Available online at www.sciencedirect.com
ScienceDirect
Journal of Magnetic Resonance 187 (2007) 293–305
JMR Journal of Magnetic Resonance
www.elsevier.com/locate/jmr

Optimal estimation of the diffusion coefficient from non-averaged and averaged noisy magnitude data

Anders Kristoffersen *

Klinikk for bildediagnostikk, St. Olavs Hospital HF, 7006 Trondheim, Norway

Received 23 January 2007; revised 23 April 2007
Available online 22 May 2007

Abstract

The magnitude operation changes the signal distribution in MRI images from Gaussian to Rician. This introduces a bias that must be taken into account when estimating the apparent diffusion coefficient. Several estimators are known in the literature. In the present paper, two novel schemes are proposed. Both are based on simple least squares fitting of the measured signal, either to the *median* (MD) or to the *maximum probability* (MP) value of the Probability Density Function (PDF). Fitting to the *mean* (MN) or a high signal-to-noise ratio approximation to the mean (HS) is also possible. Special attention is paid to the case of averaged magnitude images. The PDF, which cannot be expressed in closed form, is analyzed numerically. A scheme for performing maximum likelihood (ML) estimation from averaged magnitude images is proposed. The performance of several estimators is evaluated by Monte Carlo (MC) simulations. We focus on typical clinical situations, where the number of acquisitions is limited. For non-averaged data the optimal choice is found to be MP or HS, whereas uncorrected schemes and the power image (PI) method should be avoided. For averaged data MD and ML perform equally well, whereas uncorrected schemes and HS are inadequate. MD provides easier implementation and higher computational efficiency than ML. Unbiased estimation of the diffusion coefficient allows high resolution diffusion tensor imaging (DTI) and may therefore help solving the problem of crossing fibers encountered in white matter tractography.

Keywords: Rician distribution; Averaged magnitude data; Diffusion coefficient estimation; Signal-to-noise ratio; Diffusion tensor imaging

1. Introduction

In vivo diffusion measurements are often performed using a single shot echo-planar imaging (EPI) technique. The diffusion gradients needed in the pulse sequence prolong the echo time. Hence the mean signal intensity is reduced by transverse relaxation. Furthermore, signal is lost due to the diffusion weighting. As a result the inherent signal-to-noise ratio (SNR) in diffusion measurements is relatively low. Using low diffusion weighting factors improves the SNR, but we may also want to sensitize slow diffusion components. The value of the apparent diffusion coefficient (ADC) normal to the fiber direction in white matter can be as low as 0.2×10^{-3} mm^2/s. In cerebrospinal fluid the ADC can be as high as 3×10^{-3} mm^2/s. If we apply a wide range of diffusion weighting factors, the diffusion measurement is inevitably going to be noise biased. This is true in particular for the fast diffusion components.

Many applications require high-resolution imaging, i.e. large matrix acquisition and thin slices. This applies to traditional methods like ADC mapping as well as modern techniques such as diffusion tensor imaging (DTI). However, high resolution further decreases the SNR.

In order to avoid the problem of motion-induced phase shifts one usually discards the phase information and calculates the diffusion coefficient from magnitude data. The magnitude operation changes the signal distribution from Gaussian to Rician [1] and signal intensities at low SNR are biased by the noise to a value that (on average) is higher than the true value of the magnitude. If we do not properly account for this effect, the estimates of the diffusion coefficient will be biased. In the magnetic resonance literature,

* Fax: +47 73 55 13 55.
E-mail address: Anders.Kristoffersen@stolav.no

1090-7807/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmr.2007.05.004

Henkelman [2] was the first to discuss this problem for a single receiver system. This topic has later been addressed by several authors in a variety of contexts [3–13].

It is the purpose of the present paper to compare the performance of a number of possible diffusion estimators. The means by which we compare the estimators is Monte Carlo (MC) simulations. The simulations are performed with physical parameters and acquisition parameters that are typical for human brain imaging.

In addition to estimators proposed in the literature, we introduce two novel schemes which are based on least squares fitting of the measured signal; either to the *median* (MD) value of the Rician distribution or to the *maximum probability* (MP) value. Fitting to the *mean* (MN) value is also possible [11]. For convenience we have summarized the abbreviations used for various estimators in Table 1.

In the presence of poor SNR, it is a common strategy to average the data. This introduces some problems since the Probability Density Function (PDF) of the averaged data cannot be expressed in closed form. Dietrich et al. [12] proposed to use an empirical correction scheme that requires the acquisition of phantom data. Our approach is to evaluate the PDF of averaged magnitude data by efficient numerical methods. Apart from allowing the construction of efficient least squares estimators, this can also be used for "brute-force" maximum likelihood (ML) estimation.

The MD and MP methods are related to the mean (MN) method but their performance is better. Uncorrected schemes and traditional methods [3–5] are somewhat easier to implement but their performance is significantly lower. Maximum likelihood, although often considered to be the gold-standard, does not perform better than MP and MD, and the computational cost is higher.

Unbiased estimation of the diffusion coefficient allows high-resolution DTI and may therefore help solving the problem of crossing fibers encountered in white matter tractography. We find efficient methods for non-averaged data (MP and HS), as well as for averaged data (MD and ML). The numerical efficiency of the MP, MD and HS methods is adequate for quick image processing, which is required in clinical situations. Our results apply equally well to any situation where the task is to extract data from a noisy magnitude dataset where the true magnitude is known to be mono-exponential. Other examples are measurements of the transverse relaxation rate, R_2 and effective transverse relaxation rate, R_2^*.

Table 1
Abbreviations used for the estimators that are discussed in this manuscript

Abbreviation	Used for
LR	Linear regression
UC	Uncorrected
MN	Mean
MD	Median
MP	Maximum probability
PI	Power image
ML	Maximum likelihood
HS	High SNR approximation

2. Theory

2.1. Probability density function for the image data

After quadrature detection, the raw data are complex valued. Applying the inverse Fourier transform, we obtain the voxel signal, which is also a complex quantity, i.e. $s = x + \mathrm{i}y = m\exp(\mathrm{i}\varphi)$, where $m = \sqrt{x^2 + y^2}$ is the magnitude and $\varphi = \arctan(y/x)$ is the phase. Since imaged objects are real, the phase should be constant throughout the image, however flow and motion will introduce phase variations.

