Published in final edited form as:

Magn Reson Med. 2013 March 1; 69(3): 891–902. doi:10.1002/mrm.24324.

Assessment of Bias in Experimentally Measured Diffusion Tensor Imaging Parameters Using SIMEX

Carolyn B. Lauzon^{a,b}, Ciprian Crainiceanu^c, Brian C. Caffo^c, and Bennett A. Landman^{a,b}

^aDepartment of Electrical Engineering, Vanderbilt University, Nashville, TN, USA

^bInstitute of Imaging Science, Vanderbilt University, Nashville, TN, USA

^cDepartment of Biostatistics, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD, USA

Abstract

Diffusion tensor imaging (DTI) enables in vivo investigation of tissue cytoarchitecture through parameter contrasts sensitive to water diffusion barriers at the micrometer level. Parameters are derived through an estimation process which is susceptible to noise and artifacts. Estimated parameters (e.g., fractional anisotropy) exhibit both variability and bias relative to the true parameter value estimated from a hypothetical noise-free acquisition. Herein, we present the use of the SIMulation and EXtrapolation (SIMEX) approach for *post hoc* assessment of bias in a massively-univariate imaging setting and evaluate the potential of a SIMEX-based bias correction. Using simulated data with known truth models, spatially varying FA bias error maps are evaluated on two independent and highly differentiated case studies. The stability of SIMEX and its distributional properties are further evaluated on 42 empirical DTI datasets. Using gradient subsampling, an empirical experiment with a known true outcome is designed and SIMEX performance is compared to the original estimator. With this approach, we find SIMEX bias estimates to be highly accurate offering significant reductions in parameter bias for individual datasets and greater accuracy in averaged population-based estimates.

Keywords

DTI; bias; SIMEX; FA; MD; parameter estimation

INTRODUCTION

Thermal energy drives in vivo water to randomly explore its anatomical surroundings. Patterns in water diffusion therefore contain information regarding the cytoarchitecture and cellular organization of the surrounding tissue. The introduction of diffusion tensor imaging (DTI) enabled MRI to measure diffusion patterns (1,2) and non-invasively quantify anatomical connectivity information that was once only available through histology (3). Yet, the derivation of quantitative measures from observed DTI data is an estimation process, and numeric DTI parameters are only as quantitatively valuable as our ability to understand and interpret their distributional properties. Because of their importance, methods to investigate the distributional properties and ensure the quantitative accuracy of DTI parameters have been highly investigated (4–14). While the results of these studies have led to reasonable

understanding of the variability of parameters, bias at the individual voxel level still remains a ubiquitous and non-trivial confound in the interpretation of empirical DTI data (15).

Much computer simulation and empirical work has been invested to understand bias in DTI (4,11,12,15). From this work, it is known that bias as a function of signal-to-noise ratio (SNR) in DTI parameters is multifaceted. It is non-monotonic, dependent on acquisition sequence, sensitive to underlying anatomy, and currently has no analytic solution (non-invertible perturbation theory based expressions have been developed (4)). Although much work has been invested to theoretically understand bias in DTI parameters, surprisingly, no methods have been investigated for quantitative *post hoc* assessment of bias in experimentally measured DTI parameters. Part of the reason, besides the complex bias functions, may be that DTI is a log linear model, complicating the use of regression calibration (16), the most well-known statistical method for bias assessment and reduction.

SIMulation EXtrapolation (SIMEX) is an emerging and highly successful alternative to regression calibration. SIMEX was developed as a “completely general method” (17) for use with complicated models like DTI. As such, SIMEX theory has proven highly adaptable, including its initial use in logistic regression and its further extension into discrete (classification) settings (18) and multiplicative error models (16,19). SIMEX has been empirically validated in diverse applications including genetics, epidemiology, periodontics, and environmental engineering (20–23). Besides adaptability, SIMEX has the advantage of being a straightforward method that is accessible to non-statisticians. In SIMEX, the trend of an estimate when *synthetic noise is added* to an at-hand empirical dataset is recorded. An extrapolation of the trend in the estimate is then used to predict the value of the estimate at zero noise. Given that SIMEX is compliant with the log-linear DTI model and that the use of synthetic noise and computer simulation to evaluate bias is already established in the DTI literature, SIMEX is a promising method to evaluate bias in DTI parameters from empirical data.

The goal of this manuscript is to assess SIMEX as a method for estimating voxel-wise bias in DTI parameter estimates. First, the theoretical application of SIMEX to DTI is presented. Second we examine the properties of SIMEX using 30 simulated repeat scan DTI datasets with known ground truth. SIMEX performance is evaluated in individual and population-based settings, and cases of SIMEX failure are investigated. Finally, SIMEX is tested on a 21-subject empirical DTI dataset. To create an empirical test with a known true outcome, the gradient directions in the empirical data are sub-sampled to form a second empirical dataset. The true difference in FA values between both datasets should therefore be zero. The performance of SIMEX compared to the original estimator is tested for both individual and averaged trials in the empirical data. From the results herein, it is seen that SIMEX produces highly accurate estimates of bias for individual datasets and the SIMEX estimator, though more variable than the original estimator, outperforms the original estimator in averaged population-based studies.

THEORY

DTI and Bias

DTI has been described in detail in several reviews (1,24,25). Briefly, DTI is a method to measure the diffusion of water in vivo. Three-dimensional diffusion is described mathematically by a 3×3 diffusion tensor (**D**) that DTI estimates through data fitting procedures. For easy extension to the SIMEX method, the following describes the tensor fitting procedure from an estimation perspective. Consider a single spatial location. The observed data of a DTI experiment consist of $J - 1$ diffusion-weighted signal intensities (S_j),

each sensitized to diffusion along an orientation indexed by $j(g_j)$ as governed by the Stejskal-Tanner relation,

$$S_j = S_0 \exp(-b g_j^T \cdot \mathbf{D} \cdot g_j). \quad [1]$$

S_0 is the signal intensity without diffusion weighting. We define a *predictor vector* of noise-free values, $\mathbf{X} = [S_0, S_1, \dots, S_{J-1}]$, consisting of all observed magnitude imaging data for a particular spatial location. The imaging data are corrupted by Rician noise. (In the case of traditional frequency reconstruction the noise is Rician (13,26). Non-central chi may be a more appropriate model for other reconstructions but is not considered herein (27).) Because of the imaging noise, $\mathbf{X}$ is unavailable. Rather, the observed data are represented by the *observation vector* $\mathbf{W}$. For element $w_j \in \mathbf{W}$ and $x_j \in \mathbf{X}$,

$$w_j \sim \text{Rician}(x_j, \sigma_E^2). \quad [2]$$

Here σ_E is the scale parameter defined by the degree of experimental noise in the data. Practically, σ_E is not indexed by j as it is assumed to be constant for the same spatial location across multiple diffusion weightings. However this assumption is not theoretically necessary. The above description is for a single spatial location and herein σ_E is uniquely determined for each spatial location. We concern ourselves with the assessment of bias in estimates of fractional anisotropy (FA). FA is calculated from the three Eigenvalues of $\mathbf{D}$; λ_1 , λ_2 , and λ_3 .

$$\text{FA} = \sqrt{\frac{3}{2} \frac{(\lambda_1 - \bar{\lambda})^2 + (\lambda_2 - \bar{\lambda})^2 + (\lambda_3 - \bar{\lambda})^2}{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}} \quad [3]$$

To distinguish true from estimated quantities, we use a hat notation. For example $\hat{\mathbf{D}}$ and $\widehat{\text{FA}}$ represent estimates of the true $\mathbf{D}$ and FA.

