ELSEVIER

NeuroImage

www.elsevier.com/locate/ynimg
NeuroImage 40 (2008) 1144–1156

An optimized wild bootstrap method for evaluation of measurement uncertainties of DTI-derived parameters in human brain

Tong Zhu,[a] Xiaoxu Liu,[b] Patrick R. Connelly,[a,1] and Jianhui Zhong[a,c,d,*]

[a]*Department of Biomedical Engineering, University of Rochester, Rochester, NY, USA*
[b]*Department of Electrical and Computer Engineering, University of Rochester, Rochester, NY, USA*
[c]*Department of Physics & Astronomy, University of Rochester, Rochester, NY, USA*
[d]*Department of Imaging Sciences, University of Rochester, Rochester, NY, USA*

Received 22 March 2007; revised 21 December 2007; accepted 5 January 2008
Available online 26 January 2008

Evaluation of measurement uncertainties (or errors) in diffusion tensor-derived parameters is essential to quantify the sensitivity and specificity of these quantities as potential surrogate biomarkers for pathophysiological processes with diffusion tensor imaging (DTI). Computational methods such as the Monte Carlo simulation have provided insights into characterization of the measurement uncertainty in DTI. However, due to the complexity of real brain data as well as different sources of variations during the image acquisition, a robust estimator for DTI measurement uncertainty in human brain is required. Recent studies have shown that wild bootstrap, a novel nonparametric statistical method, can potentially provide effective estimations of DTI measurement uncertainties in human brain DTI data. In this study, we further optimized the DTI application of the wild bootstrap method for typical clinical applications. We evaluated the validity of wild bootstrap utilizing numerical simulations with different combinations of DTI protocol parameters and wild bootstrap experimental designs, and quantitatively compared estimates of uncertainties from wild bootstrapping with those from Monte Carlo simulations. Our results demonstrate that a wild bootstrap implementation using at least 1000 wild bootstrap iterations with a type II or type III heteroskedasticity consistent covariance matrix estimator provides robust evaluations of most DTI protocols.

Keywords: Diffusion tensor imaging; Measurement uncertainty; Wild bootstrap; Monte Carlo simulation

Introduction

Diffusion tensor imaging (DTI) is a technique that visualizes random molecular motions of water in tissues such as the brain (Basser et al., 1994b). To quantify local water molecule mobility within tissue, a diffusion tensor is calculated at each voxel using linear or nonlinear regressions of at least six sets of diffusion-weighted (DW) images acquired with noncollinearly and noncoplanarly applied diffusion-weighting gradients and a set of baseline images without diffusion weighting (often referred to as non-DW or b_0 images). Further quantitative parameters derived from DTI [such as the extent of diffusion by mean diffusivity (MD) and the level of anisotropy of diffusion by fractional anisotropy (FA)], can be used to infer the underlying anatomical microstructures in tissues non-invasively. Statistical inference achieved with comparisons of these DTI-derived parameters using either region-of-interest (ROI) or voxel-based methods in different patient groups or patients under different pathophysiological conditions can yield insights into microscopic details of diseases. However, when the DTI technique is applied, imprecision due to measurement uncertainties either reduces the power of statistical inference or results in erroneous fiber tractography that in turn decreases the sensitivity and specificity of these DTI-derived parameters as surrogate biomarkers for various applications.

A number of analytical techniques have been developed to characterize the uncertainty in DTI and to guide selection and optimization of DTI protocols. In one approach, Monte Carlo (MC) simulation is used with the assumption that the source of measurement uncertainty is pure white noise (Pierpaoli and Basser, 1996; Papadakis et al., 2000; Hasan et al., 2001; Jones, 2004a). In another approach, an analytical formalism of noise propagation is considered in the tensor calculation (Skare et al., 2000; Anderson, 2001; Poonawalla and Zhou, 2004). The simplified noise models in both approaches, however, may not be adequate to describe the true noise distribution in human subjects arising from the existence of physiological noises, scanner instability and various artifacts.

* Corresponding author. Department of Imaging Sciences, University of Rochester, Box 648, 601 Elmwood Avenue, Rochester, NY 14642-8648, USA.
E-mail address: Jianhui.zhong@rochester.edu (J. Zhong).
[1] This paper is dedicated in memory of one of the authors, Dr. Patrick Raymond Connelly.
Available online on ScienceDirect (www.sciencedirect.com).

1053-8119/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2008.01.016

Nonparametric statistical methods such as bootstrap (Efron and Tibshirani, 1993) have been suggested to be advantageous in uncertainty estimation for real human data (Pajevic and Basser, 2003). Bootstrap is a nonparametric procedure for estimating uncertainties such as bias, confidence interval and probability distributions, in statistics without prior assumptions about the population distribution. For a typical bootstrap implementation in DTI (Pajevic and Basser, 2003), multiple measurements are acquired for each diffusion-weighting direction. Bootstrap samples (or replicas) are then constructed by randomly selecting DWIs with replacement from repeated measurements. Estimates of the true values are derived from each bootstrap sample, and the uncertainties of DTI measurements are quantitatively evaluated with a number of bootstrap estimations. The abovementioned bootstrap technique makes no assumptions about the noise model in DTI measurements at the cost of the need for repeated scans. This method has been referred to as repetition bootstrap (Chung et al., 2006), or regular bootstrap (Whitcher et al., 2007) in the literature. The repetition bootstrap has been used to either characterize distributions of scalar DTI measures (Pajevic and Basser, 2003; Hasan et al., 2004), or infer the measurement uncertainty in terms of the principal eigenvector orientation (Jones, 2003). Effects of other factors, such as gender and subject motion during scanning, on the obtained DTI parameters have been studied with human DTI data (Heim et al., 2004). Furthermore, the repetition bootstrap technique has been extended to estimate the probability density function (PDF) of the fiber orientation for generating fiber tracts in human white matter (Lazar and Alexander, 2005; Jones and Pierpaoli, 2005). It is worth noting, however, that a basic requirement for a reliable repetition bootstrap analysis is a sufficient number of repeated measurements, and a minimum of five repeats has been proposed in a systematic evaluation of bootstrap performance using numerical simulations (O'Gorman and Jones, 2006). The need for multiple measurements per diffusion-weighting direction often excludes extension of repetition bootstrap to clinical applications where the scan time is usually limited.

To overcome the long scan time limitation of a repetition bootstrap method, wild bootstrap, a model-based resampling technique (Liu, 1988; Flachaire, 2005a), has recently been proposed as a potential robust estimator of measurement uncertainty in DTI (Whitcher et al., 2007). Wild bootstrap requires only one complete DTI measurement. The performance of wild bootstrap was recently evaluated in comparison to repetition bootstrap as well as residual bootstrap, another model-based technique, through numerical simulations (Chung et al., 2006) and in comparison to repetition bootstrap from both numerical simulations and human DTI data (Whitcher et al., 2007).

In this work, we further optimized the DTI application of the wild bootstrap method for typical clinical applications. Measurement uncertainties associated with DTI were quantified in terms of the uncertainties of three commonly used DTI-derived parameters, FA, MD and the principal diffusion direction (the orientation of the largest eigenvalue). For the optimization purpose, the performance of the wild bootstrap method was evaluated under different combinations of DTI protocol settings including signal-to-noise ratio (SNR), the ratio of the total number of DW images and non-DW baseline images (DTI_Ratio), and the number of the unique diffusion-weighting gradients (N_{UDG}). DTI protocols used in this study were selected to be close to those used in typical clinical DTI applications. Two additional parameters related to wild bootstrap designs, the number of wild bootstrap iterations (N_{WBS}) and the type of heteroskedasticity consistent covariance matrix estimator function (HCCME), were also assessed using numerical simulations. To facilitate the evaluation, DTI measurement uncertainties estimated from Monte Carlo simulations under the same DTI protocol settings were established as the "gold standard." We then quantitatively evaluated the performance of the wild bootstrap method as a robust estimator from two perspectives: one using the bias between wild bootstrap estimations of uncertainties and those from MC simulations as the measure of accuracy; the other using the standard error of uncertainty estimations from repeated wild bootstrap procedures as the measure of precision. Results of these evaluations were used as a basis for a proposed optimized wild bootstrap implementation for typical clinical applications.

