University of North Carolina at Chapel Hill
The University of North Carolina at Chapel Hill Department of Biostatistics Technical Report Series

Year 2007 *Paper* 3

Regression Models for Identifying Noise Sources in Magnetic Resonance Images

Hongtu Zhu[*] Yimei Li[†] Joseph G. Ibrahim[‡]
Xiaoyan Shi[**] Hongyu An[††] Yasheng Chen[‡‡]
Weili Lin[§] Daniel B. Rowe[¶] Bradley G. Peterson[‖]

[*]The University of North Carolina at Chapel Hill, hzhu@bios.unc.edu
[†]The University of North Carolina at Chapel Hill, liyimei@email.unc.edu
[‡]The University of North Carolina at Chapel Hill, ibrahim@bios.unc.edu
[**]The University of North Carolina at Chapel Hill, xyshi@email.unc.edu
[††]The University of North Carolina at Chapel Hill, hongyu_an@med.unc.edu
[‡‡]The University of North Carolina at Chapel Hill, yasheng_chen@med.unc.edu
[§]The University of North Carolina at Chapel Hill, weili_lin@med.unc.edu
[¶]Medical College of Wisconsin
[‖]Columbia Medical Center and the New York State Psychiatric Institute

Regression Models for Identifying Noise Sources in Magnetic Resonance Images

Hongtu Zhu, Yimei Li, Joseph G. Ibrahim, Xiaoyan Shi, Hongyu An, Yasheng Chen, Weili Lin, Daniel B. Rowe, and Bradley G. Peterson

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 MR imaging modalities, including diffusion weighted imaging and functional MRI. 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 graphic 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 artifact. 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-world dataset to il- lustrate how our diagnostic procedure localizes subtle image artifacts by detecting intravoxel variability that is not captured by the regression models.

Regression Models for Identifying Noise Sources in Magnetic Resonance Images *

Hongtu Zhu[1,2*], Yimei Li[1], Joseph G. Ibrahim[1], Xiaoyan Shi[1], Hongyu An[2],
Yashen Chen[2], Weili Lin[2], Daniel B. Rowe[3], and Bradley G. Peterson[4]

[1]Department of Biostatistics, University of North Carolina at Chapel Hill, North Carolina, USA

[2]Department of Radiology, University of North Carolina at Chapel Hill, North Carolina, U.S.A.

[3] Department of Biophysics,Medical College of Wisconsin, Milwaukee, Wisconsin, USA

[4] Department of Psychiatry, Columbia Medical Center
and the New York State Psychiatric Institute, New York, USA

*This work was supported in part by NSF grant SES-06-43663 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, as well as by the Suzanne Crosby Murphy Endowment at Columbia University Medical Center, and by the Thomas D. Klingenstein and Nancy D. Perlman Family Fund.

SUMMARY

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 MR imaging modalities, including diffusion weighted imaging and functional MRI. 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 graphic 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 artifact. 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-world 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.

1 Introduction

Magnetic Resonance Imaging (MRI) is an non-invasive imaging technique used extensively for clinical diagnosis and medical research. MR images, 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, 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 estimations 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 Nichols 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). Then, it can be shown that Rician distribution is the model for characterizing the stochastic noise in the magnitude of MR data. 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 2005). Despite that 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 non-stochastic noise components, such as those from susceptibility artifacts and rigid body motion. These non-stochastic noise sources usually introduce statistical outliers in some or all of the picture elements ("pixels" or "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. 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, or "voxel" in the tissue that corresponds with the given pixel in the image). In diffusion tensor images, 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, Reese, Wiegell, Makris, Belliveau, Wedeen 2002; Alexander, Barker, and Arridge 2002).

The aim of this paper 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 estimation of three regression models for MRI data. We develop diagnostic procedures to assess systematically 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 artifact 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-world data set to illustrate a real-world 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 MRI signal intensity and x_i includes all covariates of interest, such as 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 follows 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

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

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

$$p(S_i|\mu_i, \sigma^2) = \frac{S_i}{\sigma^2} \exp\{-0.5\sigma^{-2}(S_i^2 + \mu_i^2)\} I_0\left(\frac{\mu_i S_i}{\sigma^2}\right) 1(S_i \geq 0), \tag{1}$$

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

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

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, etc). 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 k-th 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 be shown that the kth moment of S_i given x_i (Sijbers, den Dekker, Scheunders, and van Dyck 1998a) is calculated as

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

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,

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

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

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

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

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

Moreover, the second moment of $R(\mu_i(\beta),\sigma^2)$ equals that of $N(\sqrt{\mu_i(\beta)^2+\sigma^2},\sigma^2)$, while $\mathrm{E}(S_i|x_i)$ in (5) can be accurately approximated by $\sqrt{\mu_i(\beta)^2+\sigma^2}$ even when SNR is close to 1 (Fig. 1b). 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

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

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 $\mathrm{E}(S_i^2|x_i) = 2\sigma^2$. 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 non-stochastic 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 to detect the presence of non-stochastic 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. It has been shown that the use of Rician model leads to a substantial increase in precision of the estimated T_1 (Karlsen, Verhagen, and Bovee 1999).

If the decay of transverse magnetization is mono-exponential 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 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, functional MRI 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 $t_1, \cdots, t_n$. 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 $i = 1, \cdots, n$. The magnitude image at time t_i follows Rician distribution and $\mu_i(\beta) = x_i^T\beta$, the superscript T denotes transpose and x_i may include responses to differing stimulus types, the rest status, and various reference functions. Achieving a high SNR is often compromised with low spatial resolution (e.g., 4 mm×4 mm×6 mm). The low-resolution imaging may lead to a loss in signal information generated by microvasculature and the ambiguous identification of activation boundaries for neurosurgical planning (Yoo, Talos, Golby, Black, and Panych 2004). Moreover, SNR decreases significantly for high-resolution imaging (Yoo, Guttmann and Panych 2001). It is of great interest to develop statistical methods for detecting small intensity variations originating in neural activation based on low SNR functional MRI data (Rowe and Logan

2005; den Dekker and Sijbers 2005).

EXAMPLE 4. Diffusion tensor images (DTI) has been widely used to reconstruct the pathways of white matter fibers in the human brain in vivo (Basser, Mattiello, and LeBihan 1994 a, b; Xu et al. 2002). A single shot echo-planar imaging (EPI) technique is often used to acquire diffusion-weighted (DW) images with moderate resolution (e.g., 2.5 mm×2.5 mm× 2.5 mm), and then diffusion tensors can estimated using DW image data. In voxels with a single fiber population, a simple diffusion model assumes that

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

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 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 data set 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) (Fig. 2a and 2b). For these anisotropic tensors, SNR= S_0/σ in baseline images varied from 0 to 15 with mean close to 6 (Fig. 2c), while λ_1 varied from 0.5 (10^{-3} mm^2/s) to 2.0 (10^{-3}mm^2/s) with mean close to 1.0 (10^{-3} mm^2/s). For moderate gradient strength $b_i \approx 1000\text{s/mm}^2$, SNR= $\exp(-b_i r_i^T D r_i) \times (S_0/\sigma)$ in DW images varied from 0 to 8 with mean close to 2.5 (Fig. 2d).

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

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

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 need large b value (Tuch et al. 2002; Alexander, Barker, and Arridge 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 value, SNR in DW image can be very close to zero (Fig. 2d).

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 diffusion-weighted gradient (Kristoffersen 2007). A simple mono-exponential 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 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 maximum likelihood estimate of θ. For the normal model (6), we propose an iterative procedure to maximize its log-likelihood function given by

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

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 $(\sigma^2)^{(0)} = \sum_{i=1}^{n} \{S_i - \mu_i(\beta^{(0)})\}^2/n$. Given $(\sigma^2)^{(r)}$, we use the Levenberg-Marquardt method to calculate $\beta^{(r+1)}$ that minimize $\sum_{i=1}^{n} \{S_i - \sqrt{\mu_i(\beta)^2 + (\sigma^2)^{(r)}}\}^2$. 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 iteration algorithm stops until the absolute difference between consecutive $\theta^{(t)}$s is smaller than a predefined small number, say 10^{-4}.

