

A novel measure of reliability in Diffusion Tensor Imaging after data rejections due to subject motion

V. Sairanen^{a,b,*}, L. Kuusela^{a,b}, O. Sipilä^b, S. Savolainen^{a,b}, S. Vanhatalo^{c,d}

^a Department of Physics, University of Helsinki, P.O. Box 64, FI-00014 Helsinki, Finland

^b HUS Medical Imaging Center, University of Helsinki and Helsinki University Hospital Finland, P.O. Box 340, FI-00029 HUS, Finland

^c Departments of Children's Clinical Neurophysiology and HUS Imaging, Children's Hospital, Helsinki University Central Hospital, Helsinki, Finland

^d Department of Neurological Sciences, University of Helsinki, Finland

ARTICLE INFO

Keywords:

Diffusion Tensor Imaging
Neonatal MRI
Fractional anisotropy
Quality
Condition number
Tractography

ABSTRACT

Diffusion Tensor Imaging (DTI) is commonly challenged by subject motion during data acquisition, which often leads to corrupted image data. Currently used procedure in DTI analysis is to correct or completely reject such data before tensor estimations, however assessing the reliability and accuracy of the estimated tensor in such situations has evaded previous studies. This work aims to define the loss of data accuracy with increasing image rejections, and to define a robust method for assessing reliability of the result at voxel level. We carried out simulations of every possible sub-scheme ($N=1,073,567,387$) of Jones30 gradient scheme, followed by confirming the idea with MRI data from four newborn and three adult subjects. We assessed the relative error of the most commonly used tensor estimates for DTI and tractography studies, fractional anisotropy (FA) and the major orientation vector (V1), respectively. The error was estimated using two measures, the widely used electric potential (EP) criteria as well as the rotationally variant condition number (CN). Our results show that CN and EP are comparable in situations with very few rejections, but CN becomes clearly more sensitive to depicting errors when more gradient vectors and images were rejected. The error in FA and V1 was also found depend on the actual FA level in the given voxel; low actual FA levels were related to high relative errors in the FA and V1 estimates. Finally, the results were confirmed with clinical MRI data. This showed that the errors after rejections are, indeed, inhomogeneous across brain regions. The FA and V1 errors become progressively larger when moving from the thick white matter bundles towards more superficial subcortical structures. Our findings suggest that i) CN is a useful estimator of data reliability at voxel level, and ii) DTI preprocessing with data rejections leads to major challenges when assessing brain tissue with lower FA levels, such as all newborn brain, as well as the adult superficial, subcortical areas commonly traced in precise connectivity analyses between cortical regions.

1. Introduction

Diffusion-weighted imaging (DWI) is a widely used non-invasive MRI technique. It is often extended into voxel-wise estimation of diffusion tensors (DT) to reveal fine details of brain structures by measuring DWIs in six or more non-collinear directions (Basser et al., 1994; Pierpaoli et al., 1996; Horsfield and Jones, 2002; Lerner et al., 2014). Fractional anisotropy (FA) and other diffusion anisotropy indices can be used to describe the DT properties with a single scalar value (Basser and Pierpaoli, 1996), whereas the major orientation vector of the DT (V1) is used to guide fiber tractography for visualizing white matter (WM) pathways (Basser et al., 2000).

The accuracy and precision of these parameter estimates translate

directly into subject level reliability of the DTI findings. The DTI analysis, is commonly challenged by subject motion (Dubois et al., 2006; Jiang et al., 2009; Chen et al., 2015), which is unavoidable in clinical studies where subjects may be poorly cooperating due to disease progression, for example. One approach to handle these artefacts is to computationally post-process DWIs and diffusion gradient vectors (Rohde et al., 2004; Leemans and Jones, 2009). However, more severe artefacts require a total removal of artefactual DWIs and the corresponding gradient vectors. These rejections can be done based on a visual inspection of DWIs followed by manual exclusion, with automated tools such as DTIPrep (Oguz et al., 2014), or by using robust DT estimation algorithms (Chang et al., 2005; Tax et al., 2015), which perform rejections in voxel-wise manner.

* Correspondence to: HUS Medical Imaging Center, University of Helsinki and Helsinki University Hospital, P.O. Box 340 (Haartmaninkatu 4), 00029 Helsinki, Finland.
E-mail address: viljami.sairanen@helsinki.fi (V. Sairanen).

Knowing the impact of rejecting DWIs and corresponding gradient directions on data reliability is crucial, yet it has been incompletely resolved. Previously Bootstrap (Heim et al., 2004) and Monte Carlo (Farzinfar et al., 2013) methods have been applied to estimate the reliability of the results. A recent work by Chen et al. (2015) studied the incremental deterioration of DT estimates after DWI rejections using *a priori* defined sub-scheme groups. Despite their interesting observations, the latter approach was inherently ill-posed, because of the overwhelming number of possible permutations. For instance, from the 30 original DWI directions, one rejection can be performed only in 30 combinations, however 15 rejections can be performed in altogether $\binom{30}{15} = 155\,117\,520$ different combinations (Riordan, 1958). In addition, the study by Chen et al. (2015) employed only rotationally invariant electric potential (EP) which estimates the uniformity of the remaining gradient direction scheme (Jones et al., 1999). It has been shown, however, that error propagation in DTI derived parameters is anisotropic and rotationally variant, which can be assessed using another well-established estimate, condition number (CN) instead (Skare et al., 2000; Batchelor et al., 2003). The CN value can be perceived as an index of noise sensitivity of the given gradient scheme. In theory, the value of CN ranges from one (no error propagation) to infinity (infinite error propagation) whereas in practice the lower limit is approximately 1.6 (Batchelor et al., 2003).

This work was set out to obtain a thorough quantitative assessment of DT reliability after incremental DWI rejections by analyzing every sub-scheme of Jones30 gradient scheme. With this approach, we simulate all possible outcomes from DTI quality control tools that find and reject outliers before or during the tensor estimation. We reasoned that measuring the uniformity of gradient scheme using EP alone would not be enough, and an independent measure with ability to detect angle dependent deviations, such as CN (Skare et al., 2000; Batchelor et al., 2003), is needed to reveal error-prone sub-schemes with higher sensitivity. We then wanted to evaluate the practical effects of simulation results by studying how these metrics, EP and CN, would predict data deterioration in MRI datasets from adults and newborn subjects. We also tested the feasibility of our method on a different, clinically used gradient scheme with repeated acquisitions. Finally, the effects of DWI rejections were visualized in FA and V1 maps, as well as in a streamline tractography of one of the adult subjects.

