Orientation Prior and Consistent Model Selection Increase Sensitivity of Tract-Based Spatial Statistics in Crossing-Fiber Regions

Arkesteijn, Joor; Poot, Dirk; Ikram, M.A.; Niessen, Wiro; van Vliet, Lucas; Vernooij, M.W.; Vos, Frans

DOI
10.1109/TMI.2019.2922615

Publication date
2019

Document Version
Final published version

Published in
IEEE Transactions on Medical Imaging

Citation (APA)
Arkesteijn, J., Poot, D., Ikram, M. A., Niessen, W., van Vliet, L., Vernooij, M. W., & Vos, F. (2019). Orientation Prior and Consistent Model Selection Increase Sensitivity of Tract-Based Spatial Statistics in Crossing-Fiber Regions. *IEEE Transactions on Medical Imaging*, *39*(2), 308-319. Article 8736508. https://doi.org/10.1109/TMI.2019.2922615

TMI-2018-1329

Orientation Prior and Consistent Model Selection Increase Sensitivity of Tract-Based Spatial Statistics in Crossing-Fiber Regions

G. A. M. Arkesteijn, D. H. J. Poot, M. A. Ikram, W. J. Niessen, L. J. van Vliet, M. W. Vernooij and F. M. Vos*

***Abstract*—The goal of this paper is to increase the statistical power of crossing-fiber statistics in voxelwise analyses of diffusion-weighted magnetic resonance imaging (DW-MRI) data. In the proposed framework a fiber orientation atlas and a model complexity atlas were used to fit the ball-and-sticks model to diffusion-weighted images of subjects in a prospective population-based cohort study. Reproducibility and sensitivity of the partial volume fractions in the ball-and-sticks model were analyzed using TBSS (tract-based spatial statistics), and were compared to a reference framework. The reproducibility was investigated on two scans of 30 subjects acquired with an interval of approximately three weeks by studying the intraclass correlation coefficient (ICC). The sensitivity to true biological effects was evaluated by studying the regression with age on 500 subjects between 65 and 90 years old. Compared to the reference framework, the ICC improved significantly when using the proposed framework. Higher t-statistics indicated that regression coefficients with age could be determined more precisely with the proposed framework, and more voxels correlated significantly with age. The application of a fiber orientation atlas and a model complexity atlas can significantly improve the reproducibility and sensitivity of crossing-fiber statistics in TBSS.**

***Index Terms*—DW-MRI, TBSS, atlas orientation prior**

I. Introduction

Diffusion-weighted magnetic resonance imaging (DW-MRI) is a non-invasive imaging technique in which image contrast is determined by the (hindered) molecular diffusion of water [1]. It is frequently used to assess the brain's white matter integrity, because it provides insight into the microstructural organization of neural fibers [2]. A popular application is diffusion tensor imaging (DTI), in which the water diffusion is modeled by a single Gaussian diffusion profile [3]. From the diffusion tensor, quantitative DTI metrics such as the fractional anisotropy (FA) and mean diffusivity (MD) can be derived. These DTI metrics are often used as an imaging biomarker for white matter tract integrity, to study for example neurodegenerative diseases or brain ageing [4, 5].

It is well known that the assumption of a single diffusion tensor model to represent the underlying diffusion within the volume of a single voxel is not always valid, e.g. in voxels with more than one coherently orientated fiber population [6, 7]. In such voxels, the analysis of conventional DTI metrics has undesirable effects: spurious changes may be detected in the radial and axial diffusivity [8], FA may lack sensitivity to detect changes in the white matter microstructure [9], and FA may seem to be increased merely due to selective degeneration of a fiber population [10].

Several alternative models have been proposed to provide a more adequate description of the diffusion in fiber crossings, e.g. the ball-and-sticks model [11], multi-tensor models [7. 12], or CHARMED [13]. These models describe the diffusion signal in crossing-fiber configurations by modeling each fiber population independently, and enable fiber population-specific characterization and comparison of the underlying microstructure. However, these models typically require more extensive DW-MRI protocols, making the image acquisition lengthy. Furthermore, sophisticated routines are needed to determine the appropriate number of fiber populations by using either explicit [14] or implicit model selection [11].

Researchers frequently use voxelwise analyses of DW-MRI data to localize changes in diffusion parameters in group studies. A popular framework for such an analysis is TBSS (tract-based spatial statistics) [15], which applies an FA-driven registration to establish spatial correspondence of all subjects in a common space. Subsequently, individual FA features are projected on a mean 'tract' skeleton for subsequent statistical analysis. However, applying this approach to evaluate statistics in fiber-crossings is not straightforward. In that case, not only the spatial coordinates but also the fiber population-specific metrics need to correspond across subjects for a meaningful analysis.

Manuscript submitted December 13, 2018. This work was financially supported as part of the STW Perspectiefprogramme Population Imaging Genetics (ImaGene) supported by the Dutch Technology Foundation (STW), project 12722. *Asterisk indicates corresponding author*.

G. A. M. Arkesteijn, D. H. J. Poot, W. J. Niessen, L. J. van Vliet and F. M. Vos are with the Quantitative Imaging Group, Department of Imaging Physics, Delft University of Technology, Delft, the Netherlands (e-mail: f.m.vos@tudelft.nl).

M. A. Ikram and M. W. Vernooij are with the department of Radiology and Nuclear Medicine and department of Epidemiology, Erasmus MC, Rotterdam, the Netherlands.

G. A. M. Arkesteijn, D. H. J. Poot, and W. J. Niessen are also with the departments of Medical Informatics and Radiology, Erasmus MC, Rotterdam, the Netherlands.

F. M. Vos is also with the department of Radiology, Academic Medical Center, University of Amsterdam, the Netherlands.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2019.2922615, IEEE Transactions on Medical Imaging

Recently, a framework was proposed to analyze crossing-fiber statistics in TBSS using a front-evolution algorithm to label the fiber populations based on their estimated orientations [16]. However, the fiber orientation-based labeling problem does not necessarily have a trivial solution in every voxel. Especially for conventional DW-MRI data (single non-zero *b*-value), the estimated fiber orientations and model selection routines can be imprecise. In effect, inconsistent metrics may be obtained from fiber populations across subjects, which in turn may cause crossing-fiber statistics to lose statistical power. This may explain why crossing-fiber statistics were less sensitive to ageing-effects than conventional metrics from a single diffusion tensor in [16].

The goal of this paper is to introduce a framework that strengthens the statistical power of crossing-fiber statistics in voxelwise analyses of conventional DW-MRI data. This is achieved by reducing fluctuations in the orientations of the estimated fiber populations and preventing inconsistencies in the number of fiber populations, through two additions to the fitting procedure of a crossing-fibers model. The first addition is the introduction of an 'orientation prior' into the estimation of the model parameters. This prior promotes correspondence of estimated fiber orientations across different subjects. The second addition is a 'consistent model selection', obtained by determining the number of fiber populations in a common space instead of independently in each subject.

We evaluate the impact of this new approach on the reproducibility and sensitivity of crossing-fiber statistics. More specifically, the reproducibility is evaluated on two scans of 30 subjects, acquired with an interval of approximately three weeks. Furthermore, the sensitivity to detect ageing effects is investigated in a group of 500 community-dwelling subjects aged 65 to 90 years. In this work we restrict our investigation to the analysis of partial volume fractions (PVFs) in the well-known ball-and-sticks model using TBSS, but the proposed changes are straightforward and easy to generalize to other generative crossing-fiber models and different frameworks for voxelwise analyses. The proposed framework is compared to standard approaches based on the conventional, single tensor representation and another approach relying on the ball-and-sticks model.

II. Methods

A. Overview of the proposed framework

An overview of the proposed framework is provided in Fig. 1. Two versions of the ball-and-sticks model were fitted to the diffusion-weighted images (DWIs). The first with one stick compartment and the second with two stick compartments. The *ball-and-two-sticks* model was regularized using a fiber orientation atlas (i.e. the average fiber orientations in a large population) to enhance the precision of the fit. Next, an adjusted TBSS pipeline was used to transform and skeletonize the estimated PVFs into a common space. In the common space the different skeletonized datasets were merged according to a model complexity atlas: in two-fiber skeleton voxels the PVFs originated from the ball-and-two-sticks diffusion model, whereas in single fiber voxels the first PVF originated from the ball-and-one stick model and the second PVF was set to zero. As such, consistent model selection was enforced across the subjects of the study population. The orientation atlas, complexity atlas, and source code to fit all the diffusion models to the diffusion data are made available in an online repository (https://github.com/imphys/fit-diffusion-model).

