Published in final edited form as:
J Am Stat Assoc. 2009 June 1; 104(486): 623–637. doi:10.1198/jasa.2009.0029.

Regression Models for Identifying Noise Sources in Magnetic Resonance Images

Hongtu Zhu, **Yimei Li**, **Joseph G. Ibrahim**, **Xiaoyan Shi**, **Hongyu An**, **Yashen Chen**, **Wei Gao**, **Weili Lin**, **Daniel B. Rowe**, and **Bradley S. Peterson**

Hongtu Zhu is Associate Professor, Department of Biostatistics and Biomedical Research Imaging Center and Department of Radiology, University of North Carolina at Chapel Hill, NC 27599 (hzhu@bios.unc.edu). Yimei Li is a Ph.D. student, Department of Biostatistics and Biomedical Research Imaging Center and Department of Radiology, University of North Carolina at Chapel Hill, NC 27599 (liyimei@email.uce.edu). Joshep G. Ibrahim is Alumni Distinguished Professor, Department of Biostatistics and Biomedical Research Imaging Center and Department of Radiology, University of North Carolina at Chapel Hill, NC 27599 (ibrahim@bios.unc.edu). Xiaoyan Shi is a Ph.D. student, Department of Biostatistics and Biomedical Research Imaging Center and Department of Radiology, University of North Carolina at Chapel Hill, NC 27599 (xyshi@email.unc.edu. Hongyu An is Research Assistant Professor, Department of Radiology, University of North Carolina at Chapel Hill, NC 27599 (HongyuAn@med.unc.edu). Yashen Chen is Research Fellow, Department of Radiology, University of North Carolina at Chapel Hill, NC 27599 (yasheng.chen@med.unc.edu). Wei Gao is a Ph.D. student, Department of Radiology, University of North Carolina at Chapel Hill, NC 27599 (wgao@med.unc.edu). Weili Lin is Professor, Department of Radiology, University of North Carolina at Chapel Hill, NC 27599 (weili.lin@med.unc.edu). Daniel B. Rowe is Associate Professor, Department of Biophysics, Medical College of Wisconsin, Milwaudee, WI 53226 (dbrowe@mcw.edu). Bradley S. Peterson is Professor, Department of Psychiatry, Columbia Medical Center and the New York State Psychiatric Institiute, New York, NY 10032 (petersob@childpsych.columbia.edu)

Abstract

Stochastic noise, susceptibility artifacts, magnetic field and radiofrequency inhomogeneities, and other noise components in magnetic resonance images (MRIs) can introduce serious bias into any measurements made with those images. We formally introduce three regression models including a Rician regression model and two associated normal models to characterize stochastic noise in various magnetic resonance imaging modalities, including diffusion-weighted imaging (DWI) and functional MRI (fMRI). Estimation algorithms are introduced to maximize the likelihood function of the three regression models. We also develop a diagnostic procedure for systematically exploring MR images to identify noise components other than simple stochastic noise, and to detect discrepancies between the fitted regression models and MRI data. The diagnostic procedure includes goodness-of-fit statistics, measures of influence, and tools for graphical display. The goodness-of-fit statistics can assess the key assumptions of the three regression models, whereas measures of influence can isolate outliers caused by certain noise components, including motion artifacts. The tools for graphical display permit graphical visualization of the values for the goodness-of-fit statistic and influence measures. Finally, we conduct simulation studies to evaluate performance of these methods, and we analyze a real dataset to illustrate how our diagnostic procedure localizes subtle image artifacts by detecting intravoxel variability that is not captured by the regression models.

Keywords

Diffusion tensor; Goodness-of-fit statistic; Influence measures; Normal approximation; Rician regression; Visualization

1. INTRODUCTION

Magnetic resonance image (MRI) is an noninvasive imaging technique used extensively for clinical diagnosis and medical research. MRIs, however, contain varying amounts of noise of diverse origins, including noise from stochastic variation, numerous physiological processes, eddy currents, artifacts from the differing magnetic field susceptibilities of neighboring tissues, rigid body motion, nonrigid motion, and many others (Huettel, Song, and McCarthy 2004). Some noise components, including bulk motion from cardiac pulsation and head or body movement, generate unusual observations, or statistical "outliers," that differ substantially from most MR data that do not contain those noise sources (at least, not to the same degree). Previous studies have shown that those noise components can introduce substantial bias into measurements and estimation made from those images, such as indices for the principle direction of fiber tracts in diffusion tensor images (Skare, Li, Nordell, and Ingvar 2000; Luo and Nicholas 2003; Nowark 1999). Identifying and reducing these noise components in MR images is essential to improving the validity and accuracy of studies designed to map the structure and function of the human body.

The raw data obtained during MRI scanning are complex values that represent the Fourier transformation of a magnetization distribution of a volume of tissue at a certain point in time. An inverse Fourier transform converts these raw data into magnitude, frequency, and phase components that more directly represent the physiological and morphological features of interest in the person being scanned. The magnetic susceptibility, chemical shift, and perfusion of tissues, for example, can be represented using either the magnitude or the phase angle of these Fourier-transformed data.

The electronic noise in the real and imaginary parts of the raw MR data are usually assumed to be independently Gaussian distributed (Henkelman 1985; Gudbjartsson and Patz 1995; Macorski 1996). Then, it can be shown theoretically that the Rician distribution is the model for characterizing the stochastic noise in the magnitude of MR data. Moreover, in practice, the Rician noise distribution of MR data has been experimentally validated using MR data (Haacke, Brown, Thompson, and Venkatesan 1999). Furthermore, the Rician distribution can be reasonably approximated by normal distributions at high signal-to-noise (SNR) ratios (Gudbjartsson and Patz 1995; Rowe and Logan 2004). Despite the extensive use of Rician and normal distributions in analyzing MR images (Kristoffersen 2007; Rowe 2005; Sijbers and den Dekker 2004; Sijbers, den Dekker, Scheunders, and Van Dyck 1998a; Sijbers, den Dekker, Verhoye, Van Audekerke, and Van Dyck 1998b), a formal statistical framework for characterizing stochastic noise in various MR imaging modalities has not yet been developed.

Formal assessment of the quality of MR images should include identification of nonstochastic noise components, such as those from susceptibility artifacts and rigid body motion. These nonstochastic noise sources usually introduce statistical outliers in some or all of the volume elements, called "voxels," of the image, the elemental units from which an image is constructed. Diagnostic procedures, such as an analysis of residuals, can be useful tools for detecting discrepancies between those outliers and other observations at all voxels. Moreover, even under the sole presence of stochastic noise, diagnostic methods are valuable for detecting discrepancies between MR data and fitted models at the voxel level. Such discrepancies can be caused by partial volume effects in the MR image (i.e., the presence of multiple tissues in the same volume element). In diffusion tensor images (DTIs), for instance, modeling these effects in voxels having multiple tissue compartments can be vitally important for reconstructing complex tissue structure in the human brain in vivo (Tuch et al. 2002; Alexander, Barker, and Arridge 2002).

The aim of this article is to introduce a Rician regression model and its related normal models to characterize noise contributions in various MRI modalities and to develop its associated estimation methods and diagnostic tools. We develop the estimation algorithms for calculating the maximum likelihood (ML) estimates of three regression models for MRI data. We develop a procedure to systematically assess the quality of MR images using a variety of diagnostic techniques, including an analysis of residuals, Cook's distance, goodness-of-fit test statistics, influence measures, and graphical analyses. We use the *p*-values of test statistics to evaluate directly the goodness of fit of the fitted regression models to the MRI data. Two diagnostic measures, standardized residuals and Cook's distance, identify in each voxel of the image outliers that can be caused by motion artifacts and other noise components. Graphical tools include three-dimensional (3D) images of statistical measures that can isolate problematic voxels, as well as two-dimensional (2D) plots for assessing the compatibility of the fitted regression model with data in individual voxels. Finally, we apply these diagnostic techniques to diffusion tensor images and demonstrate that the techniques are able to identify subtle artifacts and experimental variation not captured by the Rician model.

We will next present the Rician regression model and its two related normal models and discuss some of their statistical properties. Estimation algorithms will be used to maximize the likelihood function of the regression models proposed. Then, we will develop diagnostic procedures consisting of goodness-of-fit statistics, influence measures, and graphical analyses. Simulation studies will assess the empirical performance of the estimation algorithms and goodness-of-fit statistics under different experimental conditions. Finally, we will analyze a real dataset to illustrate an application of these methods, before offering some concluding remarks.

2. THE REGRESSION MODELS FOR MR IMAGES

2.1 Model Formulation

We usually acquire n MR images for each subject. Each MRI contains N voxels, and thus each voxel contains n measurements. We use $\{(S_i, x_i): i = 1, \cdots, n\}$ to denote the n measurements at a single voxel, where S_i denotes the MRI signal intensity and x_i includes all the covariates of interest, such as the gradient directions and gradient strengths for acquiring diffusion tensor images. In MR images, $S_i = \sqrt{R_i^2 + I_i^2}$ and ϕ_i are, respectively, the magnitude and phase of a complex number (R_i, I_i) from data in the imaging domain such that $R_i = S_i \sin(\phi_i)$ and $I_i = S_i \cos(\phi_i)$ for $i = 1, \cdots, n$.

The MR signal S_i is assumed to follow a Rician distribution with parameters μ_i and σ^2, denoted by $S_i \sim \mathrm{R}(\mu_i, \sigma^2)$, under the presence solely of stochastic noise (Rice 1945). Suppose that R_i and I_i are independent and follow normal distributions with the same variance σ^2, and with means $\mu_{R,i}$ and $\mu_{I,i}$, respectively. Thus, the joint density function of (S_i, ϕ_i) can be written as

$$p(S_i, \phi_i) = \frac{S_i}{2\pi\sigma^2} \exp\left\{-0.5\sigma^{-2}\left(S_i \sin(\phi_i) - \mu_{R,i}\right)^2 - 0.5\sigma^{-2} \times \left(S_i \cos(\phi_i) - \mu_{I,i}\right)^2\right\}.$$

Integrating out ϕ_i, we obtain the density function of the Rician distribution as follows:

$$p\left(S_i | \mu_i, \sigma^2\right) = \frac{S_i}{\sigma^2} \exp\left\{-0.5\sigma^{-2}\left(S_i^2 + \mu_i^2\right)\right\} I_o\left(\frac{\mu_i S_i}{\sigma^2}\right) 1 \times (S_i \geq 0), \quad (1)$$

where $\mu_i = \sqrt{\mu_{R,i}^2 + \mu_{I,i}^2}$, $1(\cdot)$ is an indicator function, and $I_0(z) = \int_0^{2\pi} \exp(z\cos\phi)\, d\phi/(2\pi)$ denotes the 0th order modified Bessel function of the first kind (Abramowitz and Stegun 1965).

We formally define a *Rician regression* model by assuming that

$$S_i \tilde{} R\left(\mu_i(\beta), \sigma^2\right) \quad \text{and} \quad \mu_i(\beta) = f(x_i, \beta), \tag{2}$$

where β is a $p \times 1$ vector in R^p and $f(\cdot, \cdot)$ is a known link function, which depends on the particular MR imaging modalities (e.g., anatomical, functional, DTI, and so on). Because the density in (1) does not belong to the exponential family, the Rician regression model is not a special case of a generalized linear model (McCullagh and Nelder 1989).

We calculate the kth moment of S_i given x_i as follows. Let $I_k(z)$ be the kth modified Bessel function of the first kind (Abramowitz and Stegun 1965) defined by $I_k(z) = \int_0^{2\pi} \cos(k\phi)\, e^{z\cos\phi} d\phi/(2\pi)$. It can shown that the kth moment of S_i given x_i (Sijbers, den Dekker, Scheunders, and Van Dyck 1998a) is calculated as

$$\mathrm{E}\left(S_i^k | x_i\right) = \left(2\sigma^2\right)^{k/2} \Gamma\left(1 + \frac{k}{2}\right) M\left(-\frac{k}{2}; 1; -\frac{\mu_i(\beta)^2}{2\sigma^2}\right), \tag{3}$$

where $\Gamma(\cdot)$ is the Gamma function and $M(\cdot)$ is the Kummer function (or confluent hypergeometric function) (Abramowitz and Stegun, 1965). The even moments of S_i given x_i are simple polynomials. For instance,

$$\begin{aligned} \mathrm{E}\left(S_i^2 | x_i\right) &= \mu_i(\beta)^2 + 2\sigma^2 \quad \text{and} \\ \mathrm{E}\left(S_i^4 | x_i\right) &= \mu_i(\beta)^4 + 8\sigma^2\mu_i(\beta)^2 + 8\sigma^4. \end{aligned} \tag{4}$$

However, the odd moments of S_i given x_i are much more complex; for instance,

$$\begin{aligned} \mathrm{E}(S_i | x_i) = \; & \sigma\sqrt{\tfrac{\pi}{2}} \exp\left\{-\tfrac{\mu_i(\beta)^2}{4\sigma^2}\right\} \\ & \times \left[\left(1 + \tfrac{\mu_i(\beta)^2}{2\sigma^2}\right) I_0\left(\tfrac{\mu_i(\beta)^2}{4\sigma^2}\right) \right. \\ & \left. + \; \tfrac{\mu_i(\beta)^2}{2\sigma^2} I_1\left(\tfrac{\mu_i(\beta)^2}{4\sigma^2}\right)\right]. \end{aligned} \tag{5}$$

The Rician distribution can be well approximated by a normal distribution at high SNR, defined by $\mu_i(\beta)/\sigma$. When SNR ≤ 1, the Rician distribution is far from being Gaussian. When SNR ≥ 2, $R(\mu_i(\beta), \sigma^2)$ can be closely approximated by a *normal regression* model (Gudbjartsson and Patz 1995) [Fig. 1(a)], which is given by

$$S_i \tilde{} N\left(\sqrt{\mu_i(\beta)^2 + \sigma^2}, \sigma^2\right) \quad \text{and} \quad \mu_i(\beta) = f(\mathbf{x}_i, \beta). \tag{6}$$

Moreover, the second moment of $R(\mu_i(\beta), \sigma^2)$ equals that of $N\left(\sqrt{\mu_i(\beta)^2+\sigma^2}, \sigma^2\right)$, whereas E $(S_i|x_i)$ in (5) can be accurately approximated by $\sqrt{\mu_i(\beta)^2+\sigma^2}$ even when SNR is close to 1 [Fig. 1(b)]. Furthermore, if SNR is greater than 5, then $\sqrt{\mu_i(\beta)^2+\sigma^2}=\mu_i(\beta)\sqrt{1+1/\mathrm{SNR}^2} \approx \mu_i(\beta)$. Thus, $R(\mu_i(\beta), \sigma^2)$ can be approximated by another normal regression model given by

$$S_i \tilde{} N\left(\mu_i(\beta), \sigma^2\right) \quad \text{and} \quad \mu_i(\beta) = f(\mathbf{x}_i, \beta). \tag{7}$$

2.2 Examples

The regression models proposed here include statistical models for various MRI modalities, including DTI and functional MRI. For the purposes of illustration, we consider the following five examples.