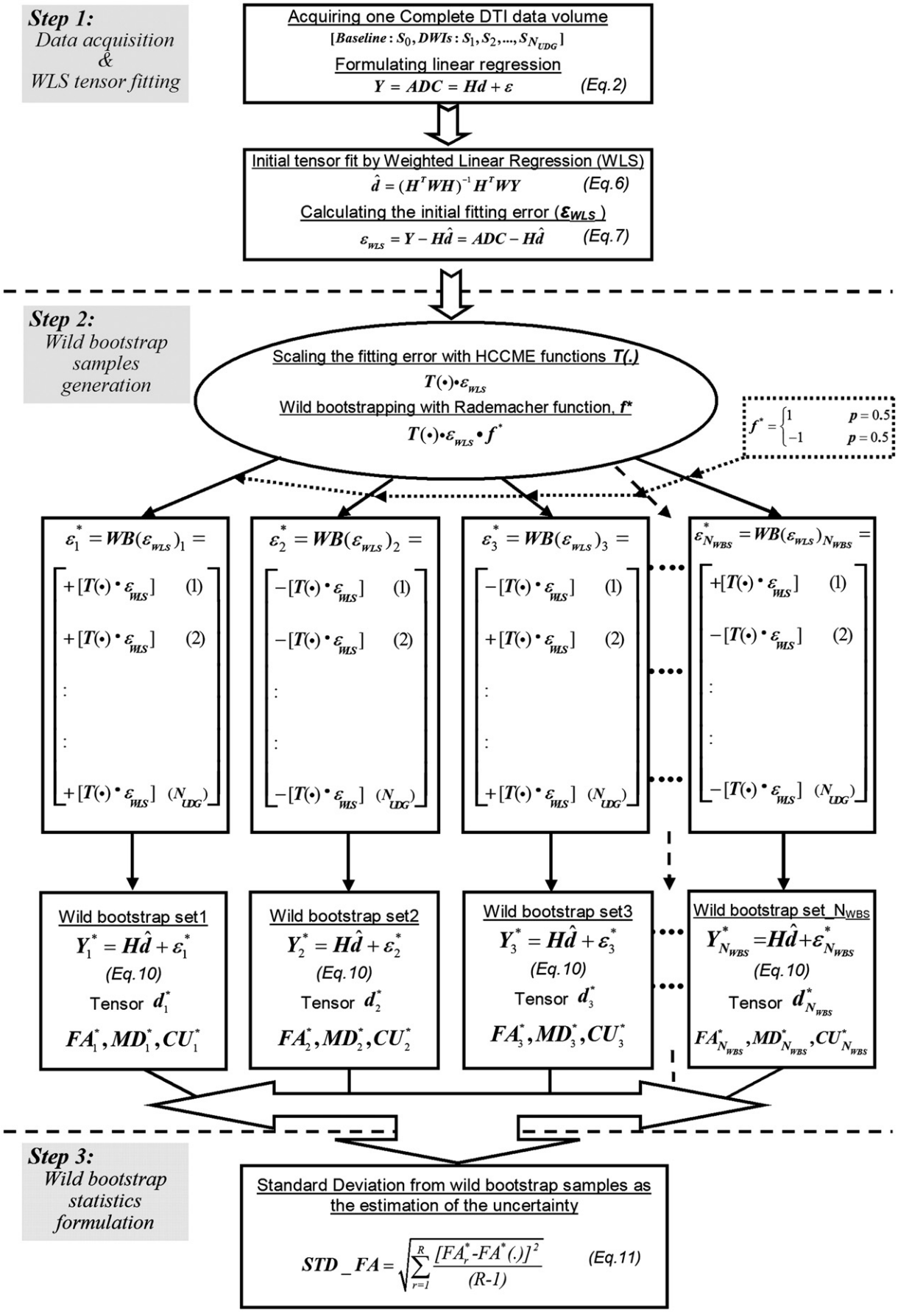
Theory

Model-based bootstrap resampling and wild bootstrap

Model-based resampling technique can be applied to estimate the measurement uncertainties in human DTI data. This approach does not require a large sample of data as with the repetition bootstrap method. Model-based resampling techniques have been commonly used in the linear regression model, where the properties of the parameters from the linear model, such as standard errors and confidence intervals for regression coefficients, are estimated by resampling methods like bootstrapping.

Depending on assumptions of the residual error structure in the linear regressing model, two different approaches, residual bootstrap and wild bootstrap, have been proposed. Under a strong assumption that the residuals ε from the original linear regression model are independently and identically distributed (I.I.D.) random variables sampled from an unknown distribution function F with an expectation value of zero (Liu, 1988; Efron and Tibshirani, 1993), residual bootstrap (Efron and Tibshirani, 1993; Chung et al., 2006; Whitcher et al., 2007), which gives equal probability to all the residuals ε from the original linear regression model and redistributes them among the diffusion gradient directions, can robustly assess statistical accuracy. Linear regression with I.I.D. residuals is also referred to as homoskedastic, while heteroskedasticity is used to describe the non-I.I.D. condition. Even under heteroskedastic conditions, a modified residual bootstrap (Chung et al., 2006) can adjust the non-constant variance structure of the residuals to achieve the abovementioned free resampling among the residuals ε, when the heteroskedasticity can be properly modeled.

In a DTI measurement, however, the assumption of a homoskedastic nature of errors is violated due to the logarithmic transformation of signal intensities of non-DW and DW images during the tensor fitting (Basser et al., 1994a; Chung et al., 2006; Whitcher et al., 2007). Moreover, there are a variety of other error sources, such as field inhomogeneity and physiological noises, during human DTI acquisition. It is difficult to model the errors that arise from these sources in a multiple linear regression. Therefore, the multiple linear regression for diffusion tensor estimation is, in general, heteroskedastic with an unknown form of error model. The wild bootstrap produces I.I.D. observations from an external probability distribution that is independent of the original data and has a mean of zero as well as a variance of 1 (Liu, 1988). Wild bootstrap samples are then generated by multiplying the fitting residuals from the original data with these exogenous I.I.D observations that are regenerated for each resampling. A typical implementation of the wild bootstrap method to estimate DTI


Step 1:
Data acquisition
&
WLS tensor fitting
Acquiring one Complete DTI data volume
[Baseline : S0, DWIs : S1, S2, ..., SNUDG]
Formulating linear regression
Y = ADC = Hd + ε (Eq.2)
Initial tensor fit by Weighted Linear Regression (WLS)
d̂ = (H^T W H)^-1 H^T W Y (Eq.6)
Calculating the initial fitting error (εWLS)
εWLS = Y − Hd̂ = ADC − Hd̂ (Eq.7)
Step 2:
Wild bootstrap
samples
generation
Scaling the fitting error with HCCME functions T(.)
T(•)•εWLS
Wild bootstrapping with Rademacher function, f*
T(•)•εWLS•f*
f* = 1 p = 0.5; −1 p = 0.5
ε1* = WB(εWLS)1 =
ε2* = WB(εWLS)2 =
ε3* = WB(εWLS)3 =
εNWBS* = WB(εWLS)NWBS =
+[T(•)•εWLS] (1)
+[T(•)•εWLS] (2)
+[T(•)•εWLS] (NUDG)
−[T(•)•εWLS] (1)
−[T(•)•εWLS] (2)
−[T(•)•εWLS] (NUDG)
−[T(•)•εWLS] (1)
+[T(•)•εWLS] (2)
+[T(•)•εWLS] (NUDG)
+[T(•)•εWLS] (1)
−[T(•)•εWLS] (2)
−[T(•)•εWLS] (NUDG)
Wild bootstrap set1
Y1* = Hd̂ + ε1*
(Eq.10)
Tensor d1*
FA1*, MD1*, CU1*
Wild bootstrap set2
Y2* = Hd̂ + ε2*
(Eq.10)
Tensor d2*
FA2*, MD2*, CU2*
Wild bootstrap set3
Y3* = Hd̂ + ε3*
(Eq.10)
Tensor d3*
FA3*, MD3*, CU3*
Wild bootstrap set_NWBS
YNWBS* = Hd̂ + εNWBS*
(Eq.10)
Tensor dNWBS*
FANWBS*, MDNWBS*, CUNWBS*
Step 3:
Wild bootstrap
statistics
formulation
Standard Deviation from wild bootstrap samples as
the estimation of the uncertainty
STD_FA = √(Σ_{r=1}^{R} [FAr* − FA*(.)]^2 / (R−1)) (Eq.11)

measurement uncertainty (Chung et al., 2006; Whitcher et al., 2007) includes "Data acquisition and initial tensor fitting" (Step 1), "Wild bootstrap sample generation" (Step 2) and "Wild bootstrap statistics formulation" (Step 3). The details of each step of wild bootstrap implementation in this study are illustrated in Fig. 1 and listed below.

Step 1: Data acquisition and initial tensor fitting

Only a single complete DTI dataset is needed for wild bootstrap implementation, and only the single tensor model (Basser et al., 1994b) was considered in this study. Intensities of the measured diffusion-weighted signals are quantified by:

$$S_i = S_0 \exp\left(-\mathrm{b} \cdot \boldsymbol{g}_i^T \cdot \boldsymbol{D} \cdot \boldsymbol{g}_i\right), \quad \text{for } i = 1, 2, \ldots, N_{\mathrm{UDG}}, \tag{1}$$

where $\boldsymbol{D}$ is a 3×3 symmetric diffusion tensor, S_0 is the signal intensity of the baseline non-DW image, and S_i is the signal intensity of the ith DW image. Both b and $\boldsymbol{g}_i$ characterize the applied diffusion-weighting gradient pulses. b is the scalar value of the "b factor" that depends on the timing and amplitude of the applied diffusion-weighting gradient, while $\boldsymbol{g}_i$ is a unit vector in the direction of the ith diffusion-weighting gradient.

