Magnetic Resonance in Medicine 68:1654–1663 (2012)

Informed RESTORE: A Method for Robust Estimation of Diffusion Tensor from Low Redundancy Datasets in the Presence of Physiological Noise Artifacts

Lin-Ching Chang,[1]* Lindsay Walker,[2,3] and Carlo Pierpaoli[3]

Physiological noise artifacts, especially those originating from cardiac pulsation and subject motion, are common in clinical Diffusion tensor-MRI acquisitions. Previous works show that signal perturbations produced by artifacts can be severe and neglecting to account for their contribution can result in erroneous diffusion tensor values. The Robust Estimation of Tensors by Outlier Rejection (RESTORE) method has been shown to be an effective strategy for improving tensor estimation on a voxel-by-voxel basis in the presence of artifactual data points in diffusion-weighted images. In this article, we address potential instabilities that may arise when using RESTORE and propose practical constraints to improve its usability. Moreover, we introduce a method, called *informed* RESTORE designed to remove physiological noise artifacts in datasets acquired with low redundancy (less than 30–40 diffusion-weighted image volumes)—a condition in which the original RESTORE algorithm may converge to an incorrect solution. This new method is based on the notion that physiological noise is more likely to result in signal dropouts than signal increases. Results from both Monte Carlo simulation and clinical diffusion data indicate that *informed* RESTORE performs very well in removing physiological noise artifacts for low redundancy diffusion-weighted image datasets. Magn Reson Med 68:1654–1663, 2012. © 2012 Wiley Periodicals, Inc.

Key words: RESTORE; diffusion tensor; robust estimation; artifact; outlier; trace; fractional anisotropy

Diffusion tensor magnetic resonance imaging (DT-MRI) is increasingly used in clinical settings for its ability to depict white matter tracts and for its sensitivity to microstructural and architectural features of brain tissue in vivo (1). Diffusion tensor maps are typically computed by fitting the signal intensities from diffusion-weighted images (DWIs) as a function of their corresponding *b*-matrices (diffusion weighting factors) (2) according to the multivariate least-squares regression model proposed by Basser et al. (3). The multivariate least-squares regression model takes into account the signal variability produced by thermal noise by including the assumed signal variance as a weighting factor in the tensor fitting. However, signal variability in DWIs is influenced not only by thermal noise but also by spatially and temporally varying artifacts (4).

In clinical DT-MRI acquisitions, artifacts are common, especially those originating from cardiac pulsation in noncardiac-gated acquisitions and from subject motion when uncooperative patients or unsedated pediatric subjects are scanned (5). Previous work demonstrated that signal perturbations produced by such artifacts can be severe, and neglecting to account for their contribution can result in erroneous diffusion tensor values and increase the variance of tensor-derived quantities (6–10). The Robust Estimation of Tensors by Outlier Rejection (RESTORE) algorithm has been shown to be an effective method for improving tensor estimation on a voxel-by-voxel basis in the presence of artifactual data points in DWIs (10). However, occasionally, the RESTORE algorithm may converge to an incorrect solution, producing inaccurate values of tensor-derived quantities, which in turn affect the statistical analysis of DTI results (11,12). This situation may occur, for example, when RESTORE excludes too many data points, leaving an inadequate set of *b*-matrices for proper tensor fitting. These "ill-defined" *b*-matrices could result from poor data quality or, simply because of an underestimation of the expected "artifact-free" signal standard deviation (SD; which is an a priori parameter needed by the algorithm). An incorrect solution also may arise when there are more bad data points than good. In this case, the iterative reweighting process may include the artifactual data points in the fitting and reject the good ones as outliers. Low data redundancy in the DWI dataset increases the chances of incorrect results from RESTORE. With a small DWI dataset, the possibility of ill-defined *b*-matrices or the likelihood of artifactual data points equaling or even outnumbering good data points increases. Although these corrupted results can often be identified and the corresponding voxels masked out (see the bright voxels in Fig. 3a), masking does introduce discontinuities that render statistical DTI analysis problematic (12).

In this article, we address RESTORE's potential pitfalls and propose practical constraints to improve its usefulness. Furthermore, we propose a new approach to remove artifacts caused by cardiac pulsation and subject motion from low redundancy DWI datasets. The new

[1]Department of Electronic Engineering and Computer Science, The Catholic University of America, Washington, District of Columbia, USA.
[2]Center for Neuroscience and Regenerative Medicine at the Uniformed Services University of the Health Sciences, Bethesda, Maryland, USA.
[3]National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland, USA.
Grant sponsors: The Department of Defense through the Center for Neuroscience and Regenerative Medicine, the Intramural Program of the Eunice Kennedy Shriver National Institute of Child Health and Human Development.
*Correspondence to: Lin-Ching Chang, D.Sc., Department of Electrical Engineering and Computer Science, The Catholic University of America, 620 Michigan Avenue, NE, Washington, DC 20064. E-mail: changl@cua.edu
Received 3 November 2011; revised 22 December 2011; accepted 3 January 2012.
DOI 10.1002/mrm.24173
Published online 27 January 2012 in Wiley Online Library (wileyonlinelibrary.com).
© 2012 Wiley Periodicals, Inc.

method, called *informed* RESTORE (*i*RESTORE), is based on the notion that physiological noise artifacts are more likely to result in signal dropouts than signal increases. We also develop a highly reliable strategy for assessing the expected artifact-free signal SD automatically from the data for the case of brain imaging. Both Monte Carlo simulation and clinical diffusion data are used to validate the proposed algorithms and to compare the performance of the nonlinear least-squares (NLS), RESTORE with added constraints, and *i*RESTORE methods.

METHODS

Constraints for the Improvement of the RESTORE Algorithm

The RESTORE method identifies voxel-wise data points as outliers and excludes them from the fitting. By removing points, it is possible to compromise the data by removing too many points with a particular gradient encoding, resulting in an imbalanced or ill-defined *b*-matrix. To improve RESTORE, a determination must be made as to how many data points or which data points can be excluded simultaneously from the tensor fitting without compromising the results. This determination is complex because it depends not only on the degree of data redundancy but also on the gradient sampling scheme (10) as well as on the final arrangement of the remaining gradient directions after the outlier rejection process. Thus, our first step in avoiding an ill-defined *b*-matrix is to compute the condition number (13) of the *b*-matrix after all identified outliers have been removed to ensure that the final *b*-matrix is still suitable for tensor fitting. However, even if not ill defined the *b*-matrix may get severely imbalanced due to the undersampling of some gradient directions. To avoid this problem, we implemented a second practical constraint. For the given gradient sampling scheme (G_{original}), we remove the directions that have been identified as outliers to obtain a new gradient scheme (G_{pruned}). We project the G_{original} and G_{pruned} directions (in unit-vector form) onto a predefined reference gradient sampling scheme (G_{ref}). G_{ref} defines the set of directions that the user chooses to have checked for potential undersampling. Given the antipodal nature of the diffusion process, we compute the vector projection after mapping all gradient directions onto the upper hemisphere. For each direction j of G_{ref}, we compute the sum d_j of the G_{original} vectors i projected on to j. The projection of direction i onto j is the vector d_{ij}, which has the same direction as j and length $|i| \times \cos\theta$, where θ is the angle between direction i and j, which can be computed using the dot product $\cos\theta = i \cdot j/(|i|\,|j|)$. The sum of projected vectors on direction j will be $d_j = \sum_{i=1}^{k} d_{ij}$, where k is the number of gradient directions used in G_{original}. It can be shown that if G_{original} samples the space uniformly, d_j is virtually constant, independent from the value of j and from the number of directions of G_{ref}, with a mean value of $k/2$. In other words, for a well-distributed gradient direction scheme, such as the repulsion model (14), d_j will be about 3 for six directions, 15 for 30 directions, 22.5 for 45 directions, etc. For other encoding schemes based on heuristic, numerically optimized, and/or regular polyhedra techniques (15), d_j should also be a constant. For a given G_{original}, we define its redundancy coefficient (RC) as $d_j/3$. With this definition, a six-direction gradient scheme has an associated RC of 1, a 30-direction scheme a RC of 5.