When the raw data is corrupted by Gaussian noise, the same holds true for the image data. If we assume the noise in the real and imaginary channels to be independently and identically distributed, then the real and imaginary parts of the voxel signal, which can be represented by the random (stochastic) variables $\underline{x}$, $\underline{y}$, follow a Gaussian (normal) distribution N, i.e. $\underline{x} \sim N(x_0, \sigma^2)$ and $\underline{y} \sim N(y_0, \sigma^2)$, where (x_0, y_0) are the true mean values and σ is the standard deviation (SD), i.e. the noise level. Underbars are used to denote random variables throughout the paper. Under these assumptions the joint PDF of the real and imaginary voxel signal is given as:

$$p_{\underline{x},\underline{y}}(x, y; x_0, y_0, \sigma) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(x - x_0)^2}{2\sigma^2} - \frac{(y - y_0)^2}{2\sigma^2}\right) \tag{1}$$

We represent the magnitude of the voxel signal by the random variable $\underline{m}$ and let m denote a specific realization of $\underline{m}$. It is straightforward to show [14] that $\underline{m}$ is *Rician* distributed [1] with PDF

$$p_{\underline{m}}(m; m_0, \sigma) = \frac{m}{\sigma^2} \exp\left(-\frac{m^2 + m_0^2}{2\sigma^2}\right) I_0\left(\frac{m m_0}{\sigma^2}\right) \Theta(m) \tag{2}$$

where I_0 is the zeroth order modified Bessel function of the first kind, Θ is the Heaviside step function and $m_0 = \sqrt{x_0^2 + y_0^2}$ is the true value of the magnitude. Eq. (2) is valid when signal is measured through a single receiver unit. The case of phased arrays has been discussed by Constantinides et al. [8]. When the true magnitude, m_0, is much greater than the noise level, σ, the Rician distribution approximates well a Gaussian distribution. When the true magnitude equals zero, the Rician distribution simplifies to a Rayleigh distribution.

The Rician distribution is shown in Fig. 1. In this figure and in what follows the term *normalized* is used for quantities that are measured in units of the noise level σ (i.e. m/σ is referred to as the *normalized* magnitude etc.). In the discussion of possible diffusion estimators, we will need a measure of the *center* of the distribution. The most obvious choice is the mean (the first moment), m_{MN}, which can

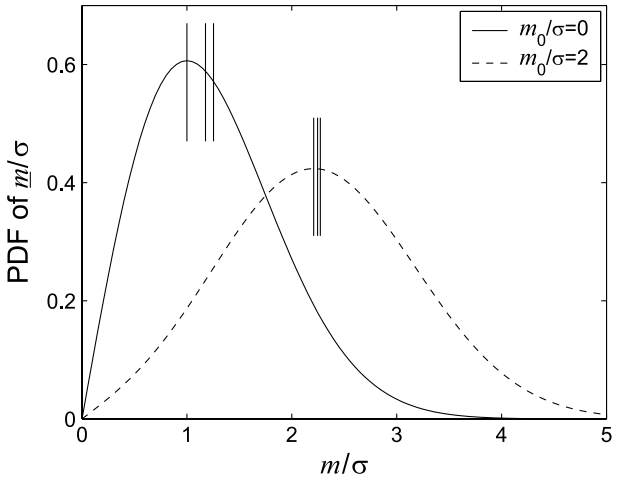

Fig. 1. The Probability Density Function (PDF) for the normalized magnitude $\underline{m}/\sigma$ (the Rician distribution). The values of the normalized true magnitude are $m_0/\sigma = 0$ and $m_0/\sigma = 2$. Left to right, vertical bars show the points of maximum probability, m_{MP}/σ, median, m_{MD}/σ, and mean, m_{MN}/σ, of the distributions.

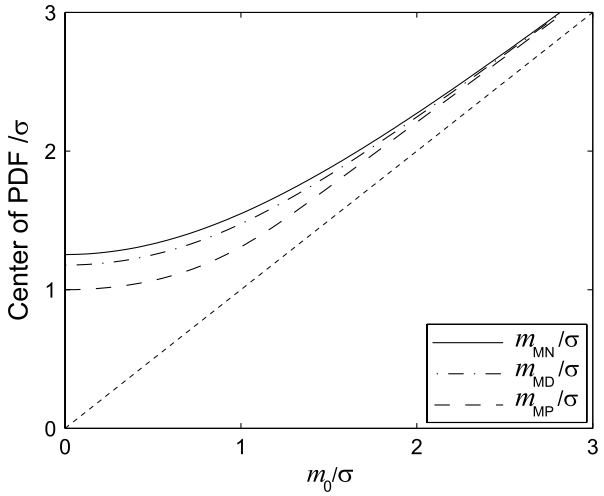

Fig. 2. Normalized measures of the center of the Rician distribution as a function of the normalized true magnitude m_0/σ: maximum probability, m_{MP}/σ, median, m_{MD}/σ and mean, m_{MN}/σ.

be expressed analytically [14] as a function of the true magnitude, m_0, as

$$m_{\mathrm{MN}}(m_0) = E[\underline{m}] = \sigma\sqrt{\frac{\pi}{2}}\,{}_1F_1\left(-\frac{1}{2}, 1, -\frac{m_0^2}{2\sigma^2}\right) \tag{3}$$

where ${}_1F_1$ is the confluent hypergeometric function [15] and $E[\cdots]$ denotes the expectation operator.

Since the distribution is non-symmetric, the mean is not the only possible measure of the center. Another measure is the median value, $m_{\mathrm{MD}}(m_0)$, which is defined by

$$\int_0^{m_{\mathrm{MD}}(m_0)} \mathrm{d}m\; p_{\underline{m}}(m; m_0, \sigma) = 1/2 \tag{4}$$

The median value cannot be expressed in closed form and must be calculated numerically. Yet another measure of the center is the maximum probability point, $m_{\mathrm{MP}}(m_0)$. Since there is only one unique maximum, $m_{\mathrm{MP}}(m_0)$ can be found by solving the equation

$$\left.\frac{\partial p_{\underline{m}}(m; m_0, \sigma)}{\partial m}\right|_{m=m_{\mathrm{MP}}(m_0)} = 0 \tag{5}$$

The three possible measures of the distribution center are indicated by vertical bars in Fig. 1 and plotted as a function of the normalized true magnitude, m_0/σ, in Fig. 2.