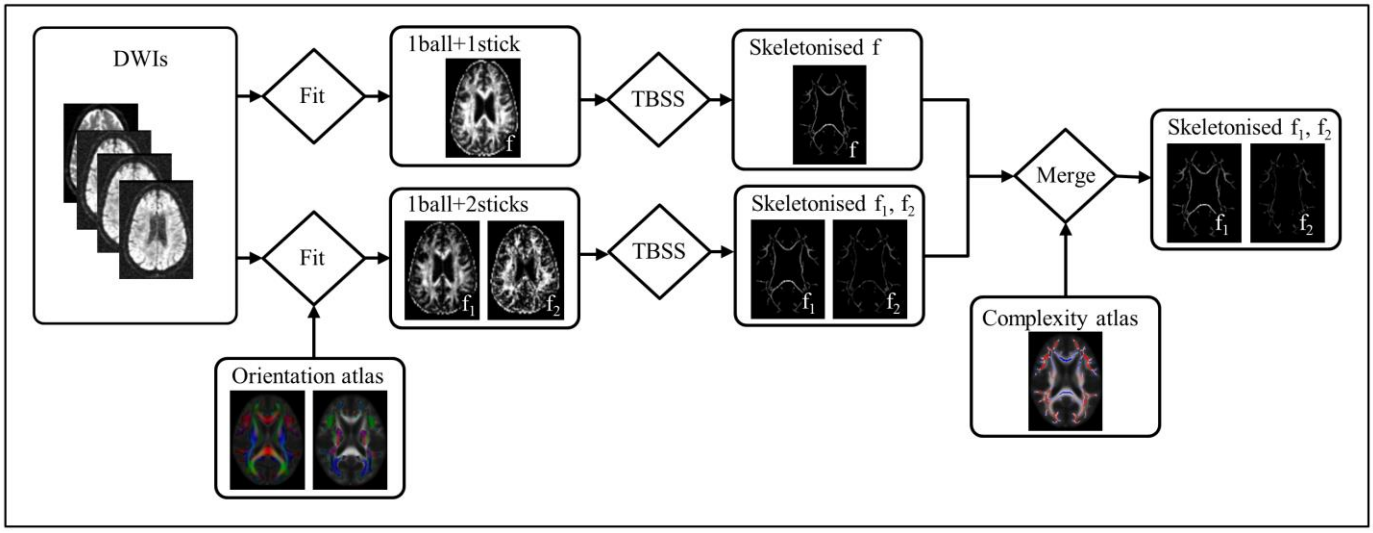

Fig. 1. Overview of the proposed framework. Two diffusion models, i.e. a ball-and-one-stick and a ball-and-two-sticks model, are fitted to the diffusion-weighted images (DWIs) per subject. An orientation prior, warped from the common space to subject space, is used to initialize the fit of the ball-and-two-sticks model and also acts as an orientation prior during fitting. The estimated partial volume fractions (PVFs) are warped and skeletonized using an adjusted TBSS pipeline. In the common space, the different skeletonized datasets are merged according to a complexity atlas, such that two skeletonized datasets remain. In single-fiber and crossing-fiber voxels (according to the model complexity atlas) the two skeletonized datasets contain PVFs from a ball-and-one-stick and a ball-and-two-sticks model respectively.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2019.2922615, IEEE Transactions on Medical Imaging

B. Study population

The proposed framework was evaluated using a subset of brain imaging data from the Rotterdam Study, a prospective population-based cohort study among middle aged and elderly subjects in a district of the city of Rotterdam, the Netherlands [17]. The Rotterdam Study has been approved by the medical ethics committee according to the Population Study Act Rotterdam Study, executed by the Ministry of Health, Welfare and Sports of the Netherlands. Written informed consent was obtained from all participants. The reproducibility of extracting crossing-fiber diffusion parameters was evaluated from two scans of 30 subjects, acquired on the same scanner with an average time interval of 19.5 days (SD 10). We will refer to this dataset as the reproducibility dataset. The subjects in this dataset were on average 76.7 (SD 4.8) years old, 50% was female. Further evaluation of the framework was performed on a group of 500 subjects from the Rotterdam Study, sampled from the entire population such that their ages were uniformly distributed between 65 and 90 years old. We will refer to this data as the ageing dataset. The mean age in this group was 77.3 (SD 7.0) years old, 45% was female. No subjects were diagnosed with dementia at the time of the MRI scans.

C. Data acquisition

All subjects were scanned on a 1.5 Tesla MRI scanner (GE Signa Excite) using an 8-channel head coil. No major hardware or software updates were performed on the scanner throughout the study [18]. DWIs were acquired with a single shot, diffusion-weighted spin echo echo-planar imaging sequence using a repetition time TR = 8575 ms, an echo time TE = 82.6 ms, a field of view FOV = 210x210 mm^2, an imaging matrix = 96x64 (zero-padded to 256x256), 35 contiguous slices with a thickness of 3.5 mm, and (hence) a native voxel size of 2.2 mm x 3.3 mm x 3.5 mm. DWIs were acquired in 25 non-collinear directions with a *b*-value of 1000 s/mm^2. Three volumes were acquired without diffusion weighting (the b_0-volumes) [18].

D. DWI preprocessing

The acquired DWIs were corrected for motion and eddy current distortion by affine coregistration to the second acquired b_0-volume with Elastix [19]. Note that FSL topup/eddy [20, 21] cannot be applied due to absence of reversed phase encoding. The DWIs were resampled to $(2.5\ mm)^3$ cubic resolution while the affine transformations were applied. After coregistration to the reference b_0-volume, the gradient directions were reoriented according to the rotation component of the transformation [22].

E. Ball-and-sticks model

In the ball-and-sticks model [11], the diffusion-weighted signal $\mathbf{S_0}$ is modeled according to:

$$S_{\mathbf{0},i} = S_0\left(\left(1-\sum_{j=1}^{N} f_j\right)\exp\left(-b_i d\right)+\sum_{j=1}^{N} f_j \exp\left(-b_i d\left(\mathbf{V}_j\cdot\mathbf{g}_i\right)^2\right)\right), \quad (1)$$

where b_i is the diffusion-weighting parameter, $\mathbf{g}_i$ is a unit vector that specifies the direction of the diffusion-encoding gradient pulses, S_0 is the non-diffusion-weighted signal, N is the number of stick compartments (0, 1 or 2 in this paper), d is a diffusivity parameter, f_j is the partial volume fraction (PVF) and $\mathbf{V}_j$ the principal eigenvector of the j^{th} stick compartment. As in [11], the eigenvectors $\mathbf{V}_j$ are parameterized using spherical coordinates ψ_j and φ_j.

F. Automatic relevance determination

The fiber orientation atlas and model complexity atlas were constructed based on automatic relevance determination (ARD). Prior work has used ARD to estimate the unknown parameters in the ball-and-sticks model [11]. In the ARD framework (available through the function `bedpostx` as part of FSL [11]), a Bayesian approach is applied to fit a ball-and-sticks model with two stick compartments in every voxel. We will refer to the stick associated with the largest volume fraction as the primary fiber population; the stick with the smallest volume fraction will be called the secondary fiber population. Overfitting is avoided by using a shrinkage prior that automatically reduces the volume fraction of the secondary fiber population to zero when it is not supported by the data.

G. Construction of the fiber orientation atlas

To construct the orientation atlas, first the diffusion tensor parameters were estimated in all 500 subjects using the fit_MRI toolbox by maximum likelihood estimation [23] assuming Rician distributed noise. Additionally, the ball-and-two-sticks parameters were estimated through ARD. DTI-TK was used to align the diffusion tensor images to a population-specific common space using the *full* diffusion tensor information [24]. Subsequently, the DTI-TK functions `deformationScalar3DVolume` and `deformationSymTensor3DVolume` were used to warp the PVFs and the corresponding stick orientations from ARD to this same common space. In the common space fiber populations with PVFs smaller than 0.05 were discarded [11], such that voxels could contain zero, one or two fiber populations.

To compute the fiber orientation atlas, in each voxel the remaining primary and secondary fiber orientations (i.e. with PVFs above 0.05) were clustered into two groups from which the average orientations were computed. The clustering was achieved using an adapted K-means clustering routine (K=2), that effectively minimizes this sum of squared angular errors (reckoning with 180^o symmetry):

$$J = \sum_{k=1}^{2}\sum_{i\in C_k} \arccos\left(\text{abs}\left(\mathbf{V}_i\cdot\mathbf{V}_{k,\text{mean}}\right)\right)^2, \quad (2)$$

where C_k is the k-th cluster, $\mathbf{V}_i$ the i-th fiber orientation in C_k, and $\mathbf{V}_{k,\text{mean}}$ the average fiber orientation of all fiber orientations in C_k. Specifically, the K-means clustering routine was initialized by assigning all largest stick compartments to the first cluster and all smallest stick compartments to the second cluster. Next, in the update step, the mean orientation of each cluster was obtained by first computing the scatter matrices $\mathbf{S}_k$ (the sum of dyadic products of the fiber orientations) [25]:

$$\mathbf{S}_k = \sum_{i\in C_k} \mathbf{V}_i\cdot\mathbf{V}_i^T. \quad (3)$$

Subsequently, the average fiber orientation $\mathbf{V}_{k,\text{mean}}$ was computed as the principal eigenvector of the matrix $\mathbf{S}_k$, i.e.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2019.2922615, IEEE Transactions on Medical Imaging

the eigenvector corresponding to the largest eigenvalue. In the assignment step each fiber orientation was assigned to the 'nearest' cluster mean in terms of angle (reckoning with 180^o symmetry). The update and assignment step were iterated until no fibers changed cluster. The eigenvector of the cluster initialized with the largest stick compartments was regarded as the primary fiber orientation, the eigenvector of the other cluster was regarded as the second fiber orientation.