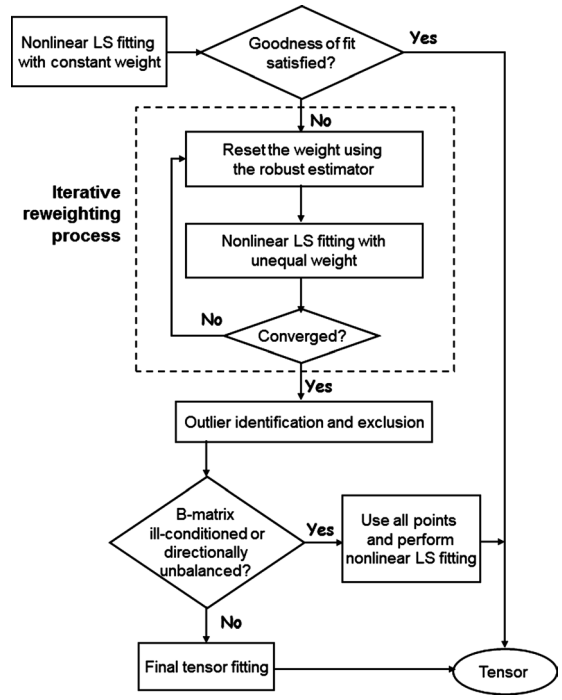

FIG. 1. Flow diagram of the RESTORE algorithm with added constraints.

With the same procedure used for G_{original}, d_j is computed for G_{pruned} for each direction j of the reference sampling scheme G_{ref}. In the case of G_{pruned}, its RC will be generally lower than the original RC and will also no longer be constant and independent of j. The user defines a threshold value for G_{pruned} RC that can range from 1 to G_{original} RC, when G_{pruned} RC is below this threshold value, the algorithm can no longer remove additional data point as outliers. This threshold RC obviously should have a value higher than 1 (otherwise the resulting *b*-matrix will be ill-defined). The choice of a correct threshold value for RC is very important. If set too high the likelihood of removing too many data points is reduced, but the elimination of outliers may be suboptimal, if it is higher than 1 but very close to 1, the risk of a very imbalanced *b*-matrix is increased. While the choice of RC threshold is very important, the choice of the set of gradients to be used as reference, G_{ref} is less relevant. The default gradient scheme that we chose as reference in this work is a six-direction scheme (16), because six is the minimum number of directions for tensor computation. One may consider choosing a 30-direction repulsion scheme (17), which ensures better rotational invariance of the statistical properties of the computed tensor quantities. A flowchart of the RESTORE algorithm with the added constraints is shown in Fig. 1.

Informed RESTORE

We propose a new method called *i*RESTORE with the goal of removing physiological noise artifacts from low

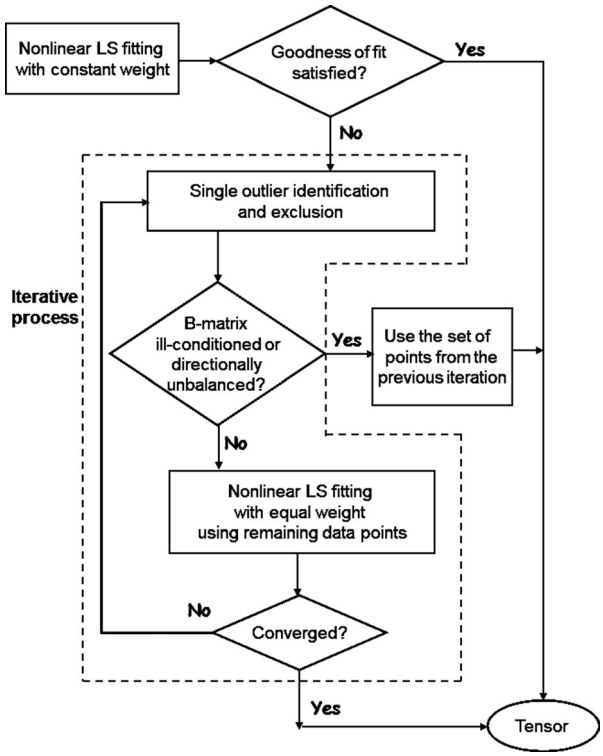

FIG. 2. Flow diagram of the *i*RESTORE algorithm.

redundancy DWI datasets. The new algorithm is based on the notion that physiological noise artifacts such as cardiac pulsation and subject motion are more likely to result in signal dropouts than signal increases. This new algorithm uses an iterative NLS method with constant weights. Outlier detection and exclusion is performed in a step-by-step approach in voxels where reduced chi-squares values are above a normal value and, therefore, the goodness of fit criterion is not satisfied (Selecting the Threshold Value of Reduced Chi-Square section). At the end of each iteration, a single data point with the maximum negative residual (i.e. the largest signal dropout from what is expected based on the fitting) is identified and excluded as an outlier from the next iteration. Then, the two constraints described in the previous section are applied to ensure that the resulting *b*-matrix at the end of the current iteration is not ill-defined or severely directionally imbalanced. The iterations continue until the fitting satisfies one of the following convergence criteria: (1) an acceptable value of reduced chi-squares is achieved (Selecting the Threshold Value of Reduced Chi-Square section), (2) the computed tensors remain unchanged compared to the previous iteration, or (3) the current iteration reaches a user-defined maximum number of excluded points. The proposed algorithm combines a greedy technique and a backtracking approach (18), i.e., making the locally optimal choice at each stage with the hope of finding the global optimum, and a minimal backtracking capability allowing the algorithm to turn back to the previous iteration to ensure a solution is found. A flow chart describing the *i*RESTORE algorithm is shown in Fig. 2.

Accurate Estimation of the Artifact-Free Signal Variability

Correctly estimating the expected signal SD due to thermal noise in artifact-free DWIs is critical to the success of the RESTORE and *i*RESTORE algorithms. This step is an important prerequisite for properly detecting outliers with RESTORE. If the value of the artifact-free signal SD, σ, is overestimated, outliers can be misidentified as good data. Conversely, if σ is underestimated, good data can be misidentified as outliers. The σ is also used in computing the value of the reduced chi-square, which is used to determine the goodness of fit and as a convergence criterion in the iterative process. For classical single-channel acquisitions, σ could be estimated from measurements of the signal SD in a ghost-free region of the image background (19). However, the signal in the background of DWIs acquired on modern clinical scanners cannot be used for this purpose because of the effects of signal processing and filtering applied during image reconstruction in data acquired with parallel imaging techniques (20).