The variance of the Rician distribution can be expressed in closed form. Since $\underline{m}^2 = \underline{x}^2 + \underline{y}^2$, we can write

$$E[\underline{m}^2] = E[\underline{x}^2] + E[\underline{y}^2] = m_0^2 + 2\sigma^2 \tag{6}$$

where we have used the fact that the second moments of the normal distributions $N(x_0, \sigma^2)$ and $N(y_0, \sigma^2)$ are given by $E[\underline{x}^2] = x_0^2 + \sigma^2$ and $E[\underline{y}^2] = y_0^2 + \sigma^2$, respectively. The variance of $\underline{m}$ is therefore given by

$$\sigma_{\underline{m}}(m_0)^2 = E[\underline{m}^2] - E[\underline{m}]^2 = m_0^2 + 2\sigma^2 - m_{\mathrm{MN}}(m_0)^2 \tag{7}$$

The Rician distribution gets narrower as the true magnitude decreases.

2.2. Effect of averaging

In order to improve the SNR we can perform repeated measurements and average the data. Averaging can be done either prior to or after the magnitude operation.

Averaging prior to the magnitude operation can be done directly on the k-space (raw) data prior to the inverse Fourier transformation or one can average the complex image data. Owing to the linearity of the inverse Fourier transform, the result is the same. In either case the effect on the PDF of the image data is trivial as long as the true phase, φ_0, is equal in all measurements. The averaging of the real and imaginary parts yields mean values that remain unchanged, whereas the variance will be inversely proportional to the number of averages, N_{AV}. The distributions of the averaged real and imaginary parts, represented by $\underline{x}_{\mathrm{ave}}$, $\underline{y}_{\mathrm{ave}}$, remain Gaussian, $\underline{x}_{\mathrm{ave}} \sim N(x_0, \sigma^2/N_{\mathrm{AV}})$ and $\underline{y}_{\mathrm{ave}} \sim N(y_0, \sigma^2/N_{\mathrm{AV}})$ respectively. Hence the distribution of the magnitude data remains Rician and it is found by substituting $\sigma^2 \to \sigma^2/N_{\mathrm{AV}}$ in Eq. (2).

Averaging complex data is not possible in diffusion measurements due to the aforementioned problem of motion-induced phase shifts (unless reliable phase-correction algorithms can be applied). Instead, we have to average the magnitude data.

This motivates the analysis of the statistical properties of averaged Rician data. We let the data be represented by the random variable $\underline{m}_{\mathrm{ave}}$ defined by

$$\underline{m}_{\mathrm{ave}} = \frac{1}{N_{\mathrm{AV}}}\underline{m}_{\Sigma}, \quad \text{where} \quad \underline{m}_{\Sigma} = \sum_{k=1}^{N_{\mathrm{AV}}} \underline{m}_k \tag{8}$$

In this equation, k is a counter from 1 to the number of averages, N_{AV}, and $\underline{m}_k$ is a random variable that represents

the magnitude of the voxel signal in the kth excitation. Some conclusions can be drawn immediately. The expectation value of $\underline{m}_{\text{ave}}$ equals the expectation value of $\underline{m}_k$ (given in Eq. (3)), and since the variables $\{\underline{m}_k\}$ are assumed to be independent and identically distributed, Eq. (8) implies that $\text{var}(\underline{m}_{\text{ave}}) = \text{var}(\underline{m}_k)/N_{\text{AV}}$. Furthermore, since the mean and variance of the PDF $p_{\underline{m}_k}$ (Eq. (2)) are finite, the *central limit theorem* [14] assures that the PDF of $\underline{m}_{\text{ave}}$ approaches a Gaussian as $N_{\text{AV}} \to \infty$.

In order to address the general problem (where N_{AV} is too small to allow use of the aforementioned Gaussian approximation to describe the composite PDF) we can use the fact that the PDF of a sum of independent and identically distributed random variables is given by the convolution of the PDFs of each variable, i.e.

$$p_{\underline{m}_\Sigma} = p_{\underline{m}_1} \otimes p_{\underline{m}_2} \otimes \cdots \otimes p_{\underline{m}_{N_{\text{AV}}}} \tag{9}$$

Performing the necessary scaling, we can express the PDF of the average value as

$$p_{\underline{m}_{\text{ave}}}(m) = N_{\text{AV}}\, p_{\underline{m}_\Sigma}(N_{\text{AV}}\, m) \tag{10}$$

Since all $\underline{m}_k$ have the same Rician PDF (given in Eq. (2)), the PDF of the sum $\underline{m}_\Sigma$ is given by the N_{AV}-fold convolution of Eq. (2). Attempts have been made to express this convolution in closed form, but the presence of the modified Bessel function and the Heaviside step-function complicates this task. Even the simpler special case where the true magnitude equals to zero (which turns the PDF of $\underline{m}_k$ into the Rayleigh distribution) has been addressed in terms of approximate schemes such as saddle-point integration [16], or infinite summation [17]. For the full Rician distribution efficient numerical schemes have been proposed [18].

Using the fact that Fourier transformation turns the convolution operation into multiplication, we can rewrite Eq. (9) as

$$p_{\underline{m}_\Sigma} = F^{-1}\left\{\left(F\{p_{\underline{m}}\}\right)^{N_{\text{AV}}}\right\} \tag{11}$$

where F and F^{-1} denote Fourier transform and inverse Fourier transform, respectively. The Fourier transform $F\{p_{\underline{m}}\}$ of the distribution, which is equal to the expectation value of $\exp(\underline{m})$, is termed the *characteristic function*. After discretization of the function $p_{\underline{m}}$, Eq. (11) is readily evaluated by the fast Fourier transformation, and the PDF of the mean value $\underline{m}_{\text{ave}}$ is found by rescaling (Eq. (10)). Examples with number of averages $N_{\text{AV}} = 1$, 2, 3, 4 and normalized true magnitude $m_0/\sigma = 0,2$ are shown in Fig. 3. The case $m_0/\sigma = 0$ gives the largest possible asymmetry, but the Gaussian approximation is reasonably accurate after averaging only four times. In Fig. 4, we show normalized mean, median and maximum probability as functions of normalized true magnitude for the PDF of the averaged magnitude with $N_{\text{AV}} = 4$. Due to increased symmetry the median and maximum probability points are nearly confluent with the mean, even for small values of m_0/σ.