In DTI, noise in the observation vector $\mathbf{W}$ is known to induce bias, B , in $\hat{\mathbf{D}}$ and $\widehat{\text{FA}}$. This bias causes the running average of tensor estimates to converge on the incorrect value.

Specifically for FA, bias is defined; $B = E(\widehat{\text{FA}}) - \text{FA}$, where E is the expectation function. Figure 1 illustrates important summary conclusions from previous studies on bias in FA. Bias in FA is known to be sensitive to noise levels and study design (e.g., gradient table). The relationship between bias and noise is known to be non-monotonic, but in the typical SNR region of clinical and research scanners ($15:1 < \text{SNR} < 50:1$) is generally expected to be monotonic. Not shown is the known sensitivity of bias to the underlying anatomy, including not only the magnitude of FA but small differences in Eigenvalues for a fixed FA.

SIMEX Bias Estimation in DTI

SIMEX is a general purpose method for estimating bias in parameter outputs of a data fitting procedure. In SIMEX, the mapping process from the observation vector domain $\mathbf{W}$ to the output parameter domain $\mathbf{y}$ is represented by a generic function $\mathbf{T}: \mathbf{W} \rightarrow \mathbf{y}$. To connect this to DTI, we associate $\mathbf{W}$ with the imaging data observation vector and $\mathbf{y}$ with the tensor parameter of interest (FA). Thus, $\mathbf{T}$ represents the concatenated effects of all processing steps, including estimating $\hat{\mathbf{D}}$ from $\mathbf{W}$ through data fitting to the diffusion model (Eq. 1), calculating the Eigenvalues of $\hat{\mathbf{D}}$, and calculating $\widehat{\text{FA}}$ (Eq. 3). Equivalently, we could have

defined $\mathbf{y}$ as any other (possibly multivariate) quantity of interest derived from the observation vector.

SIMEX models bias in the output domain ($\widehat{\mathbf{FA}}$) by adding additional synthetic noise of variance $\omega\sigma_E^2$ to $\mathbf{W}$ and observing the impact on $E(\widehat{\mathbf{FA}})$ using Monte Carlo methods. The functional form of the trend in $E(\widehat{\mathbf{FA}})$ with additional noise is modeled using approximation functions, such as linear or polynomial. The SIMEX estimate of FA, $\mathbf{FA}_{\text{SMX}}$, is the value of the approximation function extrapolated to zero noise and the SIMEX estimate of bias is defined as: $B_{\text{SMX}} = \widehat{\mathbf{FA}} - \mathbf{FA}_{\text{SMX}}$. Note that the point of zero noise occurs when $\omega = -1$. Since the variance of $\mathbf{W}$ is σ_E^2 , the variance of data simulated from $\mathbf{W}$ through the addition of synthetic noise of variance $\omega\sigma_E^2$ is $(\sigma_E^2 + \omega\sigma_E^2)$. Extrapolation of the approximation function to $\omega = -1$ yields the SIMEX estimate.

Three important properties of the SIMEX method must be considered before a novel application of SIMEX is tried. (1) SIMEX requires the experimental noise in $\mathbf{W}$ to be well understood and capable of modeling via computer simulation. (2) For approximation functions to model the true (but unknown) functional form of $E(\widehat{\mathbf{FA}})$ versus noise, the true functional form must meet certain criteria, the minimum of which is smoothness and continuity. (3) The original SIMEX theory and several theory development articles in the statistical literature present SIMEX using a Gaussian additive error model for the observation vector $\mathbf{W}$. Extension of SIMEX to other noise models should be carefully evaluated.

With these considerations, DTI is an excellent target for SIMEX. (1) Modeling DTI noise through computer simulation is well established in the literature and noise in MRI magnitude imaging data has been highly studied. Estimates of empirical noise (σ_E^2) from repeated DTI measurements are not often available, but several robust methods exist to estimate imaging noise given a single dataset (13,28–30). (2) Previous work has shown that the bias function is smooth, continuous, and even monotonic for the vast majority of cases within the SNR of clinical and research scanners. However, as seen in Figure 1, violations of monotonicity can exist at low SNR. (3) SIMEX theory, although presented in the statistical literature using a Gaussian error model, does not rely on properties singular to Gaussian distributions and SIMEX has been successfully extended to other noise models. A literature search reveals no previous SIMEX extensions to the stacked Rician noise model used herein. To be confident in our extension, we offer in the appendix a heuristic argument for the interested reader as well as a more formal mathematical proof demonstrating the stacked Rician noise model for magnitude imaging data is compliant with a Gaussian additive error model for SIMEX.

METHODS and RESULTS

A demonstration program including this simulation framework is available in open-source (GNU Lesser General Public License 2.1+, Free Software Foundation) through the Neuroimaging Informatics Tools and Resources Clearinghouse (NITRC) project id masimatlab (<http://www.nitrc.org/projects/masimatlab>). Unless otherwise stated, all processing and analysis was performed in Matlab 2010 (Mathworks, Natick, MA). Diffusion tensor estimates were calculated by fitting the model (Eq. 1) to the data using a simple log least minimum mean squared error (LLMMSE) criterion. In all cases, ‘added noise’ refers to the stacked Rician method. Diffusion-weighted images (DWI) from each dataset were registered to the S_0 image using FLIRT affine registration (FMRIB, Oxford, UK). Inter-subject registration of data was accomplished using the VABRA tool (31) for the JIST plugin (32) in MIPAV (33).

SIMEX Implementation

Because SIMEX application to DTI is novel, several SIMEX user-controlled parameters needed optimization. Based upon parameter sweeping and performance evaluation (data not shown), we used 2000, 4000, 6000, and 8000 Monte-Carlo iterations for $\omega = 2, 4, 6$, and 8 respectively. SIMEX models the functional form of the relationship between the expected FA value and noise level. Based upon several evaluations (linear, asymptotic, polynomial order 2, and 3, data not shown) we choose a polynomial of order 2.

