Constrained Maximum Likelihood Estimation of the Diffusion Kurtosis Tensor Using a Rician Noise Model

Jelle Veraart,[1]* Wim Van Hecke,[1,2,3] and Jan Sijbers[1]

A computational framework to obtain an accurate quantification of the Gaussian and non-Gaussian component of water molecules' diffusion through brain tissues with diffusion kurtosis imaging, is presented. The diffusion kurtosis imaging model quantifies the kurtosis, the degree of non-Gaussianity, on a direction dependent basis, constituting a higher order diffusion kurtosis tensor, which is estimated in addition to the well-known diffusion tensor. To reconcile with the physical phenomenon of molecular diffusion, both tensor estimates should lie within a physically acceptable range. Otherwise, clinically and artificially significant changes in diffusion (kurtosis) parameters might be confounded. To guarantee physical relevance, we here suggest to estimate both diffusional tensors by maximizing the joint likelihood function of all Rician distributed diffusion weighted images given the diffusion kurtosis imaging model while imposing a set of nonlinear constraints. As shown in this study, correctly accounting for the Rician noise structure is necessary to avoid significant overestimation of the kurtosis values. The performance of the constrained estimator was evaluated and compared to more commonly used strategies during simulations. Human brain data were used to emphasize the need for constrained estimators as not imposing the constraints give rise to constraint violations in about 70% of the brain voxels. Magn Reson Med 66:678–686, 2011. © 2011 Wiley-Liss, Inc.

Key words: DKI; maximum likelihood; Rician noise; parameter estimation; constraint

INTRODUCTION

Diffusion tensor magnetic resonance imaging (DT-MRI, or DTI) is an in vivo and noninvasive imaging modality that can provide insight into the architecture of the white matter (WM) anatomy (1). In DTI, the water molecules' diffusion through brain structures is assumed to be Gaussian and, as such, can be described by a 2nd order, 3D diffusion tensor (DT) (1,2). In biological tissues, however, the displacement probability distribution significantly deviates from a Gaussian distribution due to various factors, such as cerebral perfusion, restricted diffusion, membrane permeability, and extra- and intracellular water compartments (3–5). As a result, the DTI model fails to provide an accurate quantification of the true diffusion process (6).

Recently, diffusion kurtosis imaging (DKI), a higher order diffusion model, was introduced as a straightforward extension of the DTI model to approximate the diffusion weighted signal attenuation more accurately (6–8). In DKI, the Gaussian and non-Gaussian diffusion are quantified by the apparent diffusion coefficient (D_{APP}) and the apparent excess kurtosis (K_{APP}), respectively, on a direction-dependent basis (7–10). After applying diffusion weighting along 15 noncollinear and noncoplanar gradient directions with one or two nonzero b-values such that a total of 22 diffusion-weighted images (DWIs) are acquired, a 4th order 3D, fully symmetric tensor—the diffusion kurtosis tensor (DKT)—can be estimated in addition to the DT. From both diffusional tensors, one can derive several scalar measures such as fractional anisotropy (FA), radial- ($D_\perp$), axial- ($D_\parallel$) and mean diffusivity (MD), as well as radial- ($K_\perp$), axial- ($K_\parallel$), and mean kurtosis (MK). The recently proposed kurtosis measures have been shown to be sensitive biomarkers for the assessment of pathological or aging changes (11–14). Obviously, the interpretation and comparison of these measures are affected by the accuracy with which the diffusional tensors can be estimated.

As diffusion of water molecules is a physical property of the tissue being measured, diffusional tensor estimates must be physically meaningful. Unfortunately, conventionally estimated tensors tend to deviate from their true underlying values. In many cases, they even converge to physically irrelevant tensor values. Therefore, more false positive and false negative results might occur in group studies. The observed estimation error is caused by the following.

- Most of the current methods to estimate the DT, or the DKT, do not fully account for the noise distribution in magnetic resonance (MR) images. The noise in magnitude MR images is characterized by a Rician distribution, which stems from the nonlinear operation needed to compute the magnitude of complex MR data with real and imaginary parts corrupted by Gaussian noise (e.g., 15–17). However, the most widely used tensor estimation method, (weighted) least squares estimation, assumes the noise to be independently Gaussian distributed. An estimator based on a Gaussian noise model will result in biased tensor estimates as the mean of the Rician distributed data exceeds the underlying noise free signal magnitude.

- Due to the presence of noise and other imaging artifacts, e.g., Gibbs ringing, on the diffusion weighted images (DWIs), unconstrained estimations of diffusion and kurtosis values might lie outside a physically

[1]Vision Lab, Department of Physics, University of Antwerp, Wilrijk (Antwerp), Belgium
[2]Department of Radiology, University Hospital Antwerp, University of Antwerp, Edegem (Antwerp), Belgium
[3]Department of Radiology, University Hospitals of the Catholic University of Leuven, Leuven, Belgium

*Correspondence to: Jelle Veraart, M.S., Vision Lab, Department of Physics, University of Antwerp, Wilrijk (Antwerp), Belgium. E-mail: Jelle.Veraart@ua.ac.be

acceptable range. To be physically meaningful, the following constraints should be satisfied when estimating either of the diffusional tensors:

1. the positive definite constraint on the DT, i.e., the requirement that every eigenvalue of the DT estimate should be positive;
2. the lower bound constraint on K_{APP}, i.e., $K_{\mathrm{APP}} \geq -2$;
3. the upper bound constraint on K_{APP}, i.e., $K_{\mathrm{APP}} \leq \frac{3}{D_{\mathrm{APP}}b}$, with b the diffusion weighting strength.

The lower bound on K_{APP} as defined in #2 is a theoretical lower bound (18). In practice, however, one often increases the lower bound to 0 as it is in agreement with the compartment tissue model of diffusion, which predicts positive kurtosis (7,9). The upper bound on the K_{APP} (#3) should not be exceeded to guarantee the DKI model function to decrease with the b-value in the range of acquired b-values (9,19).

Improving the plausibility of DKI estimators by imposing constraints has previously been studied. Ardekani et al. imposed positive-definiteness on both tensors by rewriting the 4th order, 3D DKT as a 2nd order, 6D symmetric tensor with 15 unique elements (20). The Cholesky parameterization enables imposing positivity on $\boldsymbol{D}$ as well as on $\boldsymbol{K}$ to obtain, respectively, positive D_{APP} and positive K_{APP} along each direction. The method, however, lacks an upper bound on the kurtosis values as a result of which a monotonically decreasing DKI model function is not guaranteed. Tabesh et al. proposed to slightly rewrite the model equation such that the estimation problem became a special case of convex quadratic programming (19). Local minima are avoided as a convex cost function is minimized. Although the algorithm imposes all necessary constraints, it accepts inherently a Gaussian noise model, and thus the kurtosis parameters are biased (6). In this work, however, we present a constrained computational framework in which accurate estimates of both diffusional tensors are obtained by maximizing the joint likelihood function of all Rician distributed observations given the DKI model. Based on simulation and empirical experiments, we evaluate the performance of, and the need for constrained estimators that properly account for Rician distributed diffusion weighted data during DKI analyses.