Example 1—Stochastic noise in MRI data follows a $R(0, \sigma^2)$ distribution, which is a highly skewed Rayleigh distribution. The first two moments of $R(0, \sigma^2)$ are given by $\mathrm{E}(S_i|x_i) = \sigma\sqrt{0.5\pi}$ and $\mathrm{E}\left(S_i^2|x_i\right) = 2\sigma^2$. Without any other noise components present, such as ghosting artifacts, we can use the MR data in the background of the image to estimate σ^2. However, under the presence of nonstochastic noise components, such as ghosting artifacts, the background MR signals do not follow a Rician distribution, and the estimate of σ^2 is usually a biased estimate of σ^2. Therefore, testing whether the MR signal in a single voxel truly follows a Rician model is useful for detecting the presence of nonstochastic noise components.

Example 2—If we apply an inversion snapshot FLASH imaging sequence to measure T_1 relaxation times, then we have $\mu_i(\beta) = \rho\left(1 - 2 \exp\left(-t_i T_1^{-1}\right)\right)$, where x_i is time t_i and β includes a pseudo proton density ρ and spin-lattice or longitudinal relaxation constant T_1 (Karlsen, Verhagen, and Bovee 1999). It has been shown that the use of the Rician model leads to a substantial increase in precision of the estimated T_1 (Karlsen et al. 1999).

If the decay of transverse magnetization is monoexponential and conventional spin-echo imaging is used, then $f(x_i, \beta)$ is given by $\mu_i(\beta) = \rho \exp\left(-\mathrm{TE}_i \times T_2^{-1}\right)$, where x_i is the echo time TE_i and $\beta = (\rho, T_2)$, in which T_2 is the spin-spin relaxation constant.

Example 3—In a functional MRI (fMRI) session, fMRI volumes are acquired repeatedly over time while a subject performs a cognitive or behavioral task. Over the course of the experiment, n fMRI volumes are typically recorded at acquisition times $t_1, \ldots, t_n$. The standard method for computing the statistical significance of task-related activations is to use only the magnitude MR image at time t_i for $i = 1, \ldots, n$. The magnitude image at time t_i follows a Rician distribution with $\mu_i(\beta) = x_i^T\beta$, the superscript T denotes transpose and x_i may include responses to differing stimulus types, the rest status, and various reference functions (Rowe and Logan 2005; den Dekker and Sijbers 2005).

Example 4—DTIs have been widely used to reconstruct the pathways of white matter fibers in the human brain in vivo (Basser, Mattiello, and LeBihan 1994a,b; Xu, Mori, Solaiyappan, Zijl, and Davatzikos 2002). A single shot echo-planar imaging (EPI) technique is often used to acquire diffusion-weighted imagings (DWI) with moderate resolution (e.g., 2.5 mm × 2.5 mm × 2.5 mm), and then diffusion tensors can be estimated using DWI data. In voxels with a single fiber population, a simple diffusion model assumes that

$$\mu_i(\beta) = S_0 \exp\left(-b_i r_i^T D r_i\right) \tag{8}$$

for $i = 1, \cdots, n$, where $x_i = (b_i, r_i, t_i)$, in which t_i is the acquisition time for the ith image, $r_i = (r_{i,1}, r_{i,2}, r_{i,3})^T$ is an applied gradient direction and b_i is the corresponding gradient strength. In addition, S_0 is the signal intensity in the absence of any diffusion-weighted gradient and the diffusion tensor $D = (D_{i,j})$ is a 3 × 3 positive definite matrix. The three eigenvectors of D constitute the three diffusion directions and the corresponding eigenvalues define the degrees of diffusivity along each of the three spatial directions. Many tractography algorithms attempt to reconstruct fiber tracts by connecting spatially consecutive eigenvectors corresponding to the largest eigenvalues of the diffusion tensors (DTs) across adjacent voxels.

The SNRs in diffusion-weighted (DW) images are relatively low. The DW imaging acquisition scheme usually consists of few baseline images with $b = 0\text{s/mm}^2$ and many DW images with b-values greater than zero. As an illustration, we selected a representative subject from an existing DTI dataset and calculated the estimates of S_0/σ and eigenvalues of D, denoted by $\lambda_1 \geq \lambda_2 \geq \lambda_3$, in all voxels containing anisotropic tensors (λ_1 was much larger than λ_3) [Figs.2 (a) and 2(b)]. For these anisotropic tensors, the SNR = S_0/σ in baseline images varied from 0 to 15 with a mean close to 6 [Fig. 2(c)], whereas λ_1 varied from 0.5 (10^{-3} mm^2/s) to 2.0 (10^{-3}mm^2/s) with a mean close to 1.0 (10^{-3} mm^2/s). For a moderate gradient strength $b_i \approx 1000\text{s/mm}^2$, the $\text{SNR} = \exp\left(-b_i r_i^T D r_i\right) \times (S_0/\sigma)$ in DWIs varied from 0 to 8 with a mean close to 2.5 [Fig. 2(d)].

To account for the presence of multiple fibers within a single voxel, a diffusion model with M compartments may be written as

$$\mu_i(\beta) = S_0 \sum_{k=1}^{M} p_k \exp\left(-b_i r_i^T D_k r_i\right), \tag{9}$$

where p_k denotes the proportion of each compartment such that $\sum_{k=1}^{M} p_k = 1$ and $p_k \geq 0$ and where D_k is the diffusion tensor for the kth compartment. Recent studies have shown that elucidating multiple fibers require large b values (Tuch et al. 2002; Alexander et al. 2002; Jones and Basser 2004). For instance, Alexander and Barker (2005) have shown that the optimal values of b for recovering two fibers are in the range [2200, 2800]s/mm^2. For large b values, the SNR in DWIs can be very close to zero [Fig. 2(d)].

Example 5—If we are only interested in the apparent diffusion coefficient (ADC) normal to the fiber direction in white matter, then we can use a single EPI technique to acquire MR images based on multiple b_i factors in the absence of a diffusion-weighted gradient (Kristoffersen 2007). A simple monoexponential diffusion model assumes that $\mu_i(\beta) = S_0 \exp(-b_i d)$ for $i = 1, \cdots, n$. The values of ADC are in the range of [0.2, 3] ($\times 10^{-3}$mm^2/s) for the human brain. Furthermore, a diffusion model with M compartments may be written as

$$\mu_i(\beta) = S_0 \sum_{k=1}^{M} p_k \exp(-b_i d_k).$$

2.3 Estimation Methods

We consider estimation algorithms for the two normal models (6) and (7). Because the normal model (7) is a standard nonlinear regression model, we can directly use the standard Levenberg-

Marquardt method to calculate the ML estimate of θ. For the normal model (6), we propose an iterative procedure to maximize its log-likelihood function given by

$$\ell\left(\beta,\sigma^2\right) = -0.5n\log\sigma^2 - 0.5\sum_{i=1}^{n}\left\{S_i - \sqrt{\mu_i(\beta)^2+\sigma^2}\right\}^2/\left(\sigma^2\right).$$

We use the Levenberg-Marquardt method to minimize $\sum_{i=1}^{n}\{S_i-\mu_i(\beta)\}^2$, which yields an initial estimator $\beta^{(0)}$, and we subsequently calculate $\left(\sigma^2\right)^{(0)}=\sum_{i=1}^{n}\left\{S_i-\mu_i\left(\beta^{(0)}\right)\right\}^2/n$. Given $(\sigma^2)^{(r)}$, we use the Levenberg-Marquardt method to calculate the $\beta^{(r+1)}$ that minimizes $\sum_{i=1}^{n}\left\{S_i-\sqrt{\mu_i(\beta)^2+(\sigma^2)^{(r)}}\right\}^2$. Conditional on $\beta^{(r+1)}$, we use the Newton-Raphson algorithm to calculate $\sigma^{(r+1)}$ by maximizing $\ell(\beta^{(r+1)}, \sigma^2)$. This iterative algorithm stops when the absolute difference between consecutive $\theta^{(t)}$s is smaller than a predefined small number, say 10^{-4}.

We introduce an efficient expectation-maximization (EM) algorithm (Dempster, Laird, and Rubin 1977) for maximizing the likelihood function of the Rician model (2). The key idea is to introduce a latent phase variable $\boldsymbol{\phi}_i \in [0, 2\pi]$ for each S_i such that the joint density of (S_i, $\boldsymbol{\phi}_i$) is given by

$$p(S_i,\phi_i|x_i) = \frac{1}{2\pi\sigma^2}S_i \times\exp\left(-\frac{\mu_i(\beta)^2+S_i^2-2S_i\mu_i(\beta)\cos(\phi_i)}{2\sigma^2}\right).$$

Let $Y_o = (S_1, x_1, \ldots, S_n, x_n)$ denote the observed data and $Y_m = (\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_n)$ denotes the missing data. The log-likelihood function of $Y_c = (Y_o, Y_m)$, defined by $L_c(\theta|Y_c)$, can be written as

$$-n\log\left(2\pi\sigma^2\right) + \sum_{i=1}^{n}\log S_i - 0.5\sigma^{-2}\sum_{i=1}^{n}\left\{\mu_i^2(\beta)+S_i^2 - 2S_i\mu_i(\beta)\cos(\phi_i)\right\}. \tag{10}$$

A standard EM algorithm consists of two steps: the expectation (E) step and the maximization (M) step as follows. The E-step evaluates $Q(\theta|\theta^{(r)}) = \mathrm{E}\{L_c(\theta|Y_c)|Y_o, \theta^{(r)}\}$, where the expectation is taken with respect to the conditional distribution $p\left(Y_m|Y_o,\theta^{(r)}\right)=\prod_{i=1}^{n}p\left(\phi_i|S_i,\theta^{(r)}\right)$. We can show that

$$p(\phi_i|S_i,\theta) = \frac{1}{2\pi I_0(\sigma^{-2}S_i\mu_i(\beta))}\exp\left\{\sigma^{-2}S_i\mu_i(\beta)\cos(\phi_i)\right\} \times 1(\phi_i\in[0,2\pi]).$$

Thus, $Q(\theta|\theta^{(r)})$ is given by

$$-n\log\left(\sigma^2\right) - 0.5\sigma^{-2} \times\sum_{i=1}^{n}\left\{\mu_i^2(\beta)+S_i^2-2S_i\mu_i(\beta)W_i\left(\theta^{(r)}\right)\right\}, \tag{11}$$

where $W_i(\theta) = I_1(\sigma^{-2}f(x_i,\beta)S_i)/I_0(\sigma^{-2}f(x_i,\beta)S_i)$.

The M-step is to determine the $\theta^{(r+1)}$ that maximizes $Q(\theta|\theta^{(r)})$. However, because the M-step does not have a closed form, $\theta^{(r+1)}$ is obtained via two conditional maximization steps (Meng and Rubin 1993). Given $\beta^{(r)}$, we can derive

$$\left(\sigma^2\right)^{(r+1)}=0.5n^{-1}\sum_{i=1}^{n}\left\{\mu_i^2\left(\theta^{(r)}\right)+S_i^2-2S_i\mu_i\left(\theta^{(r)}\right)W_i\left(\theta^{(r)}\right)\right\}.$$

Conditional on $(\sigma^2)^{(r+1)}$, we can determine $\beta^{(r+1)}$ by minimizing $G\left(\beta|\beta^{(r)}\right)=\sum_{i=1}^{n}\left\{\mu_i(\beta)-W_i\left(\theta^{(r)}\right)S_i\right\}^2$. This is a standard nonlinear least problem, to which the Levenberg-Marquardt method can be applied. Furthermore, we may employ a generalized EM algorithm, in which the E-step is unchanged, but we replace the M-step with a generalized M-step to identify a $\beta^{(r+1)}$ such that $G(\beta^{(r+1)}|\beta^{(r)}) \leq G(\beta^{(r)}|\beta^{(r)})$. Under mild conditions, the sequence $\{\theta^{(r)}\}$ obtained from the EM algorithm converges to the ML estimate, denoted by $\hat{\theta}$ (Meng and Rubin 1993).

