Repeated Tractography of a Single Subject: How High Is the Variance?

Xuan Gu, Anders Eklund and Hans Knutsson

Book Chapter

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

Part of: Modeling, Analysis, and Visualization of Anisotropy, Thomas Schultz, Evren Özarslan, Ingrid Hotz (eds), 2017, pp. 331-354. ISBN: 978-3-319-61357-4 (Print) and 978-3-319-61358-1 (online)

Mathematics and Visualization (MATHVISUAL), 1612-3786, No.

DOI: https://doi.org/10.1007/978-3-319-61358-1_14

Series: Mathematical and Visualization (MATHVISUAL, ISSN: 1612-3786, eISSN 2197-666X)

Copyright: Springer Link

Available at: Linköping University Institutional Repository (DiVA)

<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-142047>

Repeated tractography of a single subject - How high is the variance?

Xuan Gu, Anders Eklund, Hans Knutsson

Abstract We have investigated the test-retest reliability of diffusion tractography, using 32 diffusion datasets from a single healthy subject. Preprocessing was carried out using functions in FSL (FMRIB Software Library), and tractography was carried out using FSL and Dipy. The tractography was performed in diffusion space, using two seed masks (corticospinal and cingulum gyrus tracts) created from the JHU White-Matter Tractography atlas. The tractography results were then warped into MNI standard space by a linear transformation. The reproducibility of tract metrics was examined using the standard deviation, the coefficient of variation (CV) and the Dice similarity coefficient (DSC), which all indicated a high reproducibility. Our results show that the multi-fiber model in FSL is able to reveal more connections between brain areas, compared to the single fiber model, and that distortion correction increases the reproducibility.

1 Introduction

In the past few years, a number of algorithms for reconstruction of fiber tracts from diffusion-weighted images have been proposed, collectively known as tractography (Caan, 2016). Tractography is an important neuroimaging tool which can be used for studying brain connectivity and aiding brain surgery (Bizzi, 2015; Castellano et al., 2012; Iliescu et al., 2010; Mastronardi et al., 2008). The fundamental goal of brain tumor surgery is to resect the maximum amount of tumoral tissue, while removing as little healthy tissue as possible. Figure 1 shows an example where the

Xuan Gu^{1,2}, Anders Eklund^{1,2,3}, Hans Knutsson^{1,2}

¹Division of Medical Informatics, Department of Biomedical Engineering

²Center for Medical Image Science and Visualization

³Division of Statistics and Machine Learning, Department of Computer and Information Science
Linköping University, Linköping, Sweden

e-mail: xuan.gu@liu.se, anders.eklund@liu.se, hans.knutsson@liu.se

estimated corticospinal tracts for a patient were very close to an Astrocytoma tumor in the right primary somatosensory area (Pernet et al., 2016). The corticospinal tract is one of the major nerve fiber tracts, and should be preserved as much as possible during brain surgery. Therefore, it is important to fully evaluate the reproducibility of diffusion tractography results, so that the surgical approaches can be designed to avoid damaging important tracts. Although tractography provides very nice images (Basser et al., 2000), there has been some concerns regarding the reproducibility of the results, and these concerns are especially important for clinical applications. Reconstruction of fiber tracts can, for example, differ depending on the software package being used, the specific algorithm, and different parameter settings.

Fig. 1 A mid-axial slice of the corticospinal tract, for a subject with an Astrocytoma tumor in the right primary somatosensory area. A threshold of 0.2% was used to remove less likely tracts. While tractography can aid planning of tumor surgery, it is important to know the uncertainty of the nice images.

However, only a few attempts have been made so far to quantitatively investigate the reproducibility of different methods and softwares for tractography. Heiervang et al. (2006) conducted studies to characterize the reproducibility and variability of diffusion tractography using datasets from 8 subjects scanned three times. The mean tract fractional anisotropy (FA) and the mean diffusivity (MD) along the tracts showed a very low coefficient of variation, both below 2% for inter-session and 3-5% for inter-subject. They also found that the number of diffusion directions (60 and 12) has a limited effect on the inter-session coefficient of variation (CV). Datasets with more diffusion directions, however, produced greater tract volume. Tensaouti et al. (2008) reported a maximum value of 56% of tract volume agreement for diffusion data collected in 32 directions. They drew a similar conclusion as Heiervang et al. (2006) that more directions allows to detect more tracts until a certain level between 15-32 directions. Vollmar et al. (2010) investigated the intra-site and inter-site reproducibility and reported a very low (1.6%) CV for the mean tract FA within the tracts, and 6.2-8.4% CV of the tract volume. It is also claimed that nonlinear registration between scans can be used to eliminate different distortions and improve reproducibility. Vaessen et al. (2010) evaluated the reproducibility of brain network connectivity by diffusion tractography, and reported low values (3.8%) for

the CV of the network measures. Danielian et al. (2010) assessed intra-human-rater and inter-human-rater reproducibility by measurements of tract FA, MD, axial diffusivity, transverse diffusivity and tract volume using intraclass correlation coefficient (ICC), CV and kappa (κ) statistic. They reported an ICC greater than 0.77 and a κ greater than 0.76 for inter-human-rater and an ICC greater than 0.92 and a κ greater than 0.9 for intra-human-rater. Tensaouti et al. (2011) evaluated the reproducibility of tractography in terms of data acquisitions and tractography algorithms. Similarly to (Heiervang et al., 2006; Tensaouti et al., 2008), Tensaouti et al. reported an increase in reproducibility in tractography according to the number of directions used during the scans.