2. Methods

The study consisted of two parts: simulation and MRI experiments, both starting with the construction of the same initial Jones30 gradient scheme. The workflow from data simulations to DWI rejections and tensor estimations is summarized in Fig. 1. With the simulation experiment, we aimed to examine the accuracy and precision of DTI derived FA and V1 values in every sub-scheme of the initial Jones30

scheme. Sub-schemes were collected based on two independent and conceptually distinct metrics: EP and CN. EP reflects the uniformity of the gradient direction distribution (Jones et al., 1999) and is a simplified Thomson problem (Thomson, 1904; Wales and Ulker, 2006) that quantifies the strength of repulsion force between electric charges when the charges are fixed on a spherical surface. In the DTI analysis, this can be used to find the most uniform distribution of the gradient vectors by spreading a given number of charges on a spherical surface and minimizing EP, i.e. maximizing the Euclidean distance between these charges. A mathematical property called CN is inherently included in DTI estimation procedure, and it gives relative error propagation boundaries for the ordinary linear least squares DT estimation (Skare et al., 2000; Batchelor et al., 2003). Datasets from the adult subjects were used to investigate the practical rejection effects, whereas the datasets from newborns were used to study the feasibility of our workflow on two different gradient schemes.

2.1. Gradient schemes and rotationally variant CN

The initial Jones30 scheme was generated by fixing the first gradient vector on the z -axis and minimizing EP of the system with a simplex search method (Lagarias et al., 1998). CN variations were considered around all three main axes by rotating the first gradient vector to 100 nearly equiangular points on a sphere (Fig. 2A) and applying the same rotation to the rest of the vectors. Visualizations of angular CN variations were made by interpolating areas between 100 points on the surface of the unit sphere (Fig. 2A) and then spreading the surface on azimuth – elevation map (Fig. 2B). Fig. 2C shows the level of CN variations as the function of incremental gradient vector rejections for each of four rejection criteria.

2.2. Rejections, sub-schemes, and tensor estimations

Enumeration of the sub-schemes was done in lexicographical order (Mifsud, 1963) with an in-house-developed Matlab (Mathworks, Natick, MA) program. Computation was parallelized by unranking every 100 000th combination (Buckles et al., 1977) and using those as entry points for CPUs. Computing was performed using the Alcyone cluster (138 Xeon X5650 and 4 Xeon X7550 processors) of the University of Helsinki Department of Physics, and the full workflow took approximately four days to complete. The full workflow is shown in Fig. 1. The sub-scheme refers to set of gradient directions and DWIs that are left after the rejection. For instance, rejecting one direction out of 30 can be done in 30 different ways i.e. one rejection results 30 different sub-schemes of 29 gradient directions and DWIs. Since each of these sub-schemes has rotationally invariant EP, the worst and the best EP equal to the largest and the smallest EP, respectively, found from these 30

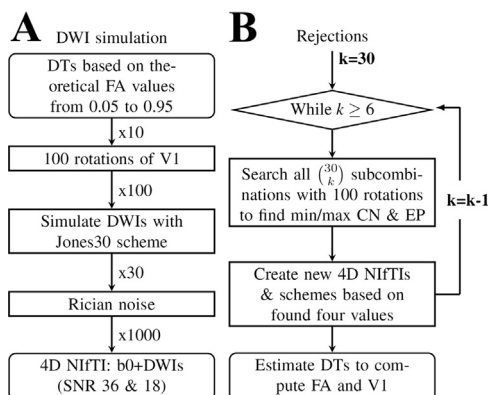


Fig. 1. Workflow of the study. Rejections were applied to the simulation and both MRI datasets independently.

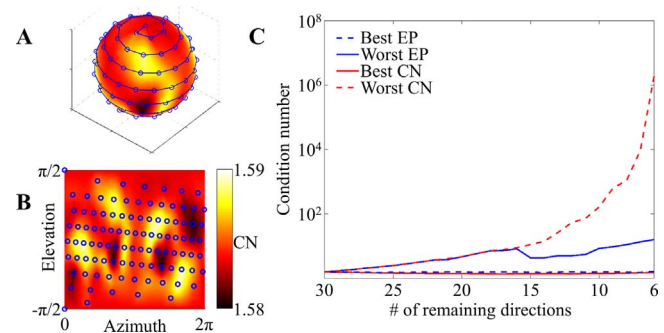


Fig. 2. Rotational variance of CN based on Jones30 scheme and its sub-schemes. (A) Angular CN variations were calculated and interpolated over the spherical surface. (B) The interpolated data was transformed into azimuth – elevation map to inspect minima and maxima areas. (C) CN of Jones30 sub-schemes as the function of remaining diffusion directions showed that CN as rejection criteria could detect up to five orders of magnitude larger relative error than EP in the worst case.

sub-schemes. However, CN is rotationally variant, so the worst and the best CN were calculated by rotating b-matrix in 100 different orientations over sphere (Fig. 2.), and the minimum and the maximum CN values were identified for each sub-scheme from this 30×100 table. For comparison of the two rejection criteria, the computational time per sub-scheme of CN using one CPU was approximately 80 times longer (3.5 ms) than the computation time of EP.

This rejection method was applied to the initial Jones30 scheme of the simulation part of this study and separately to the motion corrected schemes of the healthy adult subjects and two newborn groups referred as A and B with respect to different acquisition schemes described later. From all possible sub-schemes (N=1 073 567 387) for simulation and each subject, the final tensor estimations were performed using the ones with the smallest and the largest EP and CN values, hereafter called the best and the worst EP and CN schemes, respectively. DTs of these sub-schemes were estimated using the ordinary linear least squares algorithm as implemented in ExploreDTI (Leemans et al., 2009) and the diffusion indices (FA and V1) were exported as NIFTI images for further analysis. For all subjects, the binary brain mask of the initial case was used in tensor estimations for all sub-schemes.

2.3. Simulation experiment

Simulations were used to assess how DWI rejections lead to deviations in FA and V1 values compared with their known theoretical values, hereafter referred as FA error and V1 error, respectively. We simulated 1000 DWIs, each with 100 different tensor orientations V1 and ten equispaced FA values between 0.05 and 0.95 using two different b0 signal-to-noise ratio (SNR) values of 18 and 36 (altogether 2 million different DTs). Cylindrically symmetric DTs were simulated by computing the first eigenvalue λ_1 for each FA value using equation (1) which was derived from Eq. (2) (Basser and Pierpaoli, 1996) by setting $\lambda_2=\lambda_3$.

$$\lambda_1 = \frac{2 \cdot FA \cdot (FA \pm \sqrt{3 - 2 \cdot FA^2}) - 3}{3 \cdot (2 \cdot FA^2 - 3)} \cdot Tr, \begin{cases} - & \text{if } \lambda_1 \geq \lambda_2 \\ + & \text{otherwise} \end{cases} \quad (1)$$

$$FA = \frac{1}{\sqrt{2}} \frac{\sqrt{(\lambda_1 - \lambda_2)^2 + (\lambda_1 - \lambda_3)^2 + (\lambda_2 - \lambda_3)^2}}{\sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}} \quad (2)$$

The trace of the diffusion tensor Tr was set to $2.1 \cdot 10^{-3} \text{ mm}^2 \text{ s}^{-1}$ so the remaining eigenvalues λ_2 and λ_3 could be computed as difference between the trace and the first eigenvalue divided by two.