The next important issue is to evaluate the covariance matrix of $\hat{\theta}$, which can be obtained by inverting either the Hessian matrix or the Fisher information matrix of the observed-data log-likelihood function. For instance, for the normal model (6), it is straightforward to calculate the second derivative of $\ell(\beta, \sigma^2)$. For the Rician model (2), we use the missing information principle (Louis 1982). Calculation of the first and second derivatives of $L_c(\theta|Y_c)$ with respect to θ is straightforward and hence is omitted here for brevity.

3. A DIAGNOSTIC PROCEDURE

We propose a diagnostic procedure to identify noise components in MR images at all levels of the SNR. Our diagnostic procedure has three major components: (1) the use of goodness-of-fit test statistics to test the assumptions of the Rician model across all voxels of the image; (2) the use of influence measures to identify outliers; (3) the use of 2D and 3D graphs to search for various artifacts and to detect intravoxel variability. At a high SNR, these diagnostic measures of the Rician model reduce to those of the normal models (6) and (7). Thus, we will not specifically develop diagnostic measures of the two normal models.

3.1 Goodness-of-Fit Test Statistics

We develop test statistics to check model misspecification in the Rician model (2). These test statistics are valuable for revealing two kinds of challenges in working with MR images. The first is to identify those voxels in which the MR signal contains substantial noise components that are different from stochastic noise. The second challenge is to identify those voxels in which the signal is affected strongly by partial volume effects.

Thus, we are interested in testing whether $f(x_i, \beta)$ is correctly specified. In most statistical models, including generalized linear models, testing the specification of the link function is equivalent to testing the mean structure of the response variable (Stute 1997). However, in the Rician model (2), $E(S_i|x_i)$ does not have a simple form, and so direct testing the mean structure of the response is likely to be tedious and difficult. Let $W(\theta) = I_1(B(\theta))/I_0(B(\theta))$, where $B(\theta) = \sigma^{-2}f(x, \beta)S$. We also note the simple equality $E[W(\theta)S|x] = f(x, \beta)$, when the Rician model (2) is correctly specified. Thus, we suggest testing $h(\theta) = E[W(\theta)S|x] - f(x, \beta) = 0$, for which the null and alternative hypotheses are stated as follows:

$$H_0^{(1)}: h(\theta) = 0 \quad \text{for some} \quad \theta \in \Theta \quad \text{versus}$$
$$H_1^{(1)}: h(\theta) \neq 0 \quad \text{for all} \quad \theta \in \Theta. \tag{12}$$

Because $W(\theta)$ is close to one at a high SNR, testing $H_0^{(1)}$ is essentially testing whether $E(S|x) = f(x, \beta)$ in the normal model (7).

To test $H_0^{(1)}$, we develop two test statistics as follows. The first of these, the conditional Kolmogorov test (CK), is

$$\mathrm{CK}_1 = \sup_u |T_1(u; \widehat{\theta})|, \tag{13}$$

where $T_1(u; \hat{\theta})$ is defined as

$$T_1(u; \widehat{\theta}) = n^{-1/2} \sum_{i=1}^{n} 1\left(x_i^T \widehat{\beta} \leq u\right) \left[W_i(\widehat{\theta}) S_i - \mu_i(x_i, \widehat{\beta}) \right]. \tag{14}$$

Under the null hypothesis, $E[T_1(u;\theta_*)]$ should be close to zero, where $\theta_* = (\beta_*, \sigma_*^2)$ is the true value of θ. Therefore, a large value of CK_1 leads to rejection of the null hypothesis $H_0^{(1)}$.

We must derive the asymptotic null distribution of CK_1 to test rigorously whether $H_0^{(1)}$ is true. We regard $T_1(u; \hat{\theta})$ as a stochastic process indexed by $u \in R$. We can show that under $H_0^{(1)}$, as $n \to \infty$,

$$\begin{aligned} T_1(u; \widehat{\theta}) &= T_1(u;\theta_*) + \partial_\theta T_1(u;\theta_*)(\widehat{\theta} - \theta_*) + o_p(1) \\ &= T_1(u;\theta_*) + \Delta_1(u)\ \sqrt{n}(\widehat{\theta} - \theta_*) + o_p(1), \end{aligned}$$

where $\Delta_1(u)$ is defined by

$$\begin{aligned} \Delta_1(u) = & \int [\partial_\theta W(\theta_*) S - \partial_\theta f(x, \beta_*)] 1 \\ & \times \left(x^T \beta_* \leq u\right) p(S|x, \theta_*)\, p(x)\, dS\, dx. \end{aligned}$$

Moreover, using the central limit theorem (van der Vaart and Wellner 1996), we can show that

$$\sqrt{n}(\widehat{\theta} - \theta_*) = n^{-1/2} \sum_{i=1}^{n} \psi(S_i, x_i; \theta_*) + o_p(1), \tag{15}$$

where $\psi(\cdot, \cdot; \theta_*)$ is a known influence function depending on the likelihood function of the Rician model (2). Finally, using empirical process theory (van der Vaart and Wellner 1996), we can show that the asymptotic null distribution of CK_1 depends on the asymptotic distribution of $\left(T_1(\cdot, \theta_*),\ \sqrt{n}(\widehat{\theta} - \theta_*)^T\right)^T$, which is given in Theorem 1.

The second test statistic that we propose is based on

$$T_2\left(\alpha, u;\widehat{\theta}\right)=n^{-1/2}\sum_{i=1}^{n}\left[W_i\left(\widehat{\theta}\right)S_i-\mu_i\left(\widehat{\beta}\right)\right]1\left(x_i^T\alpha\le u\right), \tag{16}$$

where $\Pi=\{\alpha\in R^d:\alpha^T\alpha=1\}\times[-\infty,\infty]$. Following the reasoning in Escanciano (2006), we can show that $H_0^{(1)}$ is equivalent to testing

$$E\left\{\left[W_i(\theta)S_i-\mu_i(\beta)\right]1\left(x^T\alpha\le u\right)\right\}=0 \tag{17}$$

for almost every $(\alpha, u)\in\Pi$ for some $\theta_*\in\Theta$. Let $F_{n,\alpha}(u)$ be the empirical distribution function of $\{\alpha^T x_i : i=1,\ldots,n\}$. Then, we define the Cramer-von Mises test statistic as follows:

$$\mathrm{CM}_1=\int_\Pi T_2\left(\alpha, u;\widehat{\theta}\right)^2 F_{n,\alpha}(du)\,d\alpha, \tag{18}$$

where $d\alpha$ is taken with respect to the uniform density on the unit sphere. A simple algorithm for computing CM_1 can be found in Escanciano (2006). A large value of CM_1 leads to rejection of $H_0^{(1)}$. Similar to CK_1, we can show that $T_2(\alpha, u;\hat{\theta})$ is approximated as

$$T_2\left(\alpha, u;\widehat{\theta}\right)=T_2(\alpha, u;\theta_*)+\Delta_2(\alpha, u)\sqrt{n}\left(\widehat{\theta}-\theta_*\right)+o_p(1),$$

where $\Delta_2(\alpha, u)=\int[\partial_\theta W(\theta_*)S-\partial_\theta f(x,\beta_*)]\,1(\alpha^T x\le u)p(S|x,\theta_*)p(x)dSdx$. Therefore, the asymptotic null distribution of CM_1 depends on the asymptotic distribution of $\left(T_2(\alpha, u;\theta_*),\ \sqrt{n}\left(\widehat{\theta}-\theta_*\right)^T\right)^T$, which is also given in Theorem 1. The detailed proof of Theorem 1 can be found in a supplementary report and is available at www.amstat.org/publications/jasa/supplementalmaterials. We are now led to the following theorem.

Theorem 1—Under the null hypothesis $H_0^{(1)}$, we have the following results:

1. $\sqrt{n}\left(\widehat{\theta}-\theta_*\right)=n^{-1/2}\sum_{i=1}^{n}\psi_{n,i}+o_p(1)$.
2. $\left(T_1(\cdot;\theta_*),\ \sqrt{n}\left(\widehat{\theta}-\theta_*\right)^T\right)^T$ converges in distribution to $\left(G_1(\cdot;\theta_*), \nu_1^T\right)^T$, where $G_1(\cdot;\theta_*), \nu_1^T$ is a Gaussian process with mean zero and covariance function $C_1(u_1, u_2)$, which is given by

$$\begin{aligned}C_1(u_1,u_2)=\ &\int\int\begin{pmatrix}\left[W(\theta)_*S_i-f(x,\beta_*)\right]1\left(x^T\beta_*\le u_1\right)\\ \psi(S,x;\theta_*)\end{pmatrix}\\ &\times\begin{pmatrix}\left[W(\theta_*)S-f(x,\beta_*)\right]1\left(x^T\beta_*\le u_2\right)\\ \psi(S,x;\theta_*)\end{pmatrix}^T\\ &\times p(S|x;\theta_*)\,dS\,dp(x).\end{aligned} \tag{19}$$

3. CK_1 converges in distribution to $\sup_u|T_1(u;\theta_*)+\Delta_1(u)^T\nu_1|$.

4. $\left(T_2(\cdot,\cdot;\theta_*),\ \sqrt{n}\left(\widehat{\theta}-\theta_*\right)^T\right)^T$ converges in distribution to $\left(G_2(\cdot,\cdot;\theta_*),\nu_1^T\right)^T$, where $\left(G_2(\cdot,\cdot;\theta_*),\nu_1^T\right)$ is a Gaussian process with mean zero and covariance function $C_2((\alpha_1, u_1), (\alpha_2, u_2))$, which is given by

$$\begin{aligned} C_2((\alpha_1,u_1),(\alpha_2,u_2)) = & \iint \begin{pmatrix} [W(\theta_*)S - f(x,\beta_*)] \\ \times 1\left(x^T\alpha_1 \le u_1\right) \\ \psi(S,x;\theta_*) \end{pmatrix} \\ & \times \begin{pmatrix} [W(\theta_*)S - f(x,\beta_*)] \\ \times 1\left(x^T\alpha_2 \le u_2\right) \\ \psi(S,x;\theta_*) \end{pmatrix}^T \\ & \times p(S|x,\theta_*)\,dS\,dp(x). \end{aligned} \tag{20}$$

5. CM_1 converges in distribution to $\int_\Pi |T_2(\alpha,u;\theta_*) + \Delta_2(\alpha,u)\nu_1|^2 F_\alpha(du)\,d\alpha$, where $F_\alpha(u)$ is the true cumulative distribution function of $\alpha^T x$.

Theorem 1 characterizes the limiting distributions of CK_1 and CM_1 under the null hypotheses.

Because $E\left(S_i^2|x_i\right)$ has a simple form, we further use the second moment of S_i given x_i to test the specification of the link function. Specifically, the null and alternative hypotheses are given by

$$\begin{aligned} H_0^{(2)}&: \mathrm{E}\left(S^2|x\right) = f(x,\beta)^2 + 2\sigma^2 \quad \text{for some} \quad \theta \in \Theta, \\ H_0^{(2)}&: \mathrm{E}\left(S^2|x\right) \ne f(x,\beta)^2 + 2\sigma^2 \quad \text{for all} \quad \theta \in \Theta. \end{aligned}$$

Similar to testing $H_0^{(1)}$ against $H_1^{(1)}$, we introduce two other stochastic processes given by

$$\begin{aligned} T_3(u;\theta) &= n^{-1/2}\sum_{i=1}^n 1\left(x_i^T\beta \le u\right)\left[S_i^2 - \mu_i(\beta)^2 - 2\sigma^2\right] \quad \text{and} \\ T_4(\alpha,u;\theta) &= n^{-1/2}\sum_{i=1}^n \left[S_i^2 - \mu_i(\beta)^2 - 2\sigma^2\right] 1\left(x_i^T\alpha \le u\right). \end{aligned}$$

Based on $T_3(u;\theta)$ and $T_4(\alpha, u;\theta)$, we can develop two additional test statistics:

$$\begin{aligned} \mathrm{CK}_2 &= \sup|T_3\left(u;\widehat{\theta}\right)| \quad \text{and} \\ \mathrm{CM}_2 &= \int_\Pi^u T_4\left(\alpha,u;\widehat{\theta}\right)^2 F_{n,\alpha}(du)\,d\alpha. \end{aligned} \tag{21}$$

Similar to the reasoning in Theorem 1, we can establish the asymptotic null distributions of CK_2 and CM_2, which we therefore omit here. Because the normal model (6) has the same second moment as the Rician model (2), the test statistics CK_2 and CM_2 are valid for model (6) at all levels of the SNR. So far, we have introduced four test statistics CK_1, CK_2, CM_1, and CM_2, each of which may have different sensitivities in detecting the misspecification of a Rician model in various circumstances, which we will investigate with the simulation studies of Section 4.