For reproducibility of tractography in brachial plexus and kidney, evaluations were carried out by (Cutajar et al., 2011; Tagliafico et al., 2011). Apart from studies that assess the longitudinal reproducibility of fiber tractography, results regarding the accuracy of fiber tractography can be found in (Côté et al., 2013; Fillard et al., 2011; Hagmann et al., 2008; Neher et al., 2015; Pujol et al., 2015).

The aim of this study was therefore to analyze the reproducibility of tractography for diffusion data repeatedly collected from a single healthy subject. This makes it possible to investigate how the whole workflow (data collection, preprocessing and tractography) affects the final results, instead of only focusing on the tractography itself. We here present results for the popular diffusion tractography softwares FSL (Jenkinson et al., 2012) and Dipy (Garyfallidis et al., 2014) which use a probabilistic and a deterministic tractography algorithm, respectively. For the quantitative analysis of the results, we focused on the CV and the Dice similarity coefficient (DSC) which indicate how high the reproducibility is. Testing other software packages is planned for future work.

The remainder of the paper is organized as follows. In section 3, the evaluated tractography software packages and image process workflow is presented in detail. The results of the evaluations are shown in section 4, and the results are discussed in section 5.

2 Data

Diffusion datasets were acquired from the MyConnectome (Laumann et al., 2015) study (myconnectome.org), where MR imaging was performed on a fixed schedule during an entire year on a single healthy individual. In this study, we chose the 16 scan sessions (out of the total 106) containing diffusion data, obtained using a multiband EPI sequence on a Siemens Skyra 3T scanner. Each session consists of two scans, giving a total of 32 diffusion datasets, with L→R phase encoding and the other with R→L phase encoding (these two scans can be combined to correct for distortions). The following scanning parameters were used: $b=1000/2000 \text{ s} \cdot \text{mm}^{-2}$ (30 gradient directions per b-factor and 4 volumes without diffusion weighting), $1.74 \times 1.74 \times 1.7 \text{ mm}$ voxels, 72 slices, 128×128 matrix, TR = 5000 ms, TE = 108 ms, multiband factor 3.

3 Methods

3.1 FSL

Preprocessing and tractography was carried out using tools in FSL (Jenkinson et al., 2012). Susceptibility distortions were corrected for using the function *topup* (Andersson et al., 2003), while the function *eddy* (Andersson and Sotiropoulos, 2016) was used to correct for head motion and eddy-current induced distortions.

The ball and stick model (Behrens et al., 2007) is in the function *bedpostx* used to model the diffusion in each voxel, with one isotropic part (ball) and N non-isotropic parts (sticks), according to

$$\frac{S_i}{S_0} = (1 - \sum_{j=1}^N f_j) \exp(-b_i d) + \sum_{j=1}^N f_j \exp(-b_i d \mathbf{r}_i^T \mathbf{R}_j \mathbf{A} \mathbf{R}_j^T \mathbf{r}_i), \quad (1)$$

where S_i is the diffusion-weighted signal for measurement i , S_0 is the signal with no diffusion gradient applied, f_j is the fraction of signal contributed by diffusion along fiber direction j , b_i and $\mathbf{r}_i$ are the b-value and the gradient direction for measurement i , d is the diffusivity, $\mathbf{R} \mathbf{A} \mathbf{R}^T$ is the anisotropic diffusion tensor along the fiber direction (θ_j, ϕ_j) where

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Each stick is thus represented by two angles, using a spherical coordinate system. The first term represents the diffusion of free water, and the second term represents the diffusion along the different fiber orientations. The joint posterior distribution of the parameters of interest is derived using Bayesian inference. Specifically, Markov Chain Monte Carlo (MCMC) simulation is used in *bedpostx* to generate draws from the complicated posterior distribution. In the case of multiple fiber orientations, automatic relevance determination (ARD) is used to in each voxel determine the optimal number of fibers. The GPU version of *bedpostx* (Hernández et al., 2013) was used in our case, to reduce the processing time from 15 hours to 40 minutes per analysis. Each dataset was analyzed 4 times (maximum of 1, 2, 3 or 4 crossing fibers in each voxel), resulting in a total of 128 analyses for the 32 datasets.

The function *fslmaths* was used to create two seed masks in MNI space (for corticospinal and cingulum gyrus tracts), using the JHU White-Matter Tractography atlas (Hua et al., 2008). The transformation between standard space and diffusion space was achieved in three separate steps. First, the anatomical volume was linearly registered to MNI space using the function *flirt* (Jenkinson and Smith, 2001). Second, the diffusion data was linearly registered to the anatomical volume. Third, the two transformations were combined, to transform the seed masks from MNI space to diffusion space.

The probabilistic tractography was performed using the function *probtrackx2* (Behrens et al., 2007) in diffusion space. The *bedpostx* function results in draws

from the posterior distribution of the ball and stick model, which are then used by *probtrackx2* to achieve probabilistic fiber tracking. A total of 5,000 streamlines were initiated from each seed voxel, and constructed by randomly selecting one draw from the ball and stick model in each voxel, and following the main fiber orientation given by that draw. Each streamline was stopped after 2,000 steps with a step length of 0.5 mm, or terminated if the curvature exceeded 80 degrees. A volume containing the output connectivity distribution from the seed mask was finally computed, where each entry denotes the number of streamlines that passed through that voxel. The results of *probtrackx2* were transformed to MNI standard space by linear transformation, inverting the previously calculated linear transformation. The results were finally normalized by the total number of streamlines from the seed mask. By removing the less likely tracts, using a threshold, it becomes easier to see the most important connections from the seed mask to the rest of the brain.