H. Construction of the complexity atlas

In the common space the FA images, derived from the spatially-normalized diffusion tensor images, were averaged after which a skeleton was generated with the function `tbss_skeleton` (in FSL). The mean FA skeleton was thresholded at 0.2 merely to exclude voxels with large inter-subject variability and/or partial volume effects with grey matter or cerebrospinal fluid. Subsequently, the warped PVF volumes (see the previous section) were 'skeletonized' by only retaining the voxels overlapping with the mean FA skeleton mask. Note that for establishing correspondence, DTI-TK performs a high dimensional registration instead of a low dimensional warp combined with 'maximum FA skeleton projection' step of a standard TBSS analysis [15]. The former (our) approach is also taken in previous work showing that the reproducibility of FA-statistics on the TBSS skeleton increases when a single high-dimensional registration without maximum FA projection is used [26]. Eventually, each skeleton voxel may contain zero, one or two fiber populations. The model complexity atlas was defined as the average number of fiber populations in each FA skeleton voxel.

I. Model Estimation

The ball-and-*one*-stick model in the proposed framework was fit by maximum likelihood estimation as in [23], not using any prior knowledge on the stick's orientation. The ball-and-*two*-sticks model was fit using the fiber orientation atlas in two ways: (1) the atlas orientation was used to initialize the non-linear fit, and (2) a Gaussian-shaped prior $p(\boldsymbol{\theta})$ was applied to regularize the fitting of the model:

$$p(\boldsymbol{\theta}) \propto \exp\left(-\frac{\varepsilon_1^{\,2}}{2\sigma_\theta^{\,2}}\right)\exp\left(-\frac{\varepsilon_2^{\,2}}{2\sigma_\theta^{\,2}}\right) \quad \text{with} \quad \varepsilon_i = \arccos\left(\mathbf{V}_i \cdot \mathbf{V}_{i,mean}\right), \tag{4}$$

where ε_1 and ε_2 represent the angles between the principal eigenvector of both sticks and their corresponding atlas orientations, and σ_θ denotes the width of the Gaussian-shaped orientation prior. The estimate of $\boldsymbol{\theta}$ is given by:

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \log p(\mathbf{S}\,|\,\boldsymbol{\theta})\, p(\boldsymbol{\theta}), \tag{5}$$

which was implemented using the fit_MRI toolbox [23].

J. Proposed TBSS analysis

The PVFs of the ball-and-one-stick as well as the ball-and-two-sticks model parameters were warped to the common space using the computed DTI-TK transformations (see above). Subsequently, the two parameter sets were 'skeletonized' by retaining the voxels coinciding with the FA skeleton mask. Trilinear interpolation was used in warping the images to the common space. This was possible because the PVFs in the ball-and-two-sticks model are implicitly sorted in subject space by means of the orientation prior. Finally, the skeletonized datasets from the two models were merged according to the model complexity atlas: the two-fiber voxels (according to the model complexity atlas) of the skeleton received the primary and secondary PVFs from a two-stick diffusion model; the single fiber skeleton voxels obtained the primary PVF from the one-stick diffusion model and the secondary PVF was set to zero.

To facilitate comparison with the second reference pipeline (see below), we also evaluated the effect of using nearest-neighbour interpolation (instead of trilinear interpolation) to transform the PVFs to the common space. For this purpose two adapted versions of the proposed framework were considered, either simply using nearest-neighbour interpolation to warp the PVFs or using 3D Gaussian smoothing prior to using nearest-neighbour interpolation to warp the PVFs. The standard deviation of the 3D Gaussian smoothing kernel was set to 0.487 voxels, such that it approximately matched the smoothing effect of trilinear interpolation.

K. Reference frameworks

The first reference framework essentially performed a modified TBSS analysis of conventional diffusion tensor images. In the common space FA images, computed from the spatially-normalized diffusion tensor images, were 'skeletonized' by only retaining the voxels coinciding with the mean FA skeleton mask. As such, we (again) did not use the 'maximum skeleton projection' step of the standard TBSS analysis [15]. This is relevant because the projection may favor single fiber voxels since these tend to have a higher FA than crossing-fiber voxels. Subsequently, FA statistics were evaluated for each skeleton voxel as in the conventional TBSS analysis.

The second reference framework relied on the primary and secondary PVFs from the ARD modeling. The FSL routine `tbss_x` [16] was used to warp these PVFs and their corresponding stick orientations to the common space, 'skeletonize' the transformed volumes, and subsequently sort the primary and secondary PVFs based on their orientations. To do so, the routine `tbss_x` was adjusted to make it compatible with the DTI-TK transformations. Instead of the FSL routines `applywarp` and `vecreg`, DTI-TK functions `deformationScalar3DVolume` and `deformationSymTensor3DVolume` were applied, respectively. In subject space, the PVFs are still unsorted and may correspond to different fiber populations in adjacent voxels. To prevent interpolation of the PVFs of different fiber populations, similar to the original `tbss_x` routine, nearest neighbour interpolation was used to warp the PVFs to the common space. Furthermore, unless specified otherwise, the warped image volumes were 'skeletonized' by only retaining the voxels overlapping with the mean FA skeleton mask. Finally the skeletonized primary and secondary PVFs were analyzed. To further evaluate the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2019.2922615, IEEE Transactions on Medical Imaging

reproducibility we also applied a second reference framework in which the warped image volumes were skeletonized by projecting the locally maximum FA voxels onto the skeleton, i.e. the traditional skeletonization approach in TBSS [15].

L. Statistical analysis of the reproducibility

To assess the reproducibility of the proposed framework and the reference frameworks, the intraclass correlation coefficient (ICC) was computed from 30 subjects who were scanned twice. Both the proposed framework and the reference frameworks were used to estimate, warp, and skeletonize the model parameters from both scans in the common space. In the proposed framework both nearest-neighbour and trilinear interpolation were used for transforming the PVFs, and the width (spread) of the orientation prior σ_θ was varied between 1 and 90 degrees. To compute the ICC for both the primary and secondary PVFs in each skeleton voxel, a one-way random effects model was used [27].

Furthermore, to gain more insight into the conditions under which the use of the orientation atlas is appropriate we deliberately introduced biases into the orientation atlas. More specifically, all fiber orientations were rotated 15 degrees and 30 degrees along the Inferior-Superior-axis, resulting in two biased orientation atlases. The effect of using these biased orientation atlases on the ICC and the mean value of the primary and secondary PVFs was evaluated. Similarly as described above, in this evaluation the width (spread) of the orientation prior σ_θ was varied between 1 and 90 degrees.

M. Statistical analysis of the accuracy

The accuracy of the proposed framework to estimate stick fractions is evaluated on data from the ISMRM tractography challenge 2015 [28]. This dataset includes realistically simulated replication of whole brain DWIs. First, we estimated ground truth model parameters by fitting ball-and-one-stick and ball-and-two-stick models to the simulated noiseless DWIs. These models were combined such that the ball-and-two-sticks model was only used in voxels where the ball-and-one-stick model had a poor fit to the data (i.e. RMSE>$0.005S_0$). Next, Rician noise ($\sigma_{Rician} = S_0/30$) was added to the DWIs and the noisy data was estimated with the proposed framework and the reference ARD framework. Note that the proposed framework uses the co-registered orientation and complexity atlas computed from the 500 elderly subjects, rather than ground truth values. The accuracy of the stick fractions obtained with both frameworks was evaluated by comparing the sum of both stick fractions to the ground truth values in crossing-fiber regions (determined from the ground truth).

N. Statistical analysis of ageing

The proposed and reference frameworks were applied to a group of 500 subjects from the population-based Rotterdam Study [17] to study changes in diffusion measures with aging in crossing-fiber regions. The relevant diffusion statistics were analyzed using a conventional linear model analysis using age as the only covariate. Voxelwise statistics in TBSS were carried out using a permutation-based inference tool for nonparametric statistical thresholding (`randomise`, part of FSL). The number of permutations was set to 5000. The significance threshold was set at $p<0.05$ (employing Familywise Error Rate (FWE) correction for multiple comparisons) using the threshold-free cluster enhancement (TFCE) option in the randomise permutation-testing tool in FSL. The same statistical analysis of ageing was also applied to a random subset of 100 subjects to assess the performance when fewer data is available.

III. RESULTS

A. Fiber orientation atlas and model complexity atlas

The fiber orientation atlas, constructed from 500 subjects, is visualized in Fig. 2. The first row shows the orientations of the average primary fiber population. The second row in Fig. 2 visualizes the orientation of the average secondary fiber population when present in more than 50% of the population. It can be observed that the estimated average primary and secondary fiber orientations are approximately left-right symmetric, relatively smooth, and appear anatomically plausible.