In Ref. 12, Walker has proposed a method to improve the estimation of the expected signal SD due to thermal noise using a collection of reduced chi-square values measured in the object, not using background signal information. However, Walker's method requires that more than 50% of the selected voxels be artifact free. When a volume in the dataset is bad, the collected values of reduced chi-squares will all be corrupted, and the estimated signal SD severely biased. Here, we propose a new object-based method to estimate the signal SD. This method is robust to artifacts, fully automated, and is derived directly from the residual analysis of the tensor fitting of the available DWIs. For brain data, the method first selects data from regions that are known to be relatively immune from physiological noise artifacts (12), and then performs a robust assessment of the signal variability using estimators, such as the median absolute deviation (MAD) and the robust regression, that are relatively immune to the effect of outliers. Below, we briefly describe the underlying theory and the implementation of our approach.

Let us denote s_{ijk} as the standard error of the regression or the residual SD at voxel location (i, j, k). Using the residuals obtained from the fitting, we can estimate s_{ijk} based on a robust scalar estimator such as the MAD.

$$s_{ijk} = \text{median}\{|r_1 - \tilde{r}|, |r_2 - \tilde{r}|, \ldots, |r_n - \tilde{r}|\},$$

where n is the number of data points, r_1, r_2, ..., r_n are the residuals, and $\tilde{r}$ is the median of the residuals.

Based on the value of s_{ijk}, we then can estimate the value of $\hat{\sigma}_{ijk}$, the expected signal SD at voxel location (i, j ,k), using the following formula.

$$\hat{\sigma}_{ijk} = \sqrt{\frac{n}{n-p}} s_{ijk},$$

where n is the number of data points and p is 7 for the number of unknown parameters in the diffusion tensor estimation.

Using the sample median of a collection of $\hat{\sigma}_{ijk}$, we obtained the robust estimation of signal SD for the dataset. We denote the method derived from the residual

analysis using MAD as RMAD. Although RMAD improves Walker's method, it does not fully address the situation of one or more bad volumes in the dataset. To solve this problem, we propose adding a step before the residual analysis. We perform robust regression using the Geman–McClure M-estimator and remove a certain percentage of voxels based on a user-defined parameter, i.e., the user specifies the percentage of data to remove before tensor fitting and residual analysis. We denote the method using robust regression and residual analysis using MAD as RRMAD. See Appendix A for details on the steps involved in the various methods.

As indicated in the algorithms, we use the median of a collection of $\hat{\sigma}_{ijk}$ to improve the robustness of signal SD estimation. If artifacts would be randomly distributed spatially, the collection of $\hat{\sigma}_{ijk}$ could be sampled randomly from any position in the object. However, given that in the human brain, the occurrence of artifacts does not have a random spatial distribution we aim at selecting voxels from regions that have been shown to be least affected by physiological noise, specifically from white matter regions in the centrum semiovale rostral to the corpus callosum (12). The automatic segmentation method we used to select a region (or regions) of interest from a brain image is described in Appendix B.

Selecting the Threshold Value of Reduced Chi-Square

The selection of an appropriate reduced chi-square χ^2_{red} threshold (i.e., data can be considered "clean" from artifacts if the χ^2_{red} value is below the selected threshold) is important because it determines whether data from a voxel would enter the robust fitting procedure or not. The reduced chi-square value is also used as a convergence criterion in the iterative process for both the RESTORE and *i*RESTORE methods. Here, we address the subject of selecting a proper χ^2_{red} threshold for these purposes.

Chi-square, χ^2, has the probability distribution shown in Eq. 1, with mean equal to the number of degrees of freedom, $\nu = n - p$, where n is the number of data points and p is the unknown parameters, and variance equal to 2ν (21). By central limit theorem, when ν is large, the χ^2 distribution approximates a gaussian distribution with mean ν and SD $\sqrt{2\nu}$.

$$P(\chi^2, \nu) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)}(\chi^2)^{\nu/2-1}e^{-\chi^2/2},\ 0 < \chi < \infty \qquad [1]$$

Given that $\chi^2_{\text{red}} = \chi^2/\nu$, the χ^2_{red} distribution has mean 1 and variance $2/\nu$. When ν is large, the χ^2_{red} distribution also approximates a gaussian distribution, and its SD is equal to $\sqrt{2/\nu}$. Therefore, the value of χ^2_{red} from the fitting with artifacts-free data should fall within the following range if a 99% confidence interval is used:

$$1 - 3 \times \sqrt{2/\nu} \le \chi^2_{\text{red}} \le 1 + 3 \times \sqrt{2/\nu} \qquad [2]$$

In other words, the threshold value of χ^2_{red} is a function of the degrees of freedom and should be set to $1 + 3\sqrt{2/\nu}$ for a 99% confidence interval or set to $1 + 2.5\sqrt{2/\nu}$ for a 95% confidence interval. For example, for a dataset with 28 images (six gradient directions plus one b_0 image and with four repeats), the calculated threshold value of χ^2_{red} is 1.93 for a 99% confidence level, or 1.77 for a 95% confidence level; similarly, for a dataset with 120 images (110 gradient directions plus 10 b_0 images), the threshold value of χ^2_{red} is 1.40 or 1.33.

Single Tensor Simulation for Evaluation of RESTORE and *i*RESTORE

We performed Monte Carlo simulations to evaluate the effectiveness of the added constraints to the RESTORE algorithm, as well as to the newly proposed *i*RESTORE method. We simulated both an isotropic diffusion tensor and a cylindrically symmetric anisotropic diffusion tensor with diffusivity in the x direction set to five times the diffusivity in the y and z directions. The trace of both tensors was set to be representative of the trace of brain parenchyma (2100 $\mu m^2/s$). Two experimental designs were tested, both with 35 b-values (5 with $b = 0$ and 30 with $b = 1000$ s/mm^2), but with different diffusion sampling direction schemes. The first experimental design is the widely used six diffusion sampling directions scheme (16) and the second experimental design is 30 unique sampling directions (14). In this simulation, the threshold value of χ^2_{red} is set to $1 + 3\sqrt{2/28} \approx 1.80$ and the threshold value for testing the RC is set to 3. The simulation was performed similarly to what is described in Ref. 10.

For each experimental design and predefined tensor, we created synthetic diffusion-weighted signal intensity data conforming to the diffusion tensor model (3). Gaussian-distributed noise was then added in quadrature to the synthetic noise-free signal to achieve a signal-to-noise ratio of 25 in the b_0 data. The diffusion tensor was subsequently estimated using (1) NLS, (2) RESTORE with added constraints, and (3) *i*RESTORE methods. The experiment was repeated 12,288 times on these synthetic datasets to assess the distribution of tensor values in the presence of artifactual data points. The outliers were simulated by randomly corrupting two to six of the 30 DWIs by multiplying the original signal intensity values by a factor of 0.5. The distributions of Trace(D) and fractional anisotropy (FA) (17) were then computed.