2.4. MRI experiments

Subject MRI data was used to investigate the consequences of DWI rejections on the FA and V1 values of adult and newborn subjects, as well as a streamline tractography of corticospinal tract (CST) of one adult subject (Catani, 2012). The use of subject images in this work was approved by the relevant Ethics Committee of the Helsinki University Hospital. The newborn subjects were imaged as a part of a larger clinical study (Stjerna et al., 2015) for which parents had given their informed consent.

2.4.1. Adult subject data acquisitions

Three adult subjects were scanned with a Siemens Verio 3 T MR scanner (Siemens, Erlangen, Germany) using a 12-channel head matrix coil. The full brain single shot echo planar imaging (SS-EPI) sequence was imaged twice with the following parameters: repetition time (TR) 13,000 ms, echo time (TE) 87 ms, field-of-view: 240 mm, voxel size $2 \times 2 \times 2 \text{ mm}^3$, b-value 1000 s/mm^2 , number of averages: 1, and GRAPPA factor: 2. Used direction scheme was study specific Jones30 with 7 b0 volumes in AP and 9 b0 volumes in PA phase encoding directions.

2.4.2. Newborn subject data acquisitions

Neonatal MRI data was acquired using Philips Achieva 1.5 T MR scanner (Philips Medical Systems, Best, The Netherlands) with an 8-channel phased array head coil. First, we scanned our main three subjects (group A) twice with an acquisition protocol which consisted of one b0 volume and DW volumes in 32 non-collinear directions which were provided by the vendor. TR was 5075 ms; TE: 60 ms, field-of-view: 171 mm, voxel size $1.79 \times 1.79 \times 2 \text{ mm}^3$, b-value 800 s/mm^2 , number of averages: 1, and SENSE factor: 2, and phase encoding direction AP along with three b0 volumes in the AP direction and four in the PA direction. Second, we obtained additional MRI dataset (group B) from our earlier clinical research dataset (Stjerna et al., 2015) that used a different acquisition scheme. The newborn B acquisition was a SS-EPI sequence with one b0 volume and DW volumes in 15 non-collinear directions and it was repeated three times to gather at least two intact sets. The diffusion gradient scheme was provided by the manufacturer. TR was the shortest possible; TE: 58 ms, field-of-view: 224 mm, voxel size $1.75 \times 1.75 \times 2 \text{ mm}^3$, b-value 700 s/mm^2 , number of averages: 1, and SENSE factor: 2. The lower number of gradient directions made also possible to study repeated acquisitions by splitting the dataset into two cases: B1) Rejections were applied consecutively to the two 15 DWI sets and B2) rejections were applied simultaneously to one 30 DWI set. This approach would not be possible using two 30 DWI sets due to computational weight of enumeration of sub-schemes from total of 60 DWIs.

2.4.3. Preprocessing of MRI sets

The visual inspection of the raw data of the adult subjects did not show notable motion or other artefacts. Susceptibility induced distortions were estimated using the FSL's Topup tool (Smith et al., 2004; Andersson, et al., 2003) on b0 volumes. Corrections were applied using the Eddy tool which corrected for subject motion as well as eddy currents. The parameter file from the Eddy tool was used to rotate the diffusion gradient vectors. The newborns A were analyzed using the same workflow as the adult subjects whereas for the newborn B motion and eddy current artefacts were corrected using the FSL's eddycorrect (Smith et al., 2004) algorithm and the resulting parameter file was used to rotate the gradient vectors accordingly. To make the comparison straightforward with the simulations, a subset of corrected DWIs were selected for further analysis: one b0 and 30 DWIs in non-collinear directions for both the adults and the newborns. The acquisition protocol of the newborns A had Philips specific gradient directions from which the ones directly on x- and y-axis were rejected to obtain a set with 30 DWIs.

2.5. Fractional anisotropy and the first eigenvector

In the simulation experiment, the FA error (mean \pm standard deviation) and the V1 error (median \pm interquartile range, IQR) were calculated as difference between the estimated value and the theoretical, known value for the selected sub-schemes. Three theoretical FA values: 0.85 (high), 0.55 (medium), and 0.25 (low) in SNR 36 simulation were selected for visualizations in this paper due to their rough correspondence to FA values of well-known anatomical structures (see also Chen et al., 2015): splenium of the corpus callosum (SPL), posterior limb of the internal capsule (PLIC), and putamen (PU), respectively (for all visualizations with both SNRs, please see Supplementary videos).

In MRI experiments, FA and V1 errors were calculated as differences between the initial data set, which was assumed to provide the most accurate estimate. Since our interest was concentrated on the underlying FA regions (high, medium, and low) instead of any specific anatomical brain region, we applied thresholds to the brains of all subjects based on their initial FA values. Used three threshold values for the adults were: FA values above 0.66 (high), between 0.33 and 0.66 (medium), and below 0.33 (low). The newborns had overall lower FA

value distribution and thresholds were shifted accordingly to: above 0.40 (high), between 0.20 and 0.40 (medium), and below 0.20 (low).

2.6. Tractography

We visualized the effects of the incremental rejections of DWIs on the tractography of one adult CST which was traced using the DTI streamline tractography as implemented in ExploreDTI (Leemans et al., 2009). An AND region of interest (ROI) was placed superior to the brainstem, two SEED ROIs were placed symmetrically superior to it, and a second AND ROI superior to SEED ROIs. ROI locations were based on a white matter atlas (Catani, 2012). Tracts leaving to anterior and posterior parts of the brain were excluded with NOT ROIs. The same ROIs were used in all sub-schemes. Used tracking parameters were: minimum FA for SEED point and tracking 0.20, maximum FA 1, maximum angle 30°, step size 1 mm, min/max length 100/200 mm,

and seed point supersampling factor was $1 \times 1 \times 1$.

3. Results

Our overall findings with both metrics showed increasing deterioration of DTI accuracy and precision when incremental DWI and gradient vector rejections were applied. However, CN based assessment was clearly more sensitive in disclosing the FA and V1 errors after higher number of rejections (> 15 out of 30).