Applying a log-transformation to Eq. (1), the estimation of the diffusion tensor $\boldsymbol{D}$ becomes a multiple linear regression procedure described by

$$\boldsymbol{Y} = \boldsymbol{ADC} = \boldsymbol{H}\boldsymbol{d} + \boldsymbol{\varepsilon}, \tag{2}$$

where $\boldsymbol{Y}$ is a $N_{\mathrm{UDG}}\times1$ vector of apparent diffusion coefficients (ADC). Element $\boldsymbol{H}$ in Eq. (2) is a $N_{\mathrm{UDG}}\times6$ matrix, where each row, $\boldsymbol{H_i}$, is derived from an applied diffusion-weighting gradient and $\boldsymbol{d}$ is a 6×1 column vector including the six independent elements of the diffusion tensor $\boldsymbol{D}$:

$$\boldsymbol{H_i} = \left[g_{x_i}^2\ g_{y_i}^2\ g_{z_i}^2\ 2g_{x_i}g_{y_i}\ 2g_{x_i}g_{z_i}\ 2g_{y_i}g_{z_i}\right], \tag{3}$$

$$\boldsymbol{d} = \left[D_{xx}\ D_{yy}\ D_{zz}\ D_{xy}\ D_{xz}\ D_{yz}\right]^T. \tag{4}$$

We used a weighted least squares (WLS) fit that is more robust than the commonly used ordinary least squares fitting (OLS), especially when outliers exist (Basser et al., 1994a; Kingsley, 2006). More specifically, a modified WLS version (Jones and Basser, 2004b) that utilizes the standard deviations of log-transformed signal intensities instead of the WLS implementation in tensor fitting originally proposed by Basser et al. (1994a) was implemented in this study. The proposed WLS weights ADCs with the reciprocal error variances associated with the measurements of ADCs. Assuming the error variance, which was approximated by the noise variance in the numerical simulation of this study, in the baseline signal S_0 and the DW signal S_i to be σ_0^2 and σ_i^2, respectively, the error variance $\sigma_{\mathrm{WLS}\_i}^2$ in the ADC_i can be approximated by Eq. (5) through error propagation (Bevington and Robinson, 1992), and its reciprocal value is the corresponding diagonal element of the diagonal weighting matrix $\boldsymbol{W}$, i.e.

$$\boldsymbol{W} = \mathrm{diag}\left(\frac{1}{\sigma_{\mathrm{WLS}\_i}^2}\right), \quad \text{where } \sigma_{\mathrm{WLS}\_i}^2 = \frac{1}{b^2}\left(\frac{\sigma_0^2}{S_0^2} + \frac{\sigma_i^2}{S_i^2}\right), \tag{5}$$

The WLS estimate of the tensor element vector $\boldsymbol{d}$ is then expressed as

$$\hat{\boldsymbol{d}} = \left(\boldsymbol{H}^T\boldsymbol{W}\boldsymbol{H}\right)^{-1}\boldsymbol{H}^T\boldsymbol{W}\boldsymbol{Y}. \tag{6}$$

The residuals from the WLS fit used to generate the wild bootstrap samples in Step 2 are calculated from

$$\boldsymbol{\varepsilon}_{\mathrm{WLS}} = \boldsymbol{Y} - \boldsymbol{H}\hat{\boldsymbol{d}} = \boldsymbol{ADC} - \boldsymbol{H}\hat{\boldsymbol{d}}. \tag{7}$$

Step 2: Wild bootstrap sample generation: multiplying the I.I.D variables generated from an independent distribution with the adjusted residuals.

Due to its better performance over other candidate functions (Davidson and Flachaire, 2001), a two-point Rademacher function was used in this study to generate the I.I.D variables

$$\boldsymbol{f}^* = \begin{cases} 1 & p = 0.5 \\ -1 & p = 0.5 \end{cases}. \tag{8}$$

The residuals $\boldsymbol{\varepsilon}_{\mathrm{WLS}}$ from the WLS fitting of the original DTI data are modulated with a heteroskedasticity consistent covariance matrix estimator function (HCCME) to asymptotically correct the inference of multiple regression of tensor (MacKinnon and White, 1985). $\boldsymbol{T}(\bullet)$ is a $N_{\mathrm{UDG}}\times1$ vector with the elements $\boldsymbol{T}(\bullet)_i$ and can be selected from one of the four HCCME functions (MacKinnon and White, 1985) expressed as

$$T(\bullet)_i = \begin{cases} 1 & (HCCME0); \\ \sqrt{\dfrac{n}{n-k}} & (HCCME1); \\ \dfrac{1}{\sqrt{1-\Omega_i}} & (HCCME2); \\ \dfrac{1}{1-\Omega_i} & (HCCME3); \end{cases} \tag{9}$$

where $\boldsymbol{\Omega_i}$ is the diagonal element of the matrix $\Omega = \boldsymbol{H}(\boldsymbol{H}^T\mathbf{W}\mathbf{H})^{-1}\boldsymbol{H}^T\mathbf{W}$ related to the weighted linear regression procedure [Eqs. (5) and (6)]. $\boldsymbol{n}$ represents the total number of observations for linear fitting that is equal to the number of the unique diffusion gradient directions, while

Fig. 1. Schematic representation of wild bootstrap (WB) analysis of DTI measurement uncertainty. A typical WB analysis includes three steps. Step 1: Data acquisition and initial tensor fitting. Here, the weighted least square fitting is used for obtaining the initial diffusion tensor $\hat{\boldsymbol{d}}$ and the fitting residuals $\varepsilon_{\mathrm{WLS}}$. Step 2: Wild bootstrap samples generation. In this step, the fitting residuals $\varepsilon_{\mathrm{WLS}}$ are scaled by the selected HCCME function, $\boldsymbol{T}(\bullet)$, before multiplying with the random samples from a Rademacher distribution function, f^*, to generate the bootstrap resamples of the fitting residuals $\varepsilon^*(=\boldsymbol{T}(\bullet)\cdot\varepsilon_{\mathrm{WLS}}\cdot f^*)$ with the sample number to be N_{WBS}. The wild bootstrap samples of DTI datasets are formulated using the initial tensor estimation $\hat{\boldsymbol{d}}$ and bootstrap samples of ε^* and will further generate the wild bootstrap samples of tensor $\boldsymbol{d}^*$ as well as the tensor derived parameters FA*, MD*, CU*. Step 3: Wild bootstrap statistics formulation. Measurement uncertainties associated with the DTI-derived parameters are quantified as the standard deviation values from the pool of the wild bootstrap samples of those parameters. The equations listed on this figure correspond to those in the section on Theory.

$\boldsymbol{k}$ is the number of the independent variables corresponding to the six independent tensor elements.

Finally, a sufficient number, N_{WBS}, of wild bootstrap samples are generated as in Eq. (10)

$$\boldsymbol{Y}_i^* = \boldsymbol{H}\widehat{\boldsymbol{d}} + \varepsilon_i^* = \boldsymbol{H}\widehat{\boldsymbol{d}} + \boldsymbol{T}(\bullet)\cdot\varepsilon_{WLS}\cdot\boldsymbol{f}_i^*, \quad for \quad i = 1, 2, \ldots, N_{WBS}. \tag{10}$$

where $\boldsymbol{Y}_i^*$ is a wild bootstrap sample of the original response variable $\boldsymbol{Y}$, and $\boldsymbol{f}_i^*$ is a $N_{\mathrm{UDG}}\times 1$ column vector consisting of I.I.D samples from the Rademacher function $\boldsymbol{f}^*$. The residuals $\varepsilon_{\mathrm{WLS}}$ are modulated with one of HCCME functions, $\boldsymbol{T}(\bullet)$, by the dot product during wild bootstrap resampling. The residuals from WLS fit of the original DTI, $\varepsilon_{\mathrm{WLS}}$, are reused during the resampling procedure. For each wild bootstrap sample, $\boldsymbol{Y}_i^*$ is a column vector of new ADC signals and is passed on to Step 3 for calculating the new tensors and wild bootstrap statistics.

Step 3: Wild bootstrap statistics formulation: Utilizing the plug-in principle" (Efron and Tibshirani, 1993), measurement uncertainty, e.g., the uncertainty with FA, is estimated by calculating the standard deviation of all wild bootstrap samples FA_r^*, i.e. using

$$\mathrm{STD\_}FA = \sqrt{\sum_{r=1}^{R}\frac{[\mathrm{FA}_r^* - \mathrm{FA}^*(\cdot)]^2}{(R-1)}}, \quad \text{where}$$
$$\mathrm{FA}^*(\cdot) = \left(\sum_{r=1}^{R}\mathrm{FA}_r^*\right)/R, \tag{11}$$

in which R ($=N_{\mathrm{WBS}}$) is the total number of samples. The uncertainty of MD can also be estimated following the similarly defined Eq. (11) for FA. The uncertainty of the principle diffusion direction is quantified by the 95% confidence interval of the cone of uncertainty (95% CU) (Basser and Pajevic, 2000; Jones, 2003) from bootstrap samples.

Materials and methods

All numerical calculations including the Monte Carlo and wild bootstrap simulations were performed using a custom software package developed in Matlab (The Mathworks, Natick, MA). Following the definitions in Eq. (11), the wild bootstrap estimates of the measurement uncertainties of FA, MD and the principle diffusion direction will be expressed as STD_FA, STD_MD and 95% CU hereafter.

Parameter settings for wild bootstrap optimization

Numerical simulations were performed using various combinations of five parameters. Three of the parameters, SNR, N_{UDG}, and DTI_Ratio, determine the quality of the acquired DTI data, and are referred to as "DTI protocol parameters." The other two parameters, HCCME type and N_{WBS}, deal with the implementation of wild bootstrap method, and are referred to as "wild bootstrap parameters." More specifically:

(1) Signal-to-noise ratio (SNR): This is the SNR value for the baseline image (b=0 s/mm^2) with a single measurement, or when the number of excitations (NEX) is 1. In the numerical simulations of this study, we considered the perturbation in tensor solely resulting from the existence of Rician noise. Six different SNR levels, from 15 to 40 in increments of 5, were studied in the numerical simulations.

(2) N_{UDG}: The number of unique diffusion-weighting gradient directions used in the DTI scan. Three DTI gradient schemes with N_{UDG} equal to 21, 31 and 61, respectively (hereafter referred to as P21, P31 and P61), were investigated in numerical simulations. The diffusion-weighting gradient vectors, directly from a 1.5-T clinical scanner (General Electric, Milwaukee, WI), were determined with an electrostatic repulsion model that treats gradient vectors as distributed point charges on the surface of a unit sphere with minimal overall energy. To facilitate comparisons among three gradient schemes, the SNR value for each scheme was scaled by a factor $f_{\mathrm{scale}}=\sqrt{60/N_{\mathrm{UDG}}}$. The final combinations of N_{UDG} and f_{scale} for three diffusion gradient schemes in the numerical simulations were (in the form of $N_{\mathrm{UDG}}/f_{\mathrm{scale}}$): 21/1.69, 31/1.39 and 61/0.99. After adjustment, the effective SNR values in tensor calculation were approximately the same for all gradient schemes investigated in this study.