3.2 *Dipy*

To further extend this study, we performed deterministic tractography on the same datasets using *Dipy* (Garyfallidis et al., 2014). The deterministic tractography was performed on the distortion-corrected datasets from FSL. To get directions from the diffusion dataset, we fitted each voxel to a Constant Solid Angle Model (Aganj et al., 2010) using the function *CsaOdfModel*. This model will represent the orientation distribution function (ODF) in each voxel. The ODF is the distribution of water diffusion as a function of direction. The peaks of an ODF can be obtained by the function *peaks_from_model*, and they can be good estimates for the orientations of the streamlines passing through voxels. The same corticospinal and cingulum gyrus tracts were used as seed masks. One seed per voxel (in the center) was used for the tractography.

4 Results

4.1 *FSL*

Figure 3 shows one axial slice and one sagittal slice of the tractography results for the corticospinal and cingulum gyrus tracts, for all 32 datasets. The tractography results were thresholded at a connectivity value of 0.2%. The background image is the MNI template brain. The corticospinal and cingulum gyrus seed masks, as shown in Figure 4, were created with *fslmaths*, by thresholding the JHU White-Matter Tractography atlas (Mori et al., 2005). Due to noise and scan artifacts, none of the datasets give identical tractography results. Nevertheless, there is clearly a high degree of similarity between the tractography maps from the different datasets.

In the following results, we try to measure the degree of similarity using different metrics.

The reproducibility of the tractography was first examined using the CV, which is defined as the ratio of the standard deviation σ divided by the mean μ :

$$CV = \frac{\sigma}{\mu}. \quad (2)$$

The CV is a standardized measure of dispersion of a probability distribution (Brown, 1998). It gives an intuitive estimate of the measurement repeatability, expressed as a relative percentage (regardless of the absolute measurement value). The standard deviation σ and mean μ can be calculated from the 32 tractography results. Figure 5 shows the CV of the tractography results for various thresholds when using a seed mask for the corticospinal tract. By removing the voxels below the threshold, it is possible to focus on the tracts with a lower CV.

Figure 6 shows the average tractography results over the 32 datasets, when changing the maximum number of crossing fibers x in the *bedpostx* function. It has previously been reported that some 70% of the white matter voxels contain at least two crossing fibers (Jeurissen et al., 2013). Our results show that the multi-fiber model is in general able to reveal more connections between the brain areas, compared to the single fiber model, for both corticospinal and cingulum gyrus tracts. The ability to detect connections varies when different maximum number of crossing fibers is used. Nevertheless, the same cores of the tracts were found by all settings. The standard deviation, for the voxels that survived an initial threshold, was calculated over all 32 datasets and is shown in Figure 7.

Figure 8 shows the tract volume of the tractography results, for different settings of the maximum number of crossing fibers. The tract volume was estimated as the total number of voxels that survived thresholding. The average tract volume of corticospinal tracts over the 32 datasets are 3527, 4648, 4210 and 4208 for a maximum of 1, 2, 3 and 4 crossing fibers. The average tract volume of cingulum gyrus tracts over the 32 datasets are 1892, 3153, 2548 and 2583 for a maximum of 1, 2, 3 and 4 crossing fibers.

To better understand the results in Figure 8, we plotted the average tract volume over the 32 datasets before and after thresholding for different settings of the maximum number of crossing fibers, see Figure 9. For both before and after thresholding, tractography results obtained using a single fiber model ($x = 1$ in the FSL function *bedpostx*) are expected to represent a smaller tract, i.e. there are fewer connections. This is due to the fact that a multi-fiber model ($x = 2, 3, 4$ in the FSL function *bedpostx*) results in a greater variability of fiber orientations and tend to disperse the streamlines to more voxels. This prediction clearly holds for the results in Figure 8 and 9, for both the corticospinal and the cingulum gyrus tracts. However, after thresholding, the tract volume for the 3 fibers per voxel case was decreased a little compared with a maximum of 2 fibers. The reason is that when a larger number of crossing fibers are fitted, the tractography results tend to be more dispersive, i.e. the connectivity value in the voxels will be smaller. Therefore, a larger amount of voxels will be removed after the thresholding, which causes a smaller tract volume. It is

Fig. 2 Tractography results for 32 diffusion datasets (in MNI space), for a maximum of 2 fibers per voxel, when using seed masks for the corticospinal tract. A threshold of 0.2% was used to remove less likely tracts. The results represent the proportion of times a streamline passed through a voxel.

Fig. 3 Tractography results for 32 diffusion datasets (in MNI space), for a maximum of 2 fibers per voxel, when using seed masks for the cingulum gyrus tract. A threshold of 0.2% was used to remove less likely tracts. The results represent the proportion of times a streamline passed through a voxel.

Fig. 4 The corticospinal and cingulum gyrus tracts used as seed masks. From each voxel in the seed mask, 5000 streamlines were generated, using probabilistic fiber tracking, to obtain a volume where each voxel represents the number of times a streamline passed through that voxel.