We introduce an efficient EM algorithm (Dempster, Laird and Rubin 1977) for maximizing the likelihood function of Rician model (2). The key idea is to introduce a latent phase

variable $\phi_i \in [0, 2\pi]$ for each S_i such that the joint density of (S_i, ϕ_i) is given by

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

Let $Y_o = (S_1, x_1, \cdots, S_n, x_n)$ be observed data and $Y_m = (\phi_1, \cdots, \phi_n)$ 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

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

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 $p(Y_m|Y_o, \theta^{(r)}) = \prod_{i=1}^{n} p(\phi_i|S_i, \theta^{(r)})$. We can show that

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

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

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

The M-step is to determine a $\theta^{(r+1)}$ 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

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

Conditional on $(\sigma^2)^{(r+1)}$, we can determine $\beta^{(r+1)}$ by minimizing $G(\beta|\beta^{(r)}) = \sum_{i=1}^{n} \{\mu_i(\beta) - W_i(\theta^{(r)})S_i\}^2$. This is a standard nonlinear least squares 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 maximum likelihood 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 normal model (6), it is straightforward to calculate the second derivative of $\ell(\beta, \sigma^2)$. For 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.

3 A Diagnostic Procedure

We propose a diagnostic procedure to identify noise components in MR images at all levels of SNR. Our diagnostic procedure has three major components: (a) the use of goodness-of-fit test statistics to test the assumptions of the Rician model across all voxels of the image; (b) the use of influence measures to identify outliers; (c) the use of 2D and 3D graphs to search for various artifacts and to detect intravoxel variability. At 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. Furthermore, in the normal models (6) and (7), the goodness-of-fit statistics developed here are completely new.

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 other than stochastic noise. The second challenge is to identify those voxels, in which the signal is affected strongly by partial volume effects, or by the presence of multiple tissues within the same voxel. Statistically, identifying partial volume effects may be closely associated with testing whether the link function in (2) is correctly specified.

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; Lin, Wei, and Ying 2002). However, because, in the Rician model (2), $E(S_i|x_i)$ does not have a simple form, testing directly the mean structure of 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:

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

Because $W(\theta)$ is close to one at 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

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

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

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

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 $n \to \infty$,

$$T_1(u;\hat{\theta}) = T_1(u;\theta_*) + \partial_\theta T_1(u;\theta_*)(\hat{\theta} - \theta_*) + o_p(1) = T_1(u;\theta_*) + \Delta_1(u)\sqrt{n}(\hat{\theta} - \theta_*) + o_p(1),$$

where $\Delta_1(u)$ is defined by

$$\Delta_1(u) = \int [\partial_\theta W(\theta_*)S - \partial_\theta f(x,\beta_*)]\, 1(x^T\beta_* \leq u) p(S|x,\theta_*)p(x)dSdx.$$

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

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

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 $(T_1(\cdot,\theta_*), \sqrt{n}(\hat{\theta}-\theta_*)^T)^T$, which is given in Theorem 1.

The second test statistic that we propose is based on

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

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

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

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

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

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

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

where $\Delta_2(\alpha, u) = \int [\partial_\theta W(\theta_*) S - \partial_\theta f(x, \beta_*)] \, 1(\alpha^T x \leq u) p(S|x, \theta_*) p(x) dS dx$. Therefore, the asymptotic null distribution of CM_1 depends on the asymptotic distribution of $(T_2(\alpha, u; \theta_*), \sqrt{n}(\hat{\theta} - \theta_*)^T)^T$, which is also given in Theorem 1.

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

i) $(T_1(\cdot; \theta_*), \sqrt{n}(\hat{\theta} - \theta_*)^T)^T$ converges in distribution to $(G_1(\cdot; \theta_*), \nu_1^T)^T$, where $(G_1(\cdot; \theta_*), \nu_1^T)$ is a Gaussian process with a mean of zero and a covariance function $C_1(u_1, u_2)$, which is given by

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

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

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

$$C_2((\alpha_1,u_1),(\alpha_2,u_2)) = \int\int \begin{pmatrix} [W(\theta_*)S - f(x,\beta_*)]1(x^T\alpha_1 \le u_1) \\ \psi(S,x;\theta_*) \end{pmatrix} \times \begin{pmatrix} [W(\theta_*)S - f(x,\beta_*)]1(x^T\alpha_2 \le u_2) \\ \psi(S,x;\theta_*) \end{pmatrix}^T p(S|x,\theta_*)dSdp(x). \quad (20)$$

iv) 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 $x^T\alpha$.

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

Because $E(S_i^2|x_i)$ 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

$$H_0^{(2)} : E(S^2|x) = f(x,\beta)^2 + 2\sigma^2 \text{ for some } \theta \in \Theta,$$
$$H_1^{(2)} : E(S^2|x) \neq f(x,\beta)^2 + 2\sigma^2 \text{ for all } \theta \in \Theta.$$

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

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

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

$$\mathrm{CK}_2 = \sup_u |T_3(u;\hat{\theta})| \text{ and } \mathrm{CM}_2 = \int_\Pi T_4(\alpha,u;\hat{\theta})^2 F_{n,\alpha}(du)d\alpha. \quad (21)$$

Similar to the reasoning in Theorem 1, we can establish the asymptotic null distributions of CK_2 and CM_2, which we therefore omitted 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 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 using the simulation studies of Section 4.

3.2 Parametric bootstrap

We develop a parametric bootstrap method to estimate the null distribution of the statistic CK_1 in each of voxels in MRI data. Because we can develop similar methods for CK_2, CM_1, and CM_2, we avoid such repetition and simply present the four key steps in generating the stochastic processes that have the same asymptotic distribution as CK_1.

Step 1. We first estimate $\hat{\theta}$ by using the data $\{(S_i, x_i) : i = 1, \cdots, n\}$ and calculate CK_1.

Step 2. We generate random samples $\{(S_i^{(g)}, x_i) : i = 1, \cdots, n, g = 1, \cdots, G\}$. Explicitly, we simulate $S_i^{(g)}$ using the conditional density $p(S_i|x_i, \hat{\theta})$. Here, the superscript (g) represents the number of replications.

Step 3. For the gth bootstrap sample, we estimate the maximum likelihood estimate $\hat{\theta}^{(g)}$ and compute $CK_1^{(g)}$.

Step 4. We repeat Step 3 for all G bootstrap samples and obtain a realization: $\{CK_1^{(g)} : g = 1, \cdots, G\}$. The $p-$value of the CK_1 test is calculated from the realization $\{CK_1^{(g)} : g = 1, \cdots, G\}$ by computing the fraction of $CK_1^{(g)}$ values that are greater than the observed CK_1.

Furthermore, because each MR image may contain the thousands to hundreds of thousands of voxels, we develop a sequential scheme to reduce further the computational burden of parametric bootstrap method. The sequential scheme for CM_1 is implemented as follows:

Step a. Generate $G = G_0$, say $G_0 = 20$, bootstrap samples and compute a $p-$value, denoted by $\hat{p}$, based on the fraction of $\{CM_1^{(g)} : g = 1, \cdots, G\}$ values that are greater than the observed CM_1;

Step b. Compute a confidence interval $[\hat{p}_L, \hat{p}_U]$, where $p_L = \hat{p} - 2\sqrt{\hat{p}_1(1-\hat{p}_1)/G}$ and $\hat{p}_U = \hat{p} + 2\sqrt{\hat{p}_1(1-\hat{p}_1)/G}$, in which $\hat{p}_1 = \max\{\hat{p}, 0.01\}$.