METHODS

Data Acquisition

Diffusion weighted data of a healthy 25-year-old male volunteer were acquired on a Trio Scanner (3T; Siemens AG, Siemens Medical Solutions, Erlangen, Germany) with a single-channel head coil. A multi-slice, single-shot EPI, spin echo sequence (TR/TE = 7700/139 ms) was used to obtain 40 axial slices without slice gap and 2.2 mm nominal isotropic resolution (FOV = 220 × 220 mm). Diffusion weighting was applied according to an optimized diffusion gradient encoding scheme that consists out of 25, 40, and 75 diffusion weighted gradients, isotropically distributed over three shells with $b = 700, 1000, 2800\,\mathrm{s/mm^2}$, respectively (21). In addition, 10 nondiffusion weighted images (b_0) were acquired. The total scan time was 19 min. The

data was corrected, involving the b-matrix rotation (22), for eddy currents and subject motion using global affine transformations.

Diffusion Kurtosis Imaging

The natural logarithm of the diffusion weighted signal can be rewritten in function of the even order cumulants of the probability distribution function that describes the self-diffusion of protons (23); The DKI model corresponds to the fourth order cumulant expansion (7). In DKI, the diffusion weighted signal, $\mathcal{S}(b, \boldsymbol{g})$, is written in function of the b-value and the gradient direction $\boldsymbol{g}$ as:

$$S_{dki}(b, \boldsymbol{g}; \boldsymbol{\theta}) = S(0) \exp\left(-b \sum_{i,j=1}^{3} g_i g_j D_{ij} \right.$$
$$\left. + \frac{b^2}{6} \left(\sum_{i=1}^{3} \frac{D_{ii}}{3} \right)^2 \sum_{i,j,k,l=1}^{3} g_i g_j g_k g_l W_{ijkl} \right), \quad [1]$$

with g_i the ith component of $\boldsymbol{g}$ and $S(0)$ the nondiffusion weighted signal intensity. D_{ij} is the ijth element of the fully symmetric DT $\boldsymbol{D}$, characterized by 6 independent elements: $\boldsymbol{\theta}_{\mathrm{D}} = \{D_{ij}\}_{i \leq j \leq 3} = \{D_{11}, D_{12}, D_{22}, D_{13}, D_{23}, D_{33}\}$, and W_{ijkl} denotes the $ijkl$th element of the fully symmetric diffusion kurtosis tensor $\boldsymbol{W}$; $\boldsymbol{W}$ is fully parameterized by 15 independent elements $\boldsymbol{\theta}_{\mathrm{K}} = \{W_{ijkl}\}_{i \leq j \leq k \leq l \leq 3}$. The DKI model is in total parameterized by $\boldsymbol{\theta}$, consisting out of 22 elements: $[S(0), \boldsymbol{\theta}_{\mathrm{D}}, \boldsymbol{\theta}_{\mathrm{K}}]$.

Constraints

The following constraints should be satisfied to guarantee physical relevance of either of the diffusional tensors:

1. The positive definite constraint on $\boldsymbol{D}$: Each eigenvalue of the estimated $\boldsymbol{D}$ should be positive as it reconciles with the physical phenomenon of molecular diffusion. In this study, the constraint is imposed by representing the DT by its Cholesky decomposition during tensor estimation (24):

$$\boldsymbol{D} = \boldsymbol{C}^T \boldsymbol{C}, \quad [2]$$

with $\boldsymbol{C}$ an upper triangular matrix with nonzero diagonal elements.

2. The lower bound constraint on K_{APP}: The theoretical minimal kurtosis value equals -2. However, the compartment tissue model of diffusion predicts positive kurtosis (7).

Given both diffusional tensors, K_{APP} along direction $\boldsymbol{g}$ is calculated as:

$$K_{\mathrm{APP}}(\boldsymbol{g}) = \frac{\mathrm{MD}^2}{D_{\mathrm{APP}}(\boldsymbol{g})^2} \sum_{i,j,k,l=1}^{3} g_i g_j g_k g_l W_{ijkl}, \quad [3]$$

with

$$D_{\mathrm{APP}}(\boldsymbol{g}) = \boldsymbol{g} \boldsymbol{D} \boldsymbol{g}^T, \quad [4]$$

and

$$\text{MD} = \sum_{i=1}^{3} \frac{D_{ii}}{3}. \quad [5]$$

As $D_{\text{APP}}(\boldsymbol{g})$ is positive along each gradient direction due to constraint #1, $K_{\text{APP}}(\boldsymbol{g})$ is positive if

$$\sum_{i,j,k,l=1}^{3} g_i g_j g_k g_l W_{ijkl} > 0. \quad [6]$$

As a result, a positive K_{APP} along each unit direction of a densely sampled sphere was imposed as a set of linear constraints.

3. The upper bound constraint on K_{APP}: Assuming that the log-transformed diffusion weighted signal along a fixed gradient direction, $\ln \mathcal{S}(b)$, is a monotonically decreasing function of the b-value, then the upper bound on $K_{\text{APP}}(\boldsymbol{g})$ can be derived as:

$$K_{\text{APP}}(\boldsymbol{g}) \leq \frac{3}{D_{\text{APP}}(\boldsymbol{g})b}, \quad [7]$$

as a necessary condition for the validity of the DKI model. Only then, the first derivative of

$$\ln \mathcal{S}(b) \approx \ln S(0) - D_{\text{APP}}(\boldsymbol{g})b + \frac{1}{6}D_{\text{APP}}(\boldsymbol{g})^2 K_{\text{APP}}(\boldsymbol{g})b^2, \quad [8]$$

with respect to b is negative within the range of acquired b-values. For each evaluated direction, Eq. 7 was imposed as a nonlinear constraint during parameter estimation.

Parameter Estimation

Two approaches were used to estimate both diffusional tensors. The first approach assumes additive Gaussian noise on the DWIs with the noise level independent of the degree of diffusion weighting. A commonly used implementation is the weighted least squares (WLS) algorithm, because of its ease of implementation and low computational cost (25). However, the theoretical properties of the WLS are less favorable as diffusion weighted data is Rice distributed. The second approach exploits the a priori knowledge of the data statistics in an optimal way by maximizing the joint likelihood function of all Rician distributed observations. The maximum likelihood (ML) estimator based on the Rician probability density function (PDF) is known to be asymptotically unbiased (15).