3.1. Gradient schemes and initial CN variance

The descriptive EP and CN values of the initial gradient schemes were calculated to give a baseline for the rest of the sub-schemes. The EP minimum of the Jones30 scheme was found at 1544 V which was only 0.04% larger than 60 point charges on a unit sphere (Wales and

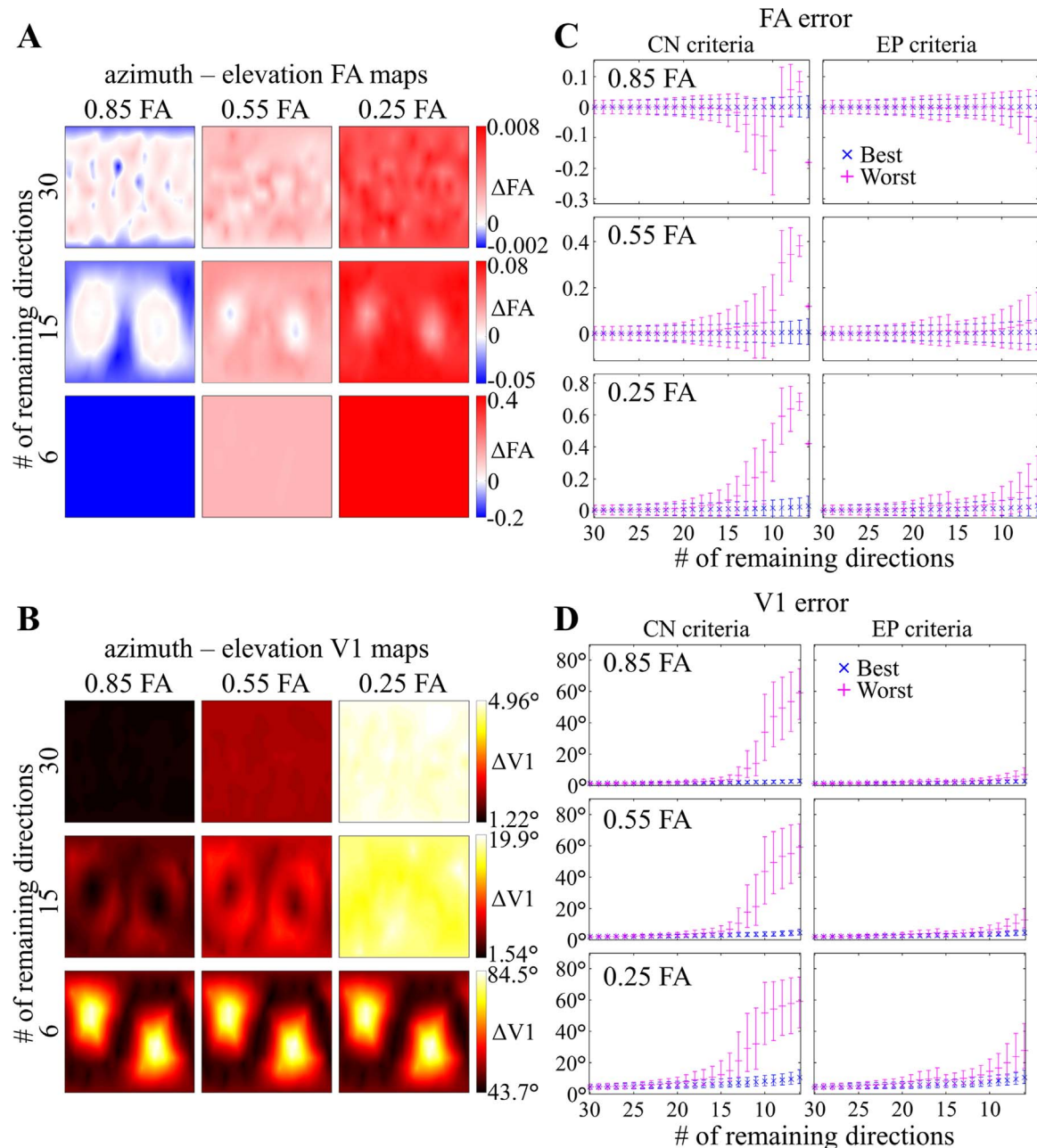


Fig. 3. Directional variance of simulated FA and V1 values and their dependency on the number of rejections. (A) Relative error in FA estimates. (B) Relative error in the V1 direction estimate. Comparison between the rejection criteria is presented in C and D. (C) mean $\pm 1\sigma$ FA values and (D) median \pm IQR V1 values calculated over all rotations.

Ulker, 2006). CN values varied between 1.58 and 1.59 when the encoding matrix was rotated to 100 nearly equiangular orientations (Fig. 2A), which was in a good agreement with previous studies (Skare et al., 2000; Batchelor et al., 2003). The rotational variance of the CN values of the initial scheme was further investigated with a similar approach but using 100,000 rotations of the encoding matrix. Increased sampling did not reveal higher variations, and both minima and maxima areas on the unit sphere remained in the same spatial locations. The EP minimum of the vendor (Philips) specific 30-direction gradient scheme used in the newborns A was 1657 V which were 7% larger than the optimal for 60 non-paired point charges (Wales and Ulker, 2006). CN of the scheme varied between 1.40 and 1.97. Respective results for the newborn B acquisition with 15-directions was 437 V which was 22% larger than the optimal for the 30 non-paired point charges (Wales and Ulker, 2006) and CN varied from 2.23 to 2.98.

3.2. Simulation experiment

FA and V1 errors were calculated for ten FA values between 0.05 and 0.95 in 100 different tensor orientations for two different SNR levels (18 and 36). A summary of the angle dependent FA errors of the worst CN case in the SNR 36 simulation is shown in Fig. 3A. An average over all orientations was computed to compare the level and the scale of these errors (Fig. 3C). The respective results for V1 errors are shown in Figs. 3B and 3D. All simulated situations had similar behavior: low and medium FA values were overestimated whereas high FA values could also be underestimated depending on the DT orientation. V1 errors in low FA simulations were greater than in high FA simulations. Both FA and V1 errors increased when more rejections were applied. All angle dependent FA and V1 errors for both SNRs are shown in Supplementary videos SV1–SV12.

Supplementary material related to this article can be found online at <http://dx.doi.org/10.1016/j.neuroimage.2016.11.061>.

3.2.1. Fractional anisotropy - simulation

Estimated FA values showed that the error is not orientationally homogeneous, and it strongly depends on the number of remaining directions after DWI rejections. The initial FA differences with the full Jones30 scheme varied between -0.002 and 0.008, after 15 rejections the variation ranged from -0.05 to 0.08, and after 24 rejections the variation ranged from -0.2 to 0.4. Importantly, the low and the medium FA values were only overestimated (red areas in Fig. 3A), whereas the high FA values could be either over- or underestimated (blue areas in Fig. 3A).

The orientational variance was averaged to allow two-dimensional comparison between EP and CN rejection criteria (Fig. 3C). The FA error (mean \pm 1 σ) of the best sub-schemes showed a comparable, moderate decrease in the accuracy and the precision in simulations with both low and medium actual FAs. However, the accuracy in the high FA simulation remained unchanged though the precision decreased. Results were similar to the SNR 18 simulation with higher

deviations. Difference of the initial scheme and best cases with maximum amount of rejections is shown in Table 1.

When the worst cases were investigated, the CN criteria clearly outperformed the EP criteria. This can be seen in the scale of FA errors in Fig. 3C, for example. To give a more precise illustration of differences between two criteria, an example with 15 rejections out of 30 is shown in Table 1 for both worst cases: In the high SNR simulation, the worst CN case showed five times larger error in accuracy (0.10), and two times larger error in precision (0.10) for the low FA than the worst EP. For the low SNR simulation, the differences were smaller. Another notable finding was that the CN criteria showed more negative errors for high FA simulations which is intuitively expected behavior when FA approaches its maximum value of one. More detailed information about directional differences in these FA errors are shown in the supplementary videos SV01–SV06.