Step c. For a given significance level α, we reject H_0 if $\hat{p}_U < \alpha$, whereas we accept H_0 if $\hat{p}_L > \alpha$. Then, we stop and report $\hat{p}$ as the final $p-$value. Otherwise, we regenerate an additional G_0 bootstrap samples and update $G \to G + G_0$. Then, if $G < G_U$, say 99, we calculate a new $\hat{p}$ based on all of the bootstrap samples and proceed to Step b; otherwise, we

stop and report $\hat{p}$ as the final $p-$value.

The key idea of this sequential scheme is to reduce the enormous effort to estimate accurately those p-values with relatively small and large magnitude, say $\hat{p} \geq 0.1$. We expect that those voxels with $\hat{p} \geq 0.1$ account for more than 85-95% of all voxels in the image. Moreover, in MR images, we are most interested in identifying the small number of voxels that have very small $p-$values (e.g., $\hat{p} < 0.01$), indicating a high degree of misspecification of the Rician model (2). In Step b, we use $[\hat{p}_L, \hat{p}_U]$ as a 95% confidence interval for the $p-$value of CM_1. We also set the upper limit for G to be 99 or 199 in order to save computational time (Davison and Hinkley 1997; He and Zhu 2003). Finally, we present a plot of the $-\log_{10}(p)$ values for our various test statistics, such as CM_1. 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, however, have not been developed previously. Therefore, we discuss now how to develop case-deletion measures for the Rician model.

Henceforth, we assume that σ^2 is a nuisance parameter and define $U(\beta) = (\mu_1(\beta), \cdots, \mu_n(\beta))^T$, $V(\theta) = \mathrm{diag}(V_1(\theta), \cdots, V_n(\theta))$, and $S_W(\theta) = (W_1(\theta)S_1, \cdots, W_n(\theta)S_n)^T$, where $V_i(\theta) = \sigma^{-2}\mathrm{Var}(S_i W_i(\theta)) = -\sigma^{-2}\mu_i(\beta)^2 + E[\sigma^{-2}S_i^2 W_i(\theta)^2]$. 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

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

To develop influence measures, we can write the maximum likelihood 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:

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

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 high SNR, model (22) just reduces to a standard linear regression.

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

i) The residuals and standardized residuals are given by

$$\hat{r}_i = u_i^T \hat{V}(\hat{\theta})^{1/2}\{\hat{Z} - D(\hat{\beta})\hat{\beta}\} \quad \text{and} \quad \hat{t}_i = \sigma^{-1}\hat{r}_i/\sqrt{1-h_{i,i}}, \tag{23}$$

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

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

Residuals are highly informative about the compatibility of a postulated model with 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 non-constant 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 residuals, we compare the standardized residuals with the conventional benchmark 2.5. In other words, we regard the $i-$th 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.

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

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

where $\hat{\beta}_{(i)}$ denotes the maximum likelihood estimate of β based on the sample size $n-1$ with the $i-$th case deleted. Instead of calculating $\hat{\beta}_{(i)}$ directly, we compute the first order

approximation of $\hat{\beta}_{(i)}$, denoted by $\hat{\beta}^I_{(i)}$, which is given by

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

where $D_i(\hat{\beta})^T$ is the $i-$th row of $D(\hat{\beta})$. Therefore, we get the first-order approximation of C_i, denoted by C_i^I, as $C_i^I = h_{i,i}\hat{t}_i^2/(1 - h_{i,i})$. 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 numbers 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. More detailed examination of the 2D graphs for these voxels is indicated. These graphs include maps of the number of outliers pre slice and per image, index plots of influcence measures, and various plots of residuals that can reveal anomalies such as non-constant variance, curvature, transformation, and outliers in the regression (Cook and Weisberg 1982; Luo and Nichols 2002). 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 data set. 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, $S_0 = 500$, $b = [0, 50, 100, \cdots, 1100]$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 SNR were in the range of [0.366, 2]. At each S_0/σ, 4,000 diffusion weighted data sets were generated. Under each model, we calculated the parameter estimates $\hat{\theta} = (\hat{d}, \hat{S}_0, \hat{\sigma}^2)$. We finally calculated the bias and the standard deviation estimate from the 4,000 estimates based on 4,000 simulated ADC data sets (Table 1). At low S_0/σ, the estimates from model (2) had smaller biases, but larger standard deviations, whereas models (6) and (7) had larger biases, but smaller standard deviations. When $S_0/\sigma \geq 10$, models (2), (6) and (7) had comparable bias and standard deviation in the parameter estimates.

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 comprised 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): $(1000, 1000)$, $(1000, 3000)$, and $(3000, 3000)$ s/mm^2 in order to examine the sensitivity of differing b factors in detecting multiple fiber directions. Finally, for each simulation, a significance level 5% was considered, and 1,000 replications were used to estimate the nominal significance level (i.e., rejection levels for the null hypothesis).

Table 2 presented estimates for the rejection rates of the four test statistics. We observed that in a single compartment, the rejection rates of CK_i and CM_i for $i = 1, 2$ were slightly greater than the nominal level for this sample of 66 images. Overall, the rejection rates in all cases were accurate, and Type I errors were not excessive. These findings suggested that the parametric bootstrap 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, $(b_1, b_2) = (1000, 3000)$ s/mm^2 provided the best performance. Under $(b_1, b_2) = (1000, 3000)$ 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 almost 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 et al. 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 $b_1 = 1000 s/mm^2$. 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. We considered a significance level of 5% and used 1,000 replications to estimate the nominal significance levels.

Table 3 presented estimates for the rejection rates of our four statistics. 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 A Real Imaging Data

We acquired diffusion weighted (DW) images of the brain of a healthy adult male subject (right-handed; age 34 years). The imaging acquisition scheme $\{(b_i, r_i) : i = 1, \cdots, 38\}$ consisted of 3 baseline images with $b = 0$ s/mm^2 and 35 directions of diffusion gradients that were arranged uniformly in the 3-dimensional space at $b = 1000$ s/mm^2 (Hardin, Sloane and Smith 1994). Each diffusion-weighted image 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 a 2- to 6-degrees of rotation, 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 $\{(m_i, e_i) : i = 1, 2, 3\}$, $\arccos|<\mathbf{e}_1, \mathbf{v}_1>|$, 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 $= \sqrt{1 - M_2(M_1^2 - 2M_2)^{-1}}$, where $m_1 \geq m_2 \geq m_3$, $M_1 = \mathrm{tr}(\hat{D})$, $M_2 = m_1m_2 + m_1m_3 + m_2m_3$, $M_3 = m_1m_2m_3$ and $\mathbf{v}_1$ is the true principal direction $\mathbf{v}_1 = (1, 0, 0)^T$. 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, Xu, Amir, Hao, Zhang, Alayar, Ravi, and Peterson 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 parametric bootstrap in Section 3 of this paper.

We plotted maps of scan summaries to identify possible artifacts and acquisition problems in the DW images. Translational and rotational parameters (Fig. 3b), obtained from SPM2 (http://www.fil.ion.ucl.ac.uk), detected rightward rotation of 2 to 6 degrees beginning in the 28th acquisition. These motion parameters also revealed head movement that began during acquisition of images from the fourth gradient direction. The DW images acquired at the 31st and 32nd acquisitions corresponded to deliberate rotation. 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 ten directions (red to white on the color spectrum in Fig. 3a).

To reduce or eliminate motion artifacts, we used the normalized mutual information method to co-register all other DW images to the first DW image in SPM2 while properly reorienting the diffusion gradients (Rohde, Barnett, Basser, Marenco, and Pierpaoli 2004). We applied a 4-order B-spline 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 non-realigned images, as well as a decline in the percentage of outliers per slice and per direction after coregistration (Fig. 4 a, b, c, and d). Furthermore, we examined voxels having 0-10 outliers and found that motion correction using coregistration significantly increased the percentage of voxels without outliers from 14.5% to 29.5% (Fig. 4e). However, despite the efficacy of this method for correcting motion artifacts, 30% of the voxels still contained at least two outliers after coregistration (Fig. 4e), and several slices in the last 10 directions (red to white on the color spectrum) contained a large number of outliers (Fig. 4c, red to white on the color spectrum), warranting further investigation.