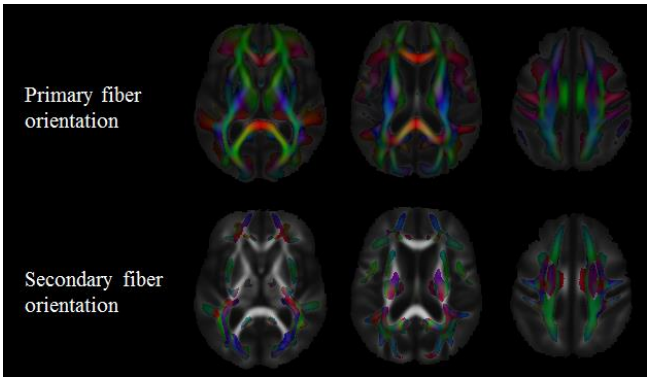

Fig. 2. Mean fiber orientations of the primary fiber tract (top row) and secondary fiber tract, if present (bottom row). The fiber orientations modulate the background FA image, such that red corresponds with left-right, green with posterior-anterior and blue with inferior-superior.

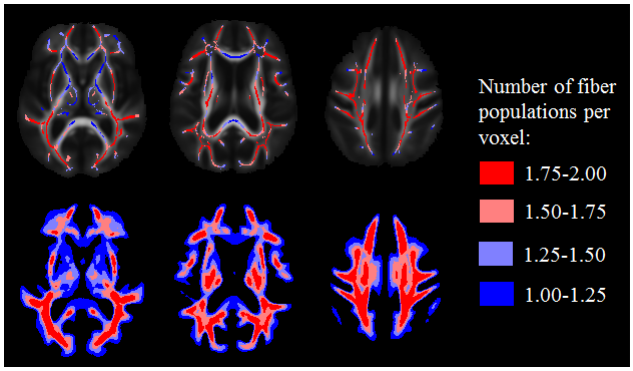

Fig. 3. The average number of fiber populations per voxel in the common space. In the top row skeletonized data is visualized with the group-mean FA as background image. The bottom row shows the average number of fiber populations per voxel in not-skeletonized data. Voxels with (on average) fewer than 1 fiber tract are masked black.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2019.2922615, IEEE Transactions on Medical Imaging

The model complexity atlas, constructed from 500 subjects, is visualized in Fig. 3. For visualization purposes, the average number of fiber populations is binned into four categories, such that regions in red, light-red, light-blue and blue reflect regions with a decreasing prevalence of crossing-fibers respectively. To obtain a discrete model complexity atlas, the average number of fiber populations was thresholded at 1.5, i.e. 2 fiber orientations in the atlas was decided by majority vote. The blue and light-blue areas in Fig. 3 are thus regarded as single-fiber areas whereas the red and light-red regions in Fig. 3 are regarded as crossing-fiber areas. Indeed, white matter structures known to contain a single fiber population, e.g. the corpus callosum, correspond with single-fiber areas in the complexity atlas. Furthermore, the crossing-fiber voxels form clusters that are approximately left-right symmetrical. In our model complexity atlas, approximately 67 percent of the white matter skeleton consists of crossing-fiber voxels. Note that the framework is not very sensitive to the particular choice of 1.5. A different choice for this parameter, i.e. 1.25 or 1.75, would include the light-blue parts, respectively discard the light-red parts of the brain in Fig. 3 to/from the crossing-fiber voxels. One may observe that these are relatively small regions.

B. Reproducibility study

The estimated primary and secondary fiber orientations of one subject from the reproducibility dataset are visualized in Fig. 4. The red and blue cylinders represent the stick orientations estimated from respectively the first and second scan in two regions-of-interest (ROIs). Only stick compartments with PVFs higher than the threshold of 0.05 are shown. For visualization purposes the proposed method was applied to the entire dataset (i.e. not only to the voxels on the FA skeleton). With the second reference (ARD) framework (middle row of Fig. 4), the estimated stick orientations show large variation (see e.g. the green circle in ROI1), or the number of stick compartments can be different (green circle in ROI2). This would decrease the effectiveness of orientation-based labeling of fiber populations used in the second reference framework. With the proposed framework (bottom row of Fig. 4), the estimated stick orientations show less variation between the first and second scan and the number of stick compartments corresponds exactly (by definition).

In Fig. 5 the ICC of the primary and secondary PVFs, obtained with the proposed and second reference framework, are shown. For the primary PVFs, the ICC was averaged over all skeleton voxels; for the secondary PVFs, the ICC was averaged over all crossing-fiber skeleton voxels (according to the complexity atlas). With respect to the second reference framework, the ICC is also shown when the PVFs from the locally maximum FA voxels are projected onto the skeleton. It can be observed that this projection step decreases the ICC for both the primary and secondary PVFs. For the proposed framework, the ICC is shown for different types of interpolation and as a function of the width of the orientation prior σ_θ. Compared to nearest-neighbour interpolation (blue line), trilinear interpolation (red line) greatly improves the ICC of both the primary and secondary PVF. Trilinear interpolation increases the ICC not only by improved spatial alignment, but also the smoothing effect (inherent to trilinear interpolation) is partly responsible for the increased ICC (green line).The results in the remainder of this paper for the proposed framework were generated using trilinear interpolation; in the second reference framework the projection step will not be applied any further.

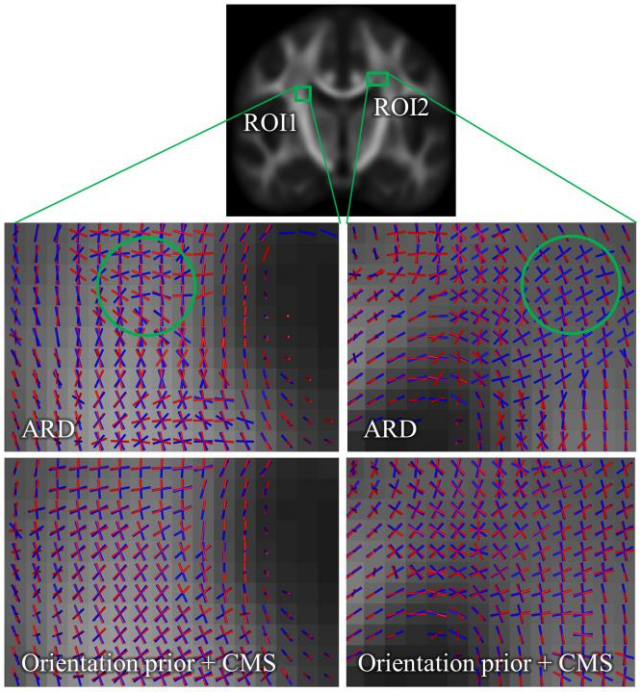

Fig. 4. Estimated stick orientations in two regions-of-interest (ROIs) of a subject from the reproducibility dataset warped to the common space. The red and blue cylinders represent the stick orientations estimated from respectively the first scan and second scan. Both automatic relevance detection (ARD) and the proposed framework with an orientation prior (σ_θ=25 degrees) and consistent model selection (CMS) were used to estimate the ball-and-sticks model parameters. The single tensor FA is visualized on the background for anatomical reference. When using ARD, the estimated stick orientations do not always correspond (green circle in ROI1). Furthermore, the number of stick compartments may deviate (green circle in ROI2). In the bottom row, the application of prior information about the orientation and number of stick compartments increases the similarity of estimated orientations across different datasets.

The ICC of the primary and secondary PVFs obtained with the proposed framework using biased and unbiased orientation atlases, are shown in the top row of Fig. 6. When the atlas orientations are biased by 15 degrees, the ICC is slightly decreased compared to using an unbiased atlas. When the atlas orientations are biased by 30 degrees, the ICC is further decreased. Note that even with a biased orientation atlas, the ICC of the proposed framework is higher than the ICC obtained with the reference frameworks shown in Fig. 5. Fig. 6 also shows that orientation priors with a width smaller than approximately 25 degrees cause shrinkage of the estimated secondary PVFs (see the bottom right graph). For the remainder of the paper, we will use a prior width σ_θ of 25 degrees. Using this value balances unwanted shrinkage of the PVFs and unwanted decrease of ICC.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2019.2922615, IEEE Transactions on Medical Imaging