3.2.2. The first eigenvector - simulation

V1 errors were angle dependent in a manner comparable with the FA errors described above. The initial scheme varied between 1.22° (the high FA) and 4.96° (the low FA), with 15 directions variation was from 1.54° (the high FA) to 19.9° (the low FA) and with 6 directions from 43.7° to 84.5° with no notable difference between low, medium, and high FA values (Fig. 3B). The V1 error (median \pm IQR) over orientations showed no difference between the best cases but CN criteria outperformed EP when the worst cases were studied (Fig. 3D). To illustrate this, the same example with 24 rejections for the best cases and 15 rejections for the worst cases as in FA - simulation is shown in Table 2. The directional differences of V1 errors can be seen in Supplementary videos SV07–SV12.

3.3. MRI experiments

The simulation findings above indicated that errors in FA and V1 estimates due to DWI rejections are sensitive to the actual level of anisotropy in the brain tissue. This implies that DWI rejections would affect data analysis in a spatially and orientationally variable manner. To analyze this effect on MRI data, we computed the FA and V1 errors with respect to the increasing numbers of rejections in all adult and newborn MRI datasets.

3.3.1. FA errors in the adult MRIs

The effect of incremental DWI rejections on the estimated FA values of adult MRI data (Fig. 4D and Supplementary Figs. S3A–S3C) were essentially similar to the findings in our FA simulations (Fig. 3C): in the low FA regions the loss of precision, i.e. higher standard deviation values and the loss of accuracy were greater than in the higher FA regions. The best CN and EP criteria resulted in the smallest errors, while the worst CN and EP criteria resulted in the largest errors. Moreover, there were no notable differences between the worst CN and EP criteria when less than half of the directions were rejected (Supplementary Figs. S3A–S3C).

The deterioration of FA estimation and its spatial variability in

Table 1
Summary of FA simulations.

Scheme	FA error					
	SNR 36			SNR 18		
	Low FA	Medium FA	High FA	Low FA	Medium FA	High FA
Initial (no rejections)	0.01 \pm 0.03	0.00 \pm 0.03	0.00 \pm 0.02	0.02 \pm 0.06	0.01 \pm 0.06	0.00 \pm 0.04
The best CN (24 rejections)	0.03 \pm 0.06	0.01 \pm 0.06	0.00 \pm 0.04	0.11 \pm 0.11	0.04 \pm 0.10	0.00 \pm 0.06
The best EP (24 rejections)	0.03 \pm 0.06	0.01 \pm 0.06	0.00 \pm 0.04	0.11 \pm 0.10	0.04 \pm 0.10	0.00 \pm 0.06
The worst CN (15 rejections)	0.10 \pm 0.10	0.02 \pm 0.08	-0.01 \pm 0.05	0.16 \pm 0.14	0.04 \pm 0.11	-0.04 \pm 0.09
The worst EP (15 rejections)	0.02 \pm 0.05	0.01 \pm 0.05	0.00 \pm 0.03	0.10 \pm 0.10	0.03 \pm 0.09	0.00 \pm 0.06

Table 2
Summary of V1 simulations.

Scheme	V1 error					
	SNR 36			SNR 18		
	Low FA	Medium FA	High FA	Low FA	Medium FA	High FA
Initial (no rejections)	5 ± 4°	2.0 ± 1.5°	1.2 ± 0.9°	9 ± 7°	4 ± 3°	2 ± 2°
The best CN (24 rejections)	11 ± 9°	5 ± 4°	3 ± 2°	21 ± 20°	9 ± 8°	5 ± 4°
The best EP (24 rejections)	10 ± 9°	5 ± 4°	3 ± 2°	21 ± 20°	9 ± 8°	5 ± 4°
The worst CN (15 rejections)	12 ± 14°	5 ± 6°	3 ± 4°	20 ± 30°	11 ± 12°	6 ± 7°
The worst EP (15 rejections)	8 ± 7°	4 ± 3°	2 ± 2°	17 ± 15°	7 ± 6°	4 ± 4°

different FA regions is visualized in the color encoded FA slices (Fig. 4A) and in the relative error maps (Fig. 4B) for four and eight rejections. Such relative error maps can be intuitively perceived as the maps of uncertainty for DTI estimates after rejections. Notably, these maps should not be confused with uncertainty maps gained using bootstrapping methods (Heim et al., 2004). The overall observation is perfectly in line with the simulations presented above and previous studies (Chen et al., 2015): Regions with high FA levels, e.g. the commonly analyzed white matter skeleton, showed markedly lower errors, while the error magnitude increases towards the cortex and in the deep nuclei. For example, after the 8th rejection the relative error increased up to 300% in the low FA regions whereas in the high FA regions remained stable.

3.3.2. V1 error in the adult MRIs

The V1 errors of the adult MRI (Fig. 4E) increased rapidly with the rejections, which were more pronounced than the observation with simulated V1 errors (Fig. 3D). There was no notable difference between the best CN vs EP rejection criteria. However, when more than half of the directions were rejected the worst EP criteria started to underestimate the error (Supplementary Figs. S3D–S3F). The V1 error maps (Fig. 4C) show that the errors were overall lower in the high FA regions than elsewhere in the brain.

3.3.3. Tractography - adult

Both the above simulations and the MRI experiments indicated that V1 is rapidly affected by the number of rejected directions. It would hence be expected that the quality, or reliability, of tractography results may substantially suffer from rejections. To visualize this effect, we traced the CST in one of our adult MRI dataset repeatedly after incremental the rejections of DWIs (Fig. 5). We modified the original idea of Leemans and Jones (2008), in which the tensor residual was used to color code the tracts and used the apex angle of the cone of uncertainty (CU) instead. This approach was chosen, since the residual would not produce meaningful results in the case of DWI rejections due to the rapid decrease of the sum of squares of the residual vector. A visually salient loss of tracts could be readily seen after five DWI rejections, and the loss was the most pronounced at the upper end where the tracts approach the cortical layers. We also examined how the total number of identified tracts in the CST (Fig. 5D) would change as the function of DWI rejections. Even in the best cases, the tracts started to decline after four or more rejections.

3.3.4. FA errors in the newborn MRIs

The results of the main newborns in the group A (Supplementary Figs. S5A–S5C) with the vendor specific 30-direction scheme were largely comparable with the adult FA errors (Fig. 4D and Supplementary Figs. S3A–S3C) as well as simulations (Fig. 3). However, the difference between the rejection criteria was notably small for the worst cases as well. For the additional newborn (case B1), we first rejected data incrementally from one set of 15 DWIs and then continued rejections on the second set of 15 DWIs. This showed an

expected loss of precision, i.e. higher standard deviation values, and a small incremental loss of accuracy (Supplementary Figs. S1A and S4B) in FA estimations. Second, we examined a common real life scenario where rejections maybe applied to both sets of DWIs (newborn case B2) which resulted in a markedly worse FA errors. The estimated FA values also behaved in a way not predicted by the simulation, which was only detected when using the worst CN criteria (Supplementary Figs. S2D and S4A): In the low FA regions that comprise much of the newborn MRI, both the accuracy (mean) and the precision (standard deviation) started to increase immediately when rejections were applied and in the regions with medium FA this effect was somewhat milder. The accuracy and precision of the high FA remained until the sixth rejection was applied, then both values showed rapid deterioration. These effects were not observed when inspected using the worst EP criteria. Color encoded FA slices and FA error maps for both rejection methods are shown in Supplementary Figs. S2A and S2B.