The 3D images of the $-\log_{10}(p)$ values for the test statistics CK_1, CM_2, and FA, and the outlier counts per voxel were more sensitive and specific in assessing the quality of the DW images (Figs. 5a-5p). A p-value of 0.01 corresponded to a $-\log_{10}(p)$ value of 2; thus a voxel having a $-\log_{10}(p)$ value greater than or equal to 2.0 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. For instance, inspecting the DW images at baseline, the $-\log_{10}(p)$ values did not reveal significant change in of CK_1 and FA values as a consequence of the coregistration procedure (Fig. 5a, 5b, 5g, 5i, 5j, and 5o). However, we found several notable changes after coregistration as follows. The number of voxels having large $-\log_{10}(p)$ values for the FA statistics increased dramatically following coregistration (Fig. 5c and 5k), whereas the number of voxels containing multiple outliers declined dramatically (Fig. 5e, 5f, 5m, and 5n). Moreover, prior to coregistration, many voxels had $-\log_{10}(p)$ values of CM_2 that were greater than or equal to 2.0 (Fig. 5h), whereas the number of those voxels declined dramatically following coregistration (Fig. 5p). Consistent with our simulation studies (Section 4), this data analysis suggested that CM_2 was a sensitive measure for detecting head motion.

Assessing the quality of DW images was crucial for further processing images. As shown above (Fig. 5), the maps of the $-\log_{10}(p)$ values of 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 DW images. Those artifacts strongly influenced the estimation of DTs, 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 (Fig. 5d, 5l). Before coregistration, we found that 59.97% were isotropic, 9.37% were oblate, 23.06% were prolate and 7.61% were nondegenerate (Fig. 5d). Following coregistration, we found that 36.43% were isotropic, 14.09% were oblate, 31.06% were prolate and 18.42% were nondegenerate (Fig. 5l). Most tractography algorithms can only track fibers across voxels containing either anisotropic or prolate DTs, which accounted for 49.48% 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 DW images after coregistration, we also examined 3D images of $-\log_{10}(p)$ values of the test statistics, standardized residuals and Cook's distances. Even after coregistration, a large number of voxels were deemed problematic (Fig. 5o, 5p), suggesting strongly that some of the assumptions of the Rician model were not satisfied in those regions and therefore that further investigation was warrented. Moreover, we searched the standardized residuals (or Cook's distance) in all voxels across all slices and directions to identify slices having large numbers of positive and negative outliers (i.e., data points of excessive influence). Most of the images of standardized residuals in the last 10 directions contained such slices. For illustration, we selected 10 slices having high concentrations of outliers from the 32nd direction (Fig. 6). Some of these slices contained many positive residuals (Figs. 6a-c), whereas others contained many negative residuals (Figs. 6d-6j), indicating that even after coregistration, motion artifacts or other unspecified problems remain in the resliced DW images. Developing methods for identifying the precise sources of non-Rician noise and correcting for them in the resliced DW images 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)). Prior to 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. In contrast, following coregistration, the p-values for CK_1, CK_2, CM_1, and CM_2 were 0.25, 0.05, 0.78, and 0.86, 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 (Fig. 7a, 7b) revealed that the 4th, 8th, and 34th observations were likely outliers. A plot of the standardized residuals against the raw MRI values (Fig. 7c) revealed a strong linear relationship between residuals and raw MRI values (Cook and Weisberg 1982). Furthermore, we observed a nonlinear relationship (Fig. 7d) of Cook's distances against raw MRI values. Together these plots (Fig. 7c, 7d) indicated that a Rician model (8) did not fit the MRI data satisfactorily. Further improvements in model specification or post-acquisition 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 DW images. Coregistration improved image quality, but substantial non-stochastic noise sources remained in the last 10 slices: the standardized residual images revealed a large concentration of outliers in several image volumes. 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 DW images. 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 SNR. The key features of our procedures include: calculating test statistics that assess the validity of the assumptions of the statistical mod-

els 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. Nevertheless, our procedure assesses the quality of MRI statistically and cannot replace various preprocessing techniques, such as registration and smoothing methods.

References

Abramowitz, M., and Stegun I., editor (1965), *Handbook of Mathmatical Functions.* New York: Dover Publications.

Alexander, D. C., and Barker, G. J. (2005), "Optimal Imaging Parameter for Fibre-Orientation Estimation in Diffusion MRI," *NeuroImage*, 27, 357-367.

Alexander, D. C., Barker, G. J., and Arridge, S. R. (2002), "Detection and Modeling of Non-Gaussian Apparent Diffusion Coefficient Profiles in Human Brain Data," *Magnetic Resonance in Medicine*, 48, 331-340.

Basser, P. J., Mattiello, J., and LeBihan, D. (1994 a), "MR Diffusion Tensor Spectroscopy and Imaging," *Biophysical Journal*, 66, 259-267.

Basser, P. J., Mattiello, J., and LeBihan, D. (1994 b), "Estimation of the Effective Self-diffusion Tensor from the NMR Spin Echo," *Journal of Magnetic Resonance, Series B.*, 103, 247-254.

Cook, R. D. and Weisberg, S. (1982), *Residuals and Influence in Regression.* London: Chapman and Hall.

Davison, A. C., and Hinkley, D. V. (1997), *Bootstrap Methods and Their Application.* London: Cambridge University Press.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977), "Maximum Likelihood from Incomplete Data via the EM Algorithm (with Discussion)," *Journal of the Royal Statistical Society, Series B.*, 39, 1-38.

den Dekker, A. J., and Sijbers, J. (2005), "Implications of the Rician Distribution for fMRI Generalized Likelihood Ratio Tests," *Magnetic Resonance Imaging* 23, 953-959.

Escanciano, J. C. (2006), "A Consistent Diagnostic Test for Regression Models using Projections," *Econometric Theory*, 22, 1030-1051.

Gudbjartsson, H., and Patz, S. (1995), "The Rician Distribution of Noisy MRI Data," *Magnetic Resonance in Medicine*, 34, 910-914.

Hardin, R. H., Sloane, N. J. A., and Smith, W. D. (1994). *Minimal energy arrangements of points on a sphere with minimal 1/r potential.* available at http://www.research.att.com/ njas/electrons/.

He, X. M., and Zhu, L. X. (2003), "A Lack-of-fit Test for Quantile Regression," *Journal of the American Statistical Association*, 98, 1013-1022.

Henkelman, R. M. (1985), "Measurement of Signal Intensities in the Presence of Noise in MR Images," *Medical Physics*, 12, 232-233.

Huettel, S.A., Song, A.W., and McCarthy, G. (2004), *Functional Magnetic Resonance Imaging.* Sinauer Associates, Inc.

Jones, D. K., and Basser, P. J. (2004), "Squashing Peanuts and Smashing Pumpkins: How Noise Distorts Diffusion-Weighted MR Data," *Magnetic Resonance in Medicine*, 52, 979-993.

Jones, D. K., Horsfield, M. A., and Simmons, A. (1999), " Optimal Strategies for Measuring Diffusion in Anisotropic Systems by Magnetic Resonance Imaging," *Magnetic Resonance in Medicine*, 42, 515-525.

Jorgensen, B. (1992), "Exponential Dispersion Models and Extension: a Review," *International Statistical Review*, **60,** 5-20.

Karlsen, O. T., Verhagen, R., and Bovee, W.M. (1999), "Parameter Estimation from Rician-distributed Data Sets using a Maximum Likelihood Estimator: Application to T_1 and Perfusion Measurements," *Magn. Reson. Med.*, 41, 614-623.

Kristoffersen, A. (2007), "Optimal Estimation of the Diffusion Coefficient from Non-averaged and Averaged Noisy Magnitude Data," *J. Magn. Reson.*, 187, 293-305.