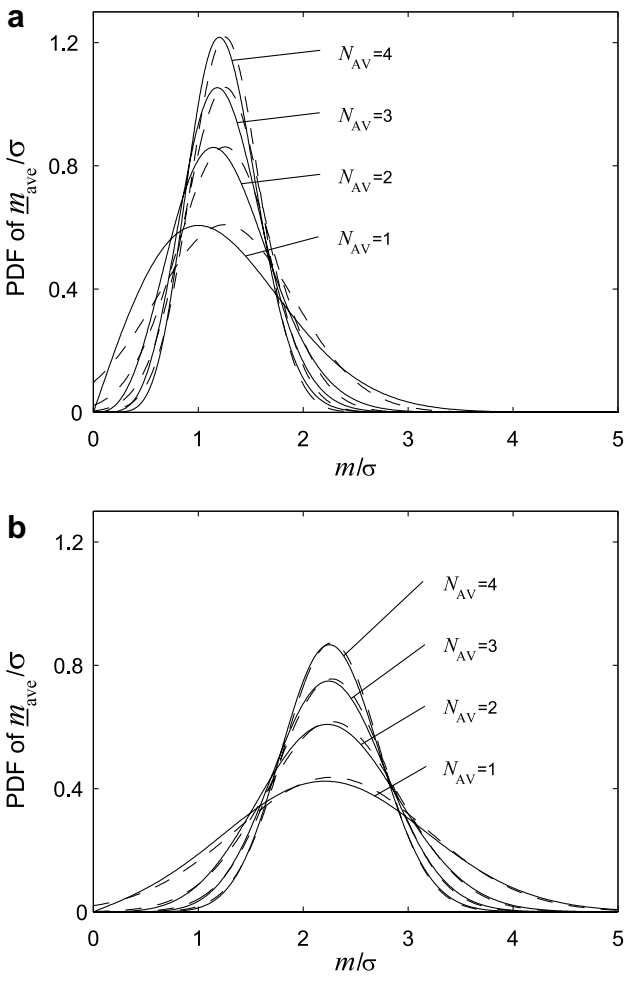

Fig. 3. The Probability Density Function (PDF) for the normalized averaged magnitude, $\underline{m}_{\text{ave}}/\sigma$ (solid lines) and the Gaussian approximation (dashed lines). The number of averages is indicated in the figure ($N_{\text{AV}} = 1,2,3,4$). The normalized true magnitude is (a): $m_0/\sigma = 0$ and (b): $m_0/\sigma = 2$.

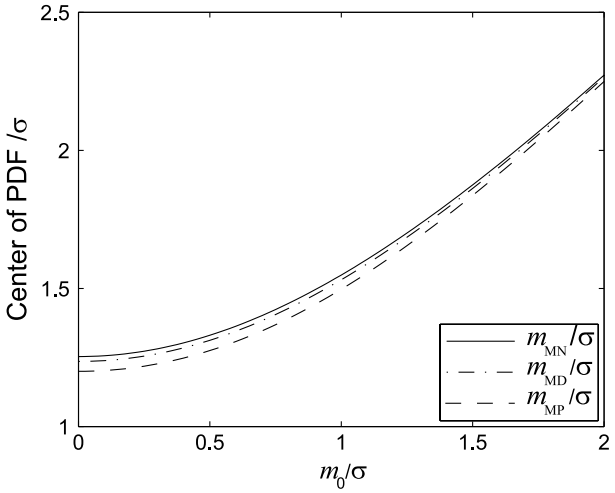

Fig. 4. Normalized measures of the distribution center for averaged magnitude data ($N_{\text{AV}} = 4$) as a function of the normalized true magnitude, m_0/σ: Maximum probability, m_{MP}/σ, median, m_{MD}/σ and mean, m_{MN}/σ.

2.3. Estimators of the diffusion coefficient

We want to estimate the diffusion coefficient, D, from an experiment in which the signal magnitude $m(b)$ is measured at several (at least two) values of the diffusion-weighting factor b. We assume a simple mono-exponential diffusion model, for which the true magnitude is given by:

$$m_0(b) = \rho \exp(-Db) \tag{12}$$

In this equation, and in what follows, ρ is used to denote the true magnitude in the absence of diffusion weighting (whereas m_0 is used for the true magnitude in a general magnitude MRI image).

In the following sections we evaluate the performance of several estimators. For this purpose we employ some concepts from the theory of statistical parameter estimation. Based on an observation vector $\mathbf{y}$ (in our case $\mathbf{y} = [m_1, m_2, ..., m_{N_b}]$, where N_b denotes the number of diffusion weighting factors) we want to find the underlying physical model parameters $\boldsymbol{\theta}$ (in our case $\boldsymbol{\theta} = [D, \rho]$). The observation vector can be viewed as a specific realization of a random variable $\underline{\mathbf{y}}$.

In order to find the model parameters we construct an *estimator*, $\hat{\underline{\boldsymbol{\theta}}}$, which is a function of $\underline{\mathbf{y}}$, and hence it is itself a random variable. (''Hat'' is used to denote estimators throughout the rest of the paper.) A given realization of the estimator is called an *estimate*. The performance of an estimator can be judged in terms of its *accuracy* and *precision*. High accuracy means that the bias is low. The bias is defined as the difference between the expectation value of the estimator and the true value of the model parameter. The precision is measured by the standard deviation (or equivalently the variance) of the estimator. In the following section we shall discuss some of the possible estimators of the diffusion coefficient.

2.3.1. Linear regression

The diffusion coefficient can be estimated by a linear fit to the logarithm of the *uncorrected* signal intensities. The advantage of the linear regression (LR) method is the fact that no numerical optimization is necessary. This gives outstanding computational efficiency. However, bias correction cannot be implemented with this method.

2.3.2. Non-linear curve fitting

We can fit a function, f, that depends on the parameters to be estimated to the observed data. This is done by minimizing the *p-norm* of the difference between the function and the data,

$$L_p(\boldsymbol{\theta}) = \sum_{k=1}^{N_b} |m_k - f(b_k, \boldsymbol{\theta})|^p \tag{13}$$

Strictly speaking, the p-norm is given by this sum raised to the power $1/p$, but this does not alter the location of the minima of L_p with respect to $\boldsymbol{\theta}$. The case $p = 2$ is equivalent to least squares minimization. Alternatively we can minimize the noise-weighted sum, L_p^w, defined by:

$$L_p^w(\boldsymbol{\theta}) = \sum_{k=1}^{N_b} |(m_k - f(b_k, \boldsymbol{\theta}))/\sigma_k|^p \tag{14}$$

For Gaussian data, L_2^w is χ^2 distributed. For this special case it can be shown that minimization of L_2^w is equivalent to maximum likelihood estimation of the parameters $\boldsymbol{\theta}$. In general, however, least squares minimization is not necessarily the optimal choice. In fact, if a dataset contains outliers, it may be better to minimize the sum of absolute values ($p = 1$) [19]. Weighting requires a priori knowledge of the variance σ_k^2. For Rician data, we do not have this information. The variance depends on the true magnitude m_0, which is unknown since it depends on the parameters that we want to estimate, see Eq. (7).