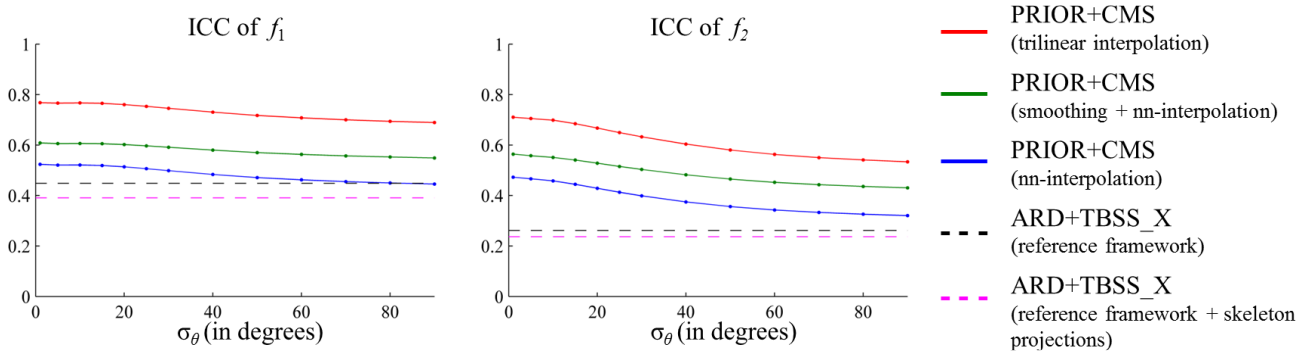

Fig. 5. Left: average intraclass correlation coefficient (ICC) of f_1 on the TBSS skeleton. Right: average ICC of f_2 in crossings-fiber regions on the TBSS skeleton. The solid colored lines denote the ICC of the proposed framework where f_i on the skeleton was obtained using trilinear interpolation (red), smoothing and nearest-neighbour interpolation (green), and nearest-neighbour interpolation (blue). The dashed lines correspond to the ICC of the second reference framework, black for the TBSS approach without projection and pink including the projection of PFVs from the locally maximum FA voxels onto the skeleton. The reference frameworks do not use the prior and hence are independent of σ_θ.

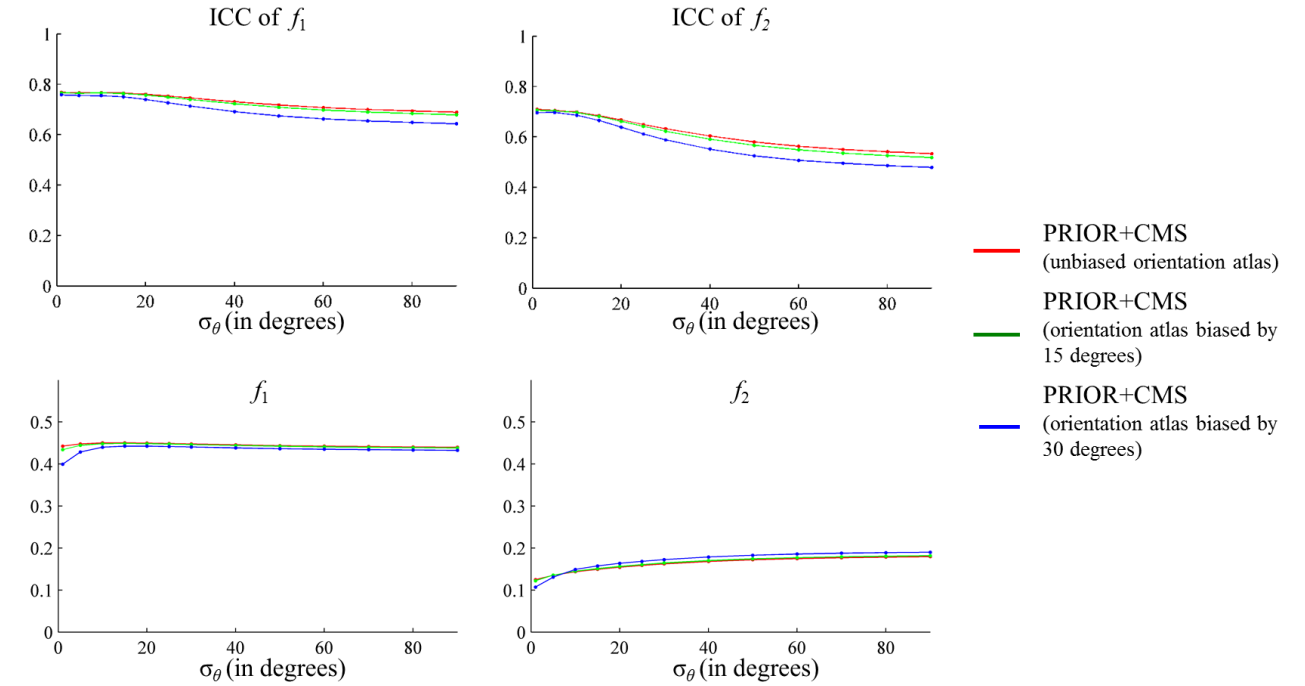

Fig. 6. Top row: The ICC of f_1 and f_2 for unbiased and (deliberately) biased atlases and different widths of the orientation prior. Bottom row: The value of f_1 and f_2 for unbiased and (deliberately) biased atlases and for different widths of the orientation prior.

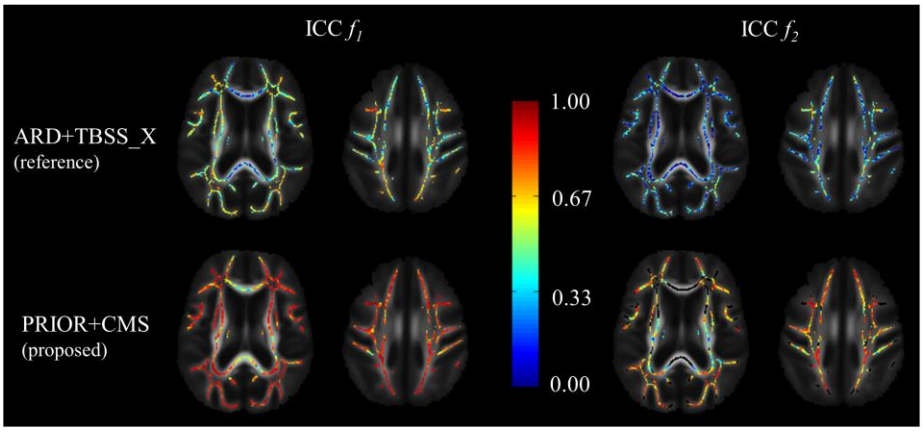

Fig. 7. Intraclass correlation coefficient (ICC) of the primary partial volume fraction f_1 (left) and the secondary partial volume fraction f_2 (right) computed from 30 subjects for whom rescan data was available. Top row: ICC of the reference framework applying ARD and orientation-based labeling of the PVFs. Bottom row: the proposed framework applying an atlas orientation prior and CMS. Note that skeleton voxels with a single fiber population have been masked black for the secondary PVFs obtained with the proposed framework.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2019.2922615, IEEE Transactions on Medical Imaging

In Fig. 7 the ICC of the primary and secondary PVFs, obtained with the second reference framework and the proposed framework, are shown on the skeleton. The ICC of both the primary and secondary PVFs is significantly higher across the whole skeleton with the proposed approach.

C. Accuracy study on simulated data

The results of the accuracy on simulated data are reported in Table 1. The positive values in the columns μ_{ARD} and $\mu_{\text{Prop.}}$ show that both frameworks on average underestimate the sum of the stick fractions. Note that the bias in the ARD framework is larger than in the proposed framework. The columns σ_{ARD} and $\sigma_{\text{Prop.}}$ reflect the standard deviation in a single voxel.

D. Ageing correlations in population data

The results of the statistical analysis of ageing are plotted in Fig. 8, Fig. 9 and

Fig. 10. In Fig. 8 the regression coefficients with age and the corresponding t-statistics of both the proposed framework and the second reference framework are shown. It can be observed that the effect size is similar, but the regression coefficients obtained with the proposed framework appear more spatially-smooth (see red arrows). Furthermore, the t-statistics for the regression coefficient with age are larger when the proposed framework is used (see red arrows), indicating that the regression coefficient can be determined more precisely.

More results of the statistical analysis of ageing are plotted in Fig. 9. The first two rows of images show skeleton voxels with significant *negative* correlations of the primary and secondary PVFs with age when only 100 subjects are included in the analysis. The bottom two rows show the results when all 500 subjects are included in the analysis. For the analysis on 500 subjects, the PVFs obtained with the reference and proposed frameworks showed very similar patterns with ageing, i.e. uniformly distributed over the brain large clusters of voxels were found in which the PVFs decreased significantly with age. Differences in the sensitivity were most apparent in the analysis on 100 subjects, e.g. see the blue arrows in Fig. 9. In the subset of 100 subjects, the percentages of skeleton voxels that showed significant decrease in PVFs with age were 30% (f_1 of reference framework), 36% (f_1 of proposed framework), 6% (f_2 of reference framework) and 12% (f_2 of proposed framework). In the full study population of 500 subjects, the proposed framework also resulted in higher percentages of skeleton voxels with significant decrease in PVFs with age than the reference framework: 76% (f_1 of reference framework), 84% (f_1 of proposed framework), 50% (f_2 of reference framework) and 51% (f_2 of proposed framework).