Weighted Least Squares Estimation. The unconstrained WLS estimation of the parameter vector, $\boldsymbol{\theta}$, based on the N log-transformed observed signal intensities $\boldsymbol{y}$, is given by a closed-form expression:

$$\hat{\boldsymbol{\theta}}_{\text{WLS}} = (\boldsymbol{X}^T \boldsymbol{\omega} \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{\omega} \ln \boldsymbol{y}. \quad [9]$$

The matrix $\boldsymbol{X}$ ($N \times 22$) encapsulates the b-value and gradient directions and is, more elaborately given in (6). The weight

matrix, $\boldsymbol{\omega}$ is a diagonal matrix for which the elements are defined as:

$$\omega_{nn} = y_n^2. \quad [10]$$

In Eq. 10, y_n is the nth observation after applying diffusion weighting with strength b_n and gradient direction $\boldsymbol{g}_n$.

The constrained WLS estimator (CWLS) optimizes the objective function:

$$f_{cwls}(\boldsymbol{\theta}_{cwls}) = \frac{1}{2} \sum_{n=1}^{N} \omega_{nn}^2 (\ln y_n - \ln S_{dki}(b_n, \boldsymbol{g}_n; \boldsymbol{\theta}_{cwls}))^2, \quad [11]$$

using the Nelder-Mead simplex algorithm (26). Constraint #1 is imposed by substituting $\boldsymbol{D}$ in Eq. 1 by Eq. 2. The (non)linear constraints #2 and #3 are imposed by setting f_{cwls} to ∞ when at least one constraint is violated for a single gradient direction.

Maximum Likelihood Estimation. As the magnitude diffusion weighted data are independently Rice distributed, the actual PDF of the magnitude of the observed diffusion weighted signal is given by (15):

$$p(y_n | S_{dki}(b_n, \boldsymbol{g}_n; \boldsymbol{\theta}), \sigma)$$
$$= \frac{y_n}{\sigma^2} e^{\left(-\frac{y_n^2 + (S_{dki}(b_n, \boldsymbol{g}_n; \boldsymbol{\theta}))^2}{2\sigma^2}\right)} I_0 \left(\frac{y_n S_{dki}(b_n, \boldsymbol{g}_n; \boldsymbol{\theta})}{\sigma^2}\right). \quad [12]$$

The underlying magnitude signal, $S_{dki}(b_n, \boldsymbol{g}_n; \boldsymbol{\theta})$, is given by Eq. 1. The noise level σ was estimated from the histogram mode of the image background (27,28). Furthermore, I_0 is the order zero modified Bessel function of the first kind. The parameter vector $\boldsymbol{\theta}$ was estimated from the independent DWIs with a maximum likelihood (ML) estimator in each voxel by substituting the observed values for the stochastic variables and maximizing over the parameters:

$$\hat{\boldsymbol{\theta}}_{\text{ML}} = \arg \max_{\boldsymbol{\theta}} \sum_{n=1}^{N} \ln p(y_n | S_{dki}(b_n, \boldsymbol{g}_n; \boldsymbol{\theta}), \sigma). \quad [13]$$

Equation 13 was solved by a numeric optimization using the Nelder-Mead simplex algorithm. Either of the constraints were imposed, analogue to the constrained WLS estimator, to obtain a constrained ML estimator (CML).

EXPERIMENTS

Simulation Study

Initially, Monte Carlo simulations were done to evaluate the effect of varying signal-to-noise ratio (SNR) on the accuracy of the (un)constrained WLS as well as the (un)constrained ML approaches for estimating either of the diffusional tensors. Simulated Rician distributed data with varying SNR within a range of [10, 40] was simulated using the biexponential diffusion model (29). The model parameters $\boldsymbol{D}_f$, $\boldsymbol{D}_s$, representing the slow and fast DT, respectively, and f, the fast diffusion volume fraction, corresponded to values observed in the corpus callosum in this study. The gradient scheme used to simulate the data was the same as the one used to acquire the human data set. The DKI
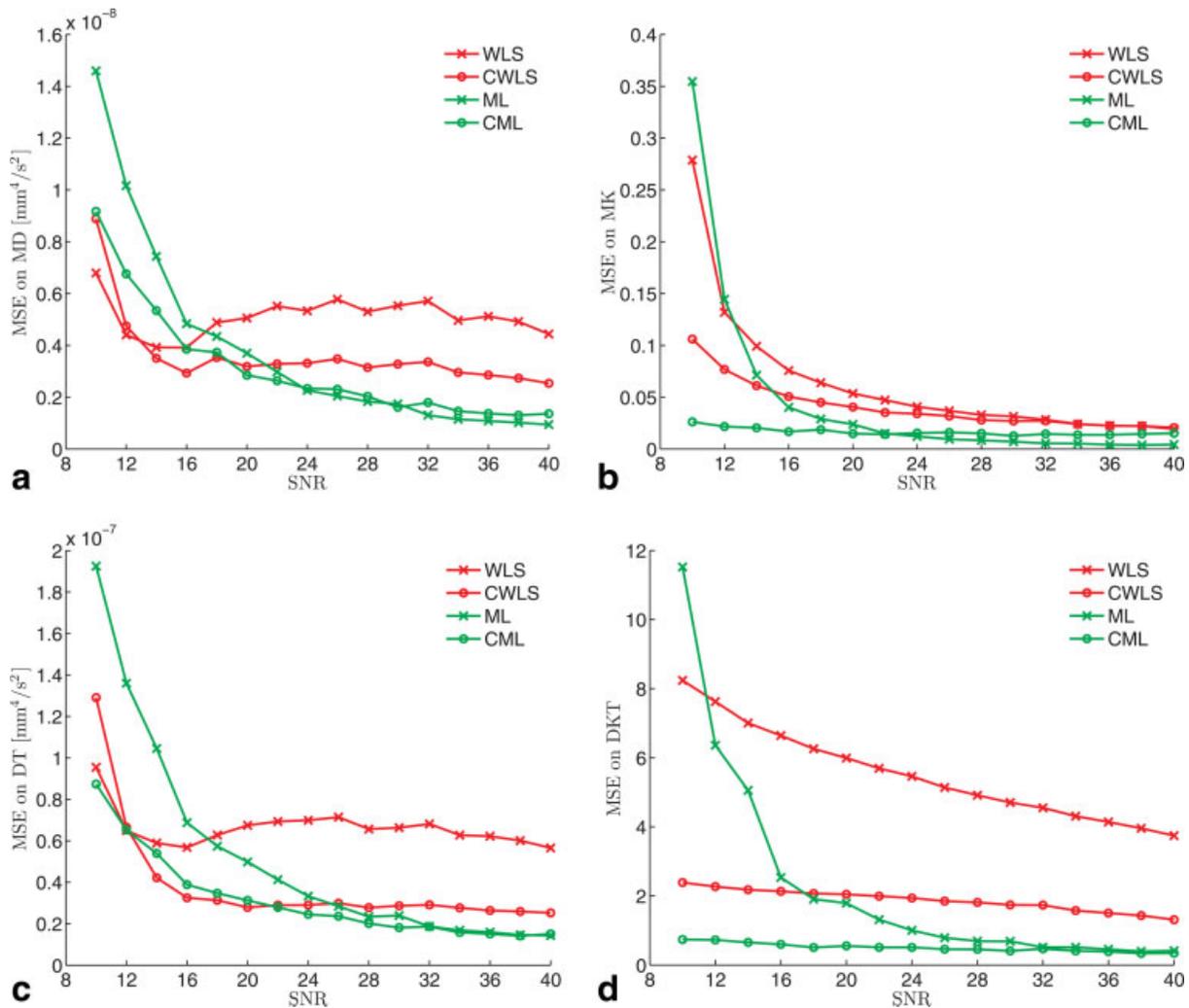
FIG. 1. The MSE based on 1000 Monte Carlo simulation trials of (**a**) MD, (**b**) MK, (**c**) DT elements, and (**d**) the DKT elements as a function of the SNR. The performance in terms of the MSE was compared between the (un)constrained WLS estimators (red) and the (un)constrained ML estimators (green). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