The simplest possible choice of the function $f(b_k, D, \rho)$ is

$$f(b_k, D, \rho) = m_{0,k} = \rho \exp(-Db_k) \tag{15}$$

This means that we minimize the difference between the measured magnitude and the true magnitude, without any correction for the noise bias. We denote this uncorrected estimator by UC.

Rather than minimizing the difference between the measured magnitude and the true magnitude, we can minimize the difference between the measured magnitude signal and the center of the Rician distribution. In this way we take the noise bias into account. As discussed, there are at least three ways to define the center of the distribution. Hence we have the following possible choices for the fitting function $f(b_k, D, \rho)$:

$$f(b_k, D, \rho) = m_{\mathrm{MN}}(m_{0,k}) \tag{16}$$

$$f(b_k, D, \rho) = m_{\mathrm{MD}}(m_{0,k}) \tag{17}$$

$$f(b_k, D, \rho) = m_{\mathrm{MP}}(m_{0,k}) \tag{18}$$

where $m_{0,k}$ is given in Eq. (15) and $m_{\mathrm{MN}}(m_0)$, $m_{\mathrm{MD}}(m_0)$ and $m_{\mathrm{MP}}(m_0)$ are defined in Eqs. (3)–(5), respectively. The corresponding estimators are denoted by MN, MD and MP respectively. Mean, median and maximum probability estimation has earlier been used for single-point bias elimination in the context of radio astronomy [20].

In Fig. 5, we show an example of non-linear curve fitting to a Rician dataset where we employ the fitting functions given in Eqs. (16)–(18). Note that we fit directly to the measured dataset. Alternatively, we could have corrected the bias at each individual data point prior to the curve fitting. For the MN method this would amount to solving the equation $m = m_{\mathrm{MN}}(m_0)$ for m_0, where m is the measured magnitude. However, this would introduce ambiguities if $m < m_{\mathrm{MN}}(0)$.

The generalization of MN, MD and MP estimation to the case of averaged data is straightforward. We now fit to the mean, median or maximum probability of the PDF for $\underline{m}_{\mathrm{ave}}$, Eq. (10). When the number of averages, N_{AV}, gets large, the median and maximum probability points are nearly confluent with the mean even for small

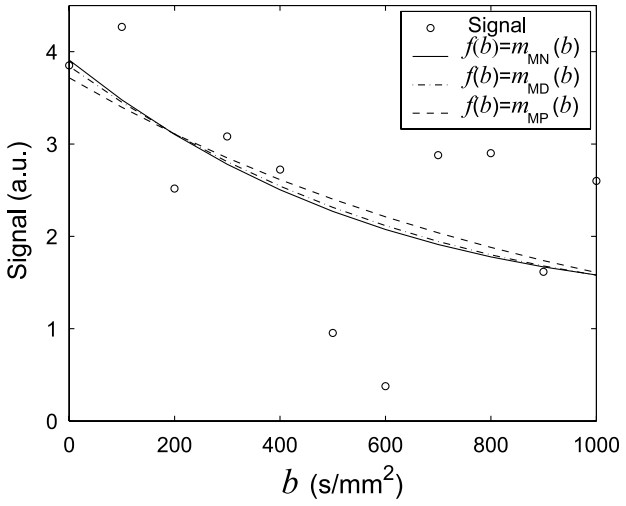

Fig. 5. Non-linear curve fitting to a Rician dataset ("Signal") using the MN, MD and MP methods (Eqs. (16)–(18), respectively). The true diffusion coefficient is $D = 1 \times 10^{-3}$ mm^2/s, the diffusion weighting factors are $\mathbf{b} = [0, 100, 200, \ldots, 1000]$ s/mm^2 and the SNR is $\rho/\sigma = 3$. The diffusion coefficient estimates obtained with this particular dataset are: MN, 1.27×10^{-3} mm^2/s; MD, 1.16×10^{-3} mm^2/s and MP, 0.97×10^{-3} mm^2/s.

values of the true magnitude, m_0. Hence the specific choice between the MN, MD or MP estimators is of minor importance in this case.

2.3.3. Power images

Miller and Joseph [3] and McGibney and Smith [4] independently proposed a bias correction scheme based on power images (PI). Due to the simple expression for the expectation value of the squared magnitude (Eq. (6)), we can introduce a variable $\underline{\mu} = \underline{m}^2 - 2\sigma^2$ that has the expectation value $E[\underline{\mu}] = m_0^2$, i.e. the true value of the squared magnitude. The true squared magnitude equals $\rho^2 \exp(-2Db)$, and the least squares PI estimate is the set of parameters (D,ρ) that minimizes the objective function.

$$L_2(D,\rho) = \sum_{k=1}^{N_b} |\mu_k - \rho^2 \exp(-2Db_k)|^2 \tag{19}$$

At high SNR, the power image method could be implemented as a linear fit to the *logarithm* of the corrected dataset [3]. However, this is not possible at low SNR, where corrected datasets occasionally contain negative numbers. The logarithm of the dataset would then contain imaginary numbers and the estimated diffusion coefficient would be complex valued.

The power image procedure cannot be applied directly to an averaged magnitude dataset. However, if we keep the N_{AV} individual datasets, we can calculate the averaged *squared* magnitude. For this dataset, we can correct the bias by subtracting $2\sigma^2$.

2.3.4. High SNR approximation to the mean of the Rician distribution

At high SNR, the mean of the Rician distribution can be approximated by $m_{MN}(m_0,\sigma) \approx \sqrt{m_0^2 + \sigma^2}$. This simple expression can be substituted for m_{MN} in Eq. (16) in order to estimate the diffusion coefficient. This approach henceforth is referred to as the High SNR (HS) approximation. It has been discussed in the context of diffusion coefficient estimation by Dietrich et al. [12] and by Jones and Basser [21].

Gudbjartsson and Patz [5] suggested a scheme inspired by the HS approximation. The bias corrected magnitude m_0 is estimated as

$$m_0 = \sqrt{|m^2 - \sigma^2|} \tag{20}$$

where m is the measured magnitude. In order to apply this method to averaged data, we need to store the N_{AV} individual datasets.