(3) DTI_Ratio: This is the ratio between the total number of DW images and the total number of non-DW images. After a survey of published studies, two different ratios, DTI_Ratio=N_{UDG}:1 and 6:1, were selected for investigation. The major difference between protocols with different DTI_Ratio values is the number of non-DW baseline images. Proportionally increased SNR value for the averaged b0 image, equivalent to obtaining extra baseline images in real DTI acquisition, was simulated for protocols with DTI_Ratio=6:1, while the number of DWIs was kept the same for protocols with either DTI_Ratio value.

(4) HCCME function type: The effects of the HCCME functions in Eq. (9) were evaluated in numerical simulations.

(5) N_{WBS}: The total number of wild bootstrap samples generated for the wild bootstrap statistics. Five different N_{WBS} values (=250, 500, 1000, 1500 and 2000) were investigated in this study.

Monte Carlo simulation

A similar approach for MC simulation as previously reported (Pierpaoli and Basser, 1996) was used in this study. For each combination of three DTI protocol parameters defined above, noise-free prolate diffusion tensors ($\lambda_1 \gg \lambda_2 = \lambda_3$) with a fixed principal diffusion direction along the z axis were produced using (Jones, 2004a)

$$\lambda_1 = \mathrm{MD}\left(1 + \frac{2\mathrm{FA}}{\sqrt{3-2\mathrm{FA}^2}}\right), \tag{12a}$$

and

$$\lambda_2 = \lambda_3 = \mathrm{MD}\left(1 - \frac{\mathrm{FA}}{\sqrt{3-2\mathrm{FA}^2}}\right), \tag{12b}$$

in which the following parameters were used: a constant MD value of 0.7×10^{-3} mm^2/s, FA values ranging between [0.1, 0.9], and a diffusion weighting of b=1000 s/mm^2. With the noise-free baseline signal intensity S_0 set to 1000, the corresponding DW signals S_i for a given DTI scheme were simulated according to the log-linear relation described in Eq. (1).

At each SNR level, the effect of thermal noise was first generated by complex random numbers with their real and imaginary parts sampled independently from a Gaussian distribution function with zero mean and a standard deviation determined by the desired SNR level. The real parts of the complex noise signals were then added to the noise-free baseline signal S_0 and DW signals S_i. The magnitude of the final complex data was then used to synthesize the noisy DTI datasets that were further used for calculations of the noisy tensors. Additional simulations using oblate diffusion tensors ($\lambda_1 = \lambda_2 \gg \lambda_3$) were also performed for FA ranges between 0.1 and 0.4 for better characterization the corresponding tissue types.

This process was performed 20,000 times to generate a sufficient number of Monte Carlo samples of noisy tensors and DTI-derived parameters. The summary statistics (the standard deviation of FA and MD, and the 95% confidence interval of the cone of uncertainty) of the Monte Carlo samples were calculated as the Monte Carlo estimates of DTI measurement uncertainties at each anisotropy level under each DTI protocol. These summary statistics from the MC simulation were used as the "gold standard" for evaluation of the performance of the wild bootstrap method under the same DTI protocol settings.

Wild bootstrap simulations and performance evaluation criteria

The wild bootstrap procedure described in Eqs. (8)–(11) was performed on the noisy tensors obtained with the same procedures as used in the MC simulation to generate N_{WBS} bootstrap samples. The principal diffusion direction was fixed along the z axis. The summary statistics, including STD_FA, STD_MD and 95% CU from the N_{WBS} bootstrap samples, are the wild bootstrap estimates of DTI measurement uncertainties. To evaluate the performance of the wild bootstrap method, the abovementioned wild bootstrap procedure was repeated for 1000 runs. The bias and standard error of wild bootstrap estimates, with respect to the "gold standard" MC simulation, were calculated based on these 1000 estimates. The equations used for the bias and standard error were

$$\mathrm{bias(E)} = \frac{|\bar{E} - E_{\mathrm{MC}}|}{E_{\mathrm{MC}}} \times 100\% \tag{13}$$

and

$$\mathrm{std}(E) = \frac{\sqrt{\sum_{i=1}^{N_{\mathrm{runs}}} \frac{[E_i - \bar{E}]^2}{(N_{\mathrm{runs}}-1)}}}{E_{\mathrm{MC}}} \times 100\%, \tag{14}$$

where $\boldsymbol{E}$ represents the wild bootstrap estimate of STD_FA, STD_MD and 95% CU for each run, while $\boldsymbol{E}_{\mathrm{MC}}$ is the counterpart estimate of $\boldsymbol{E}$ from MC simulations. $\bar{\boldsymbol{E}}$ is the mean value of $\boldsymbol{E}$ from all N_{runs} (=1000) wild bootstrap runs. To compare the performance of the wild bootstrap method under different parameter settings, both quantities were normalized as the percentage of the "gold standard" from the MC simulations.

The bias quantifies the accuracy of the wild bootstrap method, and the standard error, a second-order statistic of wild bootstrap estimates, provides a measure of the variation or precision of the method. Both quantities served as the objective criteria for characterizing the robustness of the wild bootstrap method as an estimator for DTI measurement uncertainty. A third quantity, the root-mean-square error (RMSE), i.e. the root mean square of the bias and standard error, was also adapted as a measure for the overall performance of wild bootstrap estimates (Chung et al., 2006). A smaller RMSE value indicates a better match between wild bootstrap estimates and the "gold standard" from MC simulations.

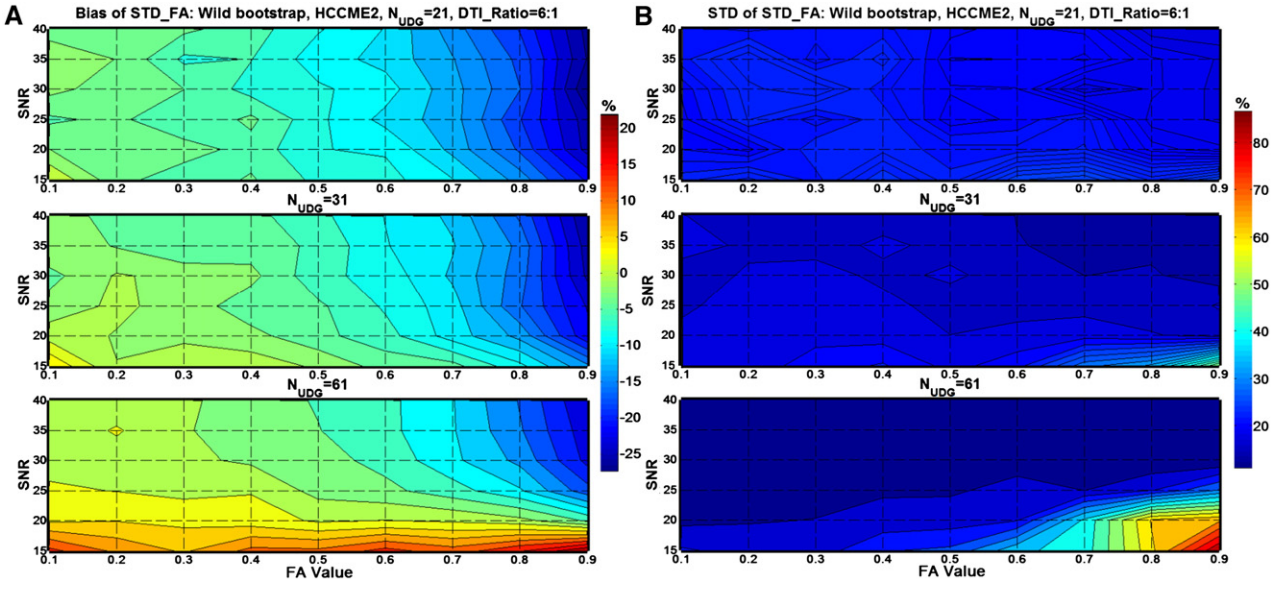


Fig. 2. Effects of the SNR level and the number of unique diffusion-weighting gradient (N_{UDG}) on the performance of the wild bootstrap method at different anisotropy levels with a prolate-shaped tensor. In this and all subsequent figures, the performance of the wild bootstrap method is quantified by bias [Eq. (13)] and std [Eq. (14)] that are wild bootstrap estimates normalized as the percentage of the gold standard MC simulation under the same experimental design. The performance of the wild bootstrap estimates is shown for FA (STD_FA) in (A, B) as the examples. The unit used in the color bars as well as those in the following figures is percentage (%), unless a special notation is added. Results from a simulation with HCCME2 and DTI_Ratio=6:1 are shown in this figure. In each subplot, there are three panels, representing the different N_{UDG} values: 21, 31 and 61, arranged in ascending order of N_{UDG} from top to bottom. Color maps are individually scaled for each subplot for better comparison of the wild bootstrap performance among different N_{UDG}. In Fig. 2A, the warm color (red) represents a less negative bias of wild bootstrap estimates while the cool color (blue) indicates a more negative bias. In Fig. 2B, the warm color (red) represents a larger standard error of repeated wild bootstrap estimates, i.e., less precision, while the cool color (blue) indicates a smaller value of standard error.

We further evaluated the dependence of wild bootstrap performance on tensor orientations by using the wild bootstrap procedure along 966 orientations uniformly distributed over a hemisphere (Jones, 2004a). Additional MC simulations along the same orientations were performed and summary statistics were then calculated. Three quantitative measures, bias($\boldsymbol{E}$) and std($\boldsymbol{E}$) given by Eqs. (13) and (14) as well as the RMSE were also calculated for the estimates from all orientations, and used for evaluation.