model parameters—D_{ref} and K_{ref}—estimated by fitting Eq. 1 to the noise-free simulated DW data were used as references values during the experiment. Both tensors did not violate any of the constraints. The derived diffusion parameters FA and MD were 0.82 and 1.2×10^3 mm^2/s, respectively, while the kurtosis parameters MK and $K_\perp$ values were 0.434 and 0.7905, respectively. For each Monte Carlo trial ($n = 1000$), D and W were estimated with either of the estimators. The difference between the tensor estimated and its reference values was quantified by the mean squared error (MSE) on (a) MD, (b) MK, (c) the DT elements, and (d) the DKT elements.

Empirical Study

Constraint Violation Using Unconstrained Estimators

The unconstrained WLS and ML estimator were run on the human data set and the violations of either of the constraints was voxelwise examined. For each voxel, constraints #2 and #3 were evaluated along the 140 acquired gradient directions. The number of directions with violated constraints were counted to determine the physical relevance of the DKTs, which were estimated with either of the unconstrained estimators. Within each tissue class [WM, gray matter (GM), and cerebrospinal fluid (CSF)]—segmented using FSL's FAST algorithm (30)—the percentage of voxels for which the constraints are not satisfied were computed and compared across both estimation strategies.

Decreased Kurtosis Parameters Using a Rician Noise Model

To demonstrate the effect of a Gaussian model on the scalar diffusion measures, average values for FA, $D_\perp$, $D_\parallel$, MD, $K_\perp$, $K_\parallel$ and MK were computed for several anatomical structures with (a) the constrained WLS and (b) constrained ML estimator. The regions (corpus callosum (CC), cingulum (Cg), internal capsule (IC), external capsule (EC) and inferior longitudinal fasciculus (ILF)) were segmented by warping the parcellation maps defined in the work of Mori et al. (31) nonlinearly to the diffusion weighted data set.
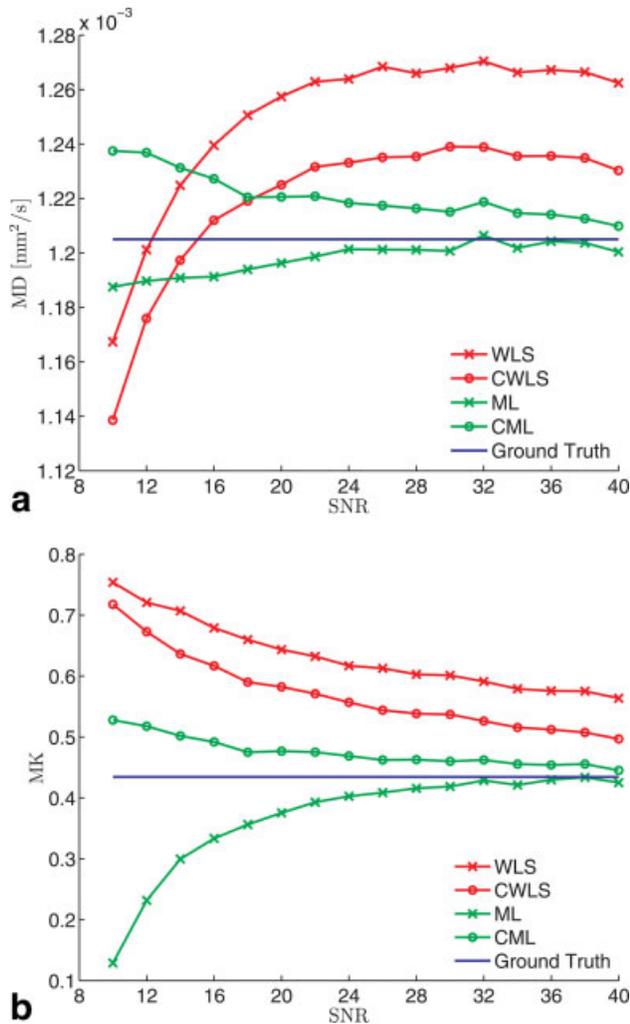
FIG. 2. The average (**a**) MD and (**b**) MK values as a function of the SNR, compared across the (un)constrained WLS estimators (red) and the (un)constrained ML estimators (green). The parameter values were averaged out over the 1000 Monte Carlo simulation trials. The ground truth value is shown by the blue line. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

A coarse, linear alignment of the atlas and the diffusion weighted data set preceded the nonlinear coregistration. The linear and nonlinear coregistration were done with FSL using FLIRT and FNIRT, respectively (32,33). The paired Mann-Whitney U test was used to compare the parameter value distributions within each ROI across both noise models; p-values < 0.01 were considered statistically significant.

RESULTS

Noise Level

The SNR was defined as the ratio of the median intensity of the b_0 image to the noise level; its estimated value was 25.8. The noise variance was constant over the entire volume due to the single coil acquisition (34).

Simulation Study

Although the difference in performance of the WLS and ML estimator vanishes with increasing SNR, the ML estimator including a Rician noise model outperforms the WLS estimator, which assumes a Gaussian noise model (see Fig. 1). Compared to the ML estimator, the accuracy of the WLS estimator appears to depend severely on the SNR. In a wide range of clinical settings (at SNR < 25), the WLS estimator resulted in significantly underestimated MD values, while the MK value was significantly overestimated.