2.3.5. Maximum likelihood estimation

With a given set of model parameters, $\boldsymbol{\theta}$, we can calculate the probability, $P_{\underline{\mathbf{y}}}(\mathbf{y};\boldsymbol{\theta})$, of a certain outcome, $\mathbf{y}$, of an experiment. In the context of parameter estimation, the problem is reversed, i.e. we want to estimate the model parameters from a given set of observations. We then use the term *likelihood*, $L(\boldsymbol{\theta};\mathbf{y})$, rather than *probability*. The likelihood hypothesis states that these quantities are proportional, i.e.

$$L(\boldsymbol{\theta};\mathbf{y}) \propto P_{\underline{\mathbf{y}}}(\mathbf{y};\boldsymbol{\theta}) \tag{21}$$

and parameter estimation amounts to maximizing the likelihood function (or equivalently the logarithm of the likelihood function) with respect to the model parameters. The maximum likelihood (ML) estimator is of special interest due to favorable asymptotic properties. It is asymptotically unbiased, i.e. unbiased as the sample size (the length of the vector $\mathbf{y}$) approaches infinity, and it can be shown to have variance equal to the Cramér-Rao lower bound in this limit [22].

The maximum likelihood estimator has been discussed by Bonny et al. [6] and by Sijbers et al. [10] in the context of T_2 estimation, by Karlsen et al. [11] in the context of T_1 and perfusion measurements, and recently by Sijbers and den Dekker [13] in the context of signal intensity and noise variance estimation.

Combining Eqs. (2), (12), (21), we can express the logarithm of the likelihood function as

$$\ln(L) = C - \sum_{k=1}^{N_b} \left\{ \frac{\rho^2 \exp(-2Db_k)}{2\sigma^2} - \ln\left[I_0\left(\frac{m_k \rho \exp(-Db_k)}{\sigma^2} \right) \right] \right\} \tag{22}$$

where m_k is the measured magnitude at the diffusion weighting b_k, and C is a constant that includes all terms independent of D and ρ. The maximum likelihood parameter estimates (given the observations $\{m_k\}$) are the values of D and ρ that maximize Eq. (22).

With averaged data the PDF cannot be expressed in closed form and we must adopt a "brute-force" numerical strategy in order to perform ML estimation of the diffusion coefficient. This is explained in Appendix A.

We can however adopt a simplified two-step approach by which we use ML principles to correct the bias of the signal intensities prior to least-squares parameter estimation. Sijbers and den Dekker [13] recently discussed the problem of ML estimation of true magnitude in homogeneous image regions consisting of several pixels (meaning that the results are applicable for averaged data as well). They found that estimation based on complex data yielded better results than estimation based on magnitude data if the true phases in the pixels were equal. As argued above, this is usually not the case for diffusion images due to motion-induced phase shifts. Hence we have to use the N_{AV} measured magnitudes to estimate the true magnitude m_0. This is achieved by maximizing the following expression with respect to m_0:

$$\ln(L) = C - \sum_{j=1}^{N_{\mathrm{AV}}} \left\{ \frac{m_0^2}{2\sigma^2} - \ln\left[I_0\left(\frac{m_j m_0}{\sigma^2}\right)\right]\right\} \tag{23}$$

where j is a counter from 1 to the number of averages, N_{AV}. Having done this at all N_b values of the diffusion-weighting factor, we end up with a bias corrected dataset.

In the limit $N_{\mathrm{AV}} \to \infty$ the PDF of averaged Rician data approaches a Gaussian PDF. Hence, the variance-weighted MN estimator, (Eqs. (14) and (16)), is equivalent to the ML estimator.

3. Methods

In order to evaluate the performance of the estimators we perform Monte Carlo (MC) simulations using datasets with one and four averages. The *Theory* section describes how the estimates are calculated. Except for the LR estimator, it is necessary to perform non-linear optimization. For the MD and MP estimators we calculate m_{MD} and m_{MP} as functions of m_0 numerically (Eqs. (4) and (5), respectively). The evaluation of these functions is then done by interpolation in a one-dimensional lookup table. This allows a very fast evaluation of the cost function. The ML estimation for averaged data is somewhat more involved. Details are given in Appendix A.

Non-linear optimization is required for all estimators except LR. We used a built-in function of Matlab® (The Mathworks Inc., Natick, MA) that implements the Nelder-Mead direct-search simplex method [23]. This algorithm is efficient in the presence of a low dimensional parameter space. Thus it is well suited for our problem, where we only have to minimize with respect to two parameters, ρ and D.

At very low SNR ($\rho/\sigma < 3$), the algorithm occasionally converges towards very large or small values of D, but the point to which it converges proves not to be a minimum, but rather a point where the cost function is flat with respect to D. This produces outliers that may severely contaminate the mean and variance of the Monte Carlo (MC) sample. This problem has to be solved in order to evaluate the estimators at low SNR, which is our main interest. In order to ensure that the point to which the algorithm converges really is a minimum, we calculate the Hessian matrix (i.e. the matrix of second derivatives), and accept the point only when the Hessian proves to be positive definite, i.e. when both eigenvalues are positive (above a cut-off value of 10^{-5}). Details about the calculation of the Hessian matrix are given in Appendix B.

If the search algorithm fails (either because it does not converge or because the Hessian is not positive definite) we alter the starting values and repeat the search. If a minimum cannot be found in fifty tries we reject the dataset.

Having run the MC simulations we can estimate the PDF for each estimator. The properties of the approximate PDF can be expressed by its moments. If the distribution is normal, all information is contained in the first two moments (mean an variance), but in general there can also be unique information in higher order moments. However, we shall only consider the sample mean and sample variance.

If the number of rejected datasets is very large we cannot use the sample mean and sample variance to determine the true mean and variance of the estimator. The bias and variance of the MC sample will be too low since rejected datasets usually correspond to extreme values of the fitting parameters ρ and D. At the rather low SNR of $\rho/\sigma = 2$ on the order of 1% of the datasets were rejected. With $\rho/\sigma \geqslant 4$ no datasets were rejected.

4. Results and discussion

We have run MC simulations for non-averaged ($N_{\mathrm{AV}} = 1$) and averaged ($N_{\mathrm{AV}} = 4$) data. In the former case the MC sample size was 10000, in the latter it was 5000. We calculate sample mean and SD for the estimators listed in Table 1. With the term SNR we refer to the ratio ρ/σ, i.e. the normalized true magnitude in the absence of diffusion weighting.