In

Fig. 10 the correlation with age is shown for the FA generated with the first reference framework and PVFs obtained with the proposed framework in a coronal slice. FA showed a very similar pattern of correlation with ageing as the primary PVFs, i.e. large clusters of voxels across the white matter skeleton were found in which the FA decreased significantly with age. In some of the voxels containing the corticospinal tract, though, both the FA and primary PVF did not have a significant correlation with age. However, in those voxels the secondary PVF often did significantly decrease with age.

Table 1. Results of accuracy evaluation on simulated data. The columns μ_{ARD} and $\mu_{\text{Prop.}}$ report the mean difference between the ground truth and estimated sum of the primary and secondary PVFs. The columns σ_{ARD} and $\sigma_{\text{Prop.}}$ reflect the standard deviation in a single voxel. Voxels with ground truth $\sum_j f_j < 0.2$ are excluded as these mostly contain isotropic tissue or CSF. Voxels with ground truth $\sum_j f_j > 0.7$ are too few to include.

		$\sum_j f_{j,\text{ground truth}} - \sum_j f_{j,\text{estimated}}$			
Ground truth	N_{voxels}	μ_{ARD}	σ_{ARD}	$\mu_{\text{Prop.}}$	$\sigma_{\text{Prop.}}$
$0.2 \leq \sum_j f_j < 0.3$	1453	0.047	0.047	0.016	0.050
$0.3 \leq \sum_j f_j < 0.4$	1866	0.043	0.056	0.015	0.060
$0.4 \leq \sum_j f_j < 0.5$	1676	0.034	0.055	0.011	0.060
$0.5 \leq \sum_j f_j < 0.6$	746	0.030	0.054	0.006	0.052
$0.6 \leq \sum_j f_j < 0.7$	128	0.021	0.048	0.010	0.054

IV. Discussion

In this paper we presented a framework to improve the analysis of MR diffusion data with a ball-and-sticks model. Two novelties were introduced. The first novelty was the application of an atlas orientation prior to guarantee correct labeling in crossing-fiber regions. The second novelty was a 'consistent model selection', obtained by determination of the number of fiber populations in a common space instead of for each subject independently. We demonstrated that the reproducibility of measuring PVFs was improved both in single fiber regions and crossing-fiber regions. Furthermore, a proof-of-principle analysis of the effect of age on white matter diffusion properties showed enhanced statistical power to detect age-related changes in white matter.

A. Fiber orientation atlas and model complexity atlas

The fiber orientation atlas and the model complexity atlas were constructed using 500 subjects from the Rotterdam Scan study. Alternatively, these atlases could have been constructed from data acquired with a more advanced DW-MRI protocol, but we deliberately did not do this. The ability to accurately model properties of a secondary fiber population does not only depend on the anatomical presence of a fiber crossing, but also on DW-MRI acquisition parameters such as the voxel size, the signal-to-noise ratio, the number of diffusion-encoding gradient directions and the used *b*-values [11]. The constructed atlases should therefore not only reflect the brain anatomy in a subject group, but also the DW-MRI protocol used in the acquisition. As such, the model complexity and fiber orientation atlas are ideally constructed from a representative population using a similar DW-MRI protocol as the study population.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2019.2922615, IEEE Transactions on Medical Imaging

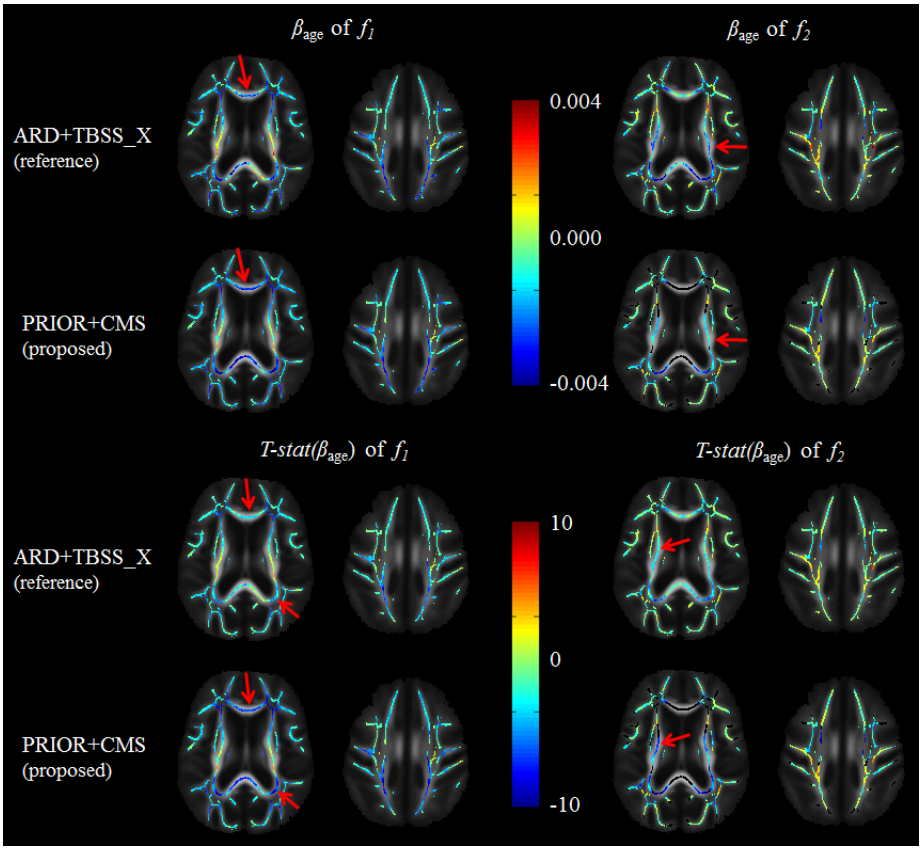

Fig. 8. The regression coefficient and its t-statistic of the primary and secondary PVFs with age. PVFs were generated using the proposed and second reference framework. Note that skeleton voxels with a single fiber population have been masked black for the secondary PVFs obtained with the proposed framework. In the top two rows, the red arrows point at examples of regions where the proposed framework appears to provide more spatially-smooth estimates of the regression coefficient. In the bottom two rows, the red arrows point at examples of regions where the proposed framework where the t-statistic is higher.

The construction of the fiber orientation atlas and model complexity atlas were both based on the ARD framework. This facilitated a comparison with the second reference framework [16] that also relied on the ARD framework for the estimation of fiber orientations and(implicit) model selection on a per-subject basis.

The discrete model complexity atlas was obtained by thresholding the average number of fiber population per voxel. The threshold value should balance between over-fitting (i.e. using too many stick compartments that are not supported by the data) and under-fitting (i.e. using too few stick compartments that cannot describe the data adequately). In this paper a threshold value of 1.5 was used to choose between fitting a ball-and-*one* stick model and ball-and-*two*-sticks model.

B. Reproducibility study

The first novelty, the application of an orientation prior, promoted a more consistent labeling of the fiber populations. In the proposed framework this labeling effectively takes place during estimation in subject space, such that the PVFs can be interpolated when transformed to the common space. Both the application of an orientation prior and the use of (trilinear) interpolation were shown to result in an improved ICC.

The second novelty, the determination of the number of fiber populations in a common space, was also shown to improve the ICC, particularly for measuring the secondary PVFs. The poorer performance of the second reference, ARD framework might be due to noise. Since the noise realization in the first and second scan is different, this could result in a different (effective) number of anisotropic compartments in the diffusion model, especially in configurations with unbalanced volume fractions or a small angular divergence between the two fiber populations. Analyzing the 'same' voxel with a different number of anisotropic compartments in the diffusion model would increase the apparent variation in the estimated parameters. Effectively, it behaves as noise that is added onto the PVFs, which is avoided in the proposed framework by imposing a consistent model selection.

C. Accuracy study

The accuracy evaluation on phantom data showed that estimated stick fraction may be slightly underestimated, but that this bias is (much) smaller than the bias introduced by using an ARD shrinkage prior. The results on accuracy are in agreement with the reproducibility study, where orientation priors with a small width resulted shrinkage of the stick fractions. It should be noted that the orientation atlas and complexity atlas in this experiment were computed from 500 elderly subjects, whereas the phantom was based on fiber tracts from a relatively young individual. The use of more representative atlases may further reduce the bias.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2019.2922615, IEEE Transactions on Medical Imaging

D. Sensitivity

A skeleton-based linear regression analysis between age and the relevant diffusion parameters was performed, primarily to gain insight into the sensitivity (reflected by higher t-statistics) of the framework. The proposed framework using both an orientation prior and consistent model selection appeared to be more sensitive compared to the second reference framework using ARD and orientation-based labeling of fiber populations in the common space. In the full sample of 500 subjects as well as in a smaller subset of 100 subjects, the number of voxels that correlated significantly with age was higher when using the proposed framework. An increased sensitivity will facilitate the detection of smaller effects, or enable the use of smaller populations while retaining similar levels of statistical significance.