Lin, D. Y., Wei, L. J., and Ying, Z. L. (2002), "Model-checking Techniques based on Cumulative Residuals," *Biometrics*, 58, 1-12.

Louis, T. A. (1982), "Finding the Observed Information Matrix when Using the EM Algorithm," *Journal of the Royal Statistical Society Series B* **44,** 190-200.

Luo, W. L., and Nicholas, T. N. (2003), "Diagnosis and Exploration of Massively Univariate fMRI Models," *NeuroImage*, 19, 1014-1032.

McCullagh, P., and Nelder, J. A. (1989), *Generalized Linear Models.* London: Chapman and Hall.

Meng, X. L., and Rubin, D. B. (1993), "Maximum Likelihood Estimation via the ECM Algorithm: A General Framework," *Biometrika*, 80, 267-278.

Nowark, R. D. (1999), "Wavelet-based Rician Noise Removal for Magnetic Resonance Imaging," *IEEE Transactions on Image Processing* **8**, 1408-1419.

Rice, S. O. (1945), "Mathematical Analysis of Random Noise," *Bell System Technical Journal*, 24, 46-156.

Rohde, G. K., Barnett, A. S., Basser, P. J., Marenco, S., and Pierpaoli, C. (2004), "Comprehensive Approach for Correction of Motion and Distortion in Diffusion-weighted MRI," *Magnetic Resonance in Medicine*, 51, 103-114.

Rowe, D. R. (2005), "Parameter Estimation in the Magnitude-only and Complex-valued fMRI Data Models," *NeuroImage*, 25, 1310-1324.

Rowe, D. R., and Logan, B. R. (2004), "A Complex Way to Compute fMRI Activation," *NeuroImage*, 23, 1078-1092.

Rowe, D. R., and Logan, B. R. (2005), "Complex fMRI Analysis with Unrestricted Phase is Equivalent to a Magnitude-only Model," *NeuroImage*, 24, 603-606.

Sijbers, J., and den Dekker, A. J. (2004), "Maximum Likelihood Estimation of Signal Amplitude and Noise Variance from MR Data," *Magnetic Resonance in Medicine*, 51, 586-594.

Sijbers, J., den Dekker, A. J., Scheunders, P., and Van Dyck, D. (1998a), "Maximum-likelihood Estimation of Rician Distribution Parameters," *IEEE Transactions on Image Processing*, 17, 357-361.

Sijbers, J., den Dekker, A. J., Verhoye, M., Van Audekerke, J., and Van Dyck, D. (1998b), "Estimation of Noise from Magnitude MR Images," *Magnetic Resonance Imaging*, 16, 87-90.

Skare, S., Li, T., Nordell, B., and Ingvar, M. (2000), "Noise Considerations in the Determination of Diffusion Tensor Anisotropy," *Magnetic Resonance Imaging*, 18, 659-669.

Stute, W. (1997), "Nonparametric Model Checks for Regression," *Annals of Statistics*, 25, 613-641.

Tuch, D. S., Reese, T. G., Wiegell, M. R., Makris, N., Belliveau, J. W., and Wedeen, V. J. (2002), "High Angular Resolution Diffusion Imaging Reveals Intravoxel White Matter Fiber Heterogeneity," *Magnetic Resonance in Medicine*, 48, 577-582.

van der Vaart, A. W., and Wellner, J. A. (1996), *Weak Convergence and Empirical Processes with Applications to Statistics.* New York: Springer.

Wei, B. C. (1998), *Exponential Family Nonlinear Models.* Singapore: Springer.

Xu, D., Mori, S., Solaiyappan, M., van Zijl, P. C., and Davatzikos, C. (2002), "A Framework for Callosal Fiber Distribution Analysis," *NeuroImage*, 17, 1131-1143.

Yoo, S.S., Guttmann, C. R. G., and Panych, L. P. (2001), "Multiresolution Data Acquisition and Detection in Functional MRI," *NeuroImage*, 14, 1476-1485.

Yoo, S.S., Talos, I.F., Golby, A.J., Black, P. M., and Panych, L. P. (2004), "Evaluating Requirements for Spatial Resolution of fMRI for Neurosurgical Planning," *Human Brain Mapping*, 21, 34-43.

Zhu, H. T., Xu, D., Amir, R., Hao, X., Zhang, H., Alayar, K., Ravi, B., and Peterson, B. (2006), "A Statistical Framework for the Classification of Tensor Morphologies in Diffusion Tensor Images," *Magnetic Resonance Imaging*, 24, 569-582.

Zhu, H. T., and Zhang, H. P. (2004), "A Diagnostic Procedure based on Local Influence Measure," *Biometrika*, 91, 579-589.

Table 1: ADC imaging: Bias and SD of three components of $\hat{\theta}$. TRUE denotes the true value of the regression parameters; BIAS denotes the bias of the mean of the regression estimates; SD denotes the standard deviations of the regression estimates. Five different S_0/σ $\{2, 4, 6, 10, 15\}$ and 4,000 simulated datasets were used for each case.

		$R(\mu_i, \sigma^2)$			$N(\sqrt{\mu_i^2 + \sigma^2}, \sigma^2)$			$N(\mu_i, \sigma^2)$		
	S_0/σ	σ^2	S_0	d	σ^2	S_0	d	σ^2	S_0	d
TRUE	2	62500	500.00	2.000	62500	500.00	2.000	62500	500.00	2.000
BIAS	2	-15048	49.63	0.263	-27524	7.73	-0.985	-24989	16.83	-1.403
SD	2	18587	174.72	1.960	13303	133.94	1.216	15232	102.71	0.364
TRUE	4	15625	500.00	2.000	15625	500.00	2.000	15625	500.00	2.000
BIAS	4	-1938	-0.46	0.023	-4976	-9.32	-0.323	-4922	-19.70	-0.711
SD	4	5454	80.73	0.963	3420	77.28	0.647	3452	65.95	0.332
TRUE	6	6944	500.00	2.000	6944	500.00	2.000	6944	500.00	2.000
BIAS	6	-694	-1.25	0.006	-1746	-5.28	-0.135	-1738	-12.39	-0.371
SD	6	2221	52.24	0.400	1704	50.59	0.325	1707	47.07	0.265
TRUE	10	2500	500.00	2.000	2500	500.00	2.000	2500	500.00	2.000
BIAS	10	-305	-0.47	-0.024	-430	-1.08	-0.033	-422	-4.20	-0.138
SD	10	721	29.54	0.199	659	30.25	0.198	661	29.33	0.182
TRUE	15	1111	500.00	2.000	1111	500.00	2.000	1111	500.00	2.000
BIAS	15	-87	0.74	0.005	-145	-0.60	-0.015	-143	-2.03	-0.065
SD	15	322	20.02	0.132	305	20.23	0.130	307	19.94	0.128

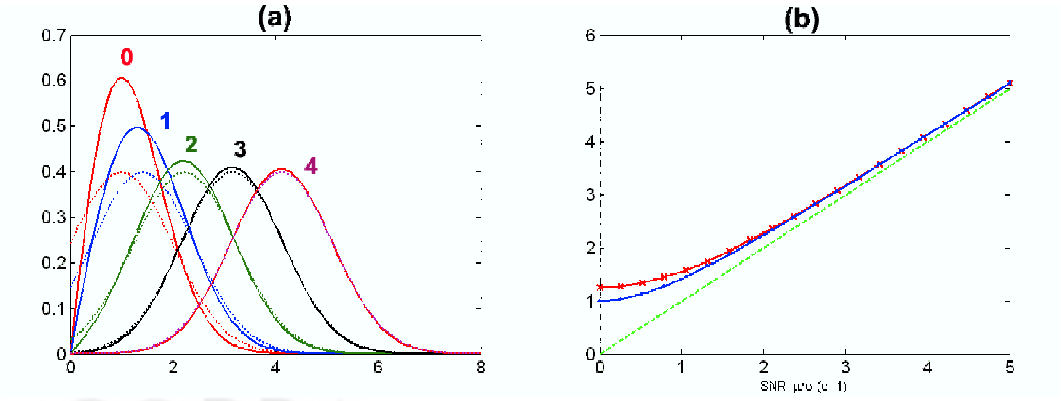

Figure 1: Rician distribution: (a) $R(\mu, 1)$ and $N(\sqrt{\mu^2 + 1}, 1)$ for $\mu = 0, 1, 2, 3, 4$; (b) the mean functions of $R(\mu, 1)$ (red), $N(\sqrt{\mu^2 + 1}, 1)$ (blue) and $N(\mu, 1)$ (green) for $\mu \in [0, 5]$.