In Fig. 6, the parameters are $D = 1 \times 10^{-3}\,\mathrm{mm^2/s}$, $\mathbf{b} = [0, 100, 200, \ldots, 1000]\,\mathrm{s/mm^2}$, the SNR ρ/σ is in the range 2–10 and $N_{\mathrm{AV}} = 1$. The HS estimator performs best in terms of bias, particularly at very low SNR. The MP estimator performs better than the other bias-correcting estimators in terms of SD. The performance of the PI estimator is poor in terms of bias as well as SD. The LR estimator performs better than UC in terms of bias. This may seem surprising, but it is readily understood from the following argument: If the random variable $\underline{m}$ has the expectation value $E[\underline{m}] = \mu$ then the expectation value of the random variable $\ln(\underline{m})$ will be lower than $\ln(\mu)$. The noise, on the other hand, introduces a positive bias for the signal intensities. Hence the two effects cancel partially. The SD of the uncorrected estimators (LR and UC) is very low compared to the bias-correcting estimators at SNR lower than 4. This is not surprising since the bias of these estimators makes the estimates assemble in a narrow region close to zero as the SNR approaches zero. At SNR above 5 the SD of the LR estimator is higher, as expected.

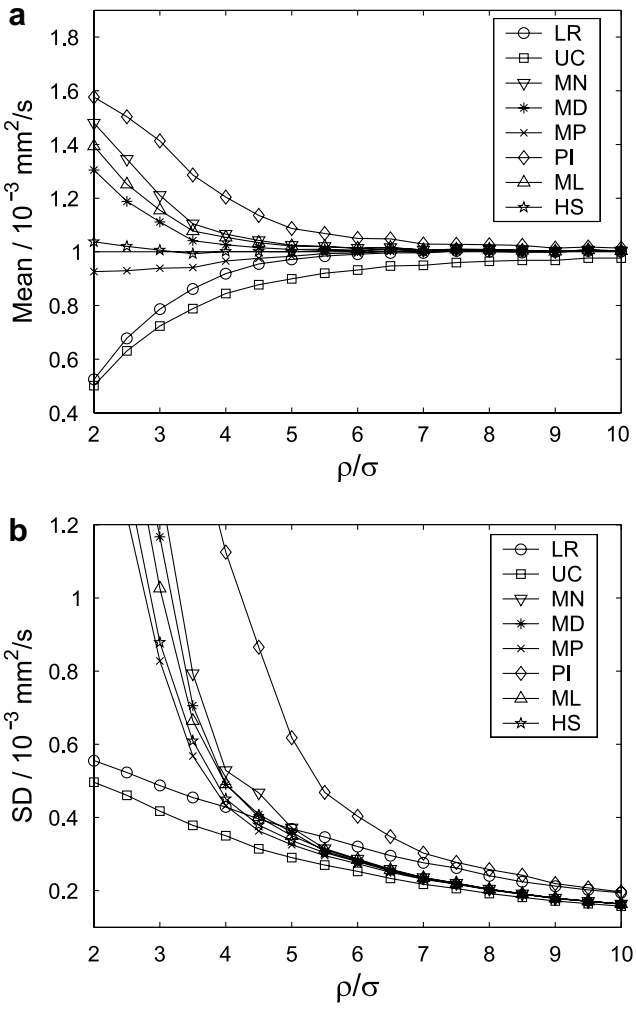

Fig. 6. Sample mean (a) and standard deviation (SD) (b) versus SNR with true diffusion coefficient $D = 1 \times 10^{-3}$ mm^2/s and diffusion weighting factors $\mathbf{b} = [0, 100, 200, \ldots, 1000]$ s/mm^2. The abbreviations in the legend are given in Table 1.

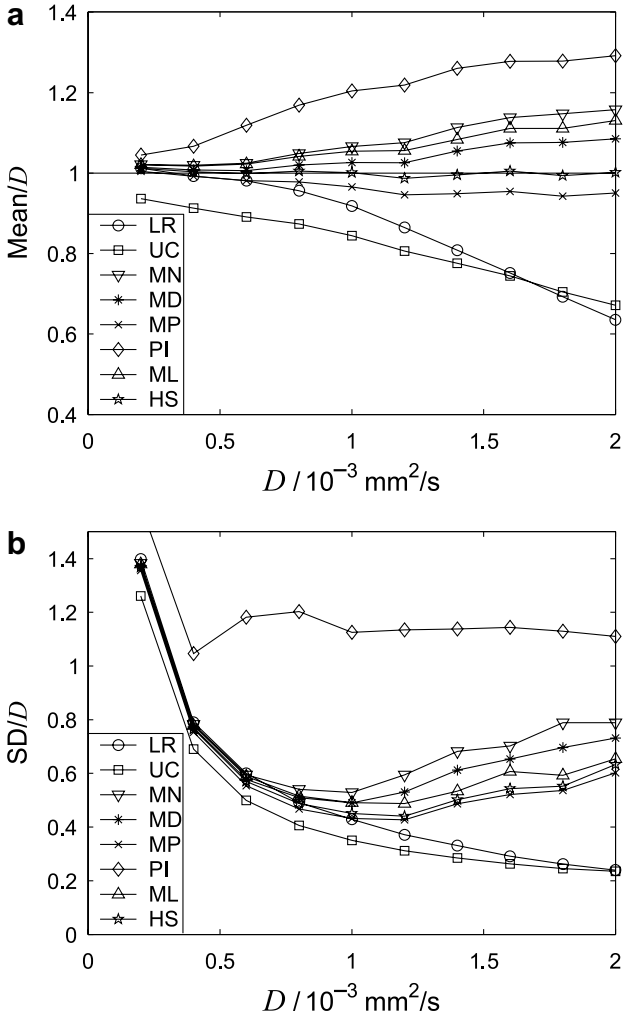

Fig. 7. Data plotted as a function of the true diffusion coefficient with a fixed SNR value of $\rho/\sigma = 4$. The figure shows the ratio of the sample mean to the true value of the diffusion coefficient (a) and the ratio of the sample standard deviation (SD) to the true value of the diffusion coefficient (i.e. the relative uncertainty) (b). The diffusion weighting factors are $\mathbf{b} = [0, 100, 200, \ldots, 1000]$ s/mm^2. The abbreviations in the legend are given in Table 1.