An optimized wild bootstrap implementation, which had the least overall estimation errors and performed equally well at different anisotropy levels, was proposed for clinical DTI applications.

Results

Although the oblate-shaped tensors were also evaluated with MC simulations for FA ranging between 0.1 and 0.4, the characteristics

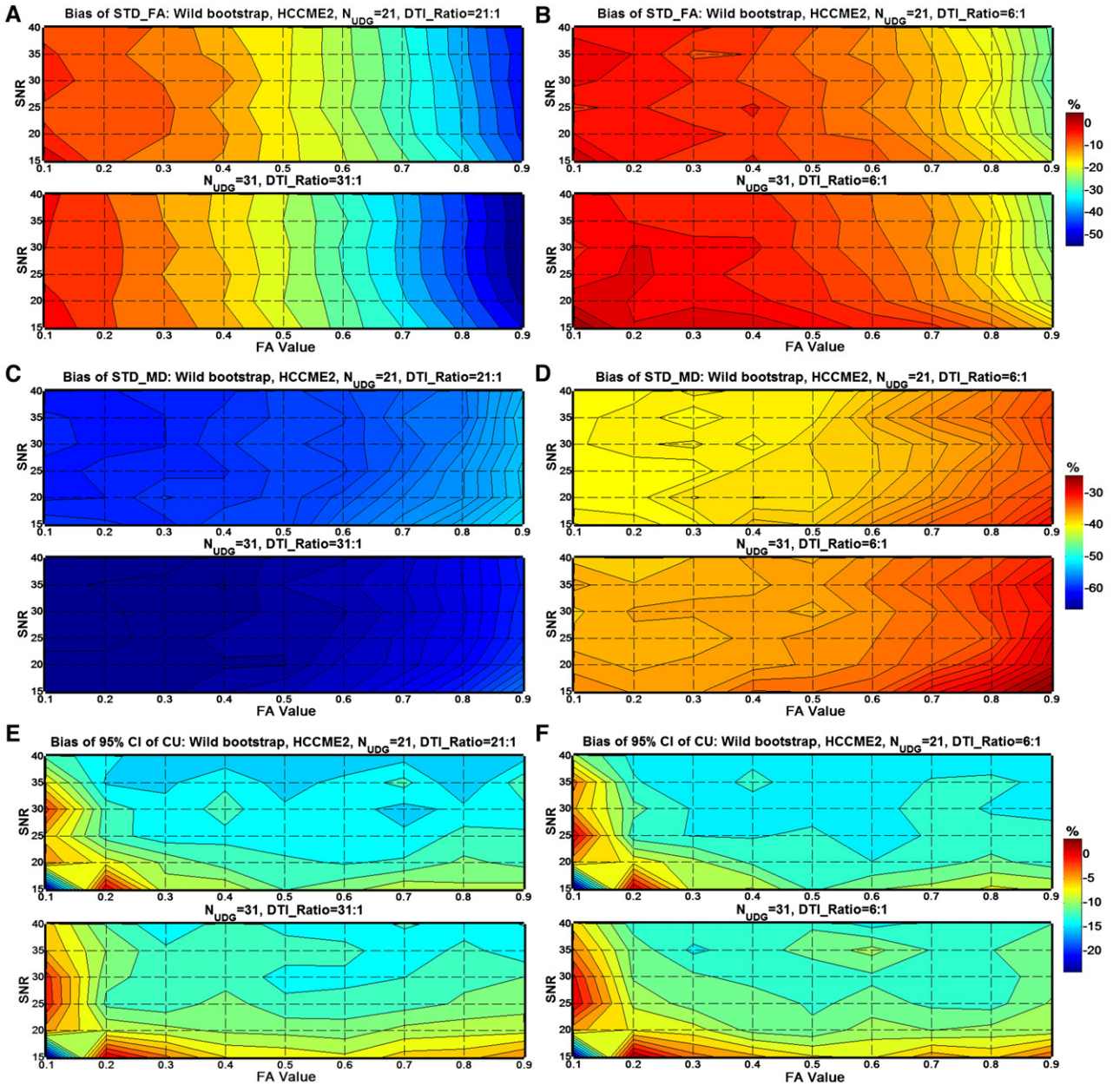


Fig. 3. Effects of the ratio of total DW images over total non-DW images (DTI_Ratio) on the performance of the wild bootstrap method at different anisotropy levels with a prolate-shaped tensor. The biases of the wild bootstrap estimates from simulations with HCCME2 and two different N_{UDG} values (=21, 31) are shown. In each subplot, there are two panels, representing the two different N_{UDG} values of 21 and 31. Plots in the left column are results with DTI_Ratio=N_{UDG}:1 while those in the right column are with DTI_Ratio=6:1. Bias values of the wild bootstrap estimates for measurement uncertainty at two DTI_Ratio settings are shown in (A, B) for FA (STD_FA), in (C, D) for MD (STD_MD), and in (E, F) for 95% CU. Color maps are individually scaled for each pair of subplots above for better illustration and comparison of the performance difference. The warm color (red) represents a less negative bias of wild bootstrap estimates while the cool color (blue) indicates a more negative bias.

of wild bootstrap performance were very similar to those of the prolate-shaped at the same anisotropy level, under the single tensor model used in this study. Therefore, only simulation results from the prolate-shaped tensors are presented here, unless a significant difference in performance existed between the two tensor types at certain anisotropy levels.

Among the five factors studied, the combined effect of SNR and N_{UDG} plays a major role affecting the precision of wild bootstrap estimate, as indicated in Fig. 2 for simulations at multiple SNR and anisotropy levels using HCCME2 function, N_{WBS} (=2000) wild bootstrap resamples and DTI_Ratio equal to 6:1. The performance of wild bootstrap estimation for the measurement uncertainty of FA (STD_FA) is selected as the example in Fig. 2B, while similar patterns of performance were also observed for STD_MD and 95% CU. The results show that a DTI protocol with either higher SNR level or more unique diffusion gradient directions yields more stable wild bootstrap estimates of DTI measurement uncertainties, as illustrated by the reduced standard errors of STD_FA in Fig. 2B. No substantial differences in precision were observed with different DTI_Ratio values or HCCME functions.

In contrast, the accuracy of wild bootstrap estimates measured by the bias is determined by multiple factors, particularly the combined effect from N_{UDG} and DTI_Ratio. Underestimation of wild bootstrap estimates for uncertainties is prominent (Figs. 2A and 3) as a negative bias (calculated from the repeated 1000 wild bootstrap runs) for all simulated protocols, except for STD_FA and 95% CU of P61 at low SNR level where an overestimation prevails. The plots in Fig. 3 demonstrate the benefit of using DTI_Ratio of 6:1. Although no substantial improvement was found for the precision of all of the three wild bootstrap estimates discussed above, the biases of wild bootstrap estimates for both STD_FA and STD_MD are reduced when DTI_Ratio of 6:1 is used. For example, at FA=0.9, the downward biases of STD_FA (Fig. 3A vs. B) and STD_MD (Fig. 3C vs. D) are reduced by approximately 20% and 25% respectively, indicated as the changes from cool color patterns (suggesting more negative values) in Figs. 3A and C to warm color patterns (suggesting less negative values) in Figs. 3B and D. The effect of N_{UDG} on the accuracy of wild bootstrap estimation is confounded by the DTI_Ratio factor, which is evident when the patterns of negative bias in Figs. 2 and 3 are compared. When DTI_Ratio is set to be 6:1, the use of a large N_{UDG} is advantageous, as indicated by the reduced level of negative bias in Figs. 2A and 3B and D. An opposite trend is observed for the N_{UDG} effect, when DTI_Ratio=N_{UDG}:1 is used in the simulation results from Figs. 3A and C. Figs. 3E and F show that improvements in accuracy of the estimation of 95% CU are always achievable with larger N_{UDG} no matter what DTI_Ratio value is