Fig. 5 CV (in MNI space) of the tractography results, for a maximum of 2 fibers per voxel, when using a seed mask for the corticospinal tract. The threshold (in percentage) is, from left to right, 0.01, 0.05, 0.1, 0.5.

interesting to note that increasing the number of maximum fibers from 3 to 4, does not change the tract volume further. This is because very few voxels in the datasets support 4 crossing fibers, see Figure 10 and 11.

Figure 10 shows an example of the multi-fiber ball and stick model fitting for one of the 32 datasets. The fraction of signal contributed by diffusion along fiber direction f_i (output of FSL function *bedpostx*, see Equation 1) was thresholded at 0.05 (Behrens et al., 2007) and the surviving voxels were able to detect more than i crossing fibers. For the chosen dataset, 40% of voxels with $FA > 0.1$ were able to detect at least two crossing fibers which is a little higher than the previous reported 33% in (Behrens et al., 2007), where a different data acquisition scheme was used. For only 4.76% of voxels with $FA > 0.1$, more than two fibers was supported by the model, and for only 0.05% of voxels with $FA > 0.1$ more than three fibers was supported.

Figure 11 shows the percentage of voxels supporting 1, 2, 3 and 4 crossing fibers in corticospinal (blue) and cingulum gyrus (yellow) tracts for the same chosen dataset as in Figure 10. For both the corticospinal and the cingulum gyrus (yellow) tracts, around 20%, 70% and 10% of the voxels supported 1, 2 and 3 crossing fibers, respectively. The number of voxels supporting more than 3 crossing fibers is negligible. In (Behrens et al., 2007) it was reported that no single voxel was able to support

Fig. 6 Averaged tractography results over the 32 datasets (in MNI space), for a maximum of 1, 2, 3 or 4 fibers per voxel (left to right), when using seed masks for the corticospinal and cingulum gyrus tracts. A threshold of 0.1% was used to remove less likely tracts. Using a maximum of 2 fibers per voxel leads to including more voxels in the tracts, compared to using a maximum of 1 fiber per voxel.

Fig. 7 Standard deviation of tractography results over the 32 datasets (in MNI space), for a maximum of 1, 2, 3 or 4 fibers per voxel (left to right), when using seed masks for the corticospinal and cingulum gyrus tracts. A threshold of 0.1% was used to remove less likely tracts.

Fig. 8 Tract volume of the tractography results, for a maximum of 1, 2, 3 or 4 fibers per voxel. The top figure shows the results for the corticospinal tract, and the bottom figure shows the results for the cingulum gyrus tract. A threshold of 0.5% was used to remove less likely tracts. Using a maximum of 2 fibers per voxel reveals most brain connections, while using a maximum of 1 fiber leads to the fewest connections.

more than 2 crossing fibers when 60 directions was used for data acquisition, and no single voxel was able to support more than 1 crossing fibers when 12 directions was used. Tuch et al. (2003) reported that it is possible to detect 3 crossing fibers in the corticospinal tract in high b-value data. In our study, the data acquisition scheme using two shells and 30 directions for each shell made it possible to detect 3 crossing fibers in 12.1% and 5.5% of the voxels in the corticospinal and cingulum gyrus tracts, respectively, as shown in Figure 11. It is reasonable to infer that data acquisition schemes using either more shells, or more directions, or higher b-value allows more crossing fibers to be supported by the data.

Figure 12 shows the probability density of the CV when the maximum number of crossing fibers varies from 1 to 4. The CV is the precision of a measure, i.e. a smaller CV is equivalent to a higher reproducibility of the fiber tracts. The probability density was estimated using the MATLAB (Version 2016b) function *ksdensity* with 100 bins. For the corticospinal case, the total number of voxels after thresholding is 48010, 172799, 143935, and 146322 for a maximum of 1, 2, 3 and 4 fibers, respectively. The corresponding number of voxels for $0 < CV < 1$ is 15556, 112227,

Fig. 9 Tract volume of the tractography results, for a maximum of 1, 2, 3 or 4 fibers per voxel. The top figure shows the results before thresholding, and the bottom figure shows the results after thresholding.

Fig. 10 Multi-fiber ball and stick model fitting. An axial slice showing where more than 1 (left), 2 (middle), and 3 (right) crossing fibers in each voxel were supported by the dataset. The fiber fractions f_2 , f_3 and f_4 were thresholded at 0.05.

Fig. 11 Percentage of voxels supporting 1, 2, 3 and 4 crossing fibers, corticospinal tract in blue and cingulum gyrus tract in yellow. A threshold of 0.5% was used to remove less likely tracts.

76359, and 77328. The mean of the CV over all voxels is 1.32, 0.99, 1.15, and 1.16, for a maximum of 1 to 4 fibers. The choice of a maximum of 2 fibers per voxel gives the highest reproducibility, at the cost of revealing fewer connections between brain areas, compared with a maximum of 3 or 4 fibers per voxel. The single-fiber model leads to the highest variation. Figure 12 shows the probability density of the CV with and without distortion correction, when the maximum number of crossing fibers is set to 3. Correcting for susceptibility effects, eddy currents and head motion clearly increases the reproducibility.