Computational time for this SIMEX implementation was recorded for dataset-16 (dataset-16 defined below). The average time for a whole volume SIMEX implementation was 17.9 ± 1.6 days on a single 2.8 GHz CPU core. There are many straightforward possibilities for decreasing this time. The longest time for SIMEX implementation on a single slice was 18 hours, so a cluster with an available CPU per slice could reduce whole-brain processing time to less than a day (as was done for the implementation presented here). Additionally, SIMEX was performed on interpolated data ($256 \times 256 \times 65$) and processing at the nominal data resolution ($96 \times 96 \times 65$) would have decreased processing time by an additional 86 % (to 2.5 hours with parallel processing). Finally, the Matlab implementation used here focused on algorithm clarity and flexibility rather than on efficiency, so additional computational gains are likely possible through improved computation methodology and possibly more efficient experimental design and computational algorithms, i.e., (34).

An example of SIMEX as implemented herein on two simulated tensors is shown in Figure 2. Figure 2A shows the two tensors have true FA values of ~ 0.9 and ~ 0.2 with smaller and larger bias respectively. Individual observations of $\widehat{FA}$ at various noise levels reveal the variance caused by the imaging noise while the $E(\widehat{FA})$ at each noise level reveals the bias caused by the imaging noise. Note from Figure 2A that owing to variance, $\widehat{FA}$ may be lower than the true FA value, even though it belongs to a positively biased population. Figure 2B displays the SIMEX procedure from a randomly selected individual $\widehat{FA}$ value at the noise level of $\sigma_E = 1000$ units. Using a stacked Rician, noise of level $\omega_r \sigma_E^2$ is added to the imaging data and $E(\widehat{FA})$ is estimated at the new noise level. The trend in $E(\widehat{FA})$ is extrapolated to zero-noise ($\omega = -1$). SIMEX correctly estimates ~ 0 bias for the FA of 0.9 and slightly positive bias for the FA of 0.2.

Empirical Datasets

Two independent and highly differentiated empirical datasets are used in this study and a third empirical dataset is created from gradient sub-sampling. All datasets used SENSE frequency-based reconstruction. Each dataset is labeled based upon its gradient table size. For the first empirical dataset, labeled “dataset-46-x-2”, a single DTI dataset was acquired in de-identified form from an ongoing local study. Briefly, DTI data were collected using echo planar imaging (EPI) with an 8 channel head coil on a Philips 3T system using 92 gradient orientations (positive and negative pairing of 46 orientations) at a b-value = 1000 s/mm². Partial Fourier imaging was used to reduce TE to 48 ms. Volumes consisted of 51 transverse slices each with 96×96 voxels (field of view = 240×240 mm, voxel size of $2.5 \times 2.5 \times 2.5$ mm). The DWI were registered to the S_0 image using FLIRT affine registration. The spatially varying noise was estimated from the repeated orientations of the empirical data (13,29) (Figure 3A and B).

For the second empirical dataset, labeled “dataset-32”, we use the DTI component of the open-access Multi-Modal MRI Reproducibility study (35). Briefly, the study consists of 42 total DTI datasets from 21 subjects, each scanned twice at 3T. Each dataset was acquired with a multi-slice, single-shot, echo planar imaging (EPI) sequence with 32 gradient

orientations at a b-value = 700 s/mm² with five signal averages used for the minimally weighted volume. The resulting images consisted of 65 transverse slices with a field of view of 212 × 212 mm, reconstructed to 256 × 256 voxels (0.83 × 0.83 × 2.2 mm). Diffusion-weighted images (DWI) from each dataset were registered to the S_0 image using FLIRT affine registration (FMRIB, Oxford, UK). The spatially varying noise was estimated using the repeated datasets for each subject (13,29) (Figure 3C and D).

A third empirical dataset, “dataset-16”, was created by a spatially optimized sub-sampling of the 32 gradients from the Multi-Modal MRI dataset. A total of 16 gradients were selected and the diffusion-weighted images corresponding to these gradients were pooled to form a new DTI dataset. The spatially varying noise estimate was identical to that for dataset-32 (Figure 3C and D).

Simulated Datasets

Three simulated datasets were constructed based upon the three empirical datasets. First, “simulated-46-x-2” was constructed from axial slice 25 of the local dataset. Second, “simulated dataset-32” was constructed from axial slice 30 of dataset KKI2009-33-DTI from the Multi-Modal MRI Reproducibility study. Finally, “simulated-16” was simulated from slice 30 of empirical dataset-16 created from subsampling dataset KKI2009-33-DTI.

For each dataset, simulated ground truth tensors were defined as the empirically observed tensors. Simulated noiseless DWI data were created by projecting the tensor models to the originally observed orientations. Thirty simulated noisy DWI datasets were synthesized by corrupting the noiseless DWI data through the addition of synthetic Rician noise. The scale parameters (σ_E , Eq. 2) used to simulate the noisy DWI datasets from the noise free DWI data were defined by the empirical estimate of noise at each voxel (Figure 3). To determine the true bias at the noise level of σ_E , 10,000 Monte-Carlo repetitions were averaged to calculate $E(\widehat{FA})$. The true bias then equals $E(\widehat{FA}) - FA$.

SIMEX Tests on Simulated Data

SIMEX is first evaluated for its ability to accurately capture bias levels. Simulated data, though not as complex as empirical data, has the advantage of known true bias values that can be compared to SIMEX bias estimates. Bias values in the simulated data match expectations from previous work, suggesting the simulated data are behaving as reasonably good models for the empirical case. SIMEX bias estimates have a magnitude and anatomical distribution similar to the true bias maps, with the noticeable exception of a region near the posterior corpus callosum for simulated-46-x-2 (Figure 4). For all three simulated test cases, the error in the SIMEX bias estimates tend to be proportional to the size of the bias, with larger errors for larger bias values. Thus the greatest errors are seen predominantly in the peripheral gray matter voxels. The magnitude of the bias and the SIMEX bias error can be simultaneously accounted for by delineating voxels using the criterion that the error in the SIMEX bias estimate should be smaller than the size of the true bias. The last column in Figure 4 shows voxels (red) that do not meet this criterion. For a vast majority of voxels, SIMEX bias correction reduces bias in the data. An improvement in bias, however, does not evaluate if SIMEX bias correction would ultimately reduce total error in the FA estimate.

To evaluate overall error, the distance from the true FA value is compared between the SIMEX estimate, FA_{SMX} , and the original estimate, $\widehat{FA}$. Interestingly, for a given single DTI study, FA_{SMX} and $\widehat{FA}$ contain about the same error (Figure 5A), but for the population averaged study (sample size of 30), FA_{SMX} significantly outperforms $\widehat{FA}$ (Figure 5B). To understand the origin of FA_{SMX} outperformance only under population average conditions three key observations are made. First, a perfect bias correction may *correctly* move an

individual estimate *further* from the true value. For example, because of variability $\widehat{FA}$ may be lower than the true FA value but belong to a positively biased group. A perfect bias correction would further lower $\widehat{FA}$, decreasing its accuracy. In this case, it is only on average that even a perfect bias correction would uniformly outperform the original estimate. Second, the level of bias in the data is small compared to the size of the standard deviation in $\widehat{FA}$ (Figure 6A). In cases of large differences (standard deviation $\gg$ the bias), even a perfect bias correction would only slightly improve the accuracy of the estimate. The third observation is that SIMEX is an additional processing step and FA_{SIMX} , though less biased, is necessarily more variable than $\widehat{FA}$ (Figure 6B). The benefit of the SIMEX bias reduction comes at the cost of increased variability. For the individual case herein, the cost benefit ratio is about equal and FA_{SIMX} performs on par with $\widehat{FA}$ (Figure 5A).