We note two types of correlation existing in CK_1, CK_2, CM_1, and CM_2 at the local and global levels. At the local level, there may be strong correlations among these four test statistics in

each voxel, because the same MRI data within the voxel are used to calculate them. At the global level, we calculate these four test statistics across multiple brain regions or across the many voxels of the imaging volume. MRI data in small spatial neighborhoods show strong similarity, whereas MRI data in voxels more distant from one another show less similarity. Thus, the same test statistics $CK_1(d)$ [or $CK_2(d)$, $CM_1(d)$, and $CM_2(d)$] are likely to be positively correlated in small spatial neighborhoods, where d denotes a particular voxel in an MRI. Finally, we need to compute the uncorrected and corrected p-values of these four test statistics at the local and global levels.

3.2 Resampling Method

Although the asymptotic distributions of $CK_1(d)$, $CK_2(d)$, $CM_1(d)$, and $CM_2(d)$ have been derived in Theorem 1, these limiting distributions usually have complicated analytic forms. To alleviate this difficulty, we develop a resampling method to estimate the null distribution of the statistic $CK_1(d)$ in each of the voxels in the MRI data. The next issue is to solve the problem of multiple testing. Because it is difficult to compute an accurate p-value of $CK_1(d)$ at each voxel, we avoid use of the false discovery rate and choose to control the family-wise error rate based on the maxima of the $CK_1(d)$ statistics defined by $CK_{1,D} = \max_{d\in D} CK_1(d)$, where D denotes the brain region. Specifically, we can easily extend the proposed resampling method to approximate the null distribution of the statistic $CK_{1,D}$. In the following, we will introduce voxel d into all of the notation, if necessary. Because we can develop similar methods for CK_2, CM_1, and CM_2, we avoid such repetition and simply present the six key steps in generating the stochastic processes that have the same asymptotic distribution as $CK_1(d)$ and $CK_{1,D}$.

Step 1—Generate independent and identically distributed random variables, $\{v_i^{(q)}: i=1,\ldots,n\}$, from a $N(0, 1)$ distriution for $q = 1, \ldots, Q$, where Q is the number of replications, say $Q = 1{,}000$.

Step 2—Calculate

$$T_1\left(u;\widehat{\theta}(d)\right)^{(q)} = n^{-1/2}\sum_{i=1}^{n} v_i^{(q)}\left\{\mathrm{E}_i\left(\widehat{\theta}(d)\right)1\left(\mathbf{x}'_i\widehat{\beta}(d)\le u\right) - \widehat{\Delta}_1(d,u)\,\psi_{ni}(d)\right\},$$

where $\mathrm{E}_i(\hat{\theta}(d)) = W_i(\hat{\theta}(d))S_i - \mu(x_i, \hat{\beta}(d))$ and $\widehat{\Delta}_1(d,u) = n^{-1}\sum_{i=1}^{n}\partial_\theta \mathrm{E}_i\left(\widehat{\theta}(d)\right)1\left(\mathbf{x}'_1\widehat{\beta}(d)\le u\right)$. Note that conditional on the observed data, $T_1(u; \hat{\boldsymbol{\theta}}(d))^{(q)}$ converges weakly to the desired Gaussian process in Theorem 1 as $n \to \infty$ (van der Vaart & Wellner, 1996).

Step 3—Calculate the test statistics $\mathrm{CK}_1^{(q)}(d) = \sup_u |T_1\left(u;\widehat{\theta}(d)\right)^{(q)}|$ and $\mathrm{CK}_{1,\mathcal{D}}^{(q)} = \sup_{d\in\mathcal{D}} \mathrm{CK}_1^{(q)}(d)$ and obtain $\left\{\mathrm{CK}_1^{(q)}(d): q=1,\ldots,Q\right\}$ and $\left\{\mathrm{CK}_{1,\mathcal{D}}^{(q)}: q=1,\ldots,Q\right\}$.

Step 4—Calculate the p-value of $CK_1(d)$ using $\left\{\mathrm{CK}_1^{(q)}(d): q=1,\ldots,Q\right\}$.

Step 5—Calculate the p-value of $CK_1(d)$ at each voxel d of the region according to $p(d) \approx Q^{-1}\sum_{q=1}^{Q} 1\left(\mathrm{CK}_1^{(q)}(d) \ge \mathrm{CK}_1(d)\right)$.

Step 6—Calculate the corrected p-value of $\mathrm{CK}_1(d)$ at each voxel d of the region using $p_{\mathcal{D}}(d) \approx Q^{-1}\sum_{q=1}^{Q} 1\left(\mathrm{CK}_{1,\mathcal{D}}^{(q)} \geq \mathrm{CK}_1(d)\right)$.

Finally, we present a plot of the uncorrected and corrected - $\log_{10}(p)$ values for our various test statistics, such as CM_1. Because the previous procedure only requires the computation of all components of $T_1(u; \hat{\boldsymbol{\theta}}(d))$ once and the repeated calculation of $\mathrm{CK}_1^{(q)}(d)$, it is computationally efficient. To identify the precise source of noise that is responsible for misspecification of the model, we need to develop influence measures to quantify the influence of each data point at each voxel.

3.3 Influence Measures

Next, we develop two influence measures that identify in each voxel of an MR image statistical "outliers," which exert undue influence on the estimation of the parameters and fitted values of the model. These influence measures are based on case-deletion diagnostics, which have been studied extensively in regression models (Cook and Weisberg 1982; Wei 1998). Influence measures for the Rician regression model, however, have not been developed previously. Therefore, we now discuss how to develop case-deletion measures for the Rician model.

Henceforth, we assume that σ^2 is a nuisance parameter and define $U(\beta) = (\mu_1(\beta), \ldots, \mu_n(\beta))^T$, $V(\theta) = \mathrm{diag}(V_1(\theta), \ldots, V_n(\theta))$, and $S_W(\theta) = (W_1(\theta)S_1, \ldots, W_n(\theta)S_n)^T$, where $V_i(\theta) = \sigma^{-2}\mathrm{Var}(S_i W_i(\theta)) = -\sigma^{-2}\mu_i(\beta)^2 + E\left[\sigma^{-2}S_i^2 W_i(\theta)^2\right]$. Thus, the score function for β is given by $SC_n(\beta) = \sigma^{-2}D(\beta)^T V(\theta)e(\beta)$, where $D(\beta) = \partial U(\beta)/\partial\beta^T$ is an $n \times p$ matrix with the ith row $\partial\mu_i(\beta)/\partial\beta^T$ and $e(\theta) = V(\theta)^{-1}[S_W(\theta) - U(\beta)]$. Furthermore, the Fisher information matrix for β takes the form

$$\begin{aligned} F_n(\beta) &= \sigma^{-2}\sum_{i=1}^{n} \frac{\partial\mu_i(\beta)}{\partial\beta} V_i(\theta) \frac{\partial\mu_i(\beta)}{\partial\beta^T} \\ &= \sigma^{-2} D(\beta)^T V(\theta) D(\beta). \end{aligned}$$

To develop influence measures, we can write the ML estimate of β as $\hat{\beta} = [D(\hat{\beta})^T V(\hat{\theta}) D(\hat{\beta})]^{-1} D(\hat{\beta})^T V(\hat{\theta})\hat{Z}$, where $\hat{Z} = Z(\hat{\beta})$ and $Z(\beta) = D(\beta)\beta + e(\beta)$ (Jorgensen 1992). Thus, $\hat{\beta}$ can be regarded as the generalized least-squares estimate of the following linear model:

$$\widehat{Z} = D\left(\widehat{\beta}\right)\beta + e \quad \text{and} \quad \mathrm{Var}(e) = \sigma^2 V\left(\widehat{\theta}\right)^{-1}. \tag{22}$$

We can extend the existing diagnostics for linear regression to Rician regression (Cook and Weisberg 1982; Jorgensen 1992; Wei 1998). Because $V(\hat{\theta})^{-1}$ reduces to an identity matrix at a high SNR, model (22) just reduces to a standard linear regression model.

We introduce two influence measures based on the representation of the linear model (22) as follows.

1. The residuals and standardized residuals are given by

$$\begin{aligned} \widehat{r_i} &= u_i^T \widehat{V}\left(\widehat{\theta}\right)^{1/2} \left\{\widehat{Z} - D\left(\widehat{\beta}\right)\widehat{\beta}\right\} \quad \text{and} \\ \widehat{t_i} &= \sigma^{-1}\widehat{r_i}/\sqrt{1 - h_{i,i}}, \end{aligned} \tag{23}$$

where u_i is an $n \times 1$ vector with ith element and all others zero, and where $\{h_{i,i}: i \leq n\}$ are the diagonal elements of the hat matrix H defined by

$$H = V(\theta)^{1/2} D(\beta) \left[D(\beta)^T V(\theta) D(\beta) \right]^{-1} D(\beta)^T V(\theta)^{1/2}. \quad (24)$$

Residuals are highly informative about the compatibility of a postulated model with the observed data. If a Rician model is correct, residuals should be centered around zero, and plots of the residuals should exhibit no systematic tendencies. Exploring residual plots may reveal nonconstant variance, curvature, and the need for transformation in the regression, and therefore the analysis of residuals has been among the most widely used tools for assessing the validity of model specification (Cook and Weisberg 1982). To assess the magnitudes of the residuals, we compare the standardized residuals with the conventional benchmark 2.5. In other words, we regard the ith data point (S_i, x_i) as having excess influence if $|\hat{t}_i|$ is larger than 2.5. We will plot the number of outliers at each voxel of the MR image. Voxels with many outliers need some further exploration.

2. Cook's distance (Cook and Weisberg 1982) can be defined as

$$C_i = (\hat{\beta} - \hat{\beta}_{(i)})^T \left[D(\hat{\beta})^T V(\theta) D(\hat{\beta}) \right] (\hat{\beta} - \hat{\beta}_{(i)}) / \sigma^2, \quad (25)$$

where $\hat{\beta}_{(i)}$ denotes the ML estimate of β based on a sample size of n - 1 with the ith case deleted. Instead of calculating $\hat{\beta}_{(i)}$ directly, we compute the first order approximation of $\hat{\beta}_{(i)}$, denoted by $\tilde{\beta}^I_{(i)}$, which is given by

$$\tilde{\beta}^I_{(i)} \approx \hat{\beta} - \left[D(\hat{\beta})^T V(\theta) D(\hat{\beta}) \right]^{-1} V_i(\theta)^{1/2} D_i(\hat{\beta}) \hat{r}_i / (1 - h_{i,i}),$$

where $D_i(\hat{\beta})^T$ is the ith row of $D(\hat{\beta})$. Therefore, we get the first-order approximation of C_i, denoted by C^I_i, as $C^I_i = h_{i,i} \tilde{t}^2_i / (1 - h_{i,i})$. Following Zhu and Zhang (2004), we compare nC^I_i with $3p$ to reveal the level of influence of (S_i, x_i) for each i at each voxel.

3.4 3D and 2D Graphics

We use 3D images of our various statistical measures to isolate all voxels in the image where specification of a Rician model is problematic. After computing the p-value of each test statistic (CM_1, CM_2, CK_1, or CK_2) at each voxel of the image, we create a 3D image of the -$\log_{10}(p)$ values for each statistic and then explore these values efficiently across all voxels. In addition, we calculate t_i and C^I_i, compute the number of outliers at each voxel, and create a 3D image for each of these influence measures (Luo and Nichols 2003). For instance, if the p-value of CK_1 in a specific voxel is smaller than a given significance level, then we have strong evidence that the noise characteristics at that voxel are non-Rician and are likely to derive from non-physiological sources that may obscure valid statistical testing in those regions. Moreover, a large number of outliers appearing in several images taken sequentially, as they are in fMRI, may indicate a problematic noise source spanning the duration over which those images are obtained, as is often true of head motion, signal drift, and other similar artifacts. In addition, we also inspect the spatial clustering behavior of the voxels, which have large values of influence measures and test statistics, such as the cluster sizes of groups of outliers. More

detailed examination of the 2D graphs for these voxels is indicated. These graphs include maps of the number of outliers per slice and per image, index plots of influence measures, and various plots of residuals that can reveal anomalies, such as nonconstant variance, curvature, transformations, and outliers in the data (Cook and Weisberg 1982; Luo and Nichols 2003). Thus, these 2D graphs of our diagnostic measures are used to help identify the nature and source of the disagreement between the Rician model and the observed MR signals at a particular voxel.

4. SIMULATION STUDIES

We conducted three sets of Monte Carlo simulations to examine the accuracy of using the Rician model, the two normal models and test statistics under differing experimental settings. The first set illustrated the performance of the Rician model and the two normal models for ADC imaging. The second set of simulations evaluated the sensitivity of the goodness-of-fit test statistics in detecting multiple tensor compartments within individual voxels of a DTI dataset. The third set of Monte Carlo simulations evaluated the sensitivity of the goodness-of-fit statistics in detecting head motion in MR images.