We derived mean and CV for tract FA, MD and volume (Table 1) across the 32 datasets for the corticospinal and cingulum gyrus tracts when a maximum of 1, 2, 3 and 4 fibers per voxel was applied. A threshold of 0.5% was used to remove less likely tracts. Tract FA and MD are defined as the average FA and MD values within a tract. The tract volume is the total number of voxels that survived thresholding. FA and MD of whole brain were obtained using the FSL function *dtifit* to fit a diffusion tensor model in each voxel. FA and MD results were then linearly registered to MNI space using the function *flirt* (Jenkinson and Smith, 2001). Mean tract FA across datasets ranged from 0.55 to 0.58 and 0.47 to 0.50 for the corticospinal and cingulum gyrus tracts, respectively. Measures of mean tract FA, MD and tract volume produced very low CVs, for both corticospinal (below 3.06%) and

Fig. 12 Probability density of CV for the tractography results, for a maximum of 1, 2, 3 or 4 fibers per voxel. The top figure shows the results for the corticospinal tract, and the bottom figure shows the results for the cingulum gyrus tract. A threshold of 0.001% was used to remove less likely tracts.

cingulum gyrus (below 6.87%) tracts. CVs for mean tract FA of cingulum gyrus across datasets (3.86 - 4.78%) are largely consistent with previous reported results (3.18 - 4.32%) (Heiervang et al., 2006). Results for the corticospinal tract show a higher degree of reproducibility than the cingulum gyrus tract. It is consistent with previous research (Heiervang et al., 2006; Vollmar et al., 2010) that larger tracts can produce a lower CV since they are less sensitive to the uncertainty of tractography, artifacts of scanning and noise.

Fig. 13 Probability density of CV for the tractography results with and without distortion correction, for a maximum of 3 fibers per voxel. The top figure shows the results for the corticospinal tract, and the bottom figure shows the results for the cingulum gyrus tract. A threshold of 0.001% was used to remove less likely tracts.

Table 1 Results for tracts analysis of mean tract FA, MD and tract volume. A threshold of 0.5% was used. Tract volume is defined as the number of voxels of the tract.

Measure	Tract	Maximum number of crossing fibers	Mean across datasets	CV across(%) datasets
FA	Corticospinal	1	0.58	1.03
		2	0.55	1.12
		3	0.57	1.04
		4	0.58	1.06
	Cingulum	1	0.50	4.29
		2	0.47	3.96
		3	0.49	3.86
		4	0.48	4.78
MD	Corticospinal	1	($\times 10^{-4}$) 5.48	1.40
		2	5.61	2.17
		3	5.52	1.78
		4	5.51	1.97
	Cingulum	1	5.86	2.01
		2	5.87	2.42
		3	5.82	1.89
		4	5.86	2.53
Tract volume	Corticospinal	1	3522	3.74
		2	4648	2.50
		3	4210	3.06
		4	4208	3.05
	Cingulum	1	1892	5.36
		2	3153	6.87
		3	2548	5.74
		4	2583	5.20

The agreement of two tracts was quantified using the DSC (Dice, 1945), which quantifies the degree of overlap using a number between 0 (no overlap) to 1 (complete overlap):

$$\text{DSC} = \frac{2n_{ab}}{n_a + n_b}, \quad (3)$$

where n_{ab} is the number of voxels common to both volumes, and n_a and n_b are the number of voxels in volume a and volume b. The average tracts over the 32 datasets was used as the benchmark, and the DSC was then calculated between each tractography result and the average one. The results are shown in Figure 14. With a threshold of 0.5%, the mean of the DSC for the corticospinal tract over the 32 datasets is 0.82, 0.85, 0.84 and 0.84 for a maximum of 1, 2, 3 and 4 fibers per voxel, respectively. Bauer et al. (2013) repeated deterministic tractography for the corticospinal tract and reported that a DSC above 0.8 can be achieved. This is largely consistent with our results. The mean of the DSC for the cingulum gyrus tract over the 32 datasets is 0.76, 0.78, 0.77 and 0.81 for a maximum of 1, 2, 3 and 4 fibers per voxel, respectively. This is very close to the previous reported 0.8 in (Besseling et al., 2012). The DSC for the cingulum gyrus tract shows a larger variance than for the corticospinal tract. The standard deviation of the DSC for the corticospinal tract is 0.035, 0.0278, 0.03 and 0.0295 for a maximum of 1, 2, 3 and 4 fibers per voxel, respectively. The standard deviation of the DSC for the cingulum gyrus tract is 0.053, 0.031, 0.036 and 0.0532 for a maximum of 1, 2, 3 and 4 fibers per voxel, respectively. Our results thereby suggest that the reproducibility of fiber tracts depends on the seed mask used. Such an effect is natural, because different parts of the brain may suffer differently from distortions and head motion. The size of the seed mask is also a factor that has an influence on the degree of reproducibility. For the DSC, the setting of the maximum of fibers per voxel did not produce a significant difference. It is hard to tell which setting provided the highest reproducibility, but in general the multi-fiber model achieved a better performance. Figure 15 shows the DSC with and without distortion correction, for a maximum of 3 crossing fibers per voxel. Together with Figure 13, we again draw the conclusion that distortion correction leads to a higher reproducibility.

From Figures 6, 8 and 14 we can see that the tractography for the different datasets resulted in very similar fiber pathways, and does not reveal more information when the maximum number of fibers is set to 4. It can be concluded that a maximum of 3 fibers per voxel may be sufficient to reveal connections between brain areas, considering the longer computation time when a larger number of maximum fibers is used.

