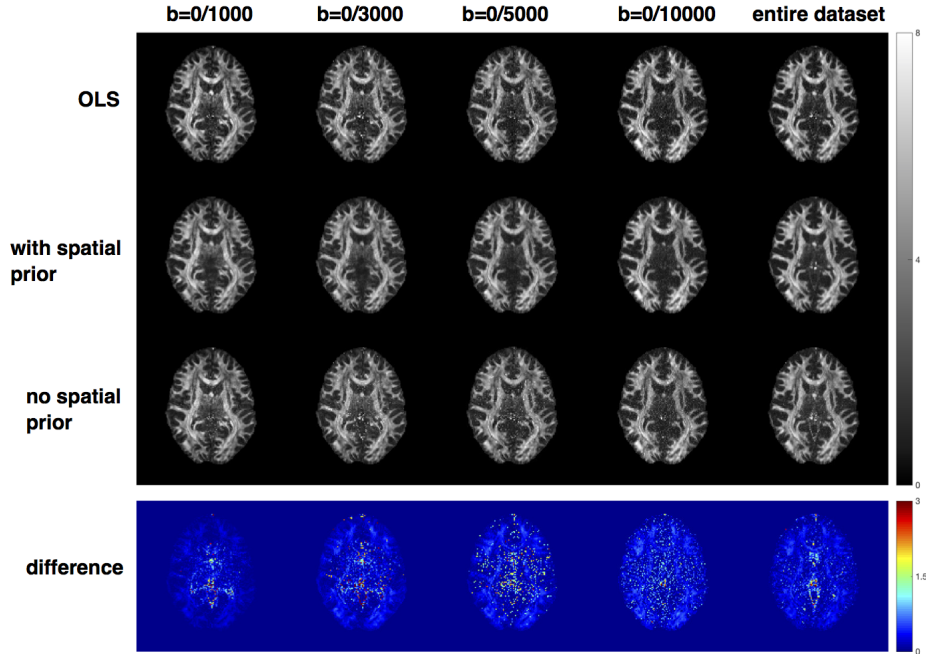


Fig. 3. FA maps from three diffusion tensor estimation methods. **First row:** least-squares method, **second row:** MCMC with 3D spatial priors, **third row:** MCMC without 3D spatial priors, **fourth row:** absolute difference of FA maps from MCMC with and without 3D spatial priors. **Columns from left to right:** $b = 0/1000, 0/3000, 0/5000, 0/10,000$ s/mm² and the entire dataset. Each FA map was normalized by its mean, to adjust the FA maps to a common scale. For the FA maps, Values no greater than 8 are mapped to the first color in the colormap, and values no less than 0 are mapped to the last color in the colormap. For the absolute difference of FA maps, the scaling interval is 0 to 3.

Figures 4 and 5. A lower uncertainty is found for some white matter voxels in the third rows, for both the 2D and 3D priors. For $b = 0/1000$ and $0/3000$ s/mm², the decrease in uncertainty was clearly higher than other combinations of b -values. Comparing the third row in Figures 4 and 5 we see that the 3D prior provides slightly better performance in lowering the uncertainty of the FA maps than the 2D prior for entire dataset.

5 Discussion

Performing Bayesian inference with spatial priors for the whole brain ($N = 294,000$) becomes a practical option with the PCG based MCMC sampling. The results presented in this paper demonstrate that the MCMC algorithm with spatial priors provides improvements by reducing the uncertainty of the MCMC samples. The presented FA maps from three diffusion tensor estimation

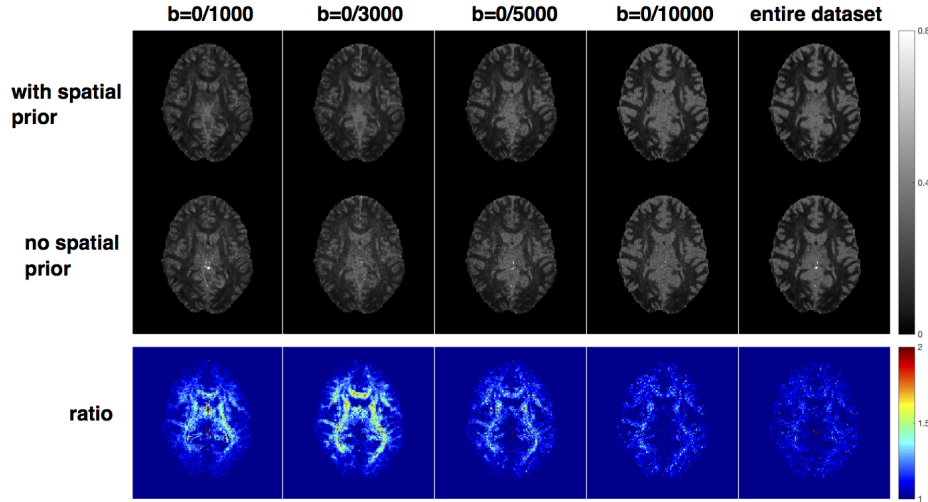


Fig. 4. CV of FA maps calculated from the MCMC samples. **First row:** MCMC with 2D spatial priors, **second row:** MCMC without 2D spatial priors, **third row:** ratios of CV of FA maps calculated from MCMC with and without 2D spatial priors. **Columns from left to right:** $b = 0/1000$, $0/3000$, $0/5000$, $0/10,000$ s/mm² and the entire dataset. For the CV, Values no greater than 0 are mapped to the first color in the colormap, and values no less than 0.4 are mapped to the last color in the colormap. For the ratio of the CV, the scaling interval is 1 to 2.