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 $f(x_i, \beta) = S_0[p_1 \exp(-b_i r_i^T D_1 r_i) + (1-p_1)\exp(-b_i r_i^T D_2 r_i)]$, at a significance level 0.05. The first DT compartment is $D_1 = \text{diag}(1.7, 0.2, 0.2)$ and the second DT compartment is $D_2 = \text{diag}(0.2, 1.7, 0.2)$. Five different S_0/σ $\{5, 10, 15, 20, 25\}$ and 1,000 simulated data sets were used.

		CK_1			CM_1			CK_2			CM_2		
		(b_1, b_2) $\times 1000 s/mm^2$											
SNR	p_1	$(1,1)$	$(1,3)$	$(3,3)$	$(1,1)$	$(1,3)$	$(3,3)$	$(1,1)$	$(1,3)$	$(3,3)$	$(1,1)$	$(1,3)$	$(3,3)$
5	1	0.066	0.081	0.089	0.072	0.08	0.068	0.074	0.086	0.082	0.094	0.08	0.064
10	1	0.069	0.078	0.088	0.071	0.076	0.074	0.072	0.081	0.074	0.068	0.078	0.059
15	1	0.059	0.072	0.081	0.071	0.078	0.069	0.07	0.069	0.086	0.074	0.086	0.061
20	1	0.058	0.062	0.078	0.072	0.082	0.074	0.068	0.062	0.086	0.072	0.098	0.056
25	1	0.05	0.064	0.076	0.068	0.078	0.064	0.066	0.064	0.074	0.05	0.076	0.058
5	0.5	0.044	0.112	0.191	0.056	0.162	0.108	0.081	0.132	0.188	0.078	0.14	0.104
10	0.5	0.136	0.398	0.204	0.072	0.344	0.05	0.098	0.126	0.074	0.06	0.154	0.038
15	0.5	0.212	0.651	0.206	0.068	0.578	0.032	0.112	0.138	0.018	0.062	0.206	0.016
20	0.5	0.332	0.796	0.276	0.062	0.734	0.022	0.198	0.152	0.008	0.054	0.246	0.006
25	0.5	0.496	0.906	0.390	0.048	0.871	0.002	0.294	0.162	0	0.031	0.256	0

Table 3: Comparison of the rejection rates for the specification statistics CK_1, CK_2, CM_1, and CM_2, under the presence of head motion at significance level 0.05. The first $[50 \times p_1]$ acquisitions were generated from a single diffusion model with $D_1 = \text{diag}(0.2, 1.7, 0.2)$ and the last $50 - [50 \times p_1]$ acquisitions were generated from a single diffusion model with $D_2 = \text{diag}(0.7, 0.7, 0.7)$. Five different S_0/σ $\{5, 10, 15, 20, 25\}$ and 1,000 simulated data sets were used.

	p_1							
	0.7				0.5			
SNR	CK_1	CK_2	CM_1	CM_2	CK_1	CK_2	CM_1	CM_2
5	0.063	0.037	0.117	0.143	0.043	0.053	0.137	0.173
10	0.170	0.117	0.187	0.376	0.083	0.177	0.180	0.403
15	0.220	0.196	0.203	0.573	0.130	0.280	0.170	0.573
20	0.263	0.226	0.200	0.696	0.203	0.430	0.170	0.677
25	0.300	0.267	0.223	0.763	0.228	0.490	0.157	0.743

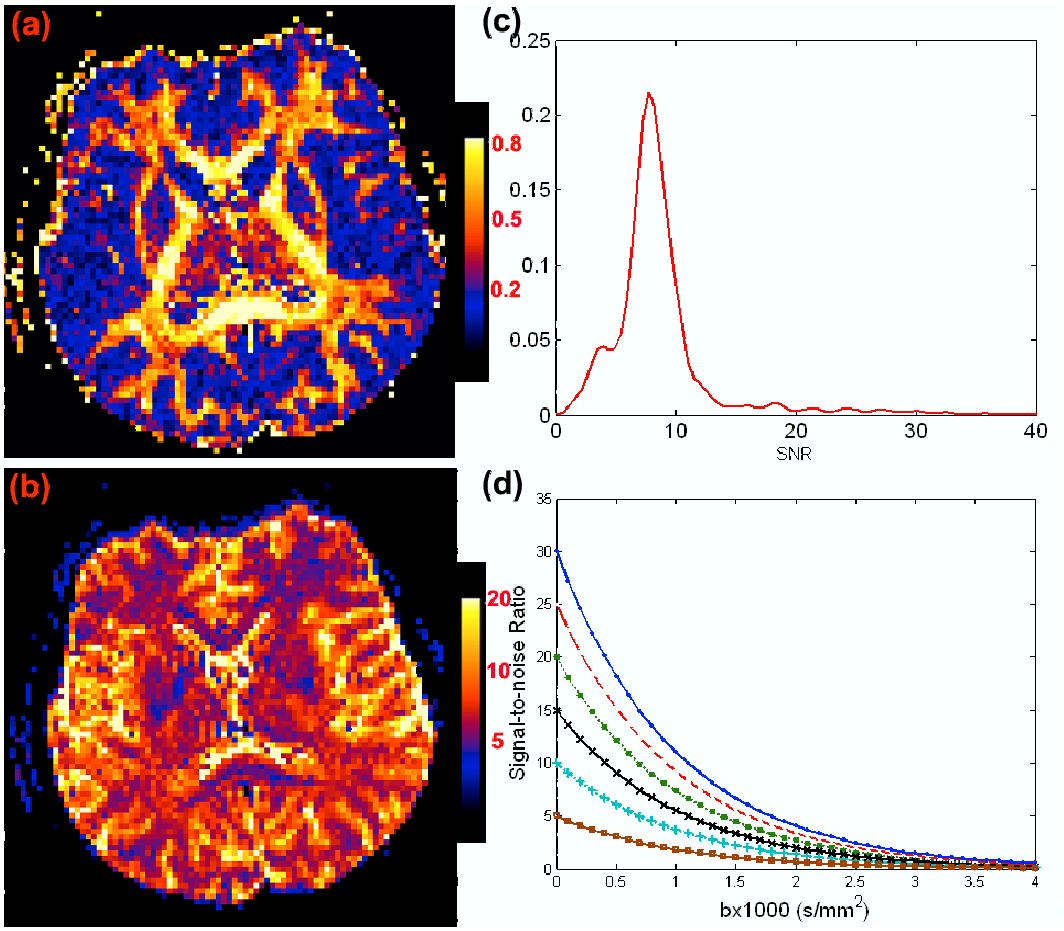

Figure 2: Maps of (a) FA; (b) S_0/σ; (c) the kernel density of S_0/σ values for anisotropic tensors having FA$\geq$ 0.5 at a selective slice from a single subject; and (d) the signal-to-noise ratio $S_0 \exp(-b_i)/\sigma$ as a function of b_i ($\times$1000 s/mm^2) at each $S_0/\sigma \in \{5, 10, 15, 20, 25, 30\}$.