In a population-based study, $\widehat{FA}$ is pooled (“averaged”) across subjects. The variability of the pooled metric is reduced but the bias remains the same. With sufficient sample size, the benefit of a SIMEX bias correction significantly outweighs the cost of the increased variability (Figure 5B). Note, for visual clarity, data in Figure 6 only includes results for simulated-32, but simulated-46-x-2 showed equivalent trends (simulated-16 was not included in the population studies).

Before proceeding to empirical test cases it is important to evaluate reasons for clear SIMEX failure seen in the simulated data and to evaluate SIMEX bias estimation accuracy when only estimates of σ_E , s_E , are available (Figure 7). SIMEX extrapolations for voxels in the failing regions around the posterior corpus callosum of simulated-46-x-2 were visualized in detail. The SIMEX trend was compared to the true trend in $E(\widehat{FA})$, which could be simulated from the true FA value. These evaluations revealed the true bias function was non-monotonic in this region of SIMEX failure. As exemplified in Figure 7A, an extrapolation could not predict the change in curvature and SIMEX bias estimation failed. Note that this region corresponds to a region of low SNR (SNR ~ 5 , Figure 3). Regions of non-monotonic bias functions and possible SIMEX failure may be predictable *a priori* from SNR measurements.

In empirical settings, only s_E is available. The tolerance of SIMEX to errors in s_E was evaluated on a randomly selected subset of voxels (2511 in simulated-32 (data not shown) and 1385 voxels in simulated-46-x-2). For both evaluations, SIMEX performance was almost unaffected by misestimations within $\pm 5\%$ of the true noise level and was robust to misestimations $\pm 10\%$ of the noise level (Figure 7B). These tolerance ranges are encouraging for empirical implementation.

SIMEX Tested On Empirical Data

Empirical data presents a more complex challenge than the simulated data. Empirical data likely does not fit the DTI tensor model and σ_E is not precisely known. Even so, a qualitative evaluation of SIMEX estimates in empirical data shows the estimates to match expectations from previous knowledge of bias distributions in FA values. Comparison of the distributions of $\widehat{FA}$ and FA_{SIMX} for all 42 DTI experiments on dataset-32 shows that SIMEX predominantly estimates a positive bias (indicated by the negative shift of FA_{SIMX} compared to $\widehat{FA}$) and that SIMEX predicts larger bias for lower FA values (Figure 8A). The variance of the distribution of FA_{SIMX} across all 42 DTI datasets is slightly larger than the variance of the distribution for $\widehat{FA}$, demonstrating the SIMEX bias estimates are stable across subjects. A small number of FA_{SIMX} values are mapped to physically impossible negative FA values. These voxels belong to a population of spatial locations with very low FA values, FA ~ 0.05 (green histogram, Figure 8A) and map to regions of CSF (Figure 8A insert), which is not a region of relevance in a DTI analysis. It is worth understanding, however, that a negative

FA_{SMX} value does not universally imply that SIMEX misestimated bias. A low initial $\widehat{FA}$ observation may belong to a highly variable population with positive bias. Subtracting the positive bias from the already low FA value may create a negative FA_{SMX} value. In this case, the negative value is a result of the variability in the data, not a misestimation of bias. Results for dataset-16 are similar (data not shown).

A more rigorous evaluation of SIMEX is made possible by construction of dataset-16 from dataset-32, where the FA values from dataset-16 in truth equal the FA values from dataset-32 ($FA_{16} = FA_{32}$). SIMEX can then be tested for accuracy by comparing the distance between FA estimates of dataset-32 and dataset-16. First we consider the variance and bias properties that were important in understanding the performance of SIMEX in the simulated results. As with the simulated data, the variance of $\widehat{FA}$ is significantly larger than the bias level (bias as estimated by SIMEX since the true bias is unknown) and the variance of $\widehat{FA}$ across the 42 DTI datasets is smaller than the variance of FA_{SMX} . These observations are true for dataset-32 (Figure 8B and C) as well as dataset-16 (data not shown). The challenge of a bias correction improving the error of the estimate: $\Delta FA = FA_{16} - FA_{32}$, is more difficult because these unfavorable ratios are magnified. The difference in bias between FA_{16} and FA_{32} is smaller than the bias in either, while the variance of ΔFA is significantly larger. The ratio of standard deviation/bias was on average twice as large for the difference estimate as for the individual estimates.

As with simulated data, we first investigate the level of bias in the SIMEX estimate of the difference (ΔFA_{SMX}). Histograms of errors, (error = $FA_{16} - FA_{32}$), shows ΔFA_{SMX} is less biased than the original estimate (Figure 9A), though some bias remains. For reference, the histogram of a perfectly unbiased estimate of ΔFA would be centered at zero in Figure 9A. The SIMEX estimate is more variable as indicated by its broader distribution curve (red versus black). If instead, the population averaged bias across all 42 sets is used to bias correct each voxel, the variance of the SIMEX estimate decreases (green curve, Figure 9A), and the net bias is further decreased by a small amount.

An evaluation of the performance of ΔFA_{SMX} shows that although ΔFA_{SMX} is less biased, its performance on an individual case is worse than the original estimate (Figure 9B). The greater variance of the SIMEX estimate is a significant contributing factor to the lower performance. Correcting bias using the population averaged bias brings the SIMEX performance to be on par with the initial estimate (green curve), mirroring the simulation test case results. Despite the underperformance of SIMEX for individual trials, as with simulated data, the SIMEX estimate still substantially outperforms the original estimate in the population averaged case (Figure 9C).

A sample mid-axial slice shows the anatomical distributions of SIMEX-estimated bias differences and regions where SIMEX outperforms the original estimator in the population study (Figure 10). The SIMEX-estimated differences in FA have clear anatomical correlation (Figure 10A). There is some correspondence between bias level and SIMEX performance, but this trend is not universal. Generally, SIMEX is seen to be more beneficial in gray matter than in white matter.