Human Brain Simulation

We performed Monte Carlo simulations using synthetic data generated from a DTI acquisition of the human brain to evaluate and compare the effectiveness of the proposed methods for estimating signal SD. We collected a very high-quality DTI dataset in the brain of a healthy male volunteer as described in Ref. 22. Images were acquired with a DW-echo-planar imaging (EPI) sequence with $2 \times 2 \times 2$ mm^3 resolution and eight b-values ranging from 3 to 1200 s/mm^2. For each b-value, different directions were sampled following the "repulsion" scheme proposed by Jones et al. (14). The b-value (s/mm^2)/number of directions scheme was 3/3, 10/6, 65/10, 113/12, 350/16, 570/18, 850/20, and 1200/22. Each direction scheme was repeated seven times and 749 DWI volumes per slice location were acquired. The total scan time was about 4 h and 40 min. The data were processed through the standard TORTOISE DTI

processing pipeline (23): DWIs were corrected for motion and eddy current distortions (24–26) and EPI distortion (26,27), and tensors were computed using the NLS method. We considered the computed diffusion tensor of this dataset as an error-free measurement of water diffusivity in the brain. Simulated human brain DWI datasets were created based on this error-free tensor and the single tensor model (3) using Jones' 30 direction scheme (14) with $b = 1000$ s/mm^2 and five $b = 0$ images for a total of 35 brain volumes. We added gaussian-distributed noise in quadrature to simulate images with the true signal SD = 500, i.e., an signal-to-noise ratio of 20 measured in the thalamus of the $b = 0$ image. We then corrupted one, two, or three images of the dataset by multiplying the original signal intensity by a factor ranging from 0.30 to 1.70. This simulation aims to investigate the robustness of the proposed signal SD estimation methods in presence of different types and severity levels of artifacts. We computed the signal SD using the (1) Walker, (2) RMAD, and (3) RRMAD methods. The user-defined parameter in RRMAD is the percentage of expected outliers in the dataset. As we have corrupted one, two, or three images out of 30 DWIs, the values of the parameter used in simulations are therefore set to be 3.3%, 6.6%, and 10%, respectively.

Evaluation on Clinical Human Brain Data Acquired with Low Data Redundancy

We visually assessed DT-MRI data acquired in the NIH MRI Study of Normal Brain Development (NIHPD; www.NIH-PediatricMRI.org). Data were acquired at 1.5 T on either a GE or a Siemens scanner, with six noncollinear directions at $b = 1000$ s/mm^2 and one $b = 0$ s/mm^2, repeated four times for a total of 28 brain volumes, with no averaging or cardiac gating. Field of view (FOV), matrix size, and slice thickness were adjusted to provide $3 \times 3 \times 3$ mm^3 voxel size, with either a matrix of 64×64 and FOV of 192 mm or a matrix of 128×128 with FOV of 384 mm. A minimum TR of 3 s was prescribed in the protocol, and TE was to be set at minimum achievable TE with full echo acquisition. These datasets are representative of clinical DTI acquisitions with low redundancy and physiological noise artifacts. We randomly selected 80 datasets from the NIHPD DTI database, 40 females and 40 males, ranging in age from 0.8 to 21.6 years (mean age 13.3 ± 4.8 years). Prior to tensor computation, DWIs were corrected for motion and eddy current distortions (24) and EPI distortion (27). Tensor fitting was performed three times on those datasets, once using NLS, once using RESTORE with added constraints, and once using the *i*RESTORE method. The threshold value of χ^2_{red} is set to $1 + 3\sqrt{2/21} \approx 1.93$ and the threshold value for testing the RC is set to 1. Directionally encoded color (DEC) maps derived from the tensor were assessed for the presence of cardiac pulsation artifacts and any difference among the NLS, RESTORE with added constraints, and *i*RESTORE methods were noted.

RESULTS

Figure 3 shows the FA map of a selected slice of clinical human brain data computed with (a) the original RESTORE and (b) RESTORE with added constraints. The bright voxels with artifactually high anisotropy, indicated by arrows in (a), are caused by excluding too many DWI data points from the fitting. By adding the proposed two new constraints to the outlier rejection process, those bright points are successfully eliminated, as seen in (b).

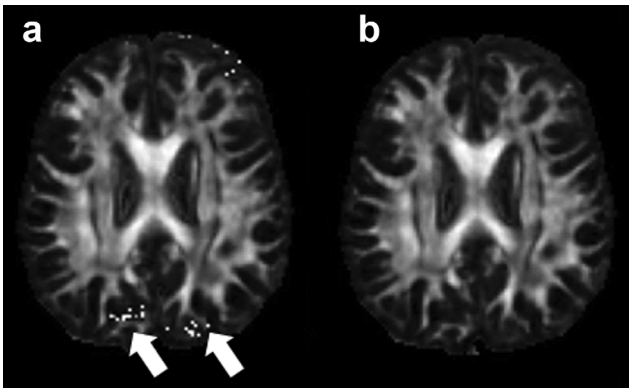

FIG. 3. The estimated FA of a selected slice of human brain using (**a**) original RESTORE algorithm and (**b**) RESTORE algorithm with added constraints. The bright voxels indicated by white arrows are caused by excluding too many DWI data points from the fitting.

Figure 4 shows the standard error of estimated signal SD using the Walker, RMAD, and RRMAD methods. The standard error is defined as the ratio of estimated signal SD minus true signal SD, divided by the true signal SD. Walker's method did not perform well in the presence of artifacts with the standard error over 50% in the case of a single corrupted volume and about 80% and 120% in the case of two and three corrupted volumes, respectively. The RMAD method improved the signal SD estimation, and RRMAD outperformed the Walker and RMAD methods regardless of the types and the severity levels of artifacts. Note that when there are no artifacts in the dataset, the Walker and RMAD methods perform better than RRMAD; RRMAD results in a slightly underestimated signal SD. However, the standard error is acceptable within about 3%, 5%, and 10% underestimation of signal SD for the case of one, two, and three corrupted volumes, respectively. This situation arises because even when the dataset is free of artifacts, the RRMAD method still identifies points in the tail of the residual distribution as outliers and removes them from the fitting.

Figures 5 and 6 show the distributions of Trace(D) and FA values for an isotropic and an anisotropic tensor, respectively, for the Jones-30 direction scheme and five b_0 images with a total of 35 images. The tensor was computed using three different methods: the widely used NLS, RESTORE with added constraints, and *i*RESTORE. The red curve here is a reference indicating the true Trace(D) and FA distributions when there are no outliers in the DWI volumes. The percentage of outliers (in DWIs) is set to 20% (six outliers) with the outliers' intensity values set to the original signal intensity values multiplying by 0.5. The true value (expected mean) of Trace(D) is 2100 μm^2/s and the true value of FA is 0.77. The results of the NLS fitting are severely affected by the