3.3.5. V1 errors in the newborn MRIs

The V1 error showed a decrease in both the precision and the accuracy with the increasing number of rejections applied for both newborns groups. The low FA regions deteriorated faster in each case than regions with higher FA values (The Supplementary Figs. S2E, S4C, S4D, and S5D–S5F). For the newborns (group A) the V1 error results were similar to the simulations (Fig. 3) and adult results (Fig. 4 and Supplementary Fig. S3). For the additional newborn (group B), the spatial variability of V1 values was stronger than what we found with FA values (Supplementary Fig. S2C), and V1 errors were found to increase up to 85° after 8th rejection, even in brain regions with high FA.

4. Discussion

Our study shows how the accuracy and the precision of DT derived parameters FA and V1 were affected when DWI and diffusion gradient vectors were rejected. Such rejections are usually required before DT estimation due to subject movements during MRI acquisition. A previous work (Chen et al., 2015) studied the same issue using only the EP criteria and selected sub-schemes of the initial gradient scheme (Landman et al., 2007; Jones et al., 1999). While our findings are fully compatible with that work, we extended those findings by computing every possible sub-scheme instead of specific sub-groups, and by applying a novel metric, CN, which allows the assessment of rotational variant error propagation during DT estimation (Skare et al., 2000; Batchelor et al., 2003).

We showed here that the rotationally variant CN (Batchelor et al., 2003) can reveal error-prone sub-schemes with higher sensitivity than the EP criteria. Also, the FA and V1 error estimates obtained with the CN criteria were always clearer than error estimates obtained with the EP criteria. This suggested that the use of the EP criteria is likely to lead to underestimation of uncertainty in DT derived parameters. In addition, we showed, using rotational variant CN, that error propagation depends on the actual orientations. It should be noted that these

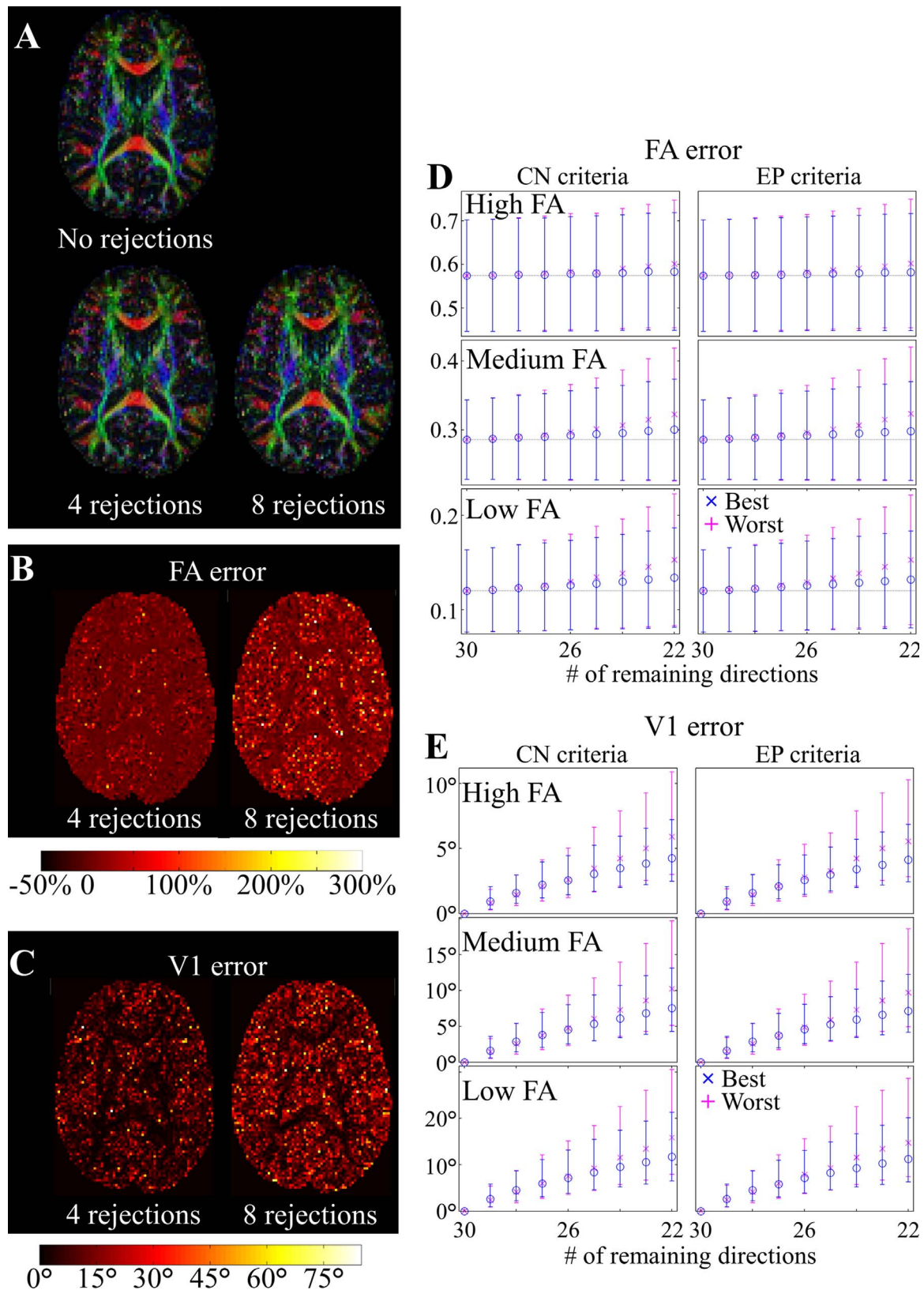


Fig. 4. Measured FA and V1 variations of one adult subject as function of incremental rejections. (A) Color encoded FA maps, (B) FA error maps, and (C) V1 error maps. (D) and (E) showed that with eight or less rejections there is no notable difference between rejection criteria with the Jones30 scheme.

results are specific to given gradient schemes, and they should be calculated only after the subject motion correction is applied to the gradient vectors. Thus, even with the same acquisition protocol, each subject or repetition of the same subject would have a unique CN map

and relative errors, respectively. This precludes a general conclusion about the exact number of rejections that would be allowed within the given limits of FA and V1 reliability. However, the converse of this is that CN offers an individual level, voxel-wise estimate for DT reliability,