The Bayesian estimation of diffusion parameters implemented in *bedposts* takes approximately 15 hours to complete for one diffusion dataset. Using a graphics card, the GPU version of *bedpostx* takes 10 to 60 minutes depending on the computer hardware, the chosen maximum number of fibers and the size of the datasets. The processing time for Bayesian estimation of the multi-fiber models are 10, 18, 27 and 40 minutes (for a maximum number of 1, 2, 3 and 4 fibers), using the graphics card NVIDIA Tesla K40c and the CPU Intel i7-5820K 3.30GHz. The processing times

Fig. 14 DSCs for the tractography results, for a maximum of 1, 2, 3 or 4 fibers per voxel. The top figure shows the results for the corticospinal tract, and the bottom figure shows the results for the cingulum gyrus tract. A threshold of 0.5% was used to remove less likely tracts.

for probabilistic tractography of the multi-fiber models are 68, 114, 112 and 115 minutes, respectively.

4.2 Dipy

Figure 17 shows one axial slice and one sagittal slice of the deterministic tractography results from Dipy, for the corticospinal and cingulum gyrus tracts for all 32 datasets. The background image is the MNI brain template. As in Figure 3, it also shows a high degree of reproducibility for the tractography results from the different datasets. In the following results, we try to measure the degree of similarity using different metrics.

Figure 18 shows the DSCs for the deterministic tractography results from Dipy. The tractography result of the first dataset was chosen as the benchmark, and the DSC was then calculated between each tractography result and the benchmark. Please note that the DSCs for the results from FSL and Dipy were calculated based on different benchmarks. Therefore the DSCs cannot be compared directly. The

Fig. 15 DSCs for the tractography results with and without distortion correction, for a maximum of 3 fibers per voxel. The top figure shows the results for the corticospinal tract, and the bottom figure shows the results for the cingulum gyrus tract. A threshold of 0.1% was used to remove less likely tracts.

mean DSC for the corticospinal and cingulum gyrus tracts over the latter 31 datasets is 0.669 and 0.590, respectively. Compared with the results in Figure 14 we can see that the deterministic tractography results from Dipy show a lower degree of reproducibility than the probabilistic tractography results from FSL. Although the deterministic tractography algorithm is very efficient, it can be sensitive to the estimated principal directions since the streamline in each voxel only follows the principal direction. The deterministic tractography algorithm fails to resolve complex fiber structure when fibers are crossing, which is one of the major limitations of the deterministic tractography algorithm. The uncertainties in the underlying fiber directions makes the tractography less reproducible than its probabilistic counterpart.

5 Discussion and conclusion

In this study, we have presumed that the brain nerve tracts do not change significantly during the scan interval (one year). There are previous studies (Nusbaum et al., 2001; Yoon et al., 2008) demonstrating changes of certain white matter tracts

Fig. 16 Deterministic tractography results for 32 diffusion datasets (in MNI space), when using seed masks for the corticospinal tract. The results represent the number of streamlines that passed through a voxel.

Fig. 17 Deterministic tractography results for 32 diffusion datasets (in MNI space), when using seed masks for the cingulum gyrus tract. The results represent the number of streamlines that passed through a voxel.

Fig. 18 DSCs for the tractography results using Dipy. The first tractography map was used as the benchmark, and the DSC was then calculated between each tractography map and the benchmark. The top figure shows the results for the corticospinal tract, and the bottom figure shows the results for the cingulum gyrus tract.

due to aging. However, it is unlikely to recognize significant age-related changes over the course of one year for a healthy subject.

We have investigated the test–retest reliability of diffusion tractography, using 32 diffusion datasets from a single healthy volunteer. A visual comparison of the results shows that the cores of the corticospinal and cingulum gyrus tracts are common over the 32 datasets, for both FSL and Dipy. We have reported inter-dataset overlaps of $DSC = 0.6 - 0.9$ for the probabilistic tractography results from FSL, and $DSC = 0.58 - 0.71$ for the deterministic tractography results from Dipy. The DSC values roughly fit with the range reported as $0.67 - 0.9$ in (Besseling et al., 2012). We also observed that the DSC differs between the corticospinal and cingulum tracts. This is because the size of the tracts can be a factor influencing the degree of reproducibility. Also, different parts of the brain may experience different distortions and head motion. The results indicate that distortions and head motion can be an important uncertainty source. It was observed that the reproducibility increases if distortion correction is used, for both the corticospinal and cingulum gyrus tracts. The results suggest that the ball and stick model representing multiple fiber orientations can reconstruct more connections, at the cost of a longer processing time. It is also demonstrated that

the tractography results do not differ much when the maximum number of crossing fibers is larger than 3 in the FSL function *bedpostx*, but a higher number of crossing fibers may be optimal for DWI data collected with a higher number of gradient directions and shells.

Based on the presented results we conclude that the tractography results obtained with different software packages, and different parameter settings, show a rather high reproducibility. It is important to note that the reproducibility of tractography by no means can be interpreted as the accuracy of tractography. Nonexisting fiber pathways can, in theory, be reconstructed with a high reproducibility. Future work will be focused on evaluating the reproducibility of other tractography softwares, such as TORTOISE (Pierpaoli et al., 2010) and DSI-Studio (Yeh et al., 2013).

Acknowledgements We thank Russell Poldrack and his colleagues for sharing the data from the MyConnectome project. We also thank Cyril Pernet and his colleagues for sharing neuroimaging data from brain tumor patients. The Nvidia Corporation is acknowledged for the donation of the Tesla K40 graphics card. 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”). Anders Eklund was also supported by Swedish Research Council Grant 2013-5229 (“Statistical Analysis of fMRI Data”).