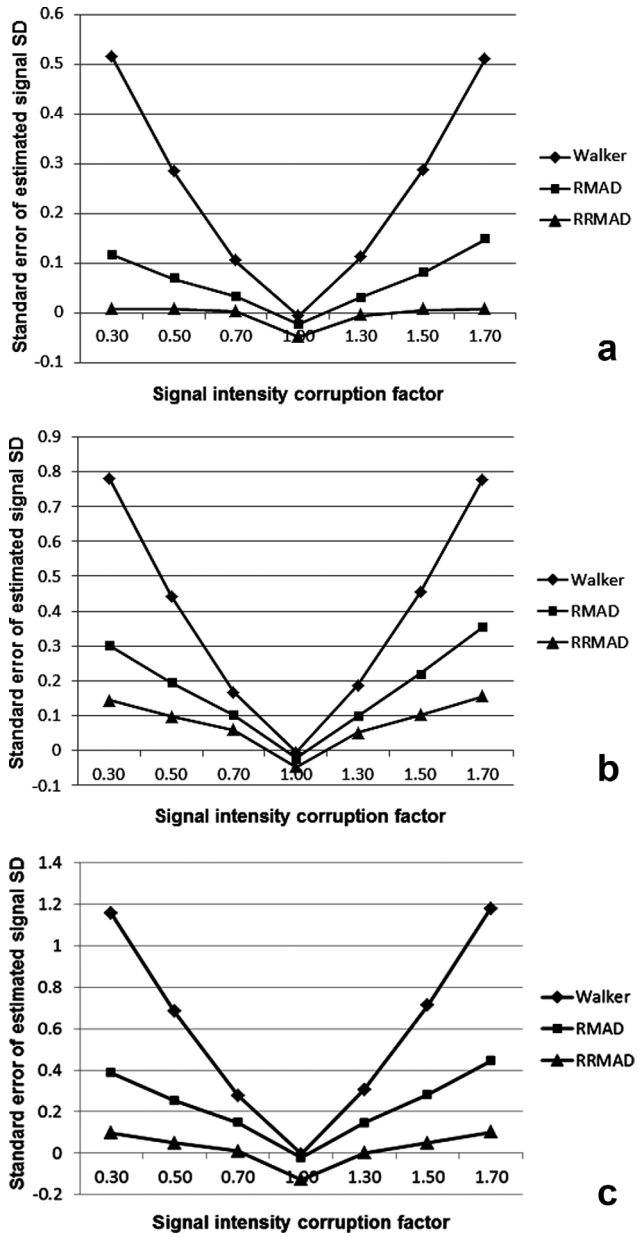

FIG. 4. The standard error of computed signal SD using the Walker, RMAD, and RRMAD methods with (**a**) one corrupted image, (**b**) two corrupted images, and (**c**) three corrupted images. The corrupted signal intensity values are set to be the original signal intensity values multiplied by the corruption factors. The different corruption factors ranging from 0.30 to 1.70 simulated different types and severity level of artifacts.

presence of outliers in the Trace(D) and FA distributions for both isotropic and anisotropic tensors (the blue curves). Using RESTORE with added constraints improves the results (light green curves), while *i*RESTORE (dark green curves) achieves a distribution, which is closest to that obtained with no artifacts (red curves).

Visual inspection of FA and DEC maps (28) of the pediatric human brain DTI datasets acquired with a low degree of data redundancy (28 DWI volumes), and no cardiac gating, from the NIHPD also indicates the effectiveness of the *i*RESTORE method for removing physiological noise artifacts. Compared with data processed with NLS, the data processed with RESTORE were scored as improved in 21% of the cases, unmodified in 74% and worsened in 5% of the cases. The data processed with *i*RESTORE were scored as improved in 51% of the cases, unmodified in 46%, and worsened in 3% of the cases. Of the 80 datasets, only 58 contained suspected physiological noise artifacts according to the criterion by Nayak et al. (29). Of these 58 datasets, *i*RESTORE improved 69%. Figures 7 and 8 show one extreme example from the NIHPD data with severe cardiac pulsation artifacts to demonstrate the superior performance of the *i*RESTORE fitting over the traditional NLS and the original RESTORE with added constraints. Figure 7 shows a slice with a region where cardiac pulsation artifacts corrupted three out of four DWIs in one of the six directions (75% corrupted data). The frequency of these severe artifacts with a low redundancy dataset causes the RESTORE algorithm to converge to an incorrect solution. Figure 8 shows the DEC maps obtained in the same slice using three different tensor fitting approaches. Cardiac pulsation is known to cause spurious anisotropy in the cerebellum (30), which is indicated by a white arrow on the DEC maps of the NLS and RESTORE with added constraints results in Fig. 8. However, the *i*RESTORE method successfully eliminates the effect of spurious anisotropy in this region, by removing voxels with signal dropouts in a targeted way.

DISCUSSION

In this article, we addressed potential pitfalls that may arise when using RESTORE on low redundancy diffusion datasets (i.e., less than ~40 DWIs), and proposed practical constraints to reduce the likelihood that RESTORE will converge to an incorrect solution. We also introduced a new method, called *i*RESTORE, which uses a priori information about the characteristics of physiological noise artifacts to effectively remove them from low redundancy diffusion datasets. The improvements introduced by both the additional constraints on the original RESTORE, as well as the new *i*RESTORE method, were demonstrated by Monte Carlo simulations and application on clinical brain DTI data.

The proposed additional constraints to the RESTORE algorithm are designed to avoid removing too many data points from the fitting, making the final b-matrix ill defined or severely directionally imbalanced. In setting these constraints, a balance needs to be found. If these constraints are too stringent, then RESTORE will behave as the nonlinear fitting and will not reject any points as outliers. If the constraints are too mild, then the likelihood of converging to an incorrect solution is still unacceptably high. First, we introduced the condition number (13) as a constraint, but this was found to be too mild, and can still result in a severely imbalanced b-matrix. For this reason, we introduced the second constraint, which consists of computing the projections on a minimal reference direction scheme of all the remaining b-matrices, and ensuring that no direction is severely undersampled. Both the direction scheme used as a

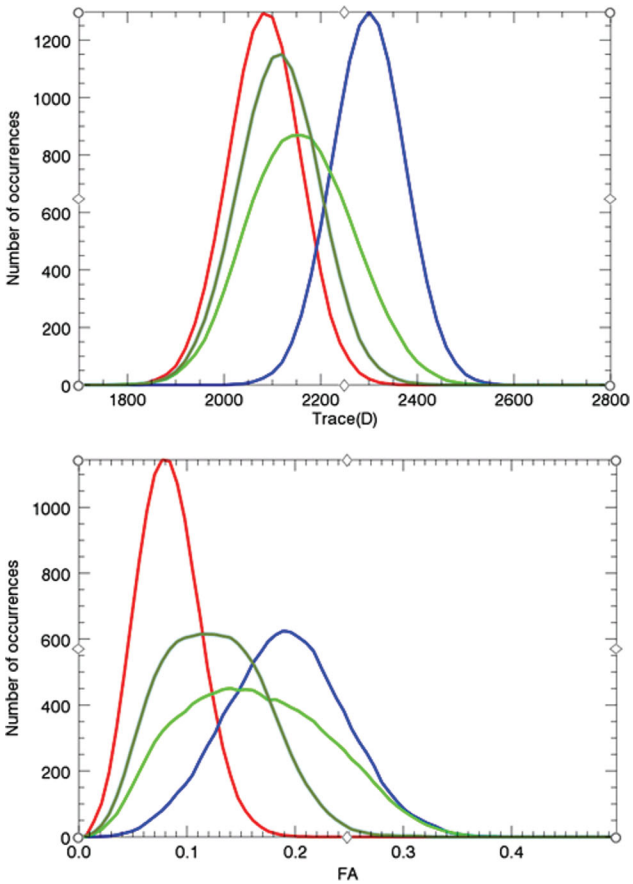

FIG. 5. Trace(*D*) and fractional anisotropy (FA) distributions for an isotropic diffusion tensor obtained from Monte Carlo simulated data with a 30-gradient direction scheme using the NLS method (blue), the RESTORE method with constraints (light green), and the *i*RESTORE method (dark green). Corrupted data points had their intensity values set to the original signal intensity values multiplied by 0.50. Outlier percentage is 20% for all plots. The red curve shows the true Trace(*D*) and FA distributions when there are no outliers in the DWIs.

reference and the threshold value of the sum of the projections can be set by the user. We found that a basic six-direction scheme (16) is appropriate as a reference, but the user may choose a larger number of directions. We provide guidance for determining reasonable values of the sum of the projections for any arbitrary gradient scheme chosen (Constraints for the Improvement of the RESTORE Algorithm section).