The effect of constraining the ML estimator on the MSE is mainly noticeable at low SNR (see Fig. 1). Although the likelihood values obtained with the unconstrained ML estimator exceeded in general the ones related to the constrained ML estimator (results not shown), the MSEs on the diffusional tensors as well as on the diffusional parameters are lower with the constrained estimator if SNR < 25. The constrained ML estimator resulted consistently in the more accurate estimation of diffusional parameters such as MD and MK (Fig. 2a), while the mean and directional kurtosis values were underestimated with the unconstrained estimator (Fig. 2b).

In Fig. 3, the fraction of Monte Carlo simulations that gave rise to constraint violations are shown. With decreasing SNR, the number of trials not satisfying constraint #3, and to a lesser degree, constraint #2 increases. Over the entire SNR range, violations of constraint #2 are more common during these simulation set-up. Negative kurtosis values tend to appear more often with the ML estimator compared to the WLS estimator, while the WLS approach clearly gave more rise to violations of constraint #3.
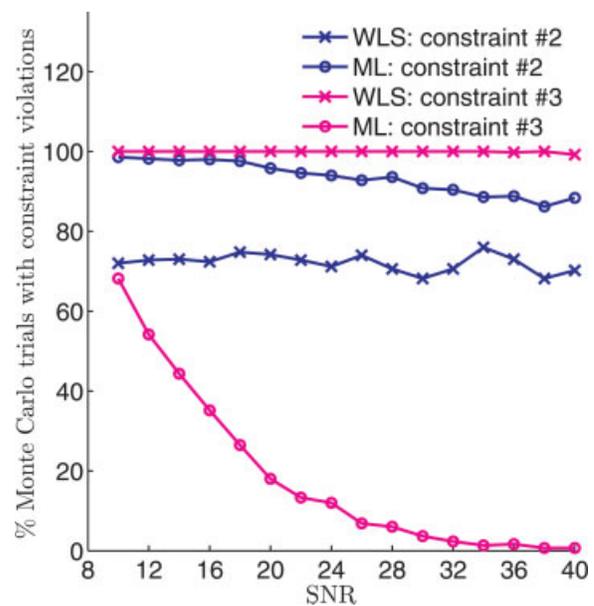

FIG. 3. The fraction of Monte Carlo simulation trials that resulted in violations of constraint #2 (blue) and constraint #3 (magenta). Both the unconstrained WLS (x) and unconstrained ML (o) estimator were compared. Violations of constraint #1 are not shown as they did not appear during simulations. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]
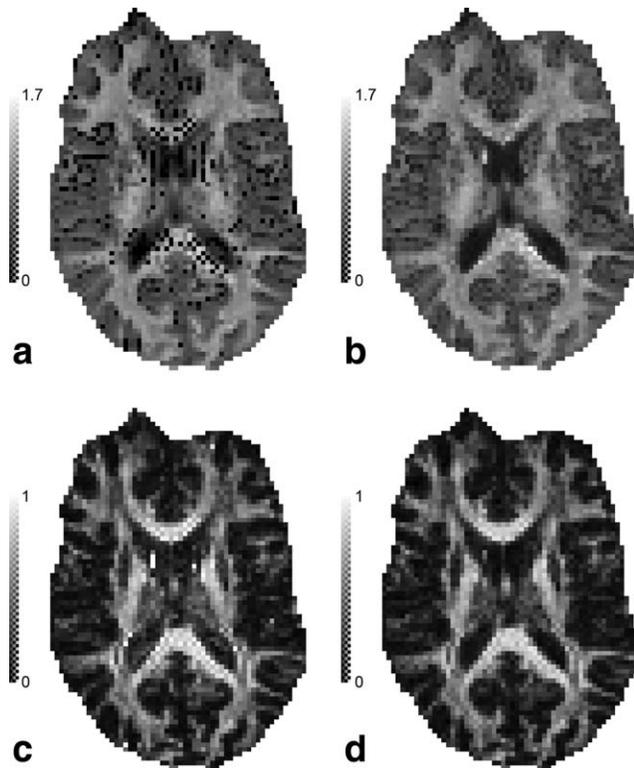
FIG. 4. The MK (**a,b**) and FA (**c,d**) maps of the same axial slice, computed with unconstrained (a,c) as well as the constrained (b,d) ML estimators. The MK maps were scaled between 0 and 1.7, the FA maps between 0 and 1.

Empirical Study: Constraint Violation Using Unconstrained Estimators

In Fig. 4, MK and FA maps of the same axial slice, computed with unconstrained as well as the constrained ML estimators, are shown. In the parameter maps estimated with the unconstrained algorithm (Fig. 4a,c), one can visually detect several outliers, which are related to constraint violations. On the one hand, the MK map (Fig. 4a) is covered with black voxels, while, on the other hand, hyper intense FA values, i.e., FA > 1, can be observed (Fig. 4c). The former indicates negative kurtosis values (constraint #2) and the latter reveals negative eigenvalues of the estimated DT (constraint #1).

An elaborate overview of the constraint violations using unconstrained estimators is given in Fig. 5. Results generated with the unconstrained WLS and ML estimator are shown in the top and bottom row, respectively. The number of constrained violations in each voxel are summarized in color encoded maps overlaying the FA map that was estimated with the unconstrained ML approach. In Fig. 5a,d, the color intensity relates to the number of negative eigenvalues with a maximum of 3. In Fig. 5b,c,d,f, the intensity relates to the number of gradient directions for which the constraints were not satisfied. The intensity is within the range [0 140] as the constraints were evaluated in all 140 gradient directions used to acquire the diffusion weighted data. Violations of constraint #1 are shown in Fig. 5a,d. All of the violations appeared in the deep WM structures such as the genu and splenium of the CC. In Fig. 5b,e, the