References

- Aganj, I., Lenglet, C., Sapiro, G., Yacoub, E., Ugurbil, K., and Harel, N. (2010). Reconstruction of the orientation distribution function in single-and multiple-shell q-ball imaging within constant solid angle. *Magnetic Resonance in Medicine*, 64(2):554–566.
- Andersson, J. L., Skare, S., and Ashburner, J. (2003). How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *Neuroimage*, 20(2):870–888.
- Andersson, J. L. and Sotiropoulos, S. N. (2016). An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. *NeuroImage*, 125:1063–1078.
- Basser, P. J., Pajevic, S., Pierpaoli, C., Duda, J., and Aldroubi, A. (2000). In vivo fiber tractography using DT-MRI data. *Magnetic resonance in medicine*, 44(4):625–632.
- Bauer, M. H., Kuhnt, D., Barbieri, S., Klein, J., Becker, A., Freisleben, B., Hahn, H. K., and Nimsch, C. (2013). Reconstruction of white matter tracts via repeated deterministic streamline tracking—initial experience. *PloS one*, 8(5):e63082.
- Behrens, T., 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.

- Besseling, R. M., Jansen, J. F., Overvliet, G. M., Vaessen, M. J., Braakman, H. M., Hofman, P. A., Aldenkamp, A. P., and Backes, W. H. (2012). Tract specific reproducibility of tractography based morphology and diffusion metrics. *PloS one*, 7(4):e34125.
- Bizzi, A. (2015). Diffusion imaging with mr tractography for brain tumor surgery. In *Clinical Functional MRI*, pages 179–228. Springer.
- Brown, C. E. (1998). Coefficient of variation. In *Applied Multivariate Statistics in Geohydrology and Related Sciences*, pages 155–157. Springer.
- Caan, M. W. (2016). DTI analysis methods: Fibre tracking and connectivity. In *Diffusion Tensor Imaging*, pages 205–228. Springer.
- Castellano, A., Bello, L., Michelozzi, C., Gallucci, M., Fava, E., Iadanza, A., Riva, M., Casaceli, G., and Falini, A. (2012). Role of diffusion tensor magnetic resonance tractography in predicting the extent of resection in glioma surgery. *Neuro-oncology*, 14(2):192–202.
- Côté, M.-A., Girard, G., Boré, A., Garyfallidis, E., Houde, J.-C., and Descoteaux, M. (2013). Tractometer: towards validation of tractography pipelines. *Medical image analysis*, 17(7):844–857.
- Cutajar, M., Clayden, J. D., Clark, C. A., and Gordon, I. (2011). Test–retest reliability and repeatability of renal diffusion tensor MRI in healthy subjects. *European journal of radiology*, 80(3):e263–e268.
- Danielian, L. E., Iwata, N. K., Thomasson, D. M., and Floeter, M. K. (2010). Reliability of fiber tracking measurements in diffusion tensor imaging for longitudinal study. *Neuroimage*, 49(2):1572–1580.
- Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302.
- Fillard, P., Descoteaux, M., Goh, A., Gouttard, S., Jeurissen, B., Malcolm, J., Ramirez-Manzanares, A., Reisert, M., Sakaie, K., Tensaouti, F., et al. (2011). Quantitative evaluation of 10 tractography algorithms on a realistic diffusion MR phantom. *Neuroimage*, 56(1):220–234.
- Garyfallidis, E., Brett, M., Amirbekian, B., Rokem, A., Van Der Walt, S., Descoteaux, M., and Nimmo-Smith, I. (2014). Dipy, a library for the analysis of diffusion MRI data. *Frontiers in neuroinformatics*, 8:8.
- Hagmann, P., Gigandet, X., Meuli, R., Kötter, R., Sporns, O., and Wedeen, V. (2008). Quantitative validation of MR tractography using the cocomac database. In *Proceedings of 16th Annual Meeting of the ISMRM*, number EPFL-CONF-135048, page 427.
- Heiervang, E., Behrens, T., Mackay, C., Robson, M., and Johansen-Berg, H. (2006). Between session reproducibility and between subject variability of diffusion MR and tractography measures. *Neuroimage*, 33(3):867–877.
- Hernández, M., Guerrero, G. D., Cecilia, J. M., García, J. M., Inuggi, A., Jbabdi, S., Behrens, T. E., and Sotiropoulos, S. N. (2013). Accelerating fibre orientation estimation from diffusion weighted magnetic resonance imaging using GPUs. *PloS one*, 8(4):e61892.
- Hua, K., Zhang, J., Wakana, S., Jiang, H., Li, X., Reich, D. S., Calabresi, P. A., Pekar, J. J., van Zijl, P. C., and Mori, S. (2008). Tract probability maps in stereo-