4.1 Apparent Diffusion Coefficient Mapping

The first set of Monte Carlo simulations was to compare the estimated ADC using the Rician model (2) and the two normal models (6) and (7). We set $d = 2 \times 10^{-3}$ mm^2/s, $S_0 = 500$, $b = [0, 50, 100, \ldots, 1100]$s/mm^2, and five different S_0/σ {2, 4, 6, 10, 15} for all Monte Carlo simulations. For $S_0/\sigma = 2$, the values of the SNR were in the range of [0.366, 2]. At each S_0/σ, 4,000 diffusion-weighted datasets were generated. Under each model, we calculated the parameter estimates $\hat{\theta} = (\hat{d}, \hat{S}_0, \hat{\sigma}^2)$. We finally calculated the biases, the empirical standard errors (SE), and the mean of the standard error estimates (SEE) based on the results from the 4,000 simulated ADC datasets (Table 1). At all S_0/σ, the estimates from model (2) had smaller biases, but larger SEs, whereas models (6) and (7) had larger biases, but smaller SEs. When $S_0/\sigma \geq 15$, models (2), (6), and (7) had comparable biases and SEs in the parameter estimates. In addition, the SE and its corresponding SEE are relatively close to each other when $S_0/\sigma \geq 4$.

4.2 Evaluating the Test Statistics for DTI Data Assuming the Presence of Fiber Crossings

We assessed the empirical performance of CK_i and CM_i for $i = 1, 2$ as our test statistics for detecting the misspecified single diffusion model (8) when two diffusion compartments were actually present in the same voxel. Simulated data were drawn from the diffusion model (9) with 2 diffusion compartments, in which $p_1 = 1 - p_2$ was set at either 0.0 or 0.5, $D_1 = \text{diag}(1.7, 0.2, 0.2)$ ($\times 10^{-3}$ mm^2/s), and $D_2 = \text{diag}(0.2, 1.7, 0.2)$ ($\times 10^{-3}$ mm^2/s). In particular, $p_1 = 0.0$ corresponded to a single diffusion compartment, whereas $p_1 = 0.5$ corresponded to two diffusion compartments. The principal directions of D_1 and D_2 were, respectively, at (1, 0, 0) and (0, 1, 0). The mean diffusivity trace(D)/3 for both D_1 and D_2 was set equal to 1×10^{-3} mm^2/s, which is typical of values for normal cerebral tissue (Skare et al. 2000). We generated the Rician noise with $S_0 = 150$ and selected S_0/σ to be 5, 10, 15, 20, and 25, respectively. Our DTI scheme included 6 baselines, 30 diffusion weighted uniformly arranged directions at b_1, and the same set of gradient directions at b_2. We chose three combinations of (b_1, b_2): (1,000, 1,000), (1,000, 3,000), and (3,000, 3,000) s/mm^2 to examine the sensitivity of differing b factors in detecting multiple fiber directions. For each simulation, 1,000 simulated datasets were used to estimate the nominal significance level (i.e., rejection levels for the null hypothesis). Finally, for each simulated dataset, we applied the resampling method with $Q = 1{,}000$ replications to calculate the four p-values of CK_i and CM_i for $i = 1, 2$ and then applied the false discovery rate procedure to correct for multiple comparisons at a significance level of 5% as suggested by a reviewer (Benjamini and Hochberg 1995).

Table 2 presents estimates for the rejection rates of the four test statistics after correction for multiple comparions based on the false discovery rate procedure. We observed that in a single compartment, the rejection rates of CK_i and CM_i for $i = 1, 2$ were smaller than the nominal level. Overall, the rejection rates in all cases were relatively accurate, and the Type I errors were not excessive. These findings suggested that the resampling method worked reasonably well under the null hypothesis. Differing (b_1, b_2) combinations strongly influenced the finite performance of the four test statistics in detecting the presence of two compartments. Specifically, compared with other (b_1, b_2) combinations, $(b_1, b_2) = (1{,}000, 3{,}000)$ s/mm^2 provided the best performance. Under $(b_1, b_2) = (1{,}000, 3{,}000)$ s/mm^2, CK_1 and CM_1 provided substantial power to detect the presence of two diffusion compartments. Compared with the other three statistics, CK_1 performed well; moreover, consistent with our expectations, increasing S_0/σ reduced the Type II errors and improved the power of the statistic CK_1 to detect the presence of two compartments. Therefore, these simulations suggested that the choice of b strongly influenced the performance of these test statistics and the test CK_1 was a useful tool for detecting the presence of multiple compartments. The selection of optimal b values in detecting multiple compartments warrants further research (Alexander et al. 2002;Jones, Horsfield, and Simmons 1999).

4.3 Evaluating the Test Statistics in the Presence of Head Motion

We also assessed the empirical performances of CK_i and CM_i for $i = 1, 2$ as test statistics for detecting the misspecified single diffusion model (8) at a single voxel in the presence of head motion. We simulated data contaminating head motion in the image as follows. We used a DTI scheme starting with 5 baselines and followed with 45 diffusion weighted uniformly arranged directions at $b_1 = 1{,}000$ s/mm^2. We simulated data from the diffusion model (8) with $D_1 = \text{diag}(0.2, 1.7, 0.2)$ ($\times 10^{-3}$ mm^2/s) in the first $[50 \times p_1]$ acquisitions, and then generated data from the diffusion model (8) with $D_2 = \text{diag}(0.7, 0.7, 0.7)$ ($\times 10^{-3}$ mm^2/s) from the last $50 - [50 \times p_1]$ acquisitions, where $[\cdot]$ denoted the largest integer smaller than $50 \times p_1$. In addition, the probability p_1 was selected to be 0.5 and 0.7, which reflected the different degrees of head motion. We also generated Rician noise from (1) with $S_0 = 150$ and set S_0/σ to be 5, 10, 15, 20, and 25, respectively. For each simulation, 1,000 simulated datasets were used to estimate the nominal significance level (i.e., rejection levels for the null hypothesis). Finally, for each simulated dataset, we applied the resampling method with $Q = 1{,}000$ replications to calculate the four p-values of CK_i and CM_i for $i = 1, 2$, and then applied the false discovery rate procedure to correct for multiple comparisons at a significance level of 5% as suggested by a reviewer.

Table 3 presented estimates for the rejection rates of our four statistics after correction for multiple comparions based on the false discovery rate procedure. Compared with the other three statistics, CM_2 was the most sensitive statistic in detecting head motion. Moreover, consistent with our expectations, increasing S_0/σ reduced the Type II errors and improved the power of the statistic CM_2 for detecting the presence of two compartments. However, the other three statistics CK_1, CM_1, and CK_2 were not particularly sensitive in detecting head motion.

5. HEAD MOTION DIFFUSION-WEIGHTED IMAGES

We acquired DWIs of the brain of a healthy adult male subject (right-handed; age 34 years). The imaging acquisition scheme $\{(b_i, r_i)$: $i = 1, \ldots, 38\}$ consisted of 3 baseline images with $b = 0$ s/mm^2 and 35 directions of diffusion gradients that were arranged uniformly in the 3D space at $b = 1{,}000$ s/mm^2 (Hardin, Sloane, and Smith 1994). Each DWI contained $256 \times 256 \times 65$ voxels. The subject was instructed to move his head deliberately during acquisition of images from the 28th to the 38th direction. Head motion varied from 2 to 6 degrees of rotation and 0 to 10 millimetres of translation, causing the diffusion-weighted images to be moderately misaligned.

We used the Rician DTI model (8) for this analysis. We subsequently calculated at each voxel the ML estimate ($\hat{D}$, $\hat{S}_0$, $\hat{\sigma}$), three eigenvalue-eigenvector pairs of $\hat{D}$, denoted by $\{(m_i, e_i)\colon i = 1, 2, 3\}$, and the invariant measures including CL = $(m_1 - m_2)/M_1$, Cp = $2(m_2 - m_3)/M_1$, RA= $\sqrt{1 - 3M_2M_1^{-2}}$ and FA= $\sqrt{1 - M_2\left(M_1^2 - 2M_2\right)^{-1}}$, where $m_1 \geq m_2 \geq m_3$, $M_1 = \mathrm{tr}(\hat{D})$, $M_2 = m_1m_2 + m_1m_3 + m_2m_3$, and $M_3 = m_1m_2m_3$. We also calculated three test statistics T_a = FA, $T_b = S(\hat{D}) + W(\hat{D})^{1.5}$, and $T_c = S(\hat{D}) - W(\hat{D})^{1.5}$, and their associated p-values, where $S(\hat{D} = (M_1/3)^3 - M_1M_2/6 + M_3/2$ and $W(\hat{D}) = (M_1/3)^2 - M_2/3$. We further set the significance level at 1% and used the p-values of T_a, T_b, and T_c to classify the morphology of the tensor at each voxel (Zhu et al. 2006).

We then assessed the quality of these diffusion-weighted images using our diagnostic methods. We searched for artifacts, scanner instability problems, and voxels that contained outliers; in addition, we obtained diagnostic measures, generated scan summaries, and applied graphical tools. We estimated the p-values of the four test statistics CK_1, CK_2, CM_1, and CM_2 using the resampling method in Section 3 of this article.

We plotted maps of scan summaries to identify possible artifacts and acquisition problems in the DW images. Translational and rotational parameters (Fig. 3), obtained from FLIRT in FSL (http://www.fmrib.ox.ac.uk), detected rightward rotation of 2 to 6 degrees and 0-10 mm translation beginning in the 28th acquisition (Jenkinson and Smith 2001;Jenkinson, Bannister, and Smith 2002). Outlier statistics detected these head motions as well. The outlier count per slice and per direction showed clearly that a large batch of outliers appeared in almost all of the slices along the last 10 directions (red to white on the color spectrum in Fig. 4). Furthermore, we used GIFT sofware (http://icatb.sourceforge.net) to perform a spatial independent component analysis (ICA) on the 16 slices covering the middle part of each directional DWI (baseline images excluded) (Calhoun, Adali, Pearlson, and Pekar 2001). Using the Bayesian information criterion (BIC), we selected 8 independent components (ICs) and plotted the associated time series from the spatial ICA. Similar to the 2D maps of scan summaries, the time series associated with the fourth, seventh, and eighth components revealed the deliberate rotation and translation from the 28th to 33rd acquisitions. The detailed information about the ICA results can be found in the supplementary report. However, the ICA cannot be used to detect nonstochastic noise components at the voxel level as detailed later.

To reduce or eliminate motion artifacts, we used the rigid-body transformation method to coregister all other DWIs to the first DWI while properly reorienting the diffusion gradients (Rohde, Barnett, Basser, Marenco, and Pierpaoli 2004). Particularly, we applied the translational and rotational parameters obtained from FLIRT and used a seventh order interpolation method to resample the DW images. After coregistration, new translational and rotational parameters (not shown here) revealed that the DW images were properly aligned. We then assessed the realigned DW images using our diagnostic procedure and used the Rician model (8) to process the reoriented DW images.

Our diagnostic procedure can be used to quantify the efficacy of the coregistration and reslicing algorithms, and to identify potential problems that remain in the DW images after registration and reslicing. We observed a substantial decline in the number of outlier counts per slice and per direction compared with the nonrealigned images, as well as a decline in the percentage of outliers per slice and per direction after coregistration [Figs. 4(a)-4(d)]. Furthermore, we examined voxels having 0-10 outliers and found that motion correction using coregistration significantly decreased the percentage of voxels having 4-10 outliers from 2.85% to 1.41%. However, despite the efficacy of this method for correcting motion artifacts, 5.7% of the voxels still contained at least three outliers after coregistration, and the 28th to 33rd acquisitions (red to white on the color spectrum) contained a number of outliers [Fig. 4(c), red to white on the

color spectrum]. This may indicate that the rigid-body transformation and the interpolation method cannot completely remove the effect of moderate and large head motions in MRIs.

The 3D images of the - $\log_{10}(p)$ values for the test statistics CK_1, CK_2, CM_1, and CM_2 were more sensitive and specific in assessing the quality of the DW images (Fig. 5). A p-value of 0.001 corresponded to a - $\log_{10}(p)$ value of 3; thus a voxel having a - $\log_{10}(p)$ value greater than or equal to 3.0 was conventionally regarded as statistically significant and in need of further investigation. In all maps of - $\log_{10}(p)$ values of the test statistics, we focused on voxels having significant p-values (white) and then searched for systematic patterns of these voxels in the brain. We found several notable changes after coregistration as follows. The number of voxels having large - $\log_{10}(p)$ values for the CK_1, CK_2, CM_1, and CM_2 statistics declined dramatically following coregistration (Fig. 5). We also used the resampling methods in Section 3.2 to calculate the corrected - $\log_{10}(p)$ values, but no significant voxel was detected for all four test statistics at the 5% significance level before and after coregistration. Moreover, compared with CK_1 and CK_2, CM_1 and CM_2 were more sensitive measures for detecting head motion.