voxels for which $K_{APP} < 0$ (constraint # 2 violations) were observed in at least one diffusion weighting gradient direction are colored. Again, most voxels violating constraint #2 were within deep WM structures. Violations of constraint #3 are shown in Fig. 5c,f. Note that in numerous voxels, this condition was not satisfied. Although most constrained violations were located in the CSF, the DKI model also resulted in too high kurtosis values within many voxels of various WM regions.
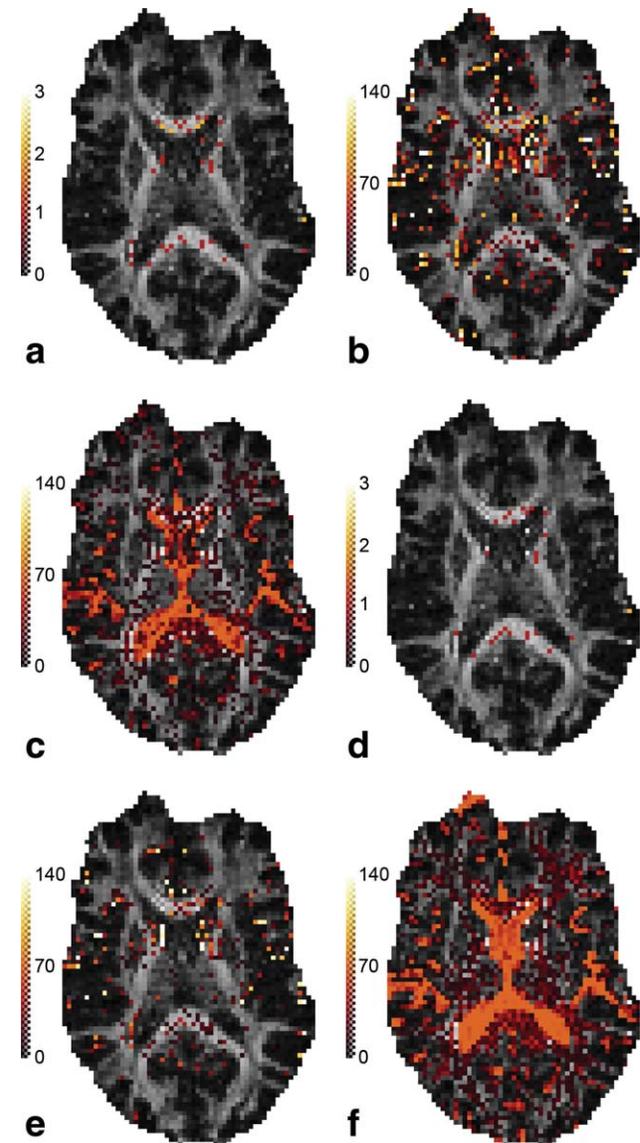

FIG. 5. Spatial distribution of violations of constraint #1 (**a,d**), constraint #2 (**b,e**), and constraint #3 (**c,f**) after estimation of the diffusional tensors with unconstrained ML estimator (a–c) or unconstrained WLS estimator (d–e). The number of constraint violations are shown as color encoded maps, overlaying an FA map. In (a,d), the color intensity relates to the number negative eigenvalues with a maximum of 3. In (b, c, e, f), the color intensity relates to the number of gradient directions for which the constraints were not satisfied. The intensity is within the range [0 140] as the constraints were evaluated in all 140 gradient directions used to acquire the diffusion weighted data.

Table 1

The percentages of Voxels Violating Constraint #1, #2 or #3 Within Each Tissue Class (WM, GM, and CSF)

	WLS			ML		
Constraint	WM	GM	CSF	WM	GM	CSF
#1	0.19	0	2.68	0.36	0.04	5.24
#2	13.51	3.87	25.50	22.51	8.49	48.61
#3	62.95	51.86	82.45	52.03	43.30	71.89
#1 ∪ #2 ∪ #3	72.71	55.53	94.68	69.95	50.48	93.95

A more quantitative overview of the number of constraint violations is given in Table 1. The percentages of voxels violating constraint #1, #2 or #3 within each tissue class (WM, GM, and CSF) are tabulated. Note that positive definiteness on the DTs was generally satisfied when estimating the DTs with the DKI model. In less than 1% of the WM voxels, the DT showed negative eigenvalues. The constraints on the DKT, however, were violated in high percentages of the voxels within each tissue class. The most violated constraint was clearly constraint #3, the upper bound on the K_{APP}; in more than 50% of the WM voxels the estimated DKTs did not satisfy the constraint when using the ML estimator. Even higher percentages (>60% in the WM) were observed when estimating the DKT with the WLS algorithm. Negative kurtosis values, thus violations of constraint #2, were observed more often with the ML estimator (±22% of the WM voxels) compared to the WLS estimator (±13.5% of the WM voxels). In general, most violations were observed in the CSF. In about 94% of the voxels classified as CSF, one of the constraints was not satisfied when using the ML estimator. Almost 70% of the WM voxels showed constraints violations and in more than 50% of the GM voxels, physically nonplausible diffusional tensors were estimated.

Empirical Study: Decreased Kurtosis Parameters Using a Rician Noise Model

Diffusion as well as kurtosis parameters were significantly different when estimated assuming a Rician noise model

Table 2

ROI Analysis of the Diffusion Parameters (FA, MD, $D_\perp$, and $D_\parallel$)

	FA			MD [×10⁻⁴ mm²/s]		
	CML	CWLS	p	CML	CWLS	p
CC	0.68	0.67	≤0.01[a]	8.67	8.60	≤0.01[a]
IC	0.60	0.60	0.099	8.19	8.15	≤0.01[a]
EC	0.42	0.42	0.023	8.54	8.52	0.021
Cg	0.53	0.54	0.984	8.53	8.39	≤0.01[a]
IFL	0.56	0.55	0.697	7.78	7.36	≤0.01[a]

	$D_\perp$ [×10⁻⁴ mm²/s]			$D_\parallel$ [×10⁻³ mm²/s]		
	CML	CWLS	p	CML	CWLS	p
CC	4.66	4.66	0.479	1.71	1.69	≤0.01[a]
IC	5.02	4.99	0.441	1.48	1.46	≤0.01[a]
EC	6.46	6.45	0.845	1.27	1.28	≤0.01[a]
Cg	5.60	5.57	0.092	1.39	1.38	≤0.01[a]
IFL	5.10	5.07	0.092	1.32	1.31	≤0.01[a]

The p-values were computed with the paired Mann-Whitney U test and used to examine the difference across the Gaussian (CWLS) and Rician (CML) noise model.

[a]Statistical differences ($p < 0.01$).

Table 3

ROI analysis of the kurtosis parameters (MK, $K_\perp$, and $K_\parallel$)

	MK			$K_\perp$			$K_\parallel$		
	CML	CWLS	p	CML	CWLS	p	CML	CWLS	p
CC	0.99	1.01	≤0.01[a]	1.68	1.68	0.477	0.44	0.47	≤0.01[a]
IC	0.95	0.97	≤0.01[a]	1.42	1.42	0.112	0.55	0.57	≤0.01[a]
EC	0.76	0.78	≤0.01[a]	0.98	1.01	≤0.01[a]	0.58	0.61	≤0.01[a]
Cg	0.89	0.91	≤0.01[a]	1.30	1.31	0.696	0.62	0.64	≤0.01[a]
IFL	1.03	1.04	≤0.01[a]	1.56	1.54	0.715	0.67	0.68	≤0.01[a]

The p-values were computed with the paired Mann-Whitney U test and used to examine the difference across the Gaussian (CWLS) and Rician (CML) noise model.