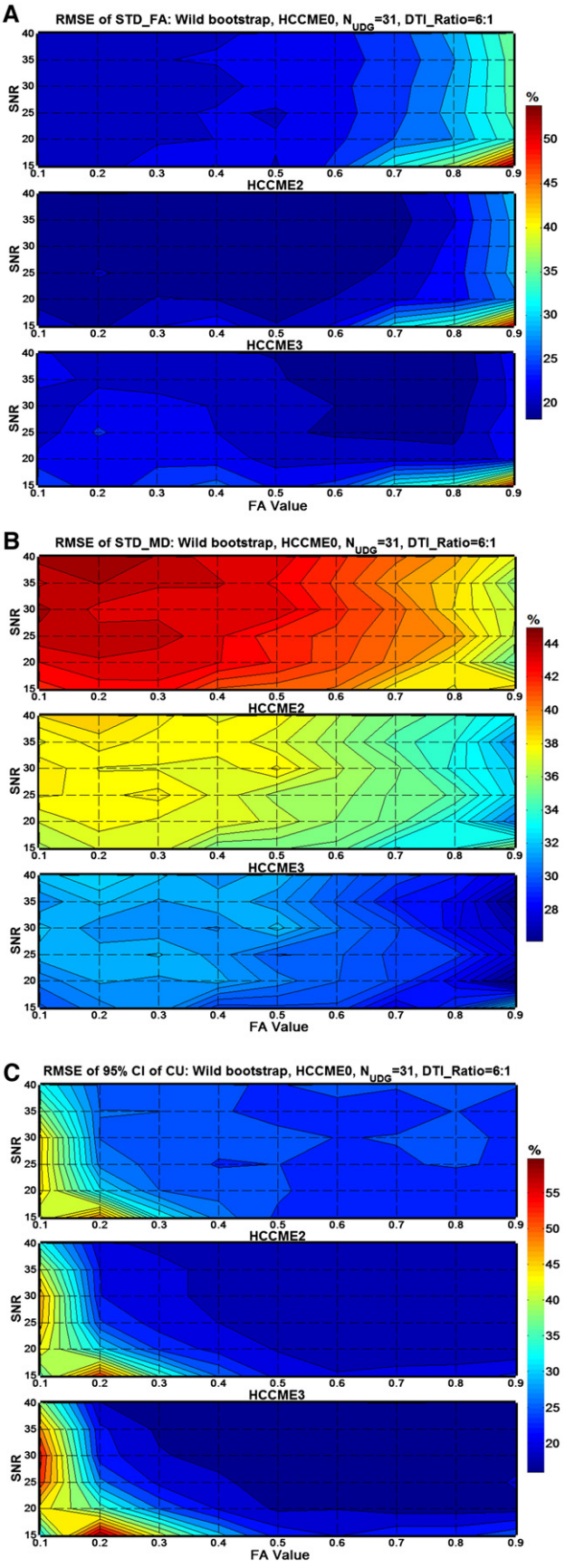


Fig. 4. Effects of different heteroskedasticity consistent covariance estimator (HCCME) functions on the performance of the wild bootstrap method at different anisotropy levels with a prolate-shaped tensor. RMSE values of the wild bootstrap estimates from simulations with DTI_Ratio=6:1 and N_{UDG}=31 are shown here as examples to demonstrate the effect of HCCME. The RMSEs of the wild bootstrap estimates are shown for FA (STD_FA) in (A), for MD (STD_MD) in (B), and for the measurement uncertainty of the principal diffusion direction (95% CU) in (C). In each subplot, there are three panels, representing the performance of three different HCCME functions. Color maps are individually scaled for each subplot for better illustration and comparison of the performance difference. The cool color (blue) represents a smaller RMSE value of wild bootstrap estimate, i.e. better overall performance, while the warm color (red) indicates a larger RMSE value.

used. Similar effects of DTI_Ratio and N_{UDG} were also observed with the other two HCCME functions (data not shown).

The performance of three HCCME functions for wild bootstrap implementation is illustrated in Fig. 4 using the RMSE as the criterion. Since no substantial difference between HCCME0 and HCCME1 was found in terms of performance, only results from HCCME0, HCCME2 and HCCME3 are compared. Results from wild bootstrap on protocol P31 with DTI_Ratio = 6:1 are shown. Wild bootstrap estimates with HCCME2 and HCCME3 similarly display less departure from the MC simulation results for all three uncertainty parameters when compared to those with HCCME0. When HCCME2 or HCMME3 instead of HCCME0 is used, improvements for estimates of STD_FA, STD_MD and 95% CU are 4%, 8% and 5%, respectively, on average. HCCME2 and HCCME3 have overall equivalent performance in wild bootstrap estimates of DTI uncertainty, although HCCME2 provides smaller RMSE values in estimation of STD_FA for less anisotropic tensors (FA < 0.6) while HCCME3 has better performance for the other two uncertainty parameters. Similar results were observed from the simulations on protocol settings with DTI_Ratio = N_{UDG}:1.

Fig. 5 demonstrates the impact of wild bootstrap iteration number, N_{WBS}, on the performance of wild bootstrap estimates using the RMSE from the estimates of 95% CU, as the example where the maximum effect from N_{WBS} was observed. Increasing the total iteration number improves the overall performance of wild bootstrap estimates, but the most substantial improvement of 9.5% for estimates of 95% CU is obtained when N_{WBS} increases from 250 to 2000 as shown in Fig. 5. A wild bootstrap resampling with a total number of 1000 gives very comparable estimates to those calculated from 2000 iterations. Thus, 1000 iterations are sufficient for most studies.

In comparison to MC simulation results from 966 tensors with uniformly distributed orientations, downward biases were also observed from the wild bootstrap method. For most of FA and SNR levels simulated, the bias [Eq. (13)] and standard error [Eq. (14)] of wild bootstrap estimations are similar among the tensors with

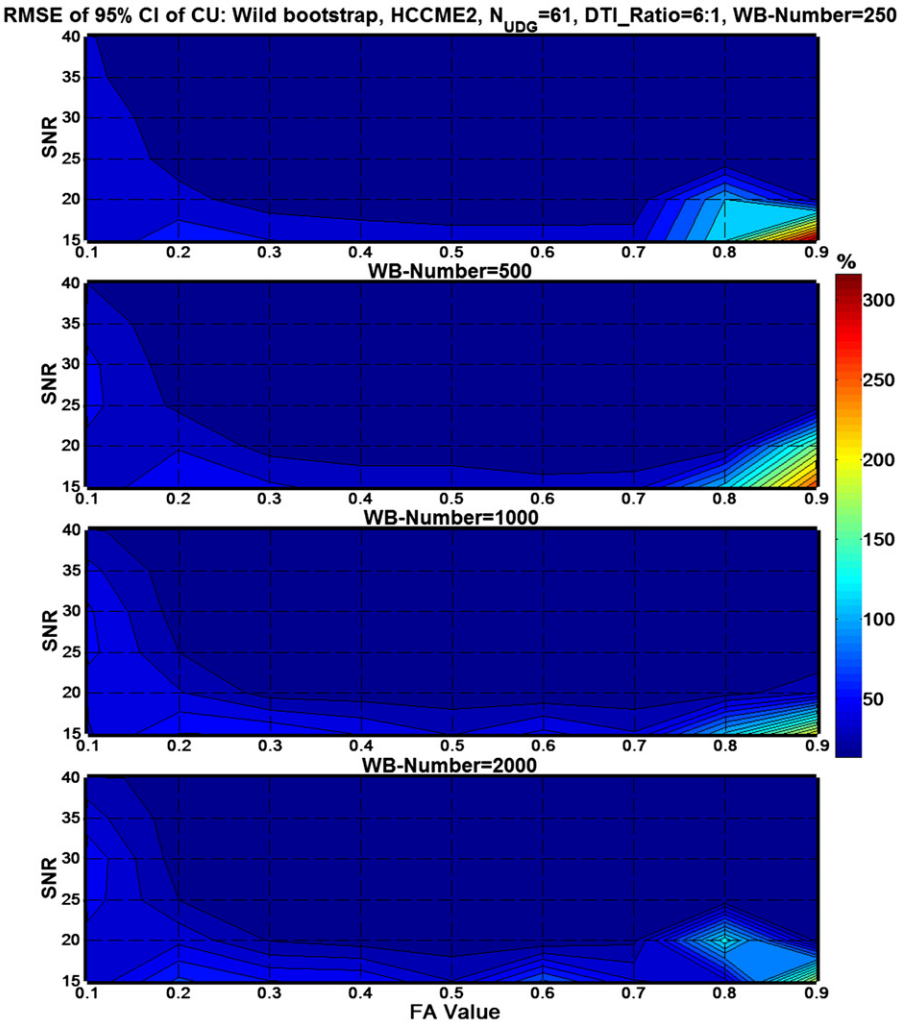


Fig. 5. Effects of the number of wild bootstrap iterations (N_{WBS}) on the estimate of measurement uncertainty of the principal diffusion direction (95% CU) at different anisotropy levels with a prolate-shaped tensor. The RMSE values of the wild bootstrap estimates from simulations with HCCME2, DTI_Ratio = 6:1 and N_{UDG} = 61 are shown. Four panels represent results from four different N_{WBS} value: 250, 500, 1000 and 2000. A common color map is used for all plots for better illustration and comparison of the performance difference. The cool color (blue) represents a smaller RMSE value of wild bootstrap estimate, i.e. better overall performance, while the warm color (red) indicates a larger RMSE value.

different tensor orientations. Therefore, there is no significant difference in the performance of wild bootstrap for different tensor orientations. This suggests that the wild bootstrap method will be suitable for uncertainty estimation in complex tissue structures for human brains where even the same structure may have different orientations in individual subjects. It should be noted that since the result from the fixed orientation simulation has shown that there is no substantial difference for N_{WBS} above 1000, in studying the tensor orientation effect, we only used a fixed value of $N_{WBS}=1000$. Additionally, in order to calculate the bias and standard error of the wild bootstrap estimation, we only repeated wild bootstrap procedure for 250 runs instead of 1000 runs for each orientation investigated.

Discussion

A robust and efficient method to evaluate the measurement uncertainty of DTI-derived parameters, such as FA and MD, can provide information to guide the design and selection of proper DTI protocols with minimal uncertainties, to achieve more accurate statistical inference by taking the measurement uncertainty level associated with different tissue types into consideration, and to provide more realistic probability distribution functions for a probabilistic fiber tracking algorithm. Evaluation of measurement uncertainties in real human DTI data obtained within a feasible clinical time is more challenging, when the conventional robust estimator such as the repetition bootstrap method is hindered by insufficient repeated DTI measurements. Following similar approaches as in the studies reported by Chung et al. (2006) and Whitcher et al. (2007), we further explored the validity of the wild bootstrap method as such an estimator for DTI measurement uncertainties under a variety of DTI protocol settings that are close to those commonly used in clinical DTI acquisition, and in the process we established the optimal implementation of the wild bootstrap analysis for DTI.