DISCUSSION and CONCLUSION

There is currently no statistical method available for investigators to estimate bias in empirical DTI data. Although the properties of bias in DTI metrics have been highly studied in controlled settings (mostly theoretical and simulated), the behavior of bias in empirical DTI data is difficult to predict because it is a function of many variables including subject characteristics, protocols, scanner stability, hardware/software changes, etc. The

experimental results herein show SIMEX to be a valuable method for bias estimation in individual DTI datasets. For example, if the true bias map was not available for the simulated data, it is clear from Figure 4 that the anatomical distribution of the SIMEX bias estimate would provide important and accurate insights to data quality and the true bias distribution. In the case of simulated-32 and simulated-16, SIMEX bias estimates from an individual subject correctly suggest that data from the collection methods are differentially biased. In comparisons of cohorts where differential bias may not be known a priori, this would be an important consideration for interpretation of FA values or for further secondary analysis.

In simulated-46-x-2, SIMEX accurately predicts the marked change in bias induced by the frontal cortex ringing artifact (Figure 3). Because the data was simulated, the artifact region was constructed to have Rician distributed noise. In empirical data the presence of the artifact would likely violate the assumption of a Rician noise distribution. An analysis of the robustness of SIMEX to incorrect noise distributional assumptions has not been published in the statistical literature and is beyond the scope of this article, but is certainly an important consideration. Even if the quantitative value of the SIMEX bias estimate would be decreased, the qualitative warning that the anatomy in the region of artifact produces substantially different bias than the surrounding anatomical region remains qualitatively useful.

In the empirical test, the decrease in bias of the SIMEX Δ FA estimate supports that SIMEX was producing accurate bias estimates in the empirical dataset-32 and dataset-16 (Figure 7A). Although the decrease in bias for Δ FA is arguably small, this task is exceedingly more difficult than accurately estimating the bias in either dataset-32 or dataset-16 alone and small improvements require appreciable accuracy. Note that with the flexibility of SIMEX, the trend in $\widehat{\Delta FA}$ as a function of added synthetic noise could be used for SIMEX extrapolation and ΔFA_{SMX} determined from a single fitting procedure rather than two separate SIMEX procedures for FA_16 and FA_32. Further investigations of such an approach may yield improved SIMEX implementation for comparative studies.

The new SIMEX estimate, FA_{SMX} , though less biased than the original estimate, offers too small an improvement in accuracy to compensate for its increased variability in individual cases. The results herein therefore do not support using FA_{SMX} in place of the original estimate for individual or small case studies. Yet, as Figure 10B illustrates, more consistent estimates may be obtained on a single subject basis outside of the core white matter regions. Clinical and research studies with a large population size, however, may significantly benefit from using FA_{SMX} over the original estimator. The SIMEX estimate outperforms the original estimate in tests of accuracy for both simulated and empirical results. These results also suggest the use of SIMEX can increase power for secondary statistical studies. Quantitative methods for evaluating at what sample size would FA_{SMX} be the better estimator are beyond the scope of this article. However, such a method would most likely include estimates of variability, bias, and an evaluation of their predicted ratio as a function of sample size. Besides improved power, the less biased SIMEX estimator would offer other statistical advantages. The presence of bias violates nearly every statistical analysis method and bias is well known to cause faulty Type-I and Type-II error rates. A more detailed analysis of the impacts of bias on secondary statistical analyses using DTI data would bring clarity to the cost of bias and benefits of a bias reduction method. From the research herein, it is clear that SIMEX can be used to provide researchers with quantitative post hoc assessment of bias levels in their data and SIMEX offers significant promise for future development as a method to remedy the problem of bias.

Acknowledgments

The authors thank Dr. Stephen Heckers for use of his data whose acquisition was funded by the Psychiatric Genotype-Phenotype Project in the Department of Psychiatry at Vanderbilt University. The authors thank Andrew Asman for general assistance and troubleshooting. This research was supported in part by a post-doctoral training grant in image science (T32 EB003817), and the Vanderbilt CTSA (UL1 RR024975-01) from NCRR/NIH. The authors thank the anonymous reviewers and editor whose feedback substantially improved this manuscript.

References

1. Basser PJ, Jones DK. Diffusion-tensor MRI: Theory, experimental design and data analysis - a technical review. *NMR Biomed.* 2002; 15:456–467. [PubMed: 12489095]
2. Basser PJ, Mattiello J, LeBihan D. MR diffusion tensor spectroscopy and imaging. *Biophys J.* 1994; 66:259–267. [PubMed: 8130344]
3. Mori, S.; Wakana, S.; Nagae-Poetscher, LM.; van Zijl, PCM. *MRI Atlas of Human White Matter.* Amsterdam: Elsevier; 2005.
4. Anderson AW. Theoretical analysis of the effects of noise on diffusion tensor imaging. *Magn Reson Med.* 2001; 46:1174–1188. [PubMed: 11746585]
5. Landman BA, Farrell JA, Jones CK, Smith SA, Prince JL, Mori S. Effects of diffusion weighting schemes on the reproducibility of DTI-derived fractional anisotropy, mean diffusivity, and principal eigenvector measurements at 1.5 T. *Neuroimage.* 2007; 36:1123–1138. [PubMed: 17532649]
6. Basser PJ, Pajevic S. Statistical artifacts in diffusion tensor MRI (DT-MRI) caused by background noise. *Magn Reson Med.* 2000; 44:41–50. [PubMed: 10893520]
7. Bastin ME, Armitage PA, Marshall I. A theoretical study of the effect of experimental noise on the measurement of anisotropy in diffusion imaging. *Magn Reson Imag.* 1998; 16:773–785.
8. Basu S, Fletcher T, Whitaker R. Rician noise removal in diffusion tensor MRI. *Int Conf Med Image Comput Comput Assist Interv.* 2006; 9:117–125.
9. Dietrich O, Heiland S, Sartor K. Noise correction for the exact determination of apparent diffusion coefficients at low SNR. *Magn Reson Med.* 2001; 45:448–453. [PubMed: 11241703]
10. Ding Z, Gore JC, Anderson AW. Reduction of noise in diffusion tensor images using anisotropic smoothing. *Magn Reson Med.* 2005; 53:485–490. [PubMed: 15678537]
11. Farrell JA, Landman BA, Jones CK, Smith SA, Prince JL, van Zijl PC, Mori S. Effects of signal-to-noise ratio on the accuracy and reproducibility of diffusion tensor imaging-derived fractional anisotropy, mean diffusivity, and principal eigenvector measurements at 1.5T. *J Magn Reson Imaging.* 2007; 26:756–767. [PubMed: 17729339]
12. Jones DK, Basser PJ. “Squashing peanuts and smashing pumpkins”: How noise distorts diffusion-weighted MR data. *Magn Reson Med.* 2004; 52:979–993. [PubMed: 15508154]
13. Landman BA, Bazin PL, Prince JL. Estimation and application of spatially variable noise fields in diffusion tensor imaging. *Magn Reson Imag.* 2008; 27:741–751.
14. 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. [PubMed: 11097823]
15. Landman BA, Farrell JA, Huang H, Prince JL, Mori S. Diffusion tensor imaging at low SNR: Nonmonotonic behaviors of tensor contrasts. *Magn Reson Imaging.* 2008; 26:790–800. [PubMed: 18499378]
16. Carroll, RJ.; Ruppert, D.; Stefanski, L.; Crainiceanu, CM. *Measurement Error in Nonlinear Models: A Modern Perspective.* Chapman and Hall/CRC Press; 2006.
17. Cook JR, Stefanski LA. Simulation-extrapolation estimation in parametric measurement error models. *J Am Stat Assoc.* 1994; 89:1314–1328.
18. Küchenhoff H, Mwalili SM, Lesaffre E. A general method for dealing with misclassification in regression: The misclassification SIMEX. *Biometrics.* 2006; 62:85–96. [PubMed: 16542233]
19. Eckert RS, Carroll RJ, Wang N. Transformations to additivity in measurement error models. *Biometrics.* 1997:262–272. [PubMed: 9147595]