[a]Statistical differences ($p < 0.01$).

compared to a Gaussian noise model in all studied anatomical regions (Tables 2 and 3). ROI analysis showed significantly ($p < 0.01$) elevated MD and $D_\parallel$ values in all the regions when the tensor elements were estimated assuming a Rician noise model (see Table 2). No significant differences in FA - with exception of the CC - and $D_\perp$ were noticed when comparing both noise models. Similar results were observed for the kurtosis parameters (Table 3). The MK and $K_\parallel$ values were significantly lower when a Rician noise model was included. Again, less differences were noticed on the kurtosis values perpendicular to the main direction of diffusivity.

DISCUSSION

Although (weighted) LS estimators are inherently biased, they are commonly used in DTI and DKI studies, probably motivated by their ease of implementation and low computational cost. Nevertheless, the assumption of a Gaussian noise model, thus not exploiting the Rician noise distributions of magnitude diffusion weighted data, will result in an SNR dependent overestimation of kurtosis parameter values as shown in simulation experiments. Since the Rician noise statistics are better approximated by a Gaussian distribution at high SNR, the results of the WLS estimator became comparable to the results of the ML estimator with increasing SNR. The latter also explains why any—with exception of the EC—significant difference was observed on the $K_\perp$ parameter when comparing the WLS and ML estimator, while many significant differences were found in the axial directions. By definition, $D_\perp$ is lower than $D_\parallel$ if the DT is anisotropic. Since the diffusion weighed signal attenuates less with a low diffusivity, the SNR is clearly higher in the radial direction. Therefore, in the radial directions the actual data PDF is more similar to a Gaussian one.

The simulation experiment as well as an empirical study demonstrated that conventional ML estimators not always result in plausible tensor estimates as the diffusion model is fitted to few and noisy DWIs, often corrupted with imaging artifacts.

First, negative mean and directional kurtosis values were observed when using one of the unconstrained estimators. Typically, negative kurtosis values are observed in the deep WM structures, such as the CC [cf. (19)]. The genu

and splenium of the CC as well as the simulated diffusion weighted data, are characterized by a low $D_\perp$. Along low diffusivity directions, the noisy diffusion weighted signal might appear being a concave function of the b-value, yielding negative estimated of directional kurtosis. In some extreme cases, the measured diffusion-weighted signal intensity may even exceed the nondiffusion weighted signal intensity. As a result, negative estimate of diffusivity (violations of constraint #1) and extremely negative kurtosis values (violations of constraint #2) arises. The latter can be observed as black voxels in Fig. 4a, while negative diffusivity yields hyper intense FA values (Fig. 4c). Obviously, negative kurtosis values did also appear within regions one expects Gaussian diffusion, e.g., CSF, due to the variance of the selected estimator (Fig. 5b,e).

Second, many violations of constraint #3 were observed. The gradient directions in which K_{APP} exceeded its theoretical upper bound are often aligned with the direction of maximal diffusivity. Since the maximal K_{APP} for which constraint #3 holds is inversely proportional to the measured D_{APP}, clearly a small overestimation of K_{APP} might be sufficient to violate constraint #3 in directions with high diffusivity. The combination of high diffusivity and high b-values yields severely attenuated signal, and thus very low SNR. At low SNR, the WLS estimator gives rise to significant overestimated kurtosis values, which explains the higher rate of contraint #3 violations during the simulation and real data experiments when using the WLS estimator. Although the ML estimator is asymptotically unbiased and asymptotically most precise, its estimates might exceed the true underlying kurtosis values, thus possibly causing violation of constraint #3, due to the variance of the estimator. The kurtosis overestimation might also be rooted in partial volume effects. A typical example is CSF contamination in which a single voxel contains both WM tissue and CSF, which is characterized by a high diffusivity. Although the quantitative effect of CSF contamination on DTI and DKI parameters needs further study, it has already been demonstrated that the apparant kurtosis values will increase due to the introduction of an additional diffusion compartment (10).

Because of these constraint violations, diffusion and kurtosis parameters might be inaccurate and unreliable, hampering statistical analyses in clinical studies. Hence, constrained estimators are crucial in DKI analyses. The estimator we proposed in this study includes either of the constraints as well as a Rician noise model. As a result, accurate, and thus physical relevant tensor estimates are obtained. Iterative optimization is possible to avoid local minima that might arise as the simplex method is not a global optimizer. However, simulations showed that the joint likelihood after a single iteration was within 0.04% of the likelihood found after 10 iterations (results not shown). Note that in a minor number (about 1%) of all voxels showing constraint violations with an unconstrained approach, the constrained estimator meets one of the constraints exactly after initialization with a boundary point. Repeating model fitting using neighbor average as an initial point is then recommended if a model parameter deviated from the average of voxel neighbors by a predefined threshold.

For human brain imaging, studies evaluating the DKI model accuracy empirically proved that DWIs acquired with b-values up to $3000\,\text{s/mm}^2$ are appropriate for accurate DKI analysis. Exceeding the maximal b-value during data acquisition might result in an increasing tail of the DKI function despite the physical relevance of the diffusion and kurtosis values. In that case, significant errors in parameter estimation will occur if constraint #3 is imposed during optimization.

A final remark concerns parallel imaging for which the noise amplitude is spatially dependent and can be described by a Rician or a noncentral χ-distribution according to the used reconstruction technique (34). If the data is still Rician distributed, the proposed estimator can be extended to cope with varying noise levels. The most straightforward way is to add the noise level, σ, as an additional unknown parameter during optimization. In case of χ-distributed data, the rician PDF as given in Eq. 12 should be replaced by a PDF of a noncentral χ-distribtion of which the equation is given in (35).

CONCLUSION

We showed with simulation and empirical experiments that unconstrained estimators fail to provide physically meaningful estimates of either of the diffusional tensors in approximately 70% of all WM voxels. Therefore, we proposed a constrained parameter estimation framework. All unknown model parameters are thereby estimated by maximizing the joint likelihood function of all Rician distributed diffusion weighted images given the DKI model. We included a Rician noise model to avoid significant overestimation of the kurtosis values.