Wild bootstrap optimization

In the wild bootstrap implementation given by Eqs. (8)–(11), the heteroskedasticity consistent covariance matrix estimator functions were applied due to the non-constant variance structure of error terms in DTI tensor fitting (Whitcher et al., 2007). Results from Chesher and Jewitt's (1987) study showed that HCCME may potentially be downward-biased in finite samples, especially with the existence of high leverage points, i.e., the data points that substantially deviate from the average behavior in observations. Meanwhile, they also pointed out that a substantial upward bias is also possible depending on the level of heteroskedasticity of the original regression model. For a typical DTI protocol, the design of a diffusion-weighting gradient scheme, with a specific selection of N_{UDG}, DTI_Ratio and the spatial orientations of diffusion gradients, directly affects the heteroskedasticity of the regression model as well as the occurrence and severity of leverage point. The performance of wild bootstrap is affected by the combined effects from all these factors. For example, the increased SNR may reduce the occurrence of high leverage points and the level of heteroskedasticity in fitting errors, while increased N_{UDG} could increase the heteroskedasticity in DTI data. Therefore, in our numerical simulations, we investigated parameters not only from wild bootstrap implementation (N_{WBS} and HCCME type) but also from the DTI protocol (SNR, DTI_Ratio and N_{UDG}) to achieve a comprehensive evaluation of wild bootstrap.

We have shown that the wild bootstrap performance, when compared to MC results under the same simulation settings, is mainly determined by the combined effect of DTI_Ratio and N_{UDG}, particularly for the accuracy of wild bootstrap estimation. In addition to the contribution from the intrinsic downward bias associated with the family of bootstrap methods (Efron and Tibshirani, 1993), the negative bias of wild bootstrap performance observed in this study may have resulted from the downward bias tendency of HCCME with the existence of high leverage points as noted above. In general, two common ways are available to reduce the influence of high leverage points on DTI tensor fitting: (1) obtaining multiple measurements per diffusion gradient direction as well as for the non-DW images and (2) using more diffusion gradient directions to achieve more uniform orientation sampling. The experimental design of DTI_Ratio and N_{UDG} in our study is an implementation of these two possible approaches. Of the two DTI_Ratio values evaluated in this study, utilizing DTI_Ratio=6:1 significantly improves the accuracy of the wild bootstrap method as indicated by the reduced negative bias in Fig. 3. The improvement may be attributed to the improved effective SNR in the baseline images under DTI_Ratio=6:1, which alleviates the influence of high leverage points. For example, with $N_{UDG}=31$, the number of baseline images obtained with DTI_Ratio=6:1 is approximately five times the number obtained with DTI_Ratio=N_{UDG}:1, when other DTI parameters and wild bootstrap parameters remain the same. Substantial improvement of SNR in the baseline signal will ultimately reduce the measurement error in ADC along each gradient direction [according to error propagation described in Eq. (5)]. Consequently, with reduced errors in log-transformed signal intensities, the biases from potential high leverage observations are reduced, which ultimately improve the downward bias.

N_{UDG} affects the performance of wild bootstrap and its effect is confounded by DTI_Ratio. In general, increasing N_{UDG} improves the accuracy of tensor fitting because the influence of a single high leverage observation is alleviated when more diffusion gradients are applied. However, our results (Fig. 3) show that this benefit from applying a larger N_{UDG} occurs only when a proper DTI_Ratio, such as DTI_Ratio=6:1, is used. As indicated in Figs. 3A and C with a non-optimal DTI_Ratio=N_{UDG}:1, the downward bias is actually smaller in the protocol with $N_{UDG}=21$ compared to the protocol with $N_{UDG}=31$. This is most likely due to a higher effective SNR value per gradient direction ($f_{scale}=1.69$), equivalent to more averages per direction in real DTI acquisition, achieved for $N_{UDG}=21$ and, consequently, the influence of high leverage observations was reduced in comparison to the protocol with $N_{UDG}=31$, where more diffusion directions were used with a lower effective SNR value at each direction ($f_{scale}=1.39$). In addition to the general downward bias pattern of the wild bootstrap method, upward bias patterns, as indicated in Fig. 2A, were also observed for all three uncertainty estimates for protocols with $N_{UDG}=61$ and DTI_Ratio=6:1. This is likely due to the increased heteroskedasticity in fitting errors with a larger N_{UDG}. The increase of heteroskedasticity with N_{UDG} is more pronounced at a low SNR level and causes the consequent upward tendency of HCCME to overwhelm the downward tendency of HCCME. In contrast to STD_MD and 95% CU whose downward bias tend to be constant at high SNR levels reflecting the intrinsic downward bias associated with all bootstrap family (Efron and Tibshirani, 1993), the bias in wild bootstrap estimation of STD_FA shows a strong dependence on anisotropy levels. This is likely due to the nonlinear nature of FA calculations.

One basic assumption of wild bootstrap is the symmetric distribution of errors with respect to the origin. In the wild bootstrap method applied to DTI, the symmetry assumption may be violated when SNR levels of the original DTI data are too low. In MR images, the noise follows Rician distribution (Gudbjartsson and Patz, 1995) and only approaches Gaussian distribution at high SNR levels. For most human white matter tissues without fiber crossing, FA ranges from 0.4 to 0.9 and the largest eigenvalue λ_1 ranges from 1.0 to 1.7×10^{-3} mm^2/s. For most clinical DTI studies, a SNR of 25 for non-DW images and a *b*-value of 1000 s/mm^2 are commonly used for image acquisition. Under these conditions, the worst SNR in a DW image is approximately 4:1 in highly anisotropic regions and the average SNR for most white matter is approximately 7:1, which makes noise distribution very close to a symmetric Gaussian distribution. The symmetry assumption could be invalid at fiber crossing regions where single tensor model fails. However, numerical simulations based on highly skewed error items from a χ^2 distribution (Davidson and Flachaire, 2001) showed that the performance of the wild bootstrap method with the Rademacher distribution [Eq. (8)], which is similar to the wild bootstrap implementation used in this study, is similar to the original wild bootstrap implementation with an intrinsic skewness correction proposed by Liu (1988). Therefore, wild bootstrap estimates are sufficiently robust for most human brain tissues.

As an additional note, the computing time is also a factor to be considered for a wild bootstrap implementation in DTI applications. The computing time increases when a larger N_{UDG} or N_{WBS} number is applied but does not differ between different HCCME functions. In our test using a Dell PC with 2G RAM, a P4 CPU at 3.2 GHz, the computing time for generating a wild bootstrap sample of $256\times256\times28$ volume data is 40 s and 60 s for protocols P21 and P31, respectively. Improvements in computing time may be achieved with parallel computing.

While the present work was in progress, Chung et al. (2006) reported a study of quantitative comparisons of four bootstrap methods based on numerical simulations, including a wild bootstrap implementation similar to the one reported here. Meanwhile, Whitcher et al. (2007) performed studies that compared wild bootstrap and repetition bootstrap analysis not only from Monte Carlo simulations but also, for the first time, from a human DTI dataset based on the empirical distributions of FA and the calculated principal eigenvector. The protocols of the three studies differ, for example, in that a constant SNR in DW signal was used for Chung's study for investigating the effect of N_{UDG} while the SNR was adjusted in the present study to keep the effective SNR values for protocols with different N_{UDG} approximately the same. Nevertheless, all three studies demonstrate the similar downward bias in wild bootstrap estimates because the bootstrap family intrinsically yields biased estimators. Other consistent results were also found in these three independent studies. For example, both Whitcher's study and this study show that utilizing larger N_{UDG} will, in general, result in more stable bootstrap estimates. These consistent results provide strong evidence for the validity of wild bootstrap as a robust estimator for measurement uncertainties with DTI-derived parameters. Besides these common findings, our study also reveals comprehensive effects of DTI protocol parameters, including SNR, DTI_Ratio and N_{UDG}, on the bias patterns of wild bootstrap estimates for STD_FA, STD_MD and 95% CU. Furthermore, we also focus on the optimization of wild bootstrap implementation for real clinical DTI applications. Therefore, parameters of the wild bootstrap implementation itself including N_{WBS} and HCCME function type were also assessed to serve the optimization purpose. In their study, Chung et al. also suggested that a modified residual bootstrap that generated bootstrap resamples from fitting errors under the assumption of similarity of all error distributions and with a proper adjustment of the non-constant variance structure of the fitting errors, was a better estimator than wild bootstrap for DTI measurement uncertainties. It will be interesting to compare these two model-based bootstrap methods using the same human data set where real error distribution functions in human DTI data are heteroskedastic with an unknown model under the influence of multiple noise sources like physiological noise and imaging artifacts. Based on the results presented in these three studies, a comprehensive picture of wild bootstrap in analysis of DTI measurement uncertainty emerges: wild bootstrap can, in general, serve as a robust estimator to provide empirical distributions of DTI-derived parameters. The reasonable robustness from a wild bootstrap implementation with 1000 bootstrap iterations makes it also practical for clinical applications. Deviation from the accuracy and precision of wild bootstrap estimates, such as the negative bias, can be further reduced by careful selections of the DTI protocol settings of the unique diffusion gradient numbers (N_{UDG}) and the proper ratio between the total number of baseline images and those of DW images (DTI_Ratio) in DTI protocols.

Limitations and future works

In this study, HCCME functions were used together with the wild bootstrap method to estimate the variance of tensor from the multiple linear regression procedure with heteroskedastic errors of unknown form (Flachaire, 2005a). The downward bias observed in our numerical simulations on wild bootstrap is similar to what was previously reported by Chesher and Jewitt (1987). A potential remedy for this bias is to use instead a Craig estimator, which has been suggested to be more robust and efficient than HCCME when used with small samples (Flachaire, 2005b). The performance of wild bootstrap together with Craig estimator merits further study.