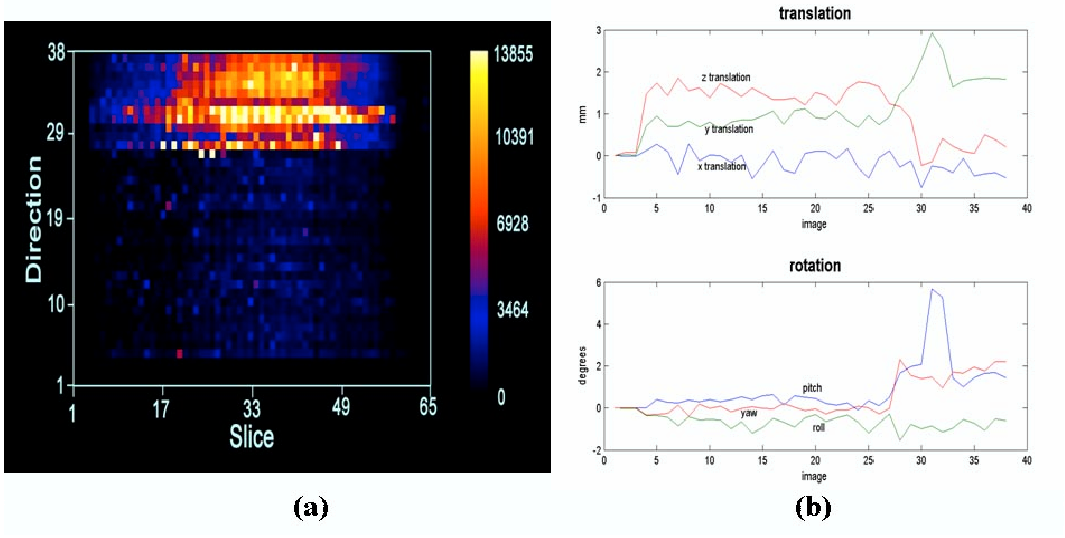

Figure 3: Scan summaries: (a) Maps of outlier count per slice (horizontal axis) and per direction (vertical axis); (b) Maps of translational and rotational parameters.

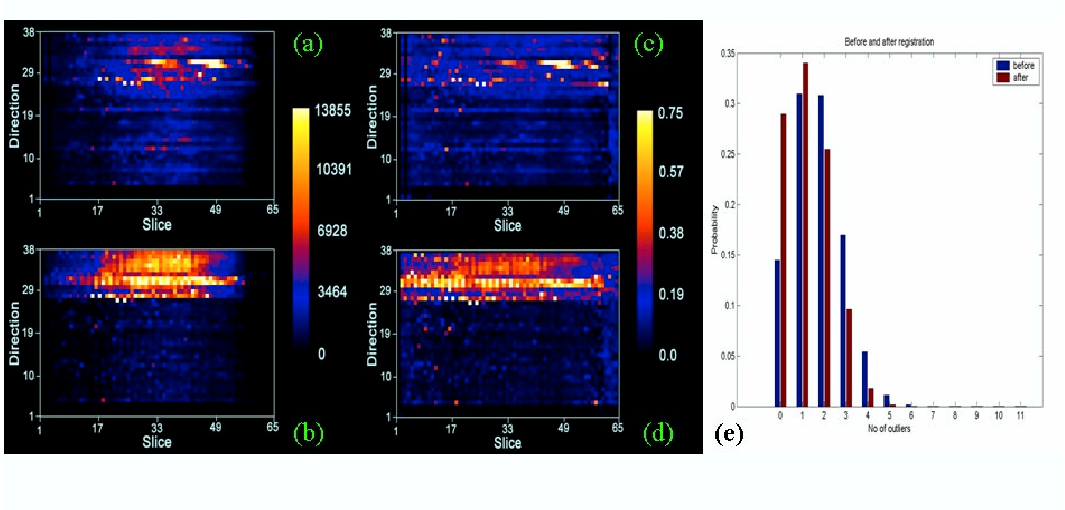

Figure 4: Assessing the effect of applying a registration algorithm to diffusion weighted images: (a) outlier count per slice and per direction after registration and (b) before registration; (c) percentages of outliers per slice and per direction after registration and (d) before registration; (e) plot of the percentages of voxels having 0-5 outliers. In (e), the blue (or red) bars represent the percentages of voxels having a given number of outliers before (or after) registration.

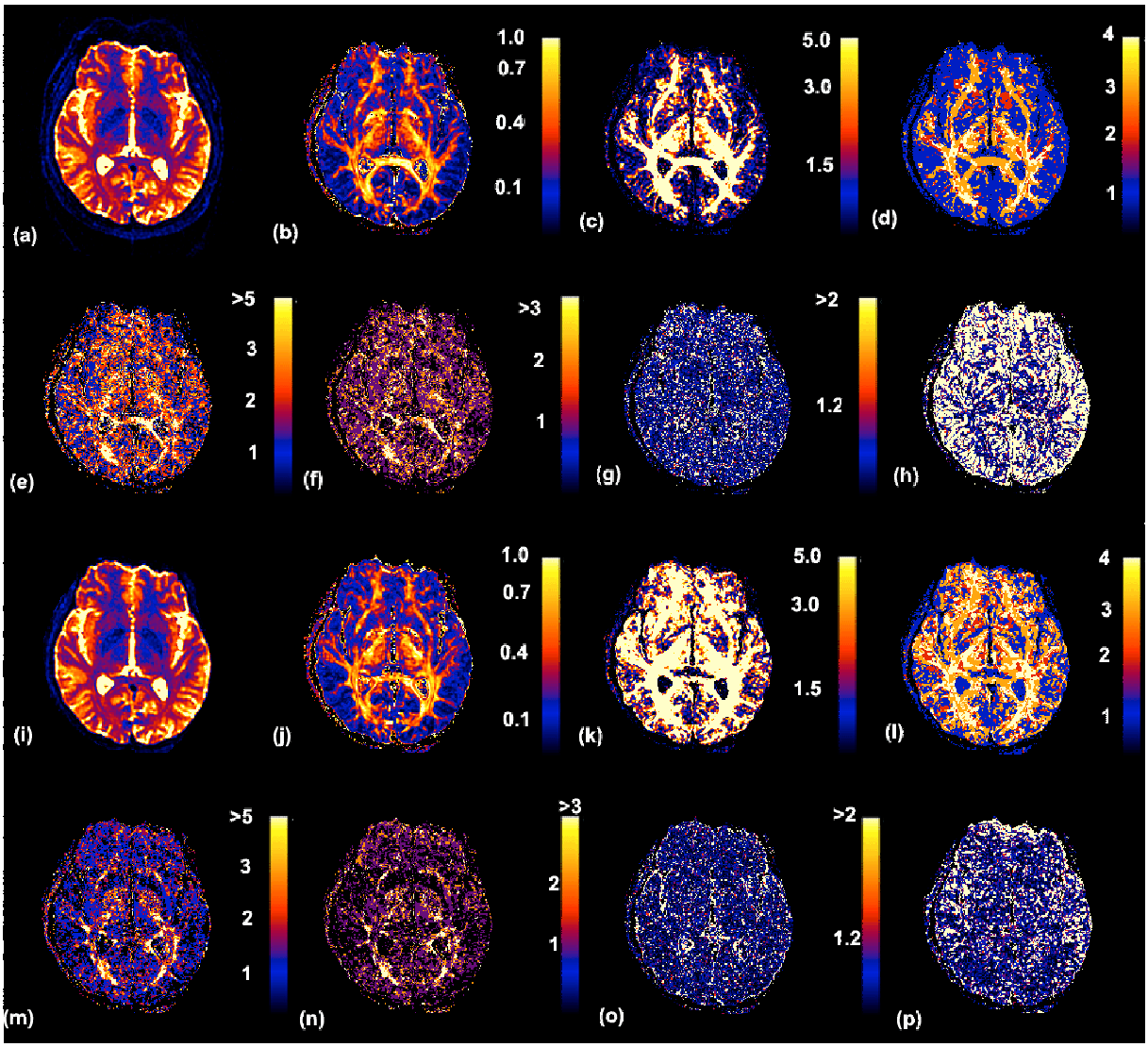

Figure 5: Maps of 3D images before registration (a-h) and after registration (i-p) in a single slice from a single subject. Before registration: (a) DW image at baseline; (b) FA value; (c) $-\log_{10}(p)$ value of FA; (d) tissue types (white=nondegenerate tensor, red=oblate tensor, yellow=prolate tensor, blue=isotropic tensor); (e) outlier count (standardized residual) in each voxel; (f) outlier count (Cook's distance) in each voxel; (g) $-\log_{10}(p)$ values of CK_1; (h) $-\log_{10}(p)$ values of CM_2. After registration: (i) DW image at baseline; (j) FA value; (k) $-\log_{10}(p)$ value of FA; (l) tissue types; (m) outlier count (standardize residual) in each voxel; (n) outlier count (Cook's distance) in each voxel; (o) $-\log_{10}(p)$ values of CK_1; (p) $-\log_{10}(p)$ values of CM_2.