ACKNOWLEDGMENTS

REFERENCES

1. Basser PJ, Mattiello J, Le Bihan D. MR diffusion tensor spectroscopy and imaging. Biophys J 1994;66:259–267.
2. Le Bihan D, Mangin JF, Poupon C, Clark C, Pappata S, Molko N, Chabriat H. Diffusion tensor imaging: concepts and applications. J Magn Reson Imaging 2001;13:534–546.
3. Melhem ER, Itho R, Jones L, Barker PB. Diffusion tensor MR imaging of the brain: effect of diffusion weighting on trace and anisotropy measurement. AJNR Am J Neuroradiol 2000;21:1813–1820.
4. Hui ES, Cheung MM, Chan KC, Wu EX. B-value dependence of DTI quantitation and sensitivity in detecting neural tissue changes. Neuroimage 2010;49:2366–2374.
5. Tuch D, Reese T, Wiegell M, Wedeen VJ. Diffusion MRI of complex neural architecture. Neuron 2003;40:885–895.
6. Veraart J, Poot DHJ, Van Hecke W, Blockx I, Van der Linden A, Verhoye M, Sijbers J. More accurate estimation of diffusion tensor parameters using diffusion kurtosis imaging. Magn Reson Med 2011;65:138–145.
7. Jensen JH, Helpern JA, Ramani A, Lu H, Kaczynski K. Diffusional kurtosis imaging: the quantification of non-gaussian water diffusion by means of magnetic resonance imaging. Magn Reson Med 2005;53:1432–1440.
8. Lu H, Jensen JH, Ramani A, Helpern JA. Three dimensional characterization of non-gaussian water diffusion in humans using diffusion kurtosis imaging. NMR Biomed 2006;19:236–247.
9. Jensen JH, Helpern JA. MRI quantification of non-Gaussian water diffusion by kurtosis analysis. NMR biomed. In press.
10. Wu E, Cheung M. MR diffusion kurtosis imaging for neural tissue characterization. NMR Biomed. In press.

11. Falangola MF, Jensen JH, Babb JS, Hu C, Castellanos FX, Martino AD, Ferris SH, Helpern JA. Age-related non-gaussian diffusion patterns in the prefrontal brain. J Magn Reson Imaging 2008;28:1345–1350.

12. Cheung MM, Hui ES, Chan KC, Helpern JA, Qi L, Wu EX. Does diffusion kurtosis imaging lead to better neural tissue characterization? A rodent brain maturation study. Neuroimage 2009;45:386–392.

13. Raab P, Hattingen E, Franz K, Zanella F, Lanfermann H. Cerebral Gliomas: diffusional kurtosis imaging analysis of microstructural differences. Radiology 2010;254:876.

14. Jansen J, Stambuk H, Koutcher J, Shukla-Dave A. Non-Gaussian analysis of diffusion-weighted MR imaging in head and neck squamous cell carcinoma: a feasibility study. AJNR Am J Neuroradiol 2010;31:741.

15. Sijbers J, den Dekker AJ, Scheunders P, Van Dyck D. Maximum likelihood estimation of Rician distribution parameters. IEEE Trans Med Imaging 1998;17:357–361.

16. Sijbers J, den Dekker AJ, Van Audekerke J, and Verhoye M, Van Dyck D. Estimation of the noise in magnitude MR images. Magn Reson Imaging 1998;16:87–90.

17. Gudbjartsson H, Patz S. The rician distribution of noisy MRI data. Magn Reson Med 1995;34:910–914.

18. Evans M, Hastings N, Peacock B. Bernoulli distribution. In: Evans M, Hastings N, Peacock B. Eds. Statistical distributions, 3rd ed. New York: Wiley; 2000. pp 31–33.

19. Tabesh A, Jensen J, Ardekani B, Helpern J. Estimation of tensors and tensor-derived measures in diffusional kurtosis imaging. Magn Res Med. In press.

20. Ardekani BA, Tabesh A, Jensen JH, Helpern JA, Bachman A, Kushner H. An improved method for diffusional kurtosis estimation. In: Proceedings of the 18th Annual Meeting of the International Society for Magnetic Resonance in Medicine, Stockholm, Sweden, 2010; 1591.

21. Poot D, den Dekker A, Achten R, Verhoye M, Sijbers J. Optimal experimental design for diffusion kurtosis imaging. IEEE Trans Med Imaging 2010;29:819–829.

22. Leemans A, Jones D. The B-matrix must be rotated when correcting for subject motion in DTI data. Magn Reson Med 2009;61:1336–1349.

23. Kiselev VG, Il'yasov KA. Is the biexponential diffusion biexponential? Magn Reson Med 2007;57:464–469.

24. Koay C, Chang L, Carew J, Pierpaoli C, Basser P. A unifying theoretical and algorithmic framework for least squares methods of estimation in diffusion tensor imaging. J Magn Reson 2006;182:115–125.

25. Andersson J. Maximum a posteriori estimation of diffusion tensor parameters using a Rician noise model: why, how and but? Neuroimage 2008;42:1340–1356.

26. Nelder JA, Mead R. A simplex method for function minimization. Comput J 1965;7:308–313.

27. Sijbers J, Poot DHJ, den Dekker AJ, Pintjens W. Automatic estimation of the noise variance from the histogram of a magnetic resonance image. Phys Med Biol 2007;52:1335–1348.

28. Rajan J, Poot D, Juntu J, Sijbers J. Noise measurement from magnitude MRI using local estimates of variance and skewness. Phys Med Biol 2010;55:441–449.

29. Niendorf T, Dijkhuizen RM, David G Norris Mv, Nicolay K. Biexponential diffusion attenuation in various states of brain tissue: implications for diffusion-weighted imaging. Magn Reson Med 1996;36:847–857.

30. Zhang Y, Brady M, Smith S. Segmentation of brain MR images through a hidden Markov random field model and the expectation maximization algorithm. IEEE Trans Med Imaging 2001;20:45–57.

31. Mori S, Oishi K, Jiang H, Jiang L, Li X, Akhter K, Hua K, Faria AV, Mahmood A, Woods R, Toga AW, Pike GB, Neto PR, Evans A, Zhang J, Huang H, Miller MI, van Zijl P, Mazziotta J. Stereotaxic white matter atlas based on diffusion tensor imaging in an ICBM template. NeuroImage 2008;40:570–582.

32. Jenkinson M, Smith S. A global optimisation method for robust affine registration of brain images. Med Image Anal 2001;5:143–156.

33. Jenkinson M, Bannister P, Brady M, Smith S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. NeuroImage 2002;17:825–841.

34. Dietrich O, Raya JG, Reeder SB, Ingrisch M, Reiser MF, Schoenberg SO. Influence of multichannel combination, parallel imaging and other reconstruction techniques on MRI noise characteristics. Magn Reson Imaging 2008;26:754–762.

35. Aja-Fernández S, Vegas-Sánchez-Ferrero G, Tristán-Vega A. About the background distribution in MR data: a local variance study. Magn Reson Imaging 2010;28:739–752.