Assessing the quality of DW images was crucial for further processing images. As shown previously (Fig. 5), the maps of the - $\log_{10}(p)$ values of the test statistics not only provided detailed information about the goodness of fit of the fitted Rician model with the DW images (Fig. 5), but also these maps indicated possible artifacts existing in the DWIs. Those artifacts strongly influenced the estimation of the diffusion tensors, the classification of tensor morphologies, the reconstruction of fiber tracts, and the quantification of uncertainty in tensor estimation and tractography. Therefore, we also assessed the prevalence of the four morphological classes of DTs (nondegenerate, oblate, prolate, and isotropic) in a single slice before and after coregistration. Before coregistration, we found that 59.97% were isotropic, 9.37% were oblate, 23.06% were prolate, and 7.61% were nondegenerate. Following coregistration, we found that 48.09% were isotropic, 11.35% were oblate, 28.11% were prolate, and 12.45% were nondegenerate. Most tractography algorithms can only track fibers across voxels containing either nondegenerate or prolate DTs, which accounted for 40.56% of the total number of voxels on this slice after coregistration, compared with 30.67% before coregistration. Moreover, we also found moderate discrepancy between the estimated principal directions before and after coregistration (not presented here).

To assess these DWIs before and after coregistration, we also examined 3D images of standardized residuals and Cook's distances. Specifically, we searched the standardized residuals (or Cook's distance) in all voxels across all slices and directions to identify voxels having large numbers of large positive and negative outliers (i.e., data points of excessive influence). For illustration, we compared the standardized residuals at the 30th slice from the 32nd acquisition before and after coregistration (Fig. 6). Before coregistration, this slice contained many large positive and negative residuals [Figs. 6(a) and 6(b)]. After coregistration, the number of large positive and negative residuals dramatically declined [Figs. 6(c) and 6(d)]. However, even after coregistration, some motion artifacts or other unspecified problems remained in the resliced DWIs. Developing methods for identifying the precise sources of non-Rician noise and correcting for them in the resliced DWIs will require further research.

For voxels having either many outliers or substantial misspecification of the Rician model, we examined multiple 2D graphs to try to identify the causes of the outliers and of model misspecifications. To illustrate this process, we considered the data at a single voxel [at location (100, 69, 30)] before coregistration. The p-values for CK_1, CK_2, CM_1, and CM_2 were 0.21, 0.13, 0.03, and 0.01, respectively. It appears that coregistration slightly improved the goodness of fit of the Rician model to the MRI signal within this voxel. The index plots of the standardized residuals and Cook's distances [Figs. 7(a) and 7(b)] revealed that the fourth, eighth, and 34th

observations were likely outliers. A plot of the standardized residuals against the raw MRI values [Fig. 7(c)] revealed a strong linear relationship between residuals and the raw MRI values (Cook and Weisberg 1982). Furthermore, we observed a nonlinear relationship [Fig. 7 (d)] of Cook's distances against raw MRI values. Together these plots [Figs. 7(c) and 7(d)] indicated that a Rician model (8) did not fit the MRI data satisfactorily. Further improvements in model specification or postacquisition processing are needed to identify and address the non-Rician sources of noise in the images.

Our diagnostic procedure effectively identified head motion artifacts in DWIs. Coregistration improved image quality, but substantial nonstochastic noise sources remained in the 28th to 33rd acquisitions. One solution is to remove these slices from the subsequent analysis; alternatively, we may resort to a robust estimate of DTs to reduce the deleterious statistical effects of these outliers. The 3D images of the test statistics further detected additional physiological noise, such as cardiac pulsation, in DWIs. Additional 2D statistical maps may identify the causes of statistically significant voxels and the location of outliers.

6. CONCLUDING REMARKS

We have developed estimation algorithms for fitting a Rician regression model and the associated two normal models, and proposed a diagnostic procedure for systematically assessing the quality of MR images at all levels of the SNR. The key features of our procedures include calculating test statistics that assess the validity of the assumptions of the statistical models for stochastic noise in MR images; use of influence measures to identify artifacts and problems with image acquisition; and multiple graphical tools for visual evaluation of the appropriateness of the model assumptions. Simulations showed the effectiveness of our test statistics in detecting the presence of multiple compartments. Moreover, an in vivo study demonstrated the effectiveness of our procedures in locating voxels that contain unreliable data due to motion artifacts or to problems with imaging acquisition. Our findings suggest that our approach to assessing the quality of MR images is both rigorous and computationally practical.

Our diagnostic procedure differs substantially from previous model-free methods, such as ICA analysis and motion correction algorithms, for detecting noise components in MRI. Most of those model-free methods cannot be used to detect nonstochastic noise components at the voxel level, because they can only provide information about MRI at the whole volume level. In addition, some of those model-free methods are limited to a specific imaging modality. For instance, although an ICA method was recently proposed to identify ICs associated with task-related motion, and then discard those ICs to reduce motion effects on realigned fMRI data (Kochiyama et al. 2005), this ICA method cannot be directly applied to other imaging modalities, such as DWI. Particularly, for DWI, we cannot discard the ICs corresponding to head motion without changing the gradient directions, which requires further research. In contrast, as shown in Section 5, our diagnostic procedure is a model-based method that uses goodness-of-fit statistics and diagnostic measures to systematically detect nonstochastic noise components at each voxel of the MRI data. Subsequently, our diagnostic procedure can combine the information from all voxels of the brain volume to identify large nonstochastic noise sources, such as head motion at the whole volume level.

Our procedure takes a further step by studying how to use existing information in the MRI data to check model assumptions and to identify imaging artifacts that may undermine applications or interpretations of the MR images. Our diagnostic procedure can also be applied to systematically check the MRI data even after these MRI data have been processed by existing noise removal methods, such as rigid-motion correction and ICA. Moreover, our diagnostic procedure can be used to detect the presence of the partial volume effect, whereas those existing methods, such as the motion correction method, cannot. However, a potential drawback of our

model-based diagnostic method is that the validity of our test statistics depends on the correct specification of model assumptions. For instance, if the model proposed is misspecified, then the maps of our diagnostic measures can reflect the confounding effects from the misspecified model. Nevertheless, our procedure assesses the quality of MRI statistically and cannot replace various preprocessing techniques, such as registration and smoothing methods.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported in part by NSF grants SES-06-43663 and BCS-08-26844 and NIH grants UL1-RR025747-01 and AG033387 to Dr. Zhu, NIDA grant DA017820 and NIMH grants MH068318 and K02-74677 to Dr. Peterson, NIH grants GM 70335 and CA 74015 to Dr. Ibrahim, and NIH grant R01NS055754 to Dr. Lin. The authors thank the editor, associate editor, and referees for their helpful comments and suggestions.

REFERENCES

Abramowitz, M.; Stegun, I., editors. Handbook of Mathmatical Functions. Dover Publications; New York: 1965.

Alexander DC, Barker GJ. Optimal Imaging Parameter for Fibre-Orientation Estimation in Diffusion MRI. NeuroImage 2005;27:357–367. [PubMed: 15921931]

Alexander DC, Barker GJ, Arridge SR. Detection and Modeling of Non-Gaussian Apparent Diffusion Coefficient Profiles in Human Brain Data. Magnetic Resonance in Medicine 2002;48:331–340. [PubMed: 12210942]

Basser PJ, Mattiello J, LeBihan D. MR Diffusion Tensor Spectroscopy and Imaging. Biophysical Journal 1994a;66:259–267. [PubMed: 8130344]

Basser PJ, Mattiello J, LeBihan D. Estimation of the Effective Self-diffusion Tensor from the NMR Spin Echo. *Journal of Magnetic Resonance:* Series B 1994b;103:247–254. [PubMed: 8019776]

Calhoun VD, Adali T, Pearlson GD, Pekar JJ. A Method for Making Group Inferences from Functional MRI Data Using Independent Component Analysis. Human Brain Mapping 2001;14:140–151. [PubMed: 11559959]

Cook, RD.; Weisberg, S. Residuals and Influence in Regression. Chapman and Hall; London: 1982.

Dempster AP, Laird NM, Rubin DB. “Maximum Likelihood from Incomplete Data via the EM Algorithm” (with Discussion). *Journal of the Royal Statistical Society:* Series B 1977;39:1–38.

den Dekker AJ, Sijbers J. Implications of the Rician Distribution for fMRI Generalized Likelihood Ratio Tests. Magnetic Resonance Imaging 2005;23:953–959. [PubMed: 16310111]

Escanciano JC. A Consistent Diagnostic Test for Regression Models Using Projections. Econometric Theory 2006;22:1030–1051.

Gudbjartsson H, Patz S. The Rician Distribution of Noisy MRI Data. Magnetic Resonance in Medicine 1995;34:910–914. [PubMed: 8598820]

Haacke, EM.; Brown, RW.; Thompson, MR.; Venkatesan, R. Magnetic Resonance imaging: Physical Principles and Sequence Design. John Wiley and Sons; New York: 1999.

Hardin, RH.; Sloane, NJA.; Smith, WD. Minimal Energy Arrangements of Points on a Sphere with Minimal 1/r Potential. 1994. available at http://www.research.att.com/njas/electrons/

Henkelman RM. Measurement of Signal Intensities in the Presence of Noise in MR Images. Medical Physics 1985;12:232–233. [PubMed: 4000083]

Huettel, SA.; Song, AW.; McCarthy, G. Functional Magnetic Resonance Imaging. Sinauer Associates, Inc.; Sunderland, MA: 2004.

Jenkinson M, Bannister JM, Smith SM. Improved Optimisation for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. NeuroImage 2002;17:825–841. [PubMed: 12377157]

Jenkinson M, Smith SM. A Global Optimisation Method for Robust Affine Registration of Brain Images. Medical Image Analysis 2001;5:143–156. [PubMed: 11516708]

Jones DK, Basser PJ. Squashing Peanuts and Smashing Pumpkins: How Noise Distorts Diffusion-Weighted MR Data. Magnetic Resonance in Medicine 2004;52:979–993. [PubMed: 15508154]

Jones DK, Horsfield MA, Simmons A. Optimal Strategies for Measuring Diffusion in Anisotropic Systems by Magnetic Resonance Imaging. Magnetic Resonance in Medicine 1999;42:515–525. [PubMed: 10467296]

Jorgensen B. Exponential Dispersion Models and Extension: A Review. International Statistical Review. Revue Internationale de Statistique 1992;60:5–20.

Karlsen OT, Verhagen R, Bovee WM. Parameter Estimation from Rician-distributed Data Sets using a Maximum Likelihood Estimator: Application to T_1 and Perfusion Measurements. Magnetic Resonance in Medicine 1999;41:614–623. [PubMed: 10204887]

Kochiyama T, Morita T, Okada T, Yonekura Y, Matsumura M, Sadato N. Removing the Effects of Task-related Motion using Independent-component Analysis. NeuroImage 2005;25:802–814. [PubMed: 15808981]

Kristoffersen A. Optimal Estimation of the Diffusion Coefficient from Non-averaged and Averaged Noisy Magnitude Data. *Journal of Magnetic Resonance* (San Diego, Calif.: 1997) 2007;187:293–305.

Louis TA. Finding the Observed Information Matrix when Using the EM Algorithm. *Journal of the Royal Statistical Society:* Ser. B 1982;44:190–200.

Luo WL, Nichols TE. Diagnosis and Exploration of Massively Univariate fMRI Models. NeuroImage 2003;19:1014–1032. [PubMed: 12880829]

Macorski A. Noise in MRI. Magnetic Resonance in Medicine 1996;36:494–497. [PubMed: 8875425]

McCullagh, P.; Nelder, JA. Generalized Linear Models. Chapman and Hall; London: 1989.

Meng XL, Rubin DB. Maximum Likelihood Estimation via the ECM Algorithm: A General Framework. Biometrika 1993;80:267–278.

Nowark RD. Wavelet-based Rician Noise Removal for Magnetic Resonance Imaging. IEEE Transactions on Image Processing 1999;8:1408–1419. [PubMed: 18267412]

Rice SO. Mathematical Analysis of Random Noise. The Bell System Technical Journal 1945;24:46–156.

Rohde GK, Barnett AS, Basser PJ, Marenco S, Pierpaoli C. Comprehensive Approach for Correction of Motion and Distortion in Diffusion-Weighted MRI. Magnetic Resonance in Medicine 2004;51:103–114. [PubMed: 14705050]

Rowe DR. Parameter Estimation in the Magnitude-only and Complex-Valued fMRI Data Models. NeuroImage 2005;25:1310–1324. [PubMed: 15850748]

Rowe DR, Logan BR. A Complex Way to Compute fMRI Activation. NeuroImage 2004;23:1078–1092. [PubMed: 15528108]

Rowe DR, Logan BR. Complex fMRI Analysis with Unrestricted Phase is Equivalent to a Magnitude-only Model. NeuroImage 2005;24:603–606. [PubMed: 15627605]

Sijbers J, den Dekker AJ. Maximum Likelihood Estimation of Signal Amplitude and Noise Variance from MR Data. Magnetic Resonance in Medicine 2004;51:586–594. [PubMed: 15004801]

Sijbers J, den Dekker AJ, Scheunders P, Van Dyck D. Maximum-Likelihood Estimation of Rician Distribution Parameters. IEEE Transactions on Image Processing 1998a;17:357–361.

Sijbers J, den Dekker AJ, Verhoye M, Van Audekerke J, Van Dyck D. Estimation of Noise from Magnitude MR Images. Magnetic Resonance Imaging 1998b;16:87–90. [PubMed: 9436952]

Skare S, Li T, Nordell B, Ingvar M. Noise Considerations in the Determination of Diffusion Tensor Anisotropy. Magnetic Resonance Imaging 2000;18:659–669. [PubMed: 10930775]

Stute W. Nonparametric Model Checks for Regression. Annals of Statistics 1997;25:613–641.