- taxic spaces: analyses of white matter anatomy and tract-specific quantification. *Neuroimage*, 39(1):336–347.
- Iliescu, B., Negru, D., and Poeata, I. (2010). MR tractography for preoperative planning in patients with cerebral tumors in eloquent areas. *Romanian Neurosurgery*, 17(4):413–420.
- Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., and Smith, S. M. (2012). FSL. *Neuroimage*, 62(2):782–790.
- Jenkinson, M. and Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Medical image analysis*, 5(2):143–156.
- Jeurissen, B., Leemans, A., Tournier, J.-D., Jones, D. K., and Sijbers, J. (2013). Investigating the prevalence of complex fiber configurations in white matter tissue with diffusion magnetic resonance imaging. *Human brain mapping*, 34(11):2747–2766.
- Laumann, T. O., Gordon, E. M., Adeyemo, B., Snyder, A. Z., Joo, S. J., Chen, M.-Y., Gilmore, A. W., McDermott, K. B., Nelson, S. M., Dosenbach, N. U., Schlaggar, B. L., Mumford, J. A., Poldrack, R. A., and Petersen, S. E. (2015). Functional system and areal organization of a highly sampled individual human brain. *Neuron*, 87:657 – 670.
- Mastronardi, L., Bozzao, A., D’Andrea, G., Romano, A., Caroli, M., Cipriani, V., Ferrante, M., and Ferrante, L. (2008). Use of preoperative and intraoperative magnetic resonance tractography in intracranial tumor surgery. *Clin Neurosurg*, 55:160–164.
- Mori, S., Wakana, S., Van Zijl, P. C., and Nagae-Poetscher, L. (2005). *MRI atlas of human white matter*. Elsevier.
- Neher, P. F., Descoteaux, M., Houde, J.-C., Stieltjes, B., and Maier-Hein, K. H. (2015). Strengths and weaknesses of state of the art fiber tractography pipelines—a comprehensive in-vivo and phantom evaluation study using tractometer. *Medical image analysis*, 26(1):287–305.
- Nusbaum, A. O., Tang, C. Y., Buchsbaum, M. S., Wei, T. C., and Atlas, S. W. (2001). Regional and global changes in cerebral diffusion with normal aging. *American Journal of Neuroradiology*, 22(1):136–142.
- Pernet, C. R., Gorgolewski, K. J., Job, D., Rodriguez, D., Whittle, I., and Wardlaw, J. (2016). A structural and functional magnetic resonance imaging dataset of brain tumour patients. *Scientific data*, 3.
- Pierpaoli, C., Walker, L., Irfanoglu, M., Barnett, A., Basser, P., Chang, L., Koay, C., Pajevic, S., Rohde, G., Sarlls, J., et al. (2010). Tortoise: an integrated software package for processing of diffusion MRI data. *Book TORTOISE: an Integrated Software Package for Processing of Diffusion MRI Data (Editor ed^eds)*, 18:1597.
- Pujol, S., Wells, W., Pierpaoli, C., Brun, C., Gee, J., Cheng, G., Vemuri, B., Commowick, O., Prima, S., Stamm, A., et al. (2015). The DTI challenge: toward standardized evaluation of diffusion tensor imaging tractography for neurosurgery. *Journal of Neuroimaging*, 25(6):875–882.
- Tagliafico, A., Calabrese, M., Puntoni, M., Pace, D., Baio, G., Neumaier, C. E., and Martinoli, C. (2011). Brachial plexus MR imaging: accuracy and reproducibility

- of DTI-derived measurements and fibre tractography at 3.0-t. *European radiology*, 21(8):1764–1771.
- Tensaouti, F., Delion, M., Lotterie, J. A., Clarisse, P., and Berry, I. (2008). Reproducibility and reliability of the DTI fiber tracking algorithm integrated in the sisyphus software. In *2008 First Workshops on Image Processing Theory, Tools and Applications*.
- Tensaouti, F., Lahlou, I., Clarisse, P., Lotterie, J. A., and Berry, I. (2011). Quantitative and reproducibility study of four tractography algorithms used in clinical routine. *Journal of Magnetic Resonance Imaging*, 34(1):165–172.
- Tuch, D. S., Reese, T. G., Wiegell, M. R., and Wedeen, V. J. (2003). Diffusion MRI of complex neural architecture. *Neuron*, 40(5):885–895.
- Vaessen, M., Hofman, P., Tijssen, H., Aldenkamp, A., Jansen, J., and Backes, W. H. (2010). The effect and reproducibility of different clinical DTI gradient sets on small world brain connectivity measures. *Neuroimage*, 51(3):1106–1116.
- Vollmar, C., O’Muircheartaigh, J., Barker, G. J., Symms, M. R., Thompson, P., Kumari, V., Duncan, J. S., Richardson, M. P., and Koepp, M. J. (2010). Identical, but not the same: intra-site and inter-site reproducibility of fractional anisotropy measures on two 3.0 t scanners. *Neuroimage*, 51(4):1384–1394.
- Yeh, F.-C., Verstynen, T. D., Wang, Y., Fernández-Miranda, J. C., and Tseng, W.-Y. I. (2013). Deterministic diffusion fiber tracking improved by quantitative anisotropy. *PLoS one*, 8(11):e80713.
- Yoon, B., Shim, Y.-S., Lee, K.-S., Shon, Y.-M., and Yang, D.-W. (2008). Region-specific changes of cerebral white matter during normal aging: a diffusion-tensor analysis. *Archives of gerontology and geriatrics*, 47(1):129–138.

Index

coefficient of variation, 1, 2

Dice similarity coefficient, 1, 3

diffusion tractography, 1–3, 21

Dipy, 1, 3, 5, 17, 18, 21

FSL, 1, 3–6, 9, 13, 17, 18, 21, 22

reproducibility, 1–3, 6, 11, 13, 14, 16–18, 21, 22

test-retest reliability, 1