Any robust method relies on data redundancy and generally requires the number of good points to be higher than the number of outliers. As such, it should be noted that the two added constraints do not preclude RESTORE from converging to an incorrect solution when outliers outnumber good data points. The *i*RESTORE method proposed in this article addresses this problem by using a priori information about the characteristics of physiological noise artifacts and therefore allows the number of outliers to be greater than the number of good data points. Using low redundancy clinical data, we demonstrated that *i*RESTORE accomplished this goal. For example, as shown in Figs. 7 and 8, when 75% of the data in one gradient direction are corrupted by physiological noise, the *i*RESTORE algorithm produced correct tensor results, whereas the RESTORE method failed to do so.

In addition to the a priori information about physiological noise, the *i*RESTORE uses a different strategy than the RESTORE for the identification of outliers. While RESTORE identifies outliers from the fitting based on a confidence interval, the *i*RESTORE identifies outliers using a sequential heuristic algorithm, which uses what in algorithm design is called a "greedy" algorithm approach with a minimum "backtracking" capability (18). One advantage of this sequential heuristic approach is that it prevents the so-called masking effect (31), where an extreme outlier can mask other smaller outliers and prevent their identification, or where a few extreme outliers can totally distort the fitting causing RESTORE to converge to an incorrect solution. One may argue that the greedy algorithm may sometimes fail to produce an optimal solution. This can occur in some applications because the greedy algorithm does not operate exhaustively on all possible cases and by committing to certain

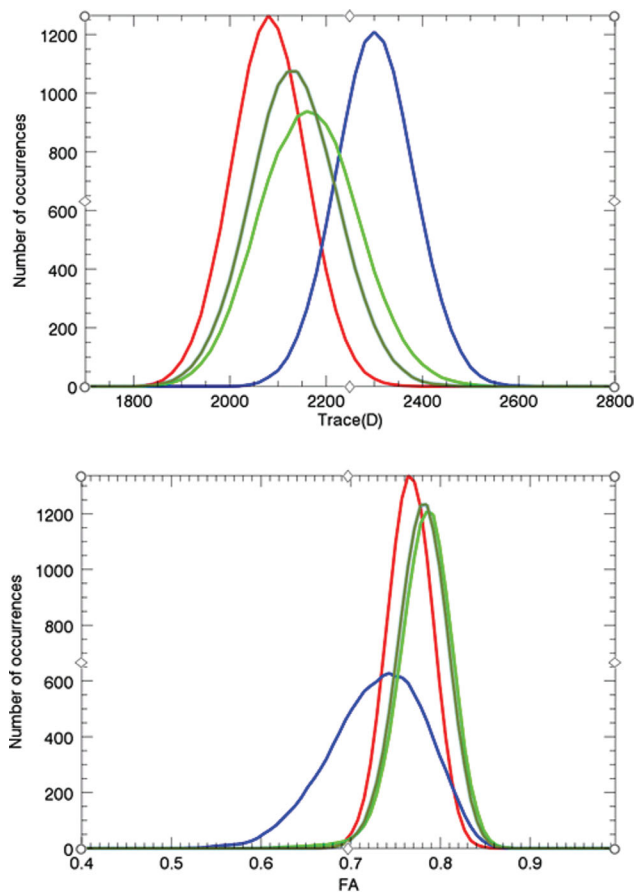

FIG. 6. Trace(*D*) and fractional anisotropy (FA) distributions for an anisotropic diffusion tensor obtained from Monte Carlo simulated data with a 30-gradient direction scheme using the NLS method (blue), the RESTORE method with constraints (light green), and the *i*RESTORE method (dark green). Corrupted data points had their intensity values set to the original signal intensity values multiplied by 0.50. Outlier percentage is 20% for all plots. The red curve shows the true Trace(*D*) and FA distributions when there are no outliers in the DWIs.

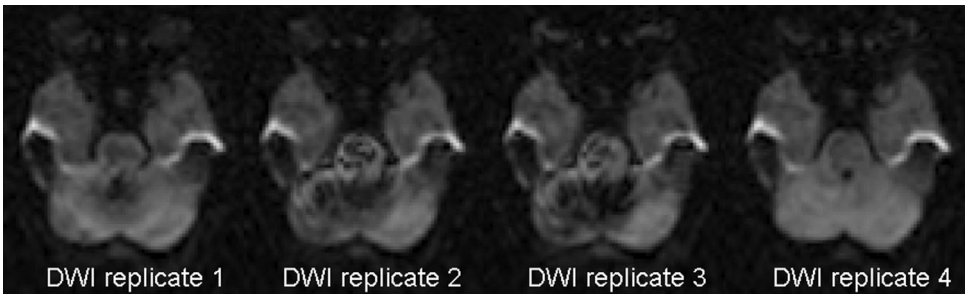

FIG. 7. A noncardiac-gated low redundancy DTI dataset using the basic six directions with four repeats. Physiological artifacts occurred in three repeats (replicate 1, 2, and 3), which is 75% of DWIs in one of the sampling directions.

choices too early, it fails to find the best overall solution later. However, we believe that, in our case, the greedy approach is likely to provide an optimal solution because our problem exhibits what is called an optimal substructure. For optimal substructure, we mean that the optimal solution for a larger subset includes already the optimal solution of a smaller subset. In our case, an optimal solution that identifies a number of outliers (say t) also represents an optimal solution that identifies $t - 1$ outliers. If the data points contain only one artifactual point, our algorithm is guaranteed to find that point. Another advantage of our algorithm is that it is very computationally effective compared to a fully backtracking strategy (18). A fully backtracking approach guarantees an optimal solution; however, the computation time grows exponentially with the number of DWIs, while our approach has a constant computation time. Unfortunately, the greedy approach, which sequentially removes the maximum absolute residual cannot be used to remove both spike and physiological noise artifacts (i.e. it cannot be used to remove artifacts both with signal increases as well as signal dropouts). Such a greedy algorithm will fail because removing spike and physiological noise artifacts iteratively does not exhibit an optimal substructure from iteration to iteration. Other outlier rejection algorithms have been proposed recently; they, too, target only specific types of artifacts (7,8).

One may wonder when it is appropriate to use RESTORE and when it is appropriate to use *i*RESTORE. RESTORE targets removal of generic artifacts, whereas *i*RESTORE targets removal of physiological noise artifacts and other artifacts that exhibit a signal dropout. The *i*RESTORE method outperforms the RESTORE method when the dataset contains a large number of physiological noise artifacts, and other artifacts resulting in signal dropouts, and when the dataset has low data redundancy (less than 30–40 DWIs). If the dataset has artifacts with increased signal intensity, such as those originating from spike noise, one should not use *i*RESTORE, because, by design, it will remove data with the most negative residual, and therefore it will not identify the voxels affected by the spikes as outliers. Even worse, this process will enhance the appearance of the spike artifact by removing the good points, and leaving the artifactual data points. From our experience, the improved RESTORE with additional constraints is always preferable to the original RESTORE method for processing clinical data, which may contain a mixture of artifacts with signal increase and/or signal dropout. In general, *i*RESTORE should be used if a visual inspection of raw images indicates that the data are primarily contaminated by artifacts that result in signal dropouts, and only in the case that the dataset has low data redundancy.