Tuch DS, Reese TG, Wiegell MR, Makris N, Belliveau JW, Wedeen VJ. High Angular Resolution Diffusion Imaging Reveals Intravoxel White Matter Fiber Heterogeneity. Magnetic Resonance in Medicine 2002;48:577–582. [PubMed: 12353272]

van der Vaart, AW.; Wellner, JA. Weak Convergence and Empirical Processes with Applications to Statistics. Springer; New York: 1996.

Wei, BC. Exponential Family Nonlinear Models. Springer; Singapore: 1998.

Xu D, Mori S, Solaiyappan M, van Zijl PC, Davatzikos C. A Framework for Callosal Fiber Distribution Analysis. NeuroImage 2002;17:1131–1143. [PubMed: 12414255]

Zhu HT, Xu D, Amir R, Hao X, Zhang H, Alayar K, Ravi B, Peterson B. A Statistical Framework for the Classification of Tensor Morphologies in Diffusion Tensor Images. Magnetic Resonance Imaging 2006;24:569–582. [PubMed: 16735178]

Zhu HT, Zhang HP. A Diagnostic Procedure based on Local Influence Measure. Biometrika 2004;91:579–589.

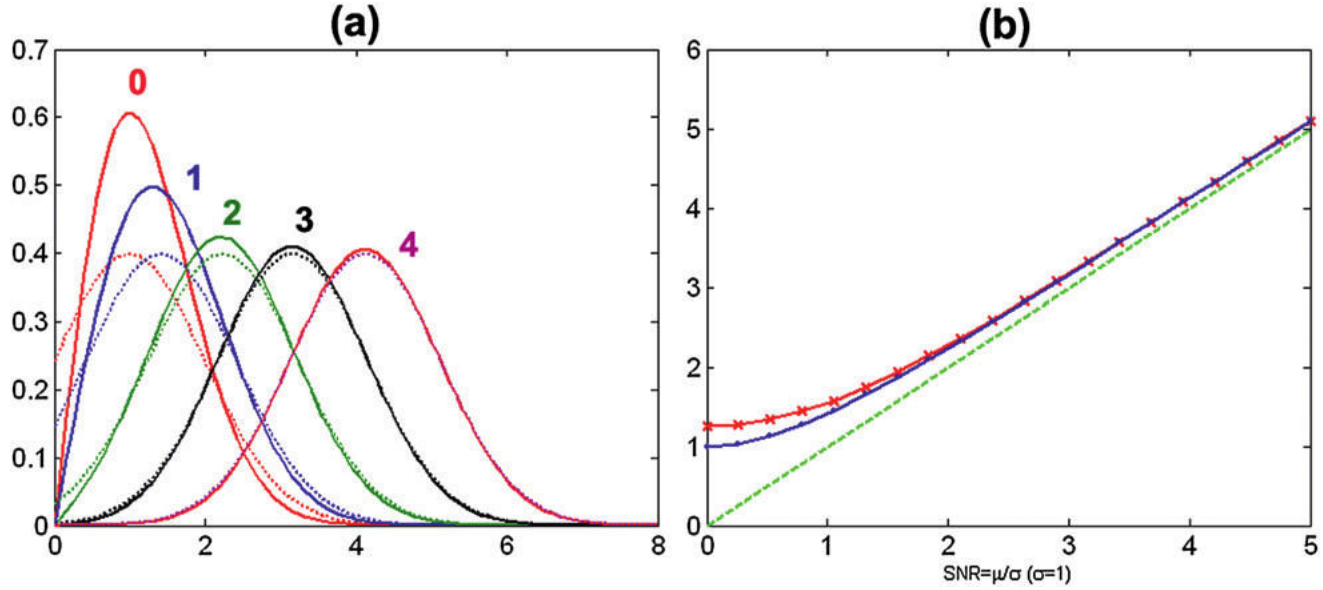

Figure 1.

Rician distribution: (a) $R(\mu, 1)$ and $N\left(\sqrt{\mu^2+1}, 1\right)$ for $\mu = 0, 1, 2, 3, 4$; (b) the mean functions of $R(\mu, 1)$ (red), $N\left(\sqrt{\mu^2+1}, 1\right)$ (blue), and $N(\mu, 1)$ (green) for $\mu \in [0, 5]$.

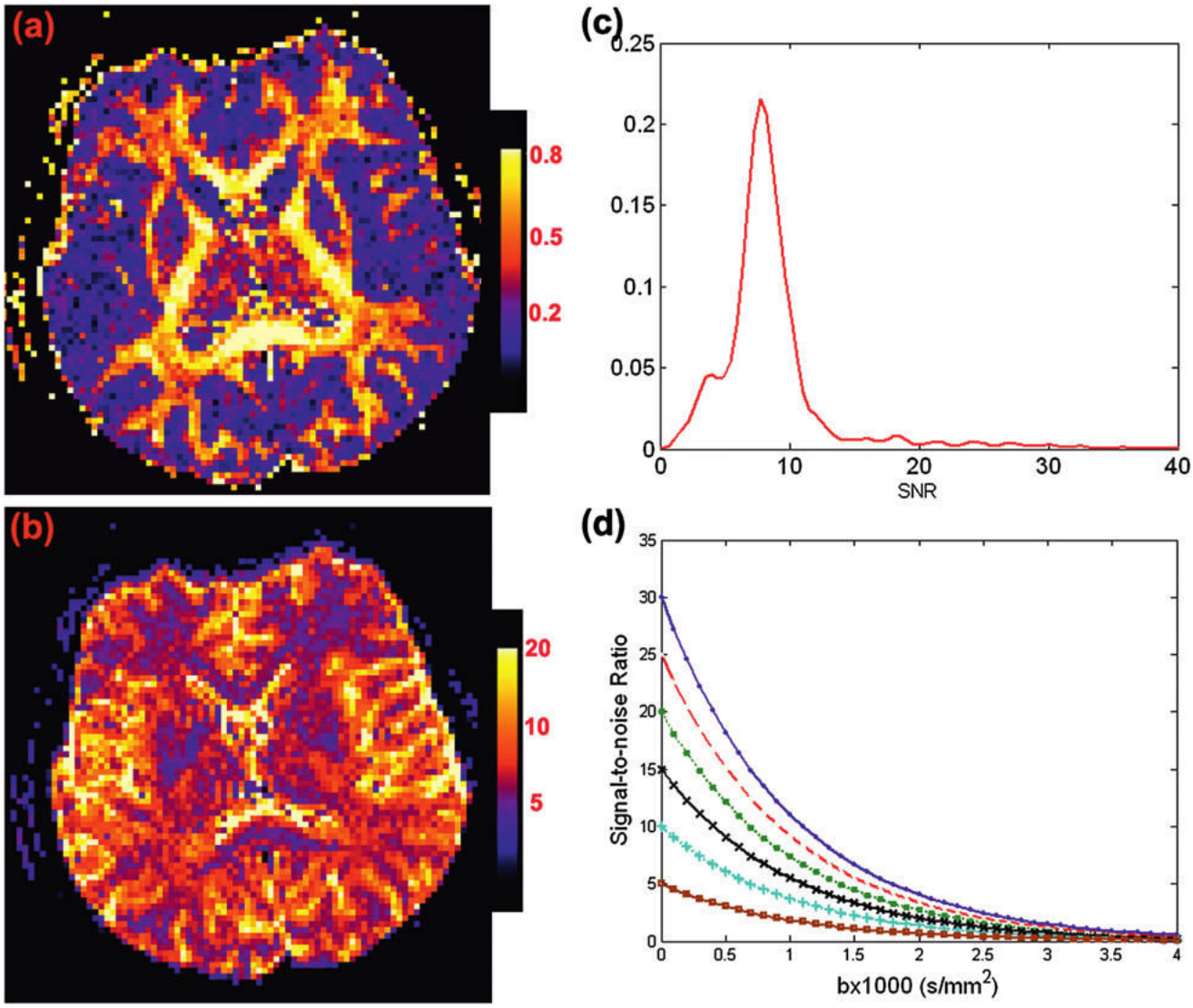

Figure 2.

Maps of (a) FA; (b) S_0/σ; (c) the kernel density of S_0/σ values for anisotropic tensors having FA ≥ 0.5 at a selective slice from a single subject; and (d) the signal-to-noise ratio $S_0 \exp(-b_i)/\sigma$ as a function of b_i ($\times$ 1,000 s/mm^2) at each $S_0/\sigma \in \{5, 10, 15, 20, 25, 30\}$.

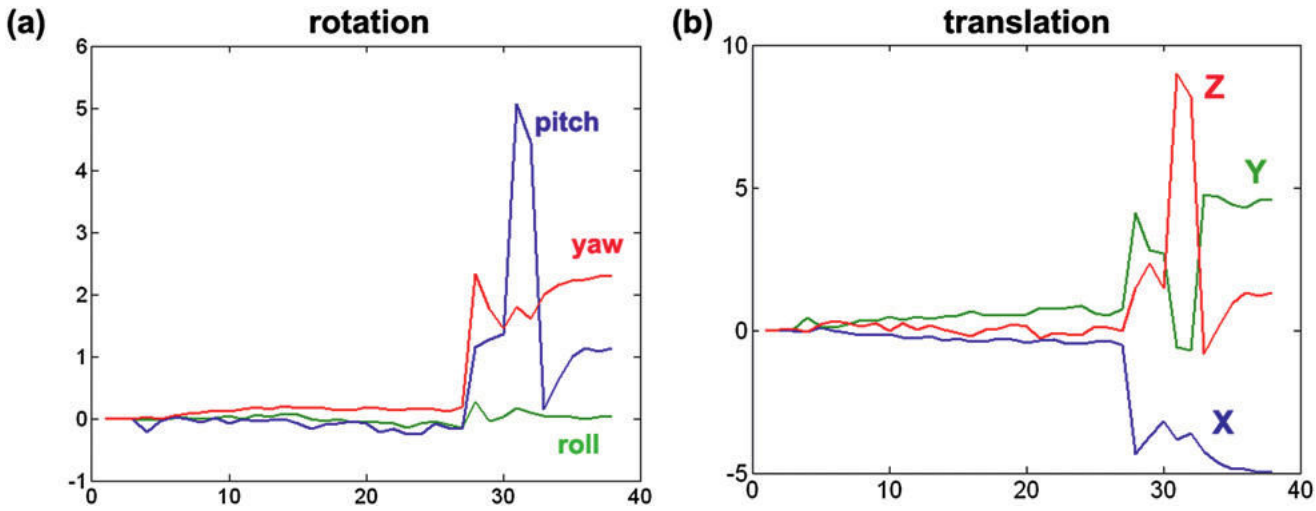

Figure 3.
Scan summaries for a set of DWIs from a single subject: (a) maps of translational parameters; (b) maps of rotational parameters.

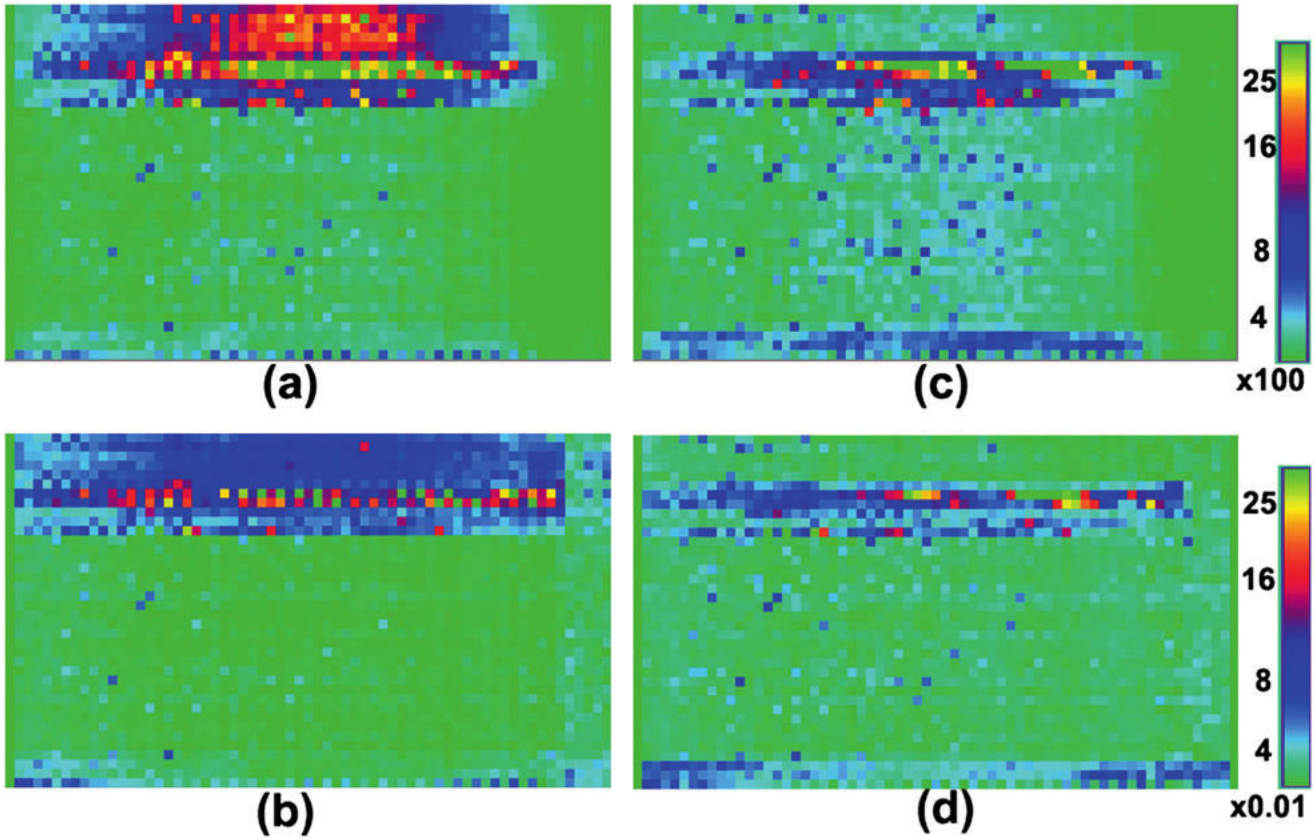

Figure 4.
Assessing the effect of applying a coregistration algorithm to diffusion-weighted images from a single subject: outlier count per slice and per direction (a) before coregistration and (c) after coregistration; percentages of outliers per slice and per direction (b) before coregistration and (d) after coregistration.