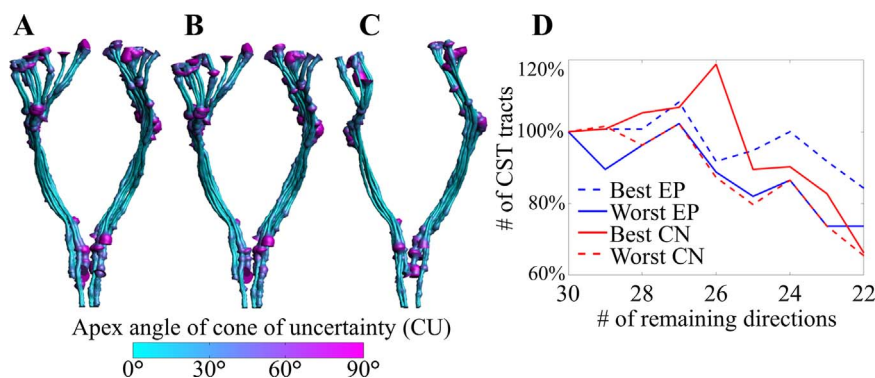


Fig. 5. Impact of incremental rejections on deterministic tractography of one adult subject. (A) 133 tracts were found in the CST when the full adult dataset was used. Colorscale and tract width correlate with the voxel-wise apex angle of CU: blue narrow tracts being the most certain and red wide tracts the most uncertain. (B) CU values are drawn from a set with 5 rejections but the visualized tracts are still the original. This is done to show that the CU alone does not necessarily reveal error prone regions. (C) Both the tracing and CU values are drawn from set with 5 rejections to show the real deterioration of tractography. (D) Relative change in the number of identified tracts as a function of rejections shows rapid decline in all cases.

including visualization to evaluate how different brain regions are affected by rejections (see Figs. 3, 4, and Supplementary Fig. S2).

It would be attractive to define universal limits for an acceptable CN values. However, the CN is specific to each gradient scheme, hence it needs to be assessed separately for each scheme after subject motion corrections. The threshold for acceptable CN levels is ultimately based on expert judgment that seeks to optimize the practical need relative to the unseen voxel-wise error indicated by the CN value. Notably, calculating voxel-wise CN from any existing DT data takes only few milliseconds, and each MRI center can readily build experience in assessing their own schemes and data.

The main limitation of the hereby used CN method relates to the way of estimating DT: CN is only exact for a linear DT estimation (Batchelor et al., 2003), and such exact CN property does not exist for non-linear DT estimation algorithms. However, this limitation is relatively less important when compared with the hereby shown limitations of the EP criteria: the worst schemes based on EP might not have monotonically increasing CN (Fig. 2C), hence the estimated FA and V1 errors with the EP criteria might even artefactually decrease when more rejections are applied (Figs. 2C, 3C, and Supplementary Fig. S3). Interestingly, this was not observed with either vendor specific gradient schemes (Supplementary Figs. S4 and S5) which suggests that distinct gradient schemes have different tolerances for error propagation.

The behavior of FA estimations due incremental rejections was highly comparable between MRI datasets from the adults and the main newborns (group A), and they closely followed our simulation findings. However, the rapid increase in V1 errors with incremental DWI rejections was observed in all MRI datasets but not in our simulations. Visualization based on voxel-wise cone of uncertainty values (Leemans and Jones, 2008) showed how the tractography tracing results deteriorated with incremental rejections, and the effect was most pronounced in the peripheral fiber branches. Such regional variance in the reliability of tractography becomes particularly relevant when studying cortico-cortical connectivity (Gong et al., 2009). Another clinically interesting question was how data with repeated acquisitions would deteriorate with incremental rejections, since such datasets are widely available from past clinical trials. This was assessed with the additional newborn (group B) experiment and showed that FA error became progressively worse when more than five out of the total thirty directions were rejected (Supplementary Fig. S2). Notably, this deviation, or loss of reliability, was not detected with the EP criteria. The performance of CN and EP was more comparable when assessing the V1 errors, however the CN method was again able to detect higher median deviations as well as larger variability.

A further application of the hereby used CN method is in situations where rejections are applied together with robust DT estimation

algorithms (Chang et al., 2005; Tax et al., 2015) that support exporting an *outlier* map. The *outlier* map is a binary volume with the same dimensions as the original MRI data but it specifies rejections from DT estimation at voxel level. In such case, CN variations should be studied in voxel-wise manner as well. Orientational variations in CN could lead into voxel-wise variations in tensor derived parameters, and this could be difficult to control in group analyses such as tract based spatial statistics (TBSS) (Smith et al., 2004; Smith et al. 2006) or even in simple ROI based analysis. For example, studies based on group comparison with healthy subjects may be confounded by different amount of movement, because healthy subjects may move less during the DWI acquisition. This could lead to situations in which error propagation in each DT orientation and voxel varies between the groups. Computation of voxel-wise CN variations could disclose this, hence holding significant promise for improving the reliability and interpretation of DTI analyses. Insufficient means or mere absence of quality control to avoid artefactual data in DTI (Jones and Cercignani, 2010) has recently raised concerns among clinical research community (Heemskerk et al. 2013; Pieterman et al., 2015; Roalf et al., 2016), and it is obvious that such procedures will need to become standard components of future guidelines of evidence based clinical research.

Author contributions

VS designed and carried out simulations and data analysis and wrote the paper. LK contributed to the study design and analysis. SO, SS, and SV contributed to study design and wrote the paper.

Disclosures

The authors have no conflicts of interest to disclose.

Acknowledgements

This work was supported financially by Helsinki University Hospital, Juselius Foundation, Finnish Cultural Foundation, the Academy of Finland Grant (#288220 and 276523).

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.neuroimage.2016.11.061>.

References

Andersson, J.L., Skare, S., Ashburner, J., 2003. How to correct susceptibility distortions