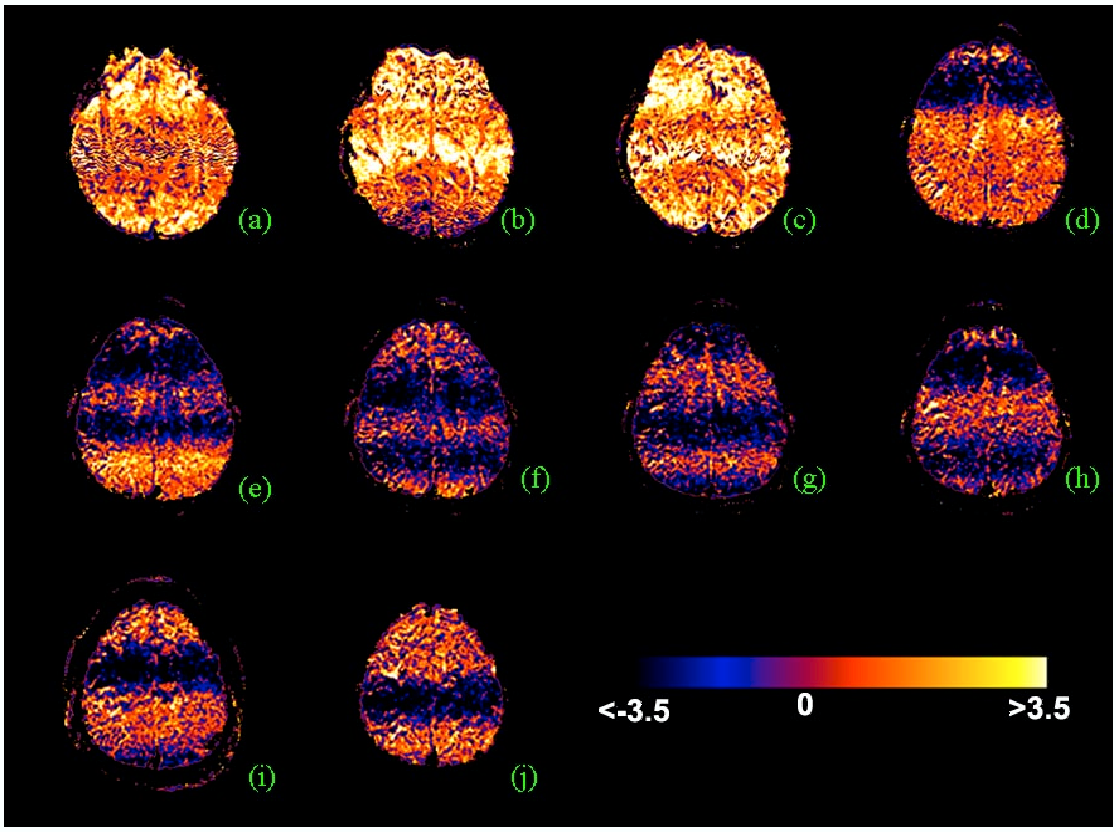

Figure 6: Problematic slices in the 32nd direction after alignment: the (a) 30th, (b) 32nd, (c) 34th, (d) 43rd, (e) 45th, (f) 46th, (g) 47th, (h) 48th, (i) 49th, and (j) 50th slices. Voxels in the black-to-blue range have large negative standardized residuals (< -3.5), while yellow to white voxels have large positive standardized residuals (> 3.5).

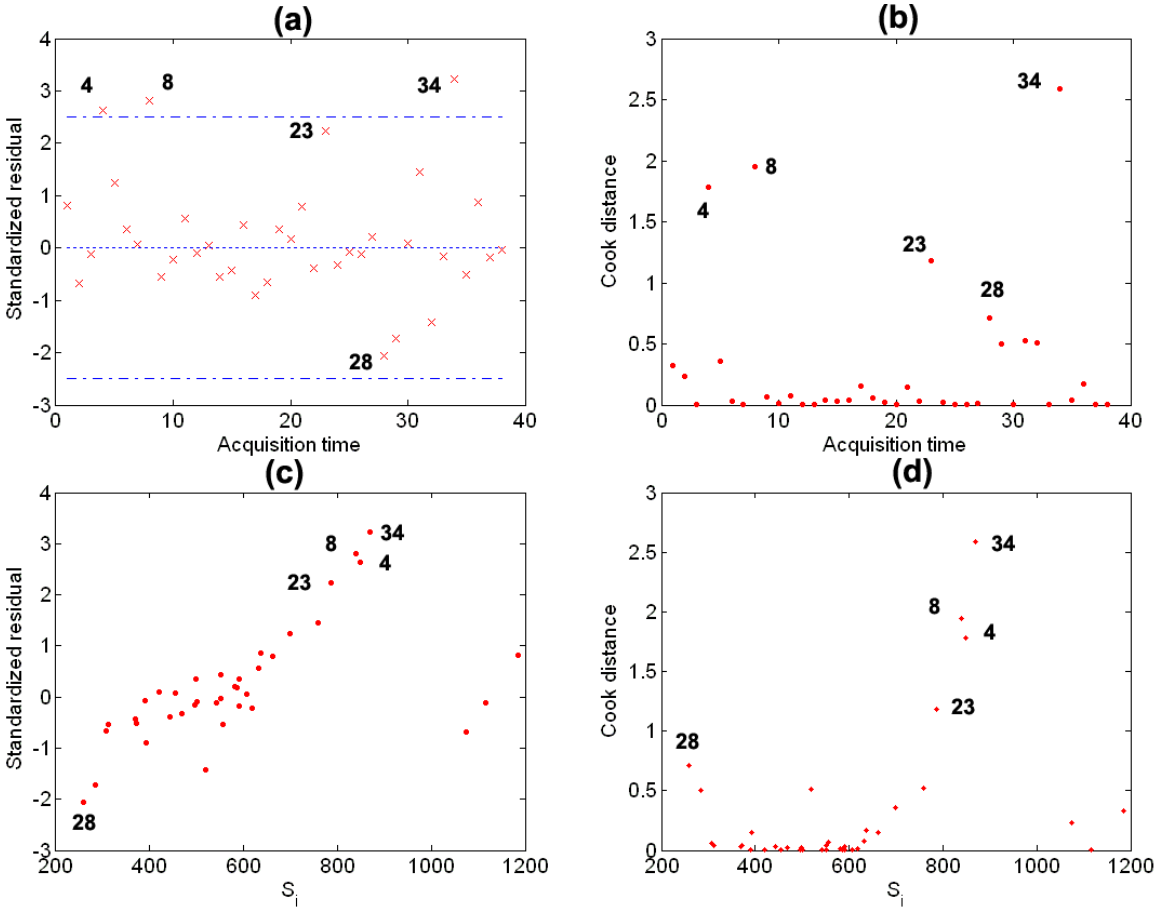

Figure 7: Multiple 2D graphs for a selected voxel (110, 69, 30): (b) index plot of standardized residuals; (b) index plot of Cook's distances; (c) standardized residuals against raw data; (d) Cook's distances against raw data.