Only a single tensor model was used in this study. This limits not only the accuracy of the interpretation of complex fiber structures when fiber crossing or partial volume effect exists within an image voxel but also the performance of the wild bootstrap estimation. Although the performance of the wild bootstrap estimation, in terms of RMSE value, is similar between the prolate tensor and the oblate tensor under the single tensor model with a b-value of 1000 s/mm^2 applied in this study, the results from the simulation in the presence of partial volume effect and with a *b*-value up to 3000 s/mm^2 (Chung et al., 2006) clearly showed the increased RMSE of wild bootstrap estimation and the better performance obtained by a modified residual bootstrap. Whether combining the abovementioned Craig estimator with wild bootstrap as well as using an advanced tensor model can improve wild bootstrap estimation in complex fiber structures will be a subject of future study.

In this study, we have evaluated the robustness and efficiency of the wild bootstrap method as an estimator for DTI measurement uncertainties from numerical simulations. A potential application of wild bootstrap is to verify the results from numerical simulations or analytical solutions for measurement uncertainties in DTI with real human data. Although studied by Jones et al. in 1999, the effect of DTI_Ratio has not been fully evaluated with different DTI protocol settings. A variety of different DTI_Ratio values have been used not only in reported studies that focused on optimizing the diffusion gradient schemes, but also in published clinical DTI

applications. Thus, it is crucial to achieve a better understanding of the effect of DTI_Ratio as an essential parameter for an optimized DTI protocol by quantitative evaluations of the performance of different clinical DTI protocols in terms of their measurement uncertainties from the wild bootstrap estimation. Another application of great interest is the integration of the wild bootstrap estimate with the probabilistic fiber tracking algorithm (Jones, 2006). In the probabilistic fiber tracking method (Parker and Alexander, 2003a), a voxel-wise probability density function (PDF) from DTI data is calculated to interpret the likely underlying multiple fiber structures. Parker et al. (2003b) utilized the Monte Carlo simulations to derive the PDF for noise-based uncertainty in the principal diffusion direction at each pixel, and successfully generated fiber tractography in brain regions with abundant fiber crossing, such as the uncinate fasciculus that connects the orbital frontal gyri with anterior portions of the temporal lobe. Instead of generating PDF from simulation, wild bootstrap can directly estimate the PDF of the principal diffusion direction from human DTI data and may potentially provide more realistic PDF for probabilistic fiber tracking.

Conclusion

The robustness and efficiency of a wild bootstrap method as an estimator for DTI measurement uncertainties has been evaluated by numerical simulations. In these simulations, the performance of wild bootstrap method has been studied for different implementation parameters and DTI protocol settings to achieve an optimized wild bootstrap implementation for DTI applications. Based on the numerical simulation results, an optimized wild bootstrap implementation for clinical DTI applications should use either the HCCME2 or the HCCME3 function together with at least 1000 wild bootstrap iterations. Further improvement with less bias can be achieved on protocols with extra averaging of non-DW baseline images, such as with a DTI_Ratio = 6:1, as verified in our study. Weighting factors from the numerical simulation results reported in this study can be applied to overcome the downward bias of wild bootstrap estimates to achieve more accurate analysis of real human DTI data.

Acknowledgments

The authors are grateful to Dr. Derek Jones and Dr. Brandon Whitcher for useful discussions of the wild bootstrap implementation in this study. The authors also appreciate the thoughtful suggestions from anonymous reviewers. The authors thank Dr. Sven Ekholm and Dr. Voyko Kavcic for useful suggestions, Dr. Robert Waag and Ms. Michelle Gaugh for editing and proofreading. This study was supported partially by the Schmitt Foundation.

References

Anderson, A.W., 2001. Theoretical analysis of the effects of noise on diffusion tensor imaging. Magn. Reson. Med. 46, 1174–1188.

Basser, P.J., Pajevic, S., 2000. Statistical artifacts in diffusion tensor MRI (DT-MRI) caused by background noise. Magn. Reson. Med. 44 (1), 41–50.

Basser, P.J., Mattiello, J., Le Bihan, D., 1994a. Estimation of the effective self-diffusion tensor from the NMR spin echo. JMR 103, 247–254.

Basser, P.J., Mattiello, J., Le Bihan, D., 1994b. MR diffusion tensor spectroscopy and imaging. Biophys. J. 66, 259–267.

Bevington, P.R., Robinson, D.K., 1992. Data Reduction and Error Analysis for the Physical Sciences. McGraw-Hill, Inc., New York, pp. 42–43.

Chesher, A., Jewitt, I., 1987. The bias of a heteroskedasticity consistent covariance matrix estimator. Econometrica 55, 1217–1222.

Chung, S., Lu, Y., Henry, R.G., 2006. Comparison of bootstrap approaches for estimation of uncertainties of DTI parameters. Neuroimage 33 (2), 531–541.

Davidson, R., Flachaire, E., 2001. The wild bootstrap, tamed at last. Working Paper IER#1000, Queen's University. http://russell-davidson.arts.mcgill.ca/articles/wild6-euro.pdf.

Efron, B., Tibshirani, R., 1993. An Introduction to the Bootstrap. Chapman and Hall, New York.

Flachaire, E., 2005a. Bootstrapping heteroskedastic regression models: wild bootstrap vs. pairs bootstrap. Computational. Statistics. &. Data. Analysis. 49 (2), 361–376.

Flachaire, E., 2005b. More efficient tests robust to heteroskedasticity of unknown form. Econometric. Reviews. 24 (2), 219–241.

Gudbjartsson, H., Patz, S., 1995. The Rician distribution of noisy MRI data. Magn. Reson. Med. 34 (6), 910–914.

Hasan, K.M., Parker, D.L., Alexander, A.L., 2001. Comparison of gradient encoding schemes for diffusion-tensor MRI. J. Magn. Reson. Imaging. 13, 769–780.

Hasan, K.M., Alexander, A.L., Narayana, P.A., 2004. Does fractional anisotropy have better noise immunity characteristics than relative anisotropy in diffusion tensor MR? An analytical approach. Magn. Reson. Med. 51, 413–417.

Heim, S., Hahn, K., Samann, P.G., Fahrmeir, L., Auer, D.P., 2004. Assessing DTI data quality using bootstrap analysis. Magn. Reson. Med. 52, 582–589.

Jones, D.K., 2003. Determining and visualization uncertainty in estimates of fiber orientation from diffusion tensor MRI. Magn. Reson. Med. 49, 7–12.

Jones, D.K., 2004a. The effect of gradient sampling schemes on measures derived from diffusion tensor MRI: a Monte Carlo study. Magn. Reson. Med. 51, 807–815.

Jones, D.K., 2006. Tractography gone wild: probabilistic tracking using the wild bootstrap. In Proceedings of the 14th Annual Meeting of ISMRM, Seattle, p. 435.

Jones, D.K., Basser, P.J., 2004b. "Squashing peanuts and smashing pumpkins": how noise distorts diffusion-weighted MR data. Magn. Reson. Med. 52, 979–993.

Jones, D.K., Pierpaoli, C., 2005. Confidence mapping in diffusion tensor magnetic resonance imaging tractography using a bootstrap approach. Magn. Reson. Med. 53, 1143–1149.

Jones, D.K., Horsfield, M.A., Simmons, A., 1999. Optimal strategies for measuring diffusion in anisotropic system by magnetic resonance imaging. Magn. Reson. Med. 42, 515–525.

Kingsley, P.B., 2006. Introduction to diffusion tensor imaging mathematics: Part III. Tensor calculation, noise, simulation and optimization. Concepts in Magn. Reson. Part. A. 28A (2), 155–179.

Lazar, M., Alexander, A.L., 2005. Bootstrap white matter tractography (BOOT-TRAC). Neuroimage 24, 524–532.

Liu, R.Y., 1988. Bootstrap procedures under some non-IID models. Ann. Stat. 16, 191–198.

MacKinnon, J.G., White, H., 1985. Some heteroskedasticity consistent covariance matrix estimator with improved finite sample properties. J. Econometrics 29, 305–325.

O'Gorman, R.L., Jones, D.K., 2006. Just how much data need to be collected for reliable bootstrap DT-MRI? Magn. Reson. Med. 56, 884–890.

Pajevic, S., Basser, P.J., 2003. Parametric and non-parametric statistical analysis of DT-MRI data. J. Magn. Reson. 161, 1–14.

Papadakis, N.G., Murrills, C.D., Hall, L.D., Huang, CL-H., Carpenter, T.A., 2000. Minimal gradient encoding for robust estimation of diffusion anisotropy. Magn. Reson. Imaging. 18, 671–679.

Parker, G.J.M., Alexander, D.C., 2003a. Probabilistic Monte Carlo based mapping of cerebral connections utilizing whole-brain crossing fiber

information. In: Taylor, C.J., Noble, A. (Eds.), Information Processing in Medical Imaging (IPMI'03). Lect. Notes. Comp. Sci., 2732, pp. 684–695.

Parker, G.J.M., Haroon, H.A., Wheeler-Kingshott, C.A.M., 2003b. A framework for a streamline-based probabilistic index of connectivity (PICo) using a structural interpretation of MRI diffusion measurements. J. Magn. Reson. Imag. 18, 242–254.

Pierpaoli, C., Basser, P.J., 1996. Toward a quantitative measurement of diffusion anisotropy. Magn. Reson. Med. 36, 893–906.

Poonawalla, A.H., Zhou, X.J., 2004. Analytical error propagation in diffusion anisotropy calculations. J. Magn. Reson. Imaging. 19, 489–498.

Skare, S., Hedehus, M., Moseley, M.E., Li, T.Q., 2000. Condition number as a measure of noise performance of diffusion tensor data acquisition schemes with MRI. J. Magn. Reson. 18, 671–679.

Whitcher, B., Tuch, D.S., Wisco, J.J., Sorenson, A.G., Wang, L., in press. Using the wild bootstrap to quantify uncertainty in DTI. Hum. Brain Mapp. doi:10.1002/hbm.20395.