20. Carroll RJ, Wang Y. Nonparametric variance estimation in the analysis of microarray data: A measurement error approach. *Biometrika*. 2008; 95:437–449. [PubMed: 19381352]
21. Lamina C, Küchenhoff H, Chang Claude J, Paulweber B, Wichmann H, Illig T, Hoehe MR, Kronenberg F, Heid IM. Haplotype misclassification resulting from statistical reconstruction and genotype error, and its impact on association estimates. *Ann Hum Genet*. 2010; 74:452–462. [PubMed: 20649529]
22. Slate EH, Bandyopadhyay D. An investigation of the MC SIMEX method with application to measurement error in periodontal outcomes. *Stat Med*. 2009; 28:3523–3538. [PubMed: 19902495]
23. Chowdhury S, Sharma A. Mitigating parameter bias in hydrological modelling due to uncertainty in covariates. *J Hydrol*. 2007; 340:197–204.
24. Le Bihan D, van Zijl P. From the diffusion coefficient to the diffusion tensor. *NMR Biomed*. 2002; 15:431–434. [PubMed: 12489093]
25. Mori S, Zhang J. Principles of diffusion tensor imaging and its applications to basic neuroscience research. *Neuron*. 2006; 51:527–539. [PubMed: 16950152]
26. Gudbjartsson H, Patz S. The Rician distribution of noisy MRI data. *Magn Reson Med*. 1995; 34:910–914. [PubMed: 8598820]
27. Aja Fernández S, Tristán Vega A, Hoge WS. Statistical noise analysis in grappa using a parametrized noncentral chi approximation model. *Magn Reson Med*. 2011; 65:1195–1206. [PubMed: 21413083]
28. Aja-Fernandez S, Tristan-Vega A, Alberola-Lopez C. Noise estimation in single- and multiple-coil magnetic resonance data based on statistical models. *Magn Reson Imaging*. 2009; 27:1397–1409. [PubMed: 19570640]
29. Landman BA, Bazin PL, Smith SA, Prince JL. Robust estimation of spatially variable noise fields. *Magn Reson Med*. 2009; 62:500–509. [PubMed: 19526510]
30. Sijbers J, Den Dekker A. Maximum likelihood estimation of signal amplitude and noise variance from MR data. *Magn Reson Med*. 2004; 51:586–594. [PubMed: 15004801]
31. Rohde GK, Aldroubi A, Dawant BM. The adaptive bases algorithm for intensity-based nonrigid image registration. *IEEE Trans Med Imaging*. 2003; 22:1470–1479. [PubMed: 14606680]
32. Lucas BC, Bogovic JA, Carass A, Bazin PL, Prince JL, Pham DL, Landman BA. The Java image science toolkit (JIST) for rapid prototyping and publishing of neuroimaging software. *Neuroinformatics*. 2010; 8:5–17. [PubMed: 20077162]
33. McAuliffe MJ, Lalonde FM, McGarry D, Gandler W, Csaky K, Trus BL. Medical image processing, analysis and visualization in clinical research. *IEEE Comp Based Med Syst*. 2001:381–386.
34. Basser PJ, Pierpaoli C. Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. *J Magn Reson SerB*. 1996; 111:209–219.
35. Landman BA, Huang AJ, Gifford A, Vikram DS, Lim IAL, Farrell JAD, Bogovic JA, Hua J, Chen M, Jarso S. Multi-parametric neuroimaging reproducibility: A 3T resource study. *Neuroimage*. 2010; 54:2854–2866. [PubMed: 21094686]
36. Henkelman RM. Measurement of signal intensities in the presence of noise in MR images. *Med Phys*. 1985; 12:232. [PubMed: 4000083]

APPENDIX

Heuristic Justification of Stacked Rician Error Model

To heuristically justify the use of a stacked Rician in SIMEX, we first explicitly present our methodology of implementing a stacked Rician. For SIMEX, noise is synthetically added to observation element $w_j \in \mathbf{W}$ to create a new observation element $w_{j,r,k}$, where $r = 1, 2, \dots, R$ indexes the amount of added noise ($\omega_r \sigma_E^2$) and $k = 1, 2, \dots, K$ indexes the Monte Carlo simulation number at each noise level. For a stacked Rician, $w_{j,r,:} \sim \text{Rician}(w_j, \omega_r \sigma_E^2)$.

A heuristic argument for the compliancy of this stacked Rician noise model with the additive Gaussian model is as follows. Magnitude imaging data is calculated from the magnitude of raw complex data. The noise for the real and imaginary parts of the complex data are independent, identically distributed, zero-mean, and Gaussian. The scale parameter defining the Rician noise level of the observation element w_j , σ_E^2 , is equal to the variance of the Gaussian noise in the complex data (σ_E^2 is not the variance of the Rician distribution). Stacking Rician noise with variance parameter $\omega_r \sigma_E^2$ on w_j to form $w_{j,r,k}$ is equivalent to first adding Gaussian noise of variance $\omega_r \sigma_E^2$ to the complex data of w_j followed by a magnitude calculation of $w_{j,r,k}$. As the function **T** in the SIMEX theory is completely general, it may or may not include a magnitude calculation step and the stacked Rician model used for magnitude imaging data herein is compliant with the original SIMEX additive Gaussian noise model.