It should be noted that the robust estimation approaches in this work are only applied to the DWIs, not to the nondiffusion-weighted (b_0) images in the DWI dataset, as it was done in the original RESTORE algorithm. The reason for this is that while physiological noise artifacts are well characterized for DWIs (1,5,12), they are not as carefully investigated for the non-DWIs. For example, it could be expected that signal from vascular in-flowing spins, which are suppressed with high-diffusion weighting, may result in increased signal variability, with outliers that may have either increased or decreased signal intensity for the b_0 images. This issue requires further investigation; however, in absence of additional information about the noise characteristics of the b_0 images, as a first step, one could take a voxel-wise median of the b_0 images and use this value in the tensor fitting.

Both RESTORE and *i*RESTORE use the reduced chi-squares, χ^2_{red}, as a goodness of fit criterion to determine whether robust estimation should be performed in a voxel, and also as a convergence criterion. Therefore, the selection of an appropriate χ^2_{red} threshold is critical to the performance of the RESTORE and *i*RESTORE methods.

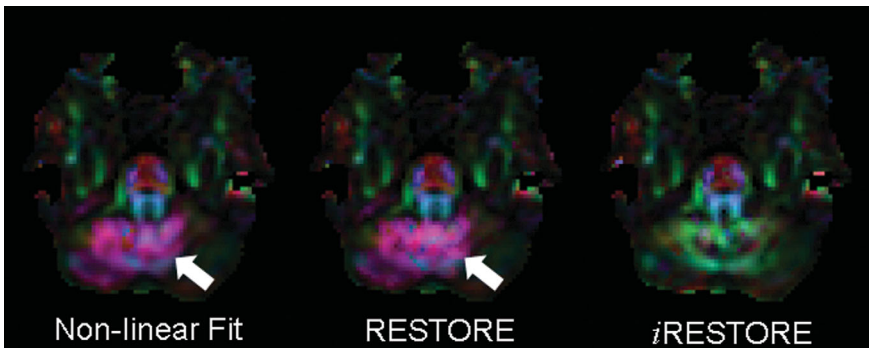

FIG. 8. The DEC map obtained from the NLS, RESTORE with added constraints and *i*RESTORE methods using the noncardiac-gated low redundancy DTI dataset shown in Fig. 7.

The computation of χ^2_{red} is intrinsically linked to a correct estimation of the expected signal SD due to thermal noise. Unfortunately, with the advent of parallel imaging and the consequent requirement of complex image reconstruction methods, it is no longer possible to extract meaningful noise information from the image background. Therefore, we resolved to extract the artifact free noise information from the object data itself (image-based methods). We tested a method previously proposed (12), and two newly proposed methods to accomplish this goal. We found that a method that does analysis of residuals in a brain region known to be free of physiological noise artifacts using the Geman–McClure iterative reweighting, the RRMAD method, proposed in the Accurate Estimation of the Artifact-Free Signal Variability section has the best performance. This method is fully automated and robust, even in the presence of DWI volumes, which are completely corrupted. The automatic region segmentation method we presented (see Appendix B) using probabilistic analysis of the spatial distribution of physiological noise artifacts in the human brain (12) is simple and straightforward, but other segmentation methods may be needed for other types of data, such as animal or body DTI data. In addition, our method assumes that the signal variability produced by thermal noise is a constant throughout the imaging volume. However, the signal variability may vary spatially with the multiple channel coils used for parallel imaging techniques; other estimation techniques (32,33) can be applied to assess the spatially varying signal SD.

Another aspect that we have reported in this work is that the distribution of reduced chi-squares, χ^2_{red}, is dependent on the number of degrees of freedom, and as such, the χ^2_{red} threshold should be computed on a case-by-case basis, based on the experimental design (i.e., the number of gradient directions used). We have derived a general formula that allows for the selection of an appropriate χ^2_{red} threshold based on the degrees of freedom of the fitting and a given confidence interval set by the user. This formula is suitable for a fully automated DTI processing pipeline.

All the methods presented in this article have been implemented in the software package TORTOISE (23) versions 1.2 and above. TORTOISE is a software package for preprocessing DWIs, for estimating the diffusion tensor in each voxel, and for computing tensor-derived quantities, which is available to the public and can be downloaded at www.tortoisedti.org.

Finally, although we applied the RESTORE and *i*RESTORE algorithms to diffusion tensor imaging, we believe both methods are also useful for other modalities of diffusion data analysis such as the HARDI type of analysis (34). Even if tensor fitting is not used to produce the final diffusion metrics with HARDI analysis, RESTORE or *i*RESTORE can be used to flag and to remove artifactual data points from the set of data prior to HARDI processing.

ACKNOWLEDGMENTS

The authors are grateful to the NIH MRI Study of Normal Pediatric Brain Development (www.NIH-PediatricMRI.org) for collecting and providing the clinical diffusion MRI data used in this paper. The authors also thank the Henry M. Jackson Foundation (HJF) for their administrative support and Ms. Liz Salac for editing the manuscript. Disclaimer: The views herein do not necessarily represent the official views of the National Institute of Child Health and Human Development, the NIH, the US Department of Health and Human Services, or any other agency of the US Government.

APPENDIX A

The details of Walker's method for estimating the signal SD due to thermal noise can be found in Ref. 12. A brief summary of the steps involved in Walker's method as well as in the RMAD and RRMAD methods proposed in this article is reported below.

Steps used in Walker's method:

- Step 1. Perform NLS tensor fitting on the selected voxels (i, j, k).
- Setp 2. Compute the sum of the residual squares of fit at each voxel location (i, j, k).

$$R^2_{ijk} = \sum_{e=1}^{n} r_e^2,$$

  where r_1, r_2, …, r_n are residuals obtained from NLS fitting.
- Step 3. Compute $\hat{\sigma}_{ijk} = \sqrt{\text{median}(R^2_{ijk})}$

The RMAD and RRMAD algorithms for estimating the signal SD due to thermal noise are described in detail in the Accurate Estimation of the Artifact-Free Signal Variability section.

Steps used in RMAD and RRMAD method:

- Step 0. (Applied to RRMAD only) Set the signal SD to a constant such as one. Perform robust regression by using the Geman–McLure M-estimator. Identify outliers based on the final residuals, and exclude a user-defined percentage of outliers. Use the remaining data points in step 1.
- Step 1. Perform NLS tensor fitting on the selected voxels (i, j, k).
- Step 2. Compute the residual SD using the robust MAD estimator $s_{ijk} = \text{median}\{|r_1 - \tilde{r}|, |r_2 - \tilde{r}|, \ldots, |r_n - \tilde{r}|\}$, where $\tilde{r} = \text{median}\{r_1, r_2, \ldots, r_n\}$ and n is the number of data points.
- Step 3. Compute $\hat{\sigma}_{ijk} = \sqrt{(n/(n-p))}s_{ijk}$, where $n - p$ is the degrees of freedom in NLS fitting.
- Step 4. Collect $\hat{\sigma}_{ijk}$ from the voxel locations (i, j, k) and compute the sample median of this collection of $\hat{\sigma}_{ijk}$.

APPENDIX B

To automatically select voxels located in the white matter of the centrum semiovale that should be relatively free from artifacts, we start by performing a brain extraction using BET (35) on one of the $b = 0$ images within the dataset. We then perform an erosion of the mask to

avoid including peripheral brain regions and we also exclude voxels from the lower two-thirds of all available slices (for axial images) to ensure that the selected voxels are from regions rostral to the corpus callosum. The inclusion of only white matter voxels is based on the fact that in the $b = 0$ images white matter has lower signal intensity than grey matter and cerebral-spinal-fluid. We compute the median value of the signal intensity of all candidate voxels and we include only voxels with signal intensities that fall between 80% and 90% of the median value. These voxels have a high likelihood of being white matter.