methods show that the MCMC method provides a denoising effect compared with the least-squares method. The denoising effect comes from the spatial priors. We have seen that MCMC sampling with spatial priors provides a denoising effect for various combinations of b-values. For $b = 0/1000$, $0/3000$, $0/5000$, $0/10000$ s/mm², MCMC with spatial priors works much better than MCMC without spatial priors. For the entire dataset, the differences of the performance between MCMC with and without spatial priors become smaller. For diffusion tensor model, MCMC with 3D spatial prior provides very close performance to MCMC with 2D spatial prior. The uncertainty of the FA maps from the posterior distributions generated with MCMC is reduced by the spatial prior, especially for the b-value combinations $b = 0/1000$ and $0/3000$ s/mm². We have noticed that most of the voxels with reduced uncertainty are located in white matter, where the FA values are relatively large.

In this paper, we have only considered the isotropic and stationary spatial prior (UGL). One potential improvement of this work is to use an anisotropic and non-stationary spatial prior, e.g. replacing the UGL prior with a weighted graph-Laplacian (WGL) prior. Wegmann et al. (2017) showed that the FA values can be greatly underestimated for methods that take the logarithm of the diffusion measurements. Thus in the formulation of the linear regression model in Equation 3, using a logarithmic link function (Wegmann et al., 2017) instead

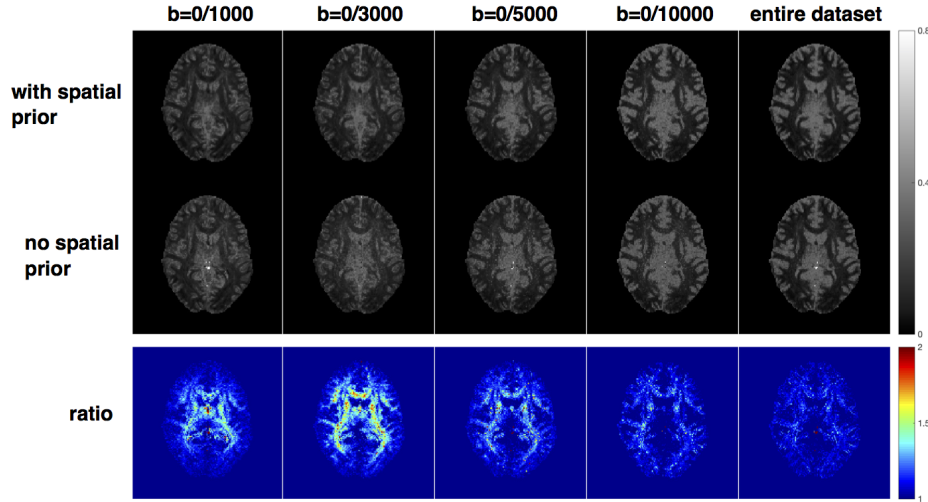


Fig. 5. CV of FA maps calculated from the MCMC samples. **First row:** MCMC with 3D spatial priors, **second row:** MCMC without 3D spatial priors, **third row:** ratios of CV of FA maps calculated from MCMC with and without 3D spatial priors. **Columns from left to right:** $b = 0/1000$, $0/3000$, $0/5000$, $0/10,000$ s/mm² and the entire dataset. For the CV, Values no greater than 0 are mapped to the first color in the colormap, and values no less than 0.4 are mapped to the last color in the colormap. For the ratio of the CV, the scaling interval is 1 to 2.

of taking the logarithm of the diffusion measurements might be a better option. Also, to make the diffusion tensor positive definite, if it is possible to impose the restriction using a log-Cholesky representation (Koay, 2010) of the diffusion tensor. Raj et al. (2011) reported that reconstruction of fiber orientation distribution function (ODF) for high angular resolution diffusion imaging (HARDI) using spatial priors provides even stronger advantages, compared with the simpler diffusion tensor model. For the latter case, only extremely noisy data benefit from spatial priors. For future work, we will extend the MCMC method in this paper to HARDI models, such as q-ball imaging and diffusion spectrum imaging, where the ODF can be estimated via a linear estimator.

Acknowledgements

This research was supported by the Information Technology for European Advancement (ITEA) 3 Project BENEFIT (better effectiveness and efficiency by measuring and modelling of interventional therapy) and the Swedish Research Council (grant 2015-05356, "Learning of sets of diffusion MRI sequences for optimal imaging of micro structures" and grant 2013-5229 "Statistical analysis of fMRI data").