E. Ageing

Besides gaining insight into the sensitivity of the different frameworks, the ageing study demonstrated the added value of analyzing the PVFs of ball-and-sticks models over single tensor FA. In large clusters throughout the brain it was found that FA as well as the primary PVFs in the ball-and-sticks model decreased linearly with age. This behavior of FA and PVFs is in agreement with prior work, where it was reported that FA and primary PVFs are highly correlated [16]. Even though the pattern with ageing appeared similar, it should be noted that analyses with FA and with primary PVFs are conceptually different. The FA is a scalar measurement and is not specific to a particular fiber population in a voxel, whereas the primary PVF is an orientation-dependent measurement of the primary fiber population in a voxel. This enables a population-specific analysis of the brain, or the testing of tract-specific hypotheses.

The most prominent difference between analyzing FA and PVFs was demonstrated in crossing fiber areas. We found that the FA in parts of the corticospinal tract did not correlate significantly with age. Previous studies also reported insignificant correlations between FA and age in the corticospinal tract [29], or even reported that the FA in parts of the corticospinal tract paradoxically increased with age in elderly subjects [30]. Our model complexity atlas and fiber orientation atlas demonstrated that the corticospinal tract crosses (or is in close proximity to) many commissural and association tracts. Analysis with ball-and-sticks models revealed that the primary PVFs, with an orientation aligned to the corticospinal tract, did not correlate significantly with age. However, the secondary PVF, aligned with various commissural and association tracts, was found to decrease significantly with age. Such a selective degeneration of the secondary fiber tract may indeed result in an estimated increase in FA, and may explain the increase in FA with age observed in the corticospinal tract [30]. This underlines the importance of using improved methods to analyze crossing fiber areas.

F. TBSS and other WM quantification techniques

In this paper we extended a fiber-orientation specific TBSS analysis pipeline. TBSS is considered a standard approach for voxel-based analysis of DW-MRI data, but has several conceptual limitations [31]. An alternative approach, known as tractometry, analyzes diffusion parameters along specific tracts reconstructed by tractography [32, 33]. An advantage of tractometry is that tracts are reconstructed and compared in subject space. Therefore, a registration of all subjects to a common space as in TBSS is not needed as illustrated with AutoPtx [26] or TRACULA [34]. However, tractography is still considered to be unreliable in tracing all physical white matter structures [28]. Therefore, tract-based frameworks typically focus on only a limited set of relatively large, well-defined tracts. Furthermore, tract-based metrics may lack the precise localization of effects along the tract as is facilitated by the proposed framework. Alternatively, tracts can also be reconstructed in a common space, e.g. based on q-ball imaging [35] or multi-tensor models [36], but these approaches require registration (just as TBSS approaches).

G. Limitations

Inherent limitations of the proposed framework are in the use of the fiber orientation and model complexity atlas. The use of an inappropriate orientation prior may bias the estimated diffusion statistics. In subjects with pathologies that drastically impact the white matter orientations (e.g. large brain tumors), the use of an orientation atlas may therefore not be adequate. At the same time, the selected width σ_θ of 25 degrees yields a rather flat orientation prior, as such having a limited effect. Furthermore, a registration algorithm like DTI-TK [24] that takes (tensor) orientation into account, may already be robust to much of the variance in white matter orientations across groups/subjects. Clearly, however, the validity of our atlases in populations deviating from ours, e.g. due to pathology, will require further research.

Similarly, the use of the complexity atlas has limitations. When the complexity atlas indicates an incorrect number to be present, this may cause under-fitting or over-fitting of the data in individual subjects. It should be noted that over-fitting of the data is reduced by the applied orientation prior, i.e. a stick compartment not supported by the data is a priori expected to point in the direction of the orientation prior; its estimated stick fraction will generally be small due to lacking support from the data.

H. Conclusion

We have developed a framework, that utilizes an atlas orientation prior and a consistent model selection, to improve the analysis of diffusion data with a ball-and-sticks model. The application of these novelties was shown to significantly improve the reproducibility and sensitivity of the ball-and-sticks model parameters in TBSS. Particularly in group studies, the proposed framework may therefore detect more subtle differences and quantify changes more precisely.

V. References

[1] E. Stejskal and J. Tanner, "Spin diffusion measurements: spin echoes in the presence of a time-dependent field gradient," *The journal of chemical physics*, vol. 42, pp. 288-292, 1965.

[2] C. Beaulieu, "The basis of anisotropic water diffusion in the nervous system–a technical review," *NMR in Biomedicine*, vol. 15, pp. 435-455, 2002.

[3] P. J. Basser, J. Mattiello, and D. LeBihan, "MR diffusion tensor spectroscopy and imaging," *Biophysical journal*, vol. 66, pp. 259-267, 1994.

[4] Y. Liu, et al., "Diffusion tensor imaging and Tract-Based Spatial Statistics in Alzheimer's disease and mild cognitive impairment," *Neurobiology of Aging*, vol. 32, pp. 1558-1571, 9// 2011.

[5] M. W. Vernooij, M. de Groot, A. van der Lugt, M. A. Ikram, G. P. Krestin, A. Hofman, W. J. Niessen, and M. M. Breteler, "White matter atrophy and lesion formation explain the loss of structural integrity of white matter in aging," *Neuroimage*, vol. 43, pp. 470-477, 2008.

[6] A. L. Alexander, K. M. Hasan, M. Lazar, J. S. Tsuruda, and D. L. Parker, "Analysis of partial volume effects in diffusion-tensor MRI," *Magnetic Resonance in Medicine*, vol. 45, pp. 770-780, 2001.

[7] D. S. Tuch, T. G. Reese, M. R. Wiegell, N. Makris, J. W. Belliveau, and V. J. Wedeen, "High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity," *Magnetic Resonance in Medicine*, vol. 48, pp. 577-582, 2002.

[8] C. A. Wheeler-Kingshott and M. Cercignani, "About “axial” and “radial” diffusivities," *Magnetic Resonance in Medicine*, vol. 61, pp. 1255-1260, 2009.

[9] C. Pierpaoli, A. Barnett, S. Pajevic, R. Chen, L. Penix, A. Virta, and P. Basser, "Water diffusion changes in Wallerian degeneration and their dependence on white matter architecture," *Neuroimage*, vol. 13, pp. 1174-1185, 2001.

[10] G. Douaud, et al., "DTI measures in crossing-fibre areas: Increased diffusion anisotropy reveals early white matter alteration in MCI and mild Alzheimer's disease," *NeuroImage*, vol. 55, pp. 880-890, 4/1/ 2011.

[11] T. Behrens, H. J. Berg, S. Jbabdi, M. Rushworth, and M. Woolrich, "Probabilistic diffusion tractography with multiple fibre orientations: What can we gain?," *Neuroimage*, vol. 34, pp. 144-155, 2007.

[12] M. W. Caan, H. Khedoe, D. H. Poot, A. J. den Dekker, S. D. Olabarriaga, K. A. Grimbergen, L. J. van Vliet, and F. M. Vos, "Estimation of diffusion properties in crossing fiber bundles," *Medical Imaging, IEEE Transactions on*, vol. 29, pp. 1504-1515, 2010.

[13] Y. Assaf and P. J. Basser, "Composite hindered and restricted model of diffusion (CHARMED) MR imaging of the human brain," *NeuroImage*, vol. 27, pp. 48-58, 8/1/ 2005.

[14] J. Yang, D. H. Poot, L. J. van Vliet, and F. M. Vos, "Estimation of diffusion properties in three-way fiber crossings without overfitting," *Physics in medicine and biology*, vol. 60, p. 9123, 2015.

[15] S. M. Smith, et al., "Tract-based spatial statistics: Voxelwise analysis of multi-subject diffusion data," *NeuroImage*, vol. 31, pp. 1487-1505, 7/15/ 2006.

[16] S. Jbabdi, T. E. Behrens, and S. M. Smith, "Crossing fibres in tract-based spatial statistics," *NeuroImage*, vol. 49, pp. 249-256, 2010.

[17] A. Hofman, et al., "The Rotterdam Study: 2016 objectives and design update," *European journal of epidemiology*, vol. 30, pp. 661-708, 2015.

[18] M. A. Ikram, A. van der Lugt, W. J. Niessen, G. P. Krestin, P. J. Koudstaal, A. Hofman, M. M. Breteler, and M. W. Vernooij, "The Rotterdam Scan Study: design and update up to 2012," *European journal of epidemiology*, vol. 26, pp. 811-24, Oct 2011.

[19] S. Klein, M. Staring, K. Murphy, M. Viergever, and J. P. Pluim, "Elastix: a toolbox for intensity-based medical image registration," *Medical Imaging, IEEE Transactions on*, vol. 29, pp. 196-205, 2010.

[20] J. L. Andersson, S. Skare, and J. Ashburner, "How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging," *NeuroImage*, vol. 20, pp. 870-888, 2003.

[21] S. M. Smith, et al., "Advances in functional and structural MR image analysis and implementation as FSL," *NeuroImage*, vol. 23, pp. S208-S219, 2004.