Proof of Compliancy for the Stacked Rician Error Model

Below is a more formal proof that the stacked Rician error model is compliant with an additive Gaussian error model for SIMEX. For the following, ‘true’ is synonymous with ‘observed without measurement error’, and all operations on vectors are elementwise. Let $\text{Rice}(v, s)$ represent a Rician distribution with magnitude parameter v and noise level parameter s . Let $N(\mu, \sigma^2)$ represent a normal distribution with mean μ and standard deviation σ . Then by definition,

$$W \sim \text{Rice}(v, s) \text{ if } W = \sqrt{R^2 + I^2} \text{ with } R \sim N(v \cos(\theta), s^2) \text{ and } I \sim N(v \sin(\theta), s^2). \quad \text{Def.1}$$

Here θ is any arbitrary but fixed angle. Also observe the following is true,

$$C = \sqrt{A^2 + B^2} \Rightarrow \exists \theta \in [0, 2\pi) \text{ s.t. } A = C \cos(\theta) \text{ and } B = C \sin(\theta). \quad \text{Obs.1}$$

Let $\mathbf{X}_T$ be the predictor vector of noise-free magnitude MRI data of length J for elements $j = 1, 2, \dots, J$ on a single spatial location. $\mathbf{X}_T$ is calculated from the true real $\mathbf{R}_T$, and true

imaginary $\mathbf{I}_T$, components of quadrature detection; $\mathbf{X}_T = \sqrt{\mathbf{R}_T^2 + \mathbf{I}_T^2}$ and by Obs.1,

$$\exists \theta \text{ s.t. } \mathbf{R}_T = \mathbf{X}_T \cos(\theta) \text{ and } \mathbf{I}_T = \mathbf{X}_T \sin(\theta). \quad [1A]$$

Here θ is a vector of length J and all operations are elementwise. Let $\mathbf{W}$ be the observation vector of $\mathbf{X}_T$ with observed real $\mathbf{R}_{\text{obs}}$, and observed imaginary $\mathbf{I}_{\text{obs}}$, components whose elements contain independently and identically distributed (iid) Gaussian noise of zero mean

and variance σ_E (36). $\mathbf{R}_{\text{obs}} = \mathbf{R}_T + \sigma_E \mathbf{Z}_R$, $\mathbf{I}_{\text{obs}} = \mathbf{I}_T + \sigma_E \mathbf{Z}_I$, and $= \sqrt{\mathbf{R}_{\text{obs}}^2 + \mathbf{I}_{\text{obs}}^2}$. The elements of $\mathbf{Z}_R$ and $\mathbf{Z}_I$ are independent standard normal random variables. Substituting Eq. 1A for $\mathbf{R}_T$ and $\mathbf{I}_T$ yields Eq. 2A and application of Obs. 1 yields Eq. 3A.

$$\mathbf{R}_{\text{obs}} = \mathbf{X}_T \cos(\theta) + \sigma_E \mathbf{Z}_R, \quad \mathbf{I}_{\text{obs}} = \mathbf{X}_T \sin(\theta) + \sigma_E \mathbf{Z}_I \quad [2A]$$

$$\exists \delta \text{ s.t. } \mathbf{R}_{\text{obs}} = \mathbf{W} \cos(\delta) \text{ and } \mathbf{I}_{\text{obs}} = \mathbf{W} \sin(\delta) \quad [3A]$$

Consider, as in SIMEX, the case of adding additional synthetic Gaussian noise of mean zero and variance $\omega\sigma_E^2$ directly to the real and imaginary components of the observed data. In equation form, for the k 'th Monte Carlo simulation;

$$\mathbf{R}_k = \mathbf{R}_{\text{obs}} + \omega^{1/2} \sigma_E \mathbf{U}_{R,k}, \quad \mathbf{I}_k = \mathbf{I}_{\text{obs}} + \omega^{1/2} \sigma_E \mathbf{U}_{I,k} \quad [4A]$$

$$\mathbf{W}_k = \sqrt{\mathbf{R}_k^2 + \mathbf{I}_k^2} \quad [5A]$$

Where the elements of $\mathbf{U}_{R,k}$ and $\mathbf{U}_{I,k}$ are simulated random drawings from $N(0, 1)$. Substituting Eq. 3A for $\mathbf{R}_{\text{obs}}$ and $\mathbf{I}_{\text{obs}}$ in Eq. 4A yields,

$$\mathbf{R}_k = \mathbf{W} \cos(\boldsymbol{\delta}) + \omega^{1/2} \sigma_E \mathbf{U}_{R,k}, \quad \mathbf{I}_k = \mathbf{W} \sin(\boldsymbol{\delta}) + \omega^{1/2} \sigma_E \mathbf{U}_{I,k}. \quad [6A]$$

It follows that $\mathbf{R}_k \sim N(\mathbf{W} \cos(\boldsymbol{\delta}), \omega\sigma_E^2)$ and $\mathbf{I}_k \sim N(\mathbf{W} \sin(\boldsymbol{\delta}), \omega\sigma_E^2)$ and by Def.1

$$\mathbf{W}_k \sim \text{Rice}(\mathbf{W}, \omega^{1/2} \sigma_E). \quad [7A]$$

Therefore, adding Gaussian noise directly to the real and complex data as in Eq. 4A followed by calculation of the magnitude as in Eq. 5A, is mathematically equivalent to sampling the elements of $\mathbf{W}_k$ directly from the Rician, as in Eq 7A. Furthermore, since substitution of $\mathbf{R}_{\text{obs}}$ and $\mathbf{I}_{\text{obs}}$ from Eq. 2A into Eq. 4A yields, $\mathbf{R}_k = \mathbf{X}_T \cos(\boldsymbol{\theta}) + \sigma_E \mathbf{Z}_R + \omega^{1/2} \sigma_E \mathbf{U}_{R,k}$ and $\mathbf{I}_k = \mathbf{X}_T \sin(\boldsymbol{\theta}) + \sigma_E \mathbf{Z}_Y + \omega^{1/2} \sigma_E \mathbf{U}_{Y,k}$ it follows that $\mathbf{R}_k$ and $\mathbf{I}_k$ are distributed according to $N(\mathbf{X}_T \cos(\boldsymbol{\theta}), \sigma_E^2 + \omega\sigma_E^2)$ and $N(\mathbf{X}_T \sin(\boldsymbol{\theta}), \sigma_E^2 + \omega\sigma_E^2)$ respectively. Then by Def.1 $\mathbf{W}_k \sim \text{Rice}(\mathbf{X}_T, \sigma_E + \omega^{1/2} \sigma_E)$. Simple substitution of $\omega = -1$ into the distribution for $\mathbf{R}_k$ and $\mathbf{I}_k$ shows that the variance of each element in $\mathbf{W}_k$ is zero when $\omega = -1$. To be clear, by 'variance of each element of $\mathbf{W}_k$ ' it is meant the variance of each element across repeated k . For example, in the case of zero variance, it is not that all elements of $\mathbf{W}_k$ are the same, it is that repeated measurements of $\mathbf{W}_k$ would produce the same vector.

Figure 1.

Examples of bias in FA. The expected value of FA as a function of SNR in the imaging data (SNR of the control S_0 image) are compared for three different DTI collection schemes. Data were simulated using the collection methods for the empirical data in this article; dataset-46-x-2 (collection-46-x-2), dataset-32 (collection-32), and dataset-16 (collection-16). Note that the curves are non-monotonic and expected FA values only converge at high SNR values (different expectation values exist even at $\sim$ zero signal). The true FA value was fixed at 0.5, MD was set to the biological value of $0.7 \times 10^{-3} \text{ mm}^2/\text{s}$, and the two lowest tensor Eigenvalues were set as equal. Expectation values are from the average of 10^6 iterations.