Data collection and sharing for this project was provided by the Human Connectome Project (HCP; Principal Investigators: Bruce Rosen, M.D., Ph.D., Arthur W. Toga, Ph.D., Van J. Weeden, MD). HCP funding was provided by the National Institute of Dental and Craniofacial Research (NIDCR), the National Institute of Mental Health (NIMH), and the National Institute of Neurological Disorders and Stroke (NINDS). HCP data are disseminated by the Laboratory of Neuro Imaging at the University of Southern California.

Bibliography

- Aboitiz, F., Scheibel, A. B., Fisher, R. S., and Zaidel, E. (1992). Fiber composition of the human corpus callosum. *Brain research*, 598(1):143–153.
- Amestoy, P. R., Davis, T. A., and Duff, I. S. (1996). An approximate minimum degree ordering algorithm. *SIAM Journal on Matrix Analysis and Applications*, 17(4):886–905.
- Behrens, T. E., Berg, H. J., Jbabdi, S., Rushworth, M., and Woolrich, M. (2007). Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? *Neuroimage*, 34(1):144–155.
- Chung, S., Lu, Y., and Henry, R. G. (2006). Comparison of bootstrap approaches for estimation of uncertainties of DTI parameters. *NeuroImage*, 33(2):531–541.
- Demiralp, C. and Laidlaw, D. H. (2011). Generalizing diffusion tensor model using probabilistic inference in Markov random fields. In *MICCAI CDMRI Workshop*.
- Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J. R., et al. (2013). The minimal preprocessing pipelines for the Human Connectome Project. *Neuroimage*, 80:105–124.
- Graessner, J. (2011). Frequently Asked Questions: Diffusion-Weighted Imaging (DWI). *MAGNETON Flash*, pages 6–9.
- King, M. D., Gadian, D. G., and Clark, C. A. (2009). A random effects modelling approach to the crossing-fibre problem in tractography. *NeuroImage*, 44(3):753–768.
- Koay, C. G. (2010). Least squares approaches to diffusion tensor estimation. *Diffusion MRI*, page 272.
- Martín-Fernández, M., San José-Estépar, R., Westin, C.-F., and Alberola-López, C. (2003). A novel Gauss-Markov random field approach for regularization of diffusion tensor maps. In *International Conference on Computer Aided Systems Theory*, pages 506–517. Springer.
- Martín-Fernández, M., Westin, C.-F., and Alberola-López, C. (2004). 3D Bayesian regularization of diffusion tensor MRI using multivariate Gaussian Markov random fields. *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2004*, pages 351–359.
- Papandreou, G. and Yuille, A. L. (2010). Gaussian sampling by local perturbations. In *Advances in Neural Information Processing Systems*, pages 1858–1866.

- Penny, W., Flandin, G., and Trujillo-Barreto, N. (2007). Bayesian comparison of spatially regularised general linear models. *Human brain mapping*, 28(4):275–293.
- Penny, W. D., Trujillo-Barreto, N. J., and Friston, K. J. (2005). Bayesian fMRI time series analysis with spatial priors. *NeuroImage*, 24(2):350–362.
- Poupon, C., Mangin, J.-F., Clark, C. A., Frouin, V., Régis, J., Le Bihan, D., and Bloch, I. (2001). Towards inference of human brain connectivity from MR diffusion tensor data. *Medical Image Analysis*, 5(1):1–15.
- Raj, A., Hess, C., and Mukherjee, P. (2011). Spatial HARDI: improved visualization of complex white matter architecture with Bayesian spatial regularization. *Neuroimage*, 54(1):396–409.
- Rue, H. and Held, L. (2005). *Gaussian Markov random fields: theory and applications*. CRC Press.
- Sidén, P., Eklund, A., Bolin, D., and Villani, M. (2017). Fast Bayesian whole-brain fMRI analysis with spatial 3D priors. *NeuroImage*, 146:211–225.
- Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E., Yacoub, E., Ugurbil, K., Consortium, W.-M. H., et al. (2013). The WU-Minn human connectome project: an overview. *Neuroimage*, 80:62–79.
- Walker-Samuel, S., Orton, M., Boulton, J. K., and Robinson, S. P. (2011). Improving apparent diffusion coefficient estimates and elucidating tumor heterogeneity using Bayesian adaptive smoothing. *Magnetic resonance in medicine*, 65(2):438–447.
- Wang, Z., Vemuri, B. C., Chen, Y., and Mareci, T. (2003). A constrained variational principle for direct estimation and smoothing of the diffusion tensor field from DWI. In *Biennial International Conference on Information Processing in Medical Imaging*, pages 660–671. Springer.
- Wegmann, B., Eklund, A., and Villani, M. (2017). Bayesian heteroscedastic regression for diffusion tensor imaging. In *Modeling, Analysis, and Visualization of Anisotropy*. Springer.