The algorithm to select a region (or regions) of interest from a brain image can be summarized as follows:

- Step 1. Extract the brain using BET.
- Step 2. Erode the brain to exclude peripheral regions.
- Step 3. Select slices from the top third of the brain (in axial images).
- Step 4.Compute the median value of b_0 signal within those slices: Median(b_0).
- Step 5. Select only voxels with signal intensities that fall between $0.8 \times$ Median(b_0) and $0.9 \times$ Median(b_0).

REFERENCES

1. Huppi P. Diffusion tensor MRI in brain development. In: Jones DK, editor. Diffusion MRI: theory, methods, and applications. Oxford University Press, USA; 2010.
2. Mattiello J, Basser PJ, Le Bihan D. The b matrix in diffusion tensor echo-planar imaging. Magn Reson Med 1997;37:292–300.
3. Basser PJ, Mattiello J, LeBihan D. Estimation of the effective self-diffusion tensor from the NMR spin echo. J Magn Reson B 1994;103: 247–254.
4. Tournier JD, Mori S, Leemans A. Diffusion tensor imaging and beyond. Magn Reson Med 2011;65:1532–1556.
5. Pierpaoli C. Artifacts in diffusion MRI. In: Jones DK, editor. Diffusion MRI: theory, methods, and applications. Oxford University Press, USA; 2010.
6. Morris D, Nossin-Manor R, Taylor MJ, Sled JG. Preterm neonatal diffusion processing using detection and replacement of outliers prior to resampling. Magn Reson Med 2011;66:92–101.
7. Zhou Z, Liu W, Cui J, Wang X, Arias D, Wen Y, Bansal R, Hao X, Wang Z, Peterson BS, Xu D. Automated artifact detection and removal for improved tensor estimation in motion-corrupted DTI data sets using the combination of local binary patterns and 2D partial least squares. Magn Reson Imaging 2011;29:230–242.
8. Zwiers MP. Patching cardiac and head motion artefacts in diffusion-weighted images. Neuroimage 2010;53:565–575.
9. Mangin JF, Poupon C, Clark C, Le Bihan D, Bloch I. Distortion correction and robust tensor estimation for MR diffusion imaging. Med Image Anal 2002;6:191–198.
10. Chang LC, Jones DK, Pierpaoli C. RESTORE: robust estimation of tensors by outlier rejection. Magn Reson Med 2005;53:1088–1095.
11. Chang LC, Walker L, Pierpaoli C. Making the robust tensor estimation approach: "RESTORE" more robust. In *ISMRM 17th Annual Meeting and Exhibition*, Honolulu, Hawaii, 2009. p 3557.
12. Walker L, Chang LC, Koay CG, Sharma N, Cohen L, Verma R, Pierpaoli C. Effects of physiological noise in population analysis of diffusion tensor MRI data. Neuroimage 2011;54:1168–1177.
13. Skare S, Hedehus M, Moseley ME, Li TQ. Condition number as a measure of noise performance of diffusion tensor data acquisition schemes with MRI. J Magn Reson 2000;147:340–352.
14. Jones DK, Horsfield MA, Simmons A. Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging. Magn Reson Med 1999;42:515–525.
15. Hasan KM, Parker DL, Alexander AL. Comparison of gradient encoding schemes for diffusion-tensor MRI. J Magn Reson Imaging 2001; 13:769–780.
16. Pierpaoli C, Jezzard P, Basser PJ, Barnett A, Di Chiro G. Diffusion tensor MR imaging of the human brain. Radiology 1996;201:637–648.
17. Basser PJ, Pierpaoli C. Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. J Magn Reson B 1996;111:209–219.
18. Cormen TH, Leiserson CE, Rivest RL. Introduction to algorithms, 3rd ed. The MIT Press; 2009.
19. Henkelman RM, Stanisz GJ, Kim JK, Bronskill MJ. Anisotropy of NMR properties of tissues. Magn Reson Med 1994;32:592–601.
20. Bammer R, Auer M, Keeling SL, Augustin M, Stables LA, Prokesch RW, Stollberger R, Moseley ME, Fazekas F. Diffusion tensor imaging using single-shot SENSE-EPI. Magn Reson Med 2002;48: 128–136.
21. Bevington P. Data reduction and error analysis for the physical sciences.New York, NY: McGraw-Hill Book Company; 1969.
22. Chang LC, Koay CG, Pierpaoli C, Basser PJ. Variance of estimated DTI-derived parameters via first-order perturbation methods. Magn Reson Med 2007;57:141–149.
23. Pierpaoli C, Walker L, Irfanoglu MO, Barnett A, Basser PJ, Chang L-C, Koay C, Pajevic S, Rohde G, Sarlls J, Wu M. TORTOISE: an integrated software package for processing of diffusion MRI data. In: ISMRM 18th annual meeting and exhibition, Stockholm, Sweden, 2010.
24. Rohde GK, Barnett AS, Basser PJ, Marenco S, Pierpaoli C. Comprehensive approach for correction of motion and distortion in diffusion-weighted MRI. Magn Reson Med 2004;51:103–114.
25. Leemans A, Jones DK. The B-matrix must be rotated when correcting for subject motion in DTI data. Magn Reson Med 2009;61:1336–1349.
26. Jones DK, Cercignani M. Twenty-five pitfalls in the analysis of diffusion MRI data. NMR Biomed 2010;23:803–820.
27. Wu M, Chang LC, Walker L, Lemaitre H, Barnett AS, Marenco S, Pierpaoli C. Comparison of EPI distortion correction methods in diffusion tensor MRI using a novel framework. Int Conf Med Image Comput Comput Assist Interv 2008;11:321–329.
28. Pajevic S, Pierpaoli C. Color schemes to represent the orientation of anisotropic tissues from diffusion tensor data: application to white matter fiber tract mapping in the human brain. Magn Reson Med 1999;42:526–540.
29. Nayak A, Walker L, Pierpaoli C, BDCG. Quality assessment in a DTI multicenter study. In: ISMRM 19th annual meeting and exhibition, Montreal, Canada, 2011. p 2342.
30. Pierpaoli C, Marenco S, Rohde GK, Jones DK, Barnett AS. Analyzing the contribution of cardiac pulsation to the variability of quantities derived from the diffusion tensor. In: ISMRM 11th annual meeting and exhibition, Toronto, Ontario, Canada, 2003. pm70.
31. Barnett V, Lewis T. Outliers in statistical data.New York: Wiley; 1985.
32. Landman BA, Bazin PL, Prince JL. Estimation and application of spatially variable noise fields in diffusion tensor imaging. Magn Reson Imaging 2009;27:741–751.
33. Ding Y, Chung YC, Simonetti OP. A method to assess spatially variant noise in dynamic MR image series. Magn Reson Med 2010;63: 782–789.
34. Tuch DS, Reese TG, Wiegell MR, Makris N, Belliveau JW, Wedeen VJ. High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity. Magn Reson Med 2002;48: 577–582.
35. Smith SM. Fast robust automated brain extraction. Hum Brain Mapp 2002;17:143–155.