Figure 2.

The SIMEX method examined at two spatial locations. (A) The $E(\widehat{FA})$ is a function of imaging noise (σ_E) for the two locations. Data were simulated from the noise free DWI data. The higher FA value has no appreciable bias, as indicated by the constant value of $E(\widehat{FA})$ with increasing noise. The lower FA has positive bias as seen from the rising value of $E(\widehat{FA})$ with increasing noise. (B) SIMEX is shown on a randomly chosen single observation of $\widehat{FA}$ at $\sigma_E = 1 \times 10^3$ from panel-A. Synthetic noise is added to the single observed DWI. The level of added noise is controlled by the variance parameter ω . The starting experimental observation is at $\omega = 0$, simulated $E(\widehat{FA})$ at higher noise levels are at $\omega = 2, 4, 6$, and 8 , and extrapolation to FA_{SMX} at zero image noise is at $\omega = -1$.

Figure 3.

Standard deviation of imaging noise and SNR levels (standard deviation/signal level of S_0) for the three empirical datasets. Dataset-16 is a sub-sampling of dataset-32 and is assumed to share the same noise structure. Axial slice 30 of dataset KKI2009-33-DTI is used as a representative example of the noise levels estimated in the empirical datasets, dataset-16 and dataset-32. DWI of dataset-46-x-2 had noticeable artifacts, including the frontal cortex ringing. Data were collected at different sites and the intensity scales differ by an order of magnitude, but the overall SNR levels are similar.

Figure 4.

Anatomical distribution of SIMEX bias estimates for each of the three simulated datasets. The true bias values for each dataset is known through controlled simulation methods. SIMEX bias estimates were made on simulated empirical observations. For simulated-32 and simulated-46-x-2, where 30 repeated observations were synthesized, only one randomly chosen observation is shown. Error is calculated as (True Bias – SIMEX Bias). Voxels whose bias would not be decreased by a SIMEX bias correction (i.e. magnitude of the error is greater than the magnitude of the bias) are labeled ‘Not Improved’. In the last column, these voxels are highlighted in red and overlaid on the designated true FA map.

Figure 5.

Distribution of errors for the original FA estimate ($\widehat{FA}$) compared to the SIMEX estimate (FA_{SMX}) for datasets simulated-32 and simulated-46-x-2. The error was calculated as the distance between the estimated FA and the true FA value. Median values for each voxel were calculated from 30 repeated observations of the single mid-axial slice. (A) The median of the absolute error values across the 30 repeated trials is used as a summary statistic to evaluate expected performance of SIMEX on an individual dataset. (B) The absolute value of the median error is used to evaluate SIMEX performance for population-based studies.

Figure 6.

Scatterplots of variance and bias ratios for dataset simulated-32. The scatterplots have a third dimension of color indicating the density of the data. (A) The standard deviation of the original estimate ($\widehat{FA}$) was calculated for each brain voxel of the mid-axial slice from the 30 repeated datasets. The standard deviation is divided by the known true bias. The condition where the ratio equals one is marked by a dashed black line. (B) The ratio of variance for $\widehat{FA}$ and FA_{SMX} is calculated at each voxel across the 30 repeated datasets. Simulated-46-x-2 is not shown but had similar results.

Figure 7.

Limitations of SIMEX. (A) An example case of SIMEX bias estimation failure. The SIMEX extrapolation (triangles and line) is plotted for a voxel from the failing region in the posterior corpus callosum of simulated-46-x-2. The true $E(\widehat{FA})$ for this voxel is calculated from the noiseless DWI data for that voxel. The true trend in $E(\widehat{FA})$ is seen to change curvature in the domain $\omega < 0$ but SIMEX observations can only be simulated for $\omega \geq 0$. The change in curvature cannot be predicted by the SIMEX extrapolant function causing a misestimation of bias. (B) SIMEX tolerance limits to misestimations of σ_E . SIMEX were repeatedly tested using ranging levels of accuracy of s_E , an estimator for σ_E . The population of absolute error values, $|B_{SMX} - B|$, from across the mid-axial slice was used to evaluate the effect of inaccuracies in s_E on the accuracy of the SIMEX bias estimate. The upper and lower edges of each box represent the 75th and 25th percentile of error values and the red line dividing the box represents the median value. The whiskers extend to the 95th and 5th percentile of the errors. The upper and lower cross marks represent the 99th percentile and 1st percentile respectively. Individual points above the 99th percentile are not shown owing to their effect on the visibility of the boxplots. Results for simulated-32 are similar.

Figure 8.

Properties of SIMEX estimates in empirical data. (A) The distribution of the original estimate ($\widehat{FA}$, black) compared to the SIMEX estimate (FA_{SMX} , red). The histograms represent the median bin count for each FA value across the 42 datasets. Error bars are the 75th and 25th percentile values. The green histogram represents the distribution of $\widehat{FA}$ values mapped to negative FA_{SMX} values. The inset is a mid-axial slice of a registered MPRAGE where green highlights indicate the anatomical location of negative FA_{SMX} values for this subject. (B) Scatterplot of the standard deviation over bias for dataset-32. The standard deviation of $\widehat{FA}$ at each voxel across the 42 datasets is divided by the averaged SIMEX bias estimate at each voxel across all 42 datasets. The dashed line represents a ratio of one. (C) Scatterplot of the variance ratio of $\widehat{FA}$ over FA_{SMX} for dataset-32. The variance was calculated across all 42 datasets at each brain voxel. The dashed line represents the ratio of 1:1. For both scatterplots, a third dimension of color is added to report the population density. Results from dataset-16 are similar as panel-A through C.

Figure 9.

Distribution of errors for the original FA estimate compared to the SIMEX estimate. Error is defined as the distance between FA values for dataset-16 and dataset-32: $\text{error} = (\text{FA}_{32} - \text{FA}_{16})$. (A) Distribution of raw error values across all 42 datasets. The median and mean values for the black, red, and green curve are (0.010, 0.012), (-0.003, -0.001), and (-0.003, 0.000), respectively. An unbiased estimator would be distributed about zero (black dashed line). (B) The median of the absolute error values for the same spatial location across all 42 datasets was used as a summary metric for individual performance. The green curve nearly completely overlays the black curve. (C) The absolute value of the median error at the same spatial location across all 42 datasets is used to evaluate the averaged performance of SIMEX. In the averaged case, the green curve is mathematically identical to the red curve.

Figure 10.

Anatomical distribution of SIMEX improvement compared to estimated bias levels. (A) The averaged SIMEX-estimated difference in bias between dataset-16 and dataset-32 is shown for a mid-axial slice. (B) A binary map representing voxels with closer FA values between dataset-16 and dataset-32 after SIMEX (red) and voxels with greater difference after SIMEX (blue). The voxels are overlaid on a registered MPRAGE.