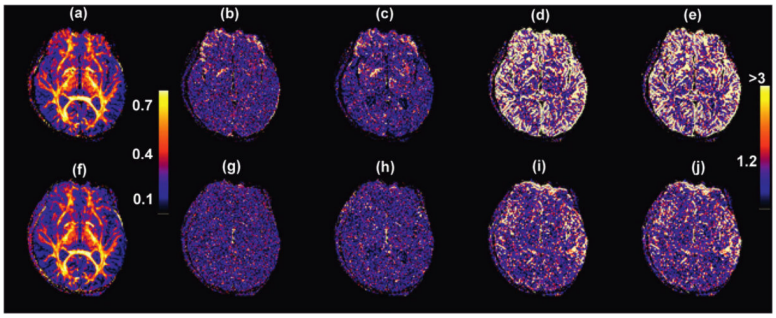

Figure 5.

(a-e) Maps of 3D images before coregistration and (f-j) after coregistration in a single slice from a single subject. Before coregistration: (a) FA value; (b) - $\log_{10}(p)$ values of CK_1; (c) - $\log_{10}(p)$ values of CK_2; (d) - $\log_{10}(p)$ values of CM_1; (e) - $\log_{10}(p)$ values of CM_2.After coregistration: (f) FA value; (g) - $\log_{10}(p)$ values of CK_1; (h) - $\log_{10}(p)$ values of CK_2; (i) - $\log_{10}(p)$ values of CM_1; (j) - $\log_{10}(p)$ values of CM_2.

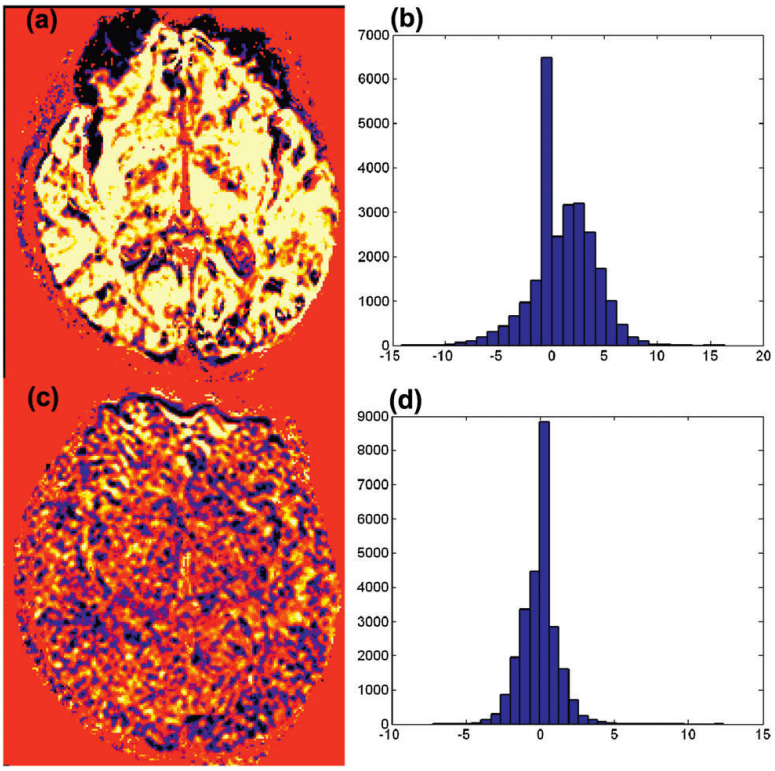

Figure 6.

Plots of standardized residuals at the 30th slice of the 32nd acquisition before and after coregistration from a single subject: standardized residuals (a) before coregistration and (c) after coregistration; histograms of standardized residuals (b) before coregistration and (d) after coregistration. Voxels in the black-to-blue range have large negative standardized residuals (< -2.5), whereas yellow to white voxels have large positive standardized residuals (>2.5).

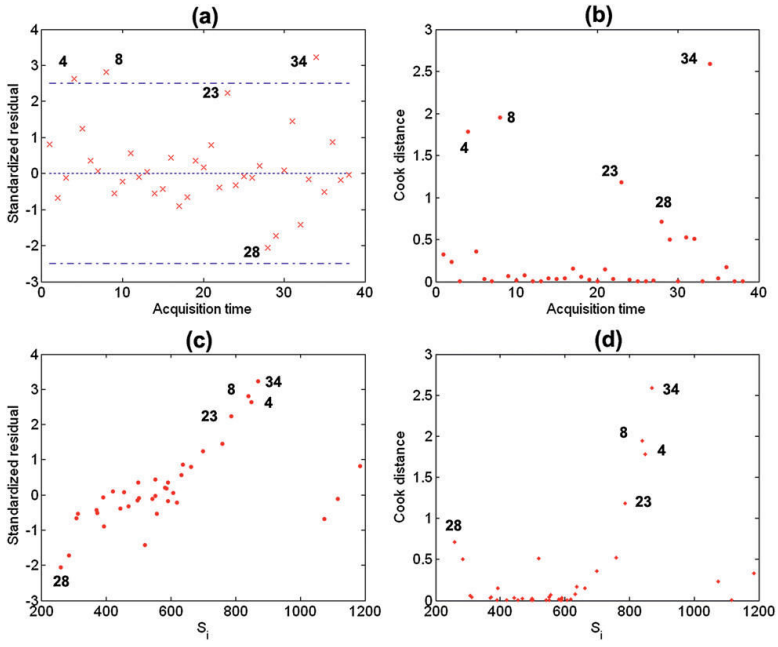

Figure 7.

Multiple 2D graphs for a selected voxel (110, 69, 30) before coregistration from a single subject: (a) index plot of standardized residuals; (b) index plot of Cook's distances; (c) standardized residuals against raw data; (d) Cook's distances against raw data.

Table 1

ADC imaging: bias and SD of three components of $\hat{\theta}$

		$R(\mu_i, \sigma^2)$			$N\left(\sqrt{\mu_i^2+\sigma^2}, \sigma^2\right)$			$N(\mu_i, \sigma^2)$		
	S_0/σ	σ^2	S_0	d	σ^2	S_0	d	σ^2	S_0	d
TRUE	2	62,500	500.00	2.000	62,500	500.00	2.000	62,500	500.00	2.000
BIAS	2	-13,413	14.31	0.249	-23,715	18.87	-0.749	-29,683	15.73	-1.403
SE	2	19,023	168.52	1.960	15,494	139.40	1.241	10,719	102.62	0.364
SEE	2	24,123	255.12	2.460	16,320	175.54	1.419	13,009	88.91	0.378
TRUE	4	15,625	500.00	2.000	15,625	500.00	2.000	15,625	500.00	2.000
BIAS	4	-1,938	-5.46	0.080	-4,542	-6.32	-0.284	-5,014	-19.73	-0.711
SE	4	5,218	82.05	0.909	3,658	76.92	0.637	3,488	65.95	0.332
SEE	4	6,106	108.88	0.998	4,285	79.48	0.611	4,040	60.23	0.343
TRUE	6	6,944	500.00	2.000	6,944	500.00	2.000	6,944	500.00	2.000
BIAS	6	-718	-2.26	0.016	-1,680	-4.55	-0.127	-1,746	-12.39	-0.371
SE	6	2,409	51.99	0.469	1,710	50.02	0.353	1,702	65.95	0.332
SEE	6	2,708	66.86	0.500	1,998	55.36	0.392	1,972	60.23	0.343
TRUE	10	2,500	500.00	2.000	2,500	500.00	2.000	2,500	500.00	2.000
BIAS	10	-230	0.43	-0.025	-414	-1.08	-0.033	-422	-4.20	-0.138
SE	10	893	31.45	0.218	651	30.68	0.204	661	29.32	181.80
SEE	10	938	37.34	0.242	683	34.65	0.228	786	32.62	196.24
TRUE	15	1,111	500.00	2.000	1,111	500.00	2.000	1,111	500.00	2.000
BIAS	15	-109	-0.23	0.008	-141	-0.60	-0.015	-143	-2.03	-0.065
SE	15	339	20.20	0.136	303	20.18	0.135	307	19.94	0.127
SEE	15	396	24.24	0.149	365	23.68	0.148	366	23.04	0.138

NOTE: TRUE denotes the true value of the regression parameters; BIAS denotes the bias of the mean of the regression estimates; SE denotes the empirical standard errors; SEE denotes the mean of the standard error estimates. Five different S_0/σ {2, 4, 6, 10, 15} and 10,000 simulated datasets were used for each case.

Table 2

Comparison of the rejection rates for the test statistics CK_1, CM_1, CK_2, and CM_2 under the two-DT model, in which $f(x_i,\beta)=S_0\left[p_1\exp\left(-b_i r_i^T D_1 r_i\right)+(1-p_1)\exp\left(-b_i r_i^T D_2 r_i\right)\right]$ at a significance level the 0.05 after correction for multiple comparisons based on the false discovery rate

		(b_1, b_2) × 1,000 s/mm^2											
		(1, 1)				(1, 3)				(3, 3)			
SNR	p_1	CK_1	CK_2	CM_1	CM_2	CK_1	CK_2	CM_1	CM_2	CK_1	CK_2	CM_1	CM_2
5	1	0.02	0.01	0.03	0.04	0.05	0.03	0.04	0.04	0.07	0.07	0.05	0.06
10	1	0.04	0.03	0.03	0.03	0.04	0.04	0.04	0.04	0.03	0.03	0.04	0.04
15	1	0.04	0.03	0.03	0.03	0.03	0.03	0.04	0.04	0.02	0.03	0.03	0.04
20	1	0.02	0.02	0.03	0.04	0.03	0.04	0.03	0.04	0.02	0.03	0.03	0.04
25	1	0.01	0.02	0.02	0.02	0.04	0.03	0.05	0.04	0.02	0.02	0.026	0.02
5	1	0.01	0.02	0.03	0.03	0.05	0.05	0.08	0.07	0.08	0.09	0.05	0.06
10	1	0.05	0.04	0.02	0.02	0.23	0.08	0.22	0.12	0.04	0.02	0.01	0.02
15	1	0.09	0.05	0.02	0.02	0.43	0.11	0.39	0.15	0.08	0.01	0.01	0.01
20	1	0.16	0.09	0.03	0.03	0.61	0.11	0.59	0.22	0.09	0	0	0
25	1	0.26	0.18	0.02	0.02	0.75	0.14	0.71	0.19	0.12	0	0	0

NOTE: The first DT compartment is D_1 = diag(1.7, 0.2, 0.2) and the second DT compartment is D_2 = diag(0.2, 1.7, 0.2). Five different S_0/σ values {5, 10, 15, 20, 25} and 1,000 simulated datasets were used for each case.

Table 3

Comparison of the rejection rates for the test statistics CK_1, CK_2, CM_1, and CM_2, under the presence of head motion at a significance level of 0.05 after correction for multiple comparisons based on the false discovery rate

	p_1							
	0.7				0.5			
SNR	CK_1	CK_2	CM_1	CM_2	CK_1	CK_2	CM_1	CM_2
5	0.02	0.02	0.05	0.05	0.03	0.01	0.05	0.05
10	0.04	0.08	0.09	0.16	0.07	0.06	0.11	0.16
15	0.09	0.14	0.10	0.23	0.08	0.09	0.10	0.23
20	0.11	0.19	0.09	0.31	0.12	0.13	0.12	0.31
25	0.12	0.23	0.08	0.32	0.13	0.13	0.10	0.31

NOTE: The first [50 × p_1] acquisitions were generated from a single diffusion model with D_1 = diag(0.2, 1.7, 0.2) and the last 50 - [50 × p_1] acquisitions were generated from asingle diffusion model with D_2 = diag(0.7, 0.7, 0.7). Five different S_0/σ values {5, 10, 15, 20, 25} and 1,000 simulated datasets were used for each case.