- in spin-echo echo-planar images: application to diffusion tensor imaging. *NeuroImage* 20 (2), 870–888.
- Basser, J., Pierpaoli, C., 1996. Microstructural and physiological features of tissues elucidated by quantitative diffusion tensor MRI. *J. Magn. Reson. B* 111, 209–219.
- Basser, P.J., Mattiello, J., LeBihan, D., 1994. Estimation of the effective self-diffusion tensor from the NMR spin echo. *J. Magn. Reson. Ser. B* 103 (3), 247–254.
- Basser, P.J., Pajevic, S., Pierpaoli, C., Duda, J., Aldroubi, A., 2000. In vivo fiber tractography using DT-MRI data. *Magn. Reson. Med.* 44 (4), 625–632.
- Batchelor, P.G., et al., 2003. Anisotropic noise propagation in diffusion tensor MRI sampling schemes. *Magn. Reson. Med.* 49 (6), 1143–1151.
- Bill P., Buckles, M. Lybanon. "Algorithm 515: Generation of a vector from the lexicographical index [G6]." *ACM Transactions on Mathematical Software (TOMS)* 3. 2 (1977): 180-182.
- Catani, Marco, de Schotten, Michel Thiebaut, 2012. *Atlas of Human Brain Connections*. Oxford University Press, Oxford, UK.
- Chang, Lin-Ching, Derek K. Jones, Pierpaoli, Carlo, 2005. RESTORE: robust estimation of tensors by outlier rejection. *Magn. Reson. Med.* 53 (5), 1088–1095.
- Chen, Yiran, et al., 2015. Effects of rejecting diffusion directions on tensor-derived parameters. *NeuroImage*.
- Dubois, Jessica, et al., 2006. Optimized diffusion gradient orientation schemes for corrupted clinical DTI data sets. *Magn. Reson. Mater. Phys. Biol. Med.* 19 (3), 134–143.
- Mahshid, Farzinfar et al., 2013. DTI quality control assessment via error estimation from Monte Carlo simulations. In: *Proceedings of the International Society for Optics and Photonics, SPIE Medical Imaging*.
- Gong, G., et al., 2009. Mapping anatomical connectivity patterns of human cerebral cortex using in vivo diffusion tensor imaging tractography. *Cereb. cortex* 19 (3), 524–536.
- Heemskerk, A.M., Leemans, A., Plaisier, A., Pieterman, K., Lequin, M.H., and Dudink Acquisition, J., 2013. Guidelines and Quality Assessment Tools for Analyzing Neonatal Diffusion Tensor MRI Data *AJNR Am. J. Neuroradiol.* 34. pp.1496–1505 originally published online on March 21, <http://dx.doi.org/10.3174/ajnr.A3465>.
- Heim, S., et al., 2004. Assessing DTI data quality using bootstrap analysis. *Magn. Reson. Med.* 52 (3), 582–589.
- Horsfield, M.A., Jones, D.K., 2002. Applications of diffusion-weighted and diffusion tensor MRI to white matter diseases—a review. *NMR Biomed.* 15 (7-8), 570–577.
- Jiang, S., Xue, H., Counsell, S., Anjari, M., Allsop, J., Rutherford, M., Hajnal, J.V., 2009. Diffusion tensor imaging (DTI) of the brain in moving subjects: application to in-utero fetal and ex-utero studies. *Magn. Reson. Med.* 62 (3), 645–655.
- Jones, D.K., Horsfield, M.A., Simmons, A., 1999. Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging. *Magn. Reson. Med.* 42.
- Jones, Derek K., Cercignani, Mara, 2010. Twenty-five pitfalls in the analysis of diffusion MRI data. *NMR Biomed.* 23 (7), 803–820.
- Lagarias, J.C., Reeds, J.A., Wright, M.H., Wright, P.E., 1998. Convergence properties of the Nelder-Mead simplex method in low dimensions. *SIAM J. Optim.* 9 (1), 112–147.
- Landman, Bennett A., et al., 2007. Effects of diffusion weighting schemes on the reproducibility of DTI-derived fractional anisotropy, mean diffusivity, and principal eigenvector measurements at 1.5 T. *Neuroimage* 36 (4), 1123–1138.
- Leemans, A., Jones, D.K., 2008. A new fiber tract color-encoding scheme based on diffusion tensor model residuals. In: *Proceedings of the 16th Scientific Meeting, International Society for Magnetic Resonance in Medicine*. p. 3309.
- Leemans, A. et al., 2009. ExploreDTI: a graphical toolbox for processing, analyzing, and visualizing diffusion MR data. In: *Proceedings of the 17th Annual Meeting of International Society for Magnetic Resonance in Medicine*.
- Leemans, Alexander, Jones, Derek K., 2009. The B-matrix must be rotated when correcting for subject motion in DTI data. *Magn. Reson. Med.* 61 (6), 1336–1349.
- Lerner, A., Mogensen, M.A., Kim, P.E., Shiroishi, M.S., Hwang, D.H., Law, M., 2014. Clinical applications of diffusion tensor imaging. *World Neurosurg.* 82 (1), 96–109.
- Charles J. Mifsud , "Algorithm 154: combination in lexicographical order." *Communications of the ACM* 6.3 (1963): 103. 2014.
- Oguz, Ipek, et al., 2014. DTIPrep: quality control of diffusion-weighted images. *Front. Neuroinformatics* 8 (4).
- Pierpaoli, C., Jezzard, P., Basser, P.J., Barnett, A., Di Chiro, G., 1996. Diffusion tensor MR imaging of the human brain. *Radiology* 201 (3), 637–648.
- Pieterman, Kay, et al., 2015. Data quality in diffusion tensor imaging studies of the preterm brain: a systematic review. *Pediatr. Radiol.* 45 (9), 1372–1381.
- RiordanJ., 2015. *An Introduction To Combinatorial Analysis [e-book]*. Princeton, New Jersey: Princeton University Press; 1980. Available from: eBook Collection (EBSCOhost), Ipswich, MA.(accessed 26.10.16).
- Roalf, David R., et al., 2016. The impact of quality assurance assessment on diffusion tensor imaging outcomes in a large-scale population-based cohort. *NeuroImage* 125, 903–919.
- Rohde, G.K., Barnett, A.S., Basser, P.J., Marengo, S., Pierpaoli, C., 2004. Comprehensive approach for correction of motion and distortion in diffusion-weighted MRI. *Magn. Reson. Med.* 51 (1), 103–114.
- Skare, Stefan, et al., 2000. Condition number as a measure of noise performance of diffusion tensor data acquisition schemes with MRI. *J. Magn. Reson.* 147 (2), 340–352.
- Smith, Smith S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E.J., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., Flitney, D.E., Niazy, R., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J.M., Matthews, P.M., 2004. Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage* 23 (S1), 208–219.
- Smith, Smith S.M., Jenkinson, M., Johansen-Berg, H., Rueckert, D., Nichols, T.E., Mackay, C.E., Watkins, K.E., Ciccarelli, O., Cader, M.Z., Matthews, P.M., Behrens, T.E.J., 2006. Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. *NeuroImage* 31, 1487–1505.
- Stjerna, et al., Susanna, 2015. Visual fixation in human newborns correlates with extensive white matter networks and predicts long-term neurocognitive development. *J. Neurosci.* 35.12, 4824–4829.
- Tax, Chantal M.W., et al., 2015. REKINDLE: robust extraction of kurtosis INDices with linear estimation. *Magn. Reson. Med.* 73 (2), 794–808.
- Thomson, Joseph John, 1904. XXIV. On the structure of the atom: an investigation of the stability and periods of oscillation of a number of corpuscles arranged at equal intervals around the circumference of a circle; with application of the results to the theory of atomic structure. *Lond. Edinb. Dublin Philos. Mag. J. Sci.* 7 (39), 237–265.
- Wales, David J., Ulker, Sidika, 2006. Structure and dynamics of spherical crystals characterized for the Thomson problem. *Phys. Rev. B* 74 (21), 212101.