[22] A. Leemans and D. K. Jones, "The B-matrix must be rotated when correcting for subject motion in DTI data," *Magnetic Resonance in Medicine*, vol. 61, pp. 1336-1349, 2009.

[23] D. H. Poot and S. Klein, "Detecting statistically significant differences in quantitative MRI experiments, applied to diffusion tensor imaging," *Medical Imaging, IEEE Transactions on*, vol. 34, pp. 1164-1176, 2015.

[24] H. Zhang, P. A. Yushkevich, D. C. Alexander, and J. C. Gee, "Deformable registration of diffusion tensor MR images with explicit orientation optimization," *Medical image analysis*, vol. 10, pp. 764-785, 2006.

[25] A. Schwartzman, R. F. Dougherty, and J. E. Taylor, "Cross-subject comparison of principal diffusion direction maps," *Magnetic Resonance in Medicine*, vol. 53, pp. 1423-1431, 2005.

[26] M. de Groot, M. W. Vernooij, S. Klein, M. A. Ikram, F. M. Vos, S. M. Smith, W. J. Niessen, and J. L. Andersson, "Improving alignment in Tract-based spatial statistics: Evaluation and optimization of image registration," *NeuroImage*, vol. 76, pp. 400-411, 2013/08/01/ 2013.

[27] K. O. McGraw and S. P. Wong, "Forming inferences about some intraclass correlation coefficients," *Psychological methods*, vol. 1, p. 30, 1996.

[28] K. H. Maier-Hein, et al., "The challenge of mapping the human connectome based on diffusion tractography," *Nature communications*, vol. 8, 1349, 2017

[29] M. E. Perry, C. R. McDonald, D. J. Hagler Jr, L. Gharapetian, J. M. Kuperman, A. K. Koyama, A. M. Dale, and L. K. McEvoy, "White matter tracts associated with set-shifting in healthy aging," *Neuropsychologia*, vol. 47, pp. 2835-2842, 11// 2009.

[30] M. de Groot, L. G. Cremers, M. A. Ikram, A. Hofman, G. P. Krestin, A. van der Lugt, W. J. Niessen, and M. W. Vernooij, "White Matter Degeneration with Aging: Longitudinal Diffusion MR Imaging Analysis," *Radiology*, vol. 279, pp. 532-541, 2015.

[31] M. Bach, F. B. Laun, A. Leemans, C. M. W. Tax, G. J. Biessels, B. Stieltjes, and K. H. Maier-Hein, "Methodological considerations on tract-based spatial statistics (TBSS)," *NeuroImage*, vol. 100, pp. 358-369, 2014/10/15/ 2014.

[32] S. Bells, M. Cercignani, S. Deoni, Y. Assaf, O. Pasternak, C. Evans, A. Leemans, and D. Jones, "Tractometry–comprehensive multi-modal quantitative assessment of white matter along specific tracts," *in Proc. ISMRM*, 2011.

[33] C. Corbin, V. Gupta, J. E. Villalon-Reina, T. M. Nir, F. M. Rashid, S. I. Thomopoulos, N. Jahanshad, and P. M. Thompson, "White Matter Alterations In Parkinson's Disease Mapped Using Tractometry," *bioRxiv*, 2017.

[34] A. Yendiki, et al., "Automated Probabilistic Reconstruction of White-Matter Pathways in Health and Disease Using an Atlas of the Underlying Anatomy," *Frontiers in Neuroinformatics*, vol. 5, 2011-October-14 2011.

[35] P. T. Yap, J. H. Gilmore, W. Lin, and D. Shen, "PopTract: Population-Based Tractography". *IEEE Transactions on Medical Imaging*, vol. 30, pp. 1829-1840, 2011.

[36] M. Taquet, B. Scherrer, O. Commowick, J. M. Peters, M. Sahin, B. Macq, and S. K. Warfield, "A mathematical framework for the registration and analysis of multi-fascicle models for population studies of the brain microstructure," *Medical Imaging, IEEE Transactions on*, vol. 33, pp. 504-517, 2014.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2019.2922615, IEEE Transactions on Medical Imaging

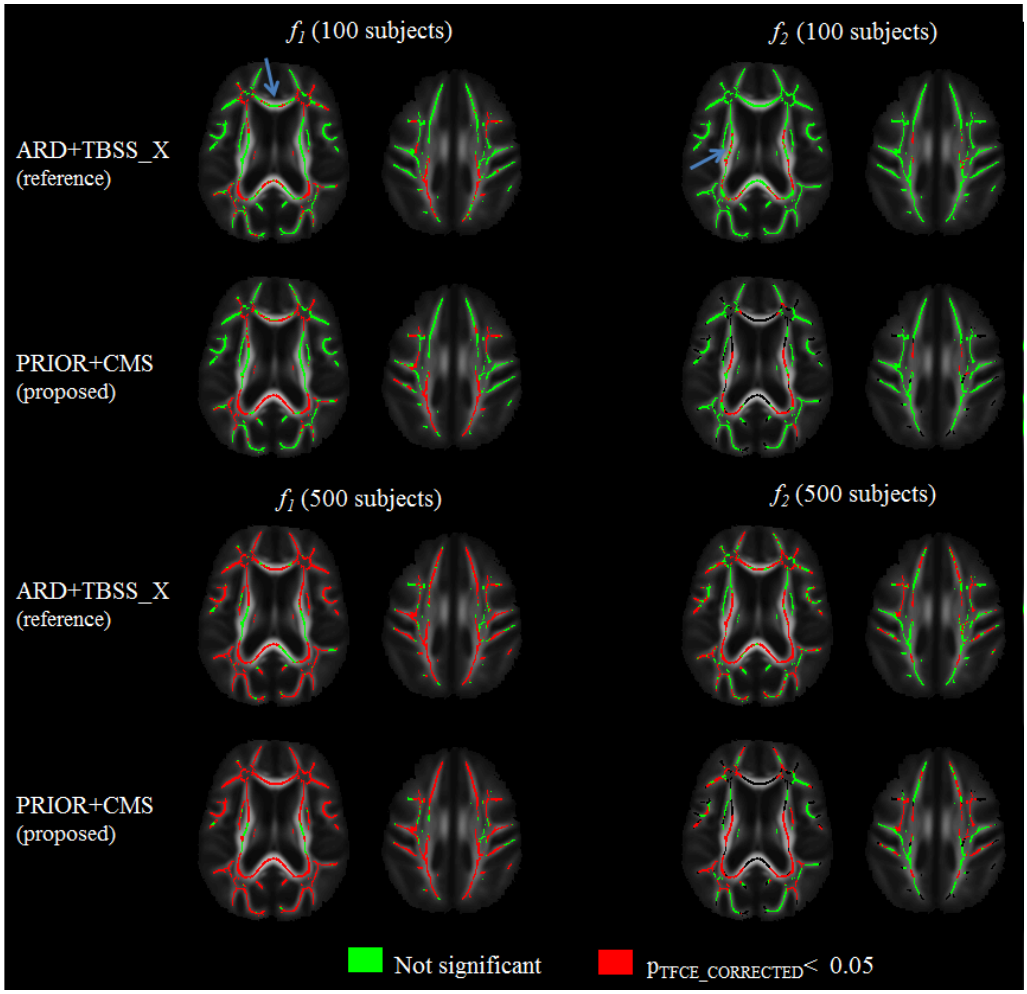

Fig. 9. Significance of the negative correlation between the primary and secondary PVFs with age. PVFs were generated using the proposed and second reference framework. Note that skeleton voxels with a single fiber population have been masked black for the secondary PVF obtained with the proposed framework. The arrows point at specific regions where the outcomes of the two frameworks differ.

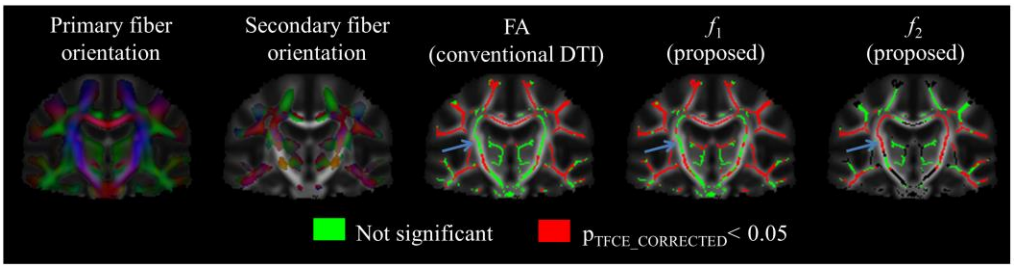

Fig. 10. Significance of the correlation between the FA, primary and secondary PVFs with age. FA was determined by means of the first reference framework, PVFs through the proposed framework. FA and primary PVF in skeleton voxels containing the corticospinal tract (arrows) show little correlation with age, whereas the secondary PVF has a significant negative correlation with age in those voxels. From the visualization of the fiber orientations it can be deduced in that region the primary fiber population reflects the corticospinal tract, and the secondary fiber population reflects a commissural tract.