In Fig. 7, we show data plotted as a function of the true diffusion coefficient with a fixed SNR of $\rho/\sigma = 4$. We plot the ratio of the sample mean to the true value of the diffusion coefficient and the ratio of the sample SD to the true value of the diffusion coefficient (i.e. the relative uncertainty). The diffusion weighting factors are $\mathbf{b} = [0, 100, 200, \ldots, 1000]$ s/mm^2. The D-range ($0.2 - 2 \times 10^{-3}$ mm^2/s) covers the values that are typically encountered in brain tissue. The HS estimator has the best overall performance in terms of bias, whereas MP performs somewhat better in terms of SD. The PI estimator performs poorly over the entire range. The LR estimator has a lower bias than UC for $D < 1.7 \times 10^{-3}$ mm^2/s. This is due to the aforementioned partial noise bias compensation introduced by the logarithm operation. Furthermore we observe that the relative uncertainties (SD/D) of the MN, MD, MP, ML and HS estimators are minimal in the range $D = 0.7 \times 10^{-3}$–1.4×10^{-3} mm^2/s. This expresses the fact that the estimators have a D-range in which the performance is optimal. The *sampling scheme* (specific distribution of the diffusion weighting factors) could be optimized rigorously by minimizing the Cramér-Rao lower bound [11,24]. However, this is beyond the scope of the present paper, where our main intention is to compare the performance of different estimators.

Fig. 8 shows the effect of reducing the number of diffusion weighting factors. We now have $\mathbf{b} = [0, 200, 400, 600, 800, 1000]$ s/mm^2. The diffusion coefficient is $D = 1 \times 10^{-3}$ mm^2/s, like in Fig. 6, but the lowest SNR value is now 3 rather than 2. The reason is that the results get unreliable at low SNR due to a larger number of rejected datasets. In this figure the MP estimator is superior, both in terms of bias and SD.

We have run simulations (data not shown) in order to investigate the effect of altering the norm from L_2 to L_1 (see Eq. (13)) for the estimators UC, MN, MD and MP. All estimators had a larger SD when we used the L_1 norm.

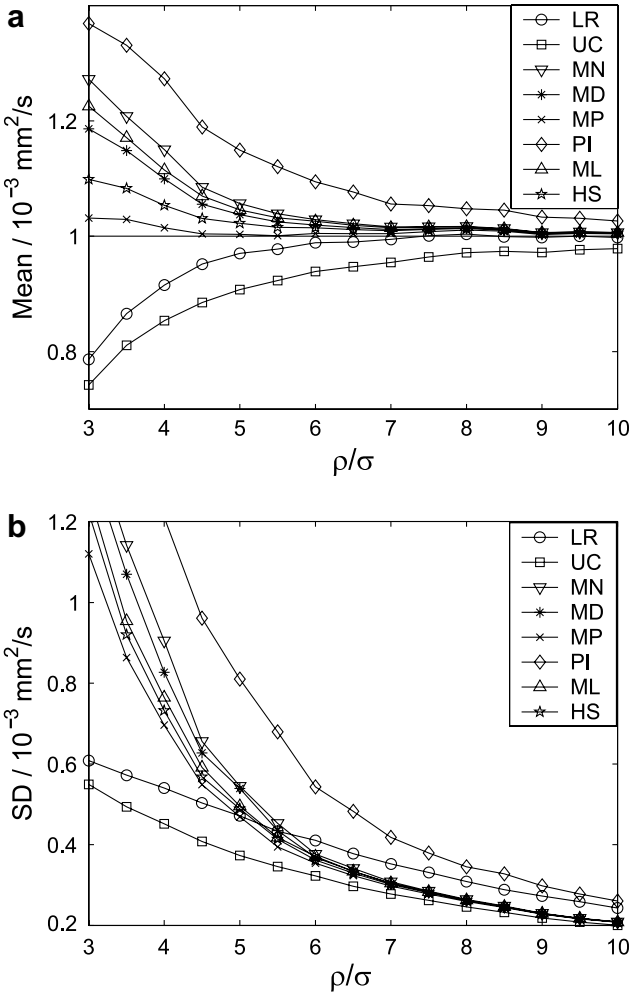

Fig. 8. Effect of altering the number of diffusion weighting factors. The figure shows the sample mean (a) and standard deviation (SD) (b) versus SNR with true diffusion coefficient $D = 1 \times 10^{-3}$ mm^2/s and diffusion weighting factors $\mathbf{b} = [0, 200, 400, 600, 800, 1000]$ s/mm^2. The abbreviations in the legend are given in Table 1.

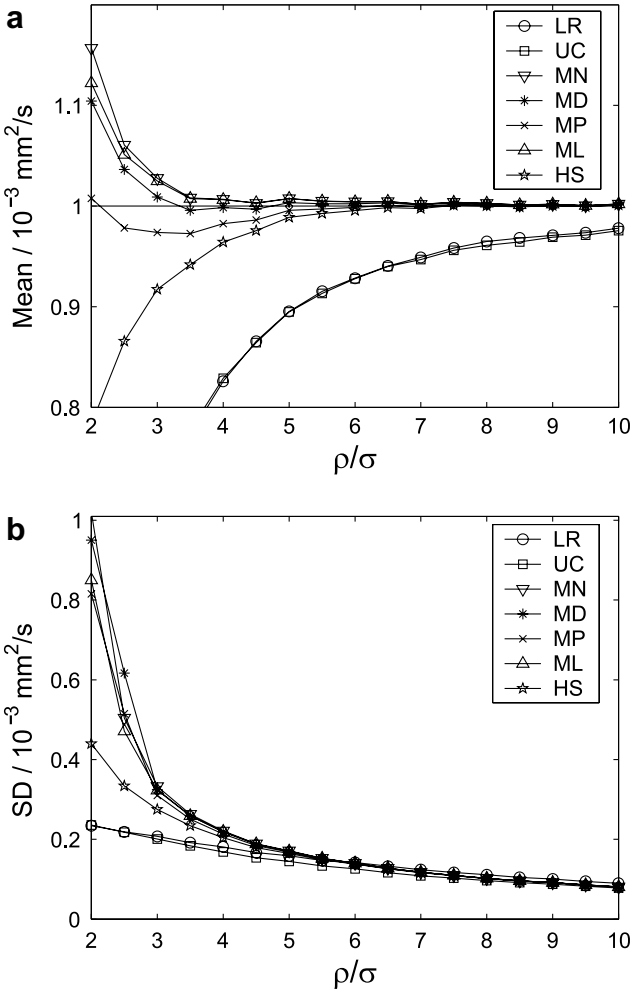

Fig. 9. Averaged data with $N_{AV} = 4$. The figure shows sample mean (a) and standard deviation (SD) (b) versus SNR with true diffusion coefficient $D = 1 \times 10^{-3}$ mm^2/s and diffusion weighting factors $\mathbf{b} = [0, 100, 200, \ldots, 1000]$ s/mm^2. Note that direct PI estimation is not possible for averaged data. The abbreviations in the legend are given in Table 1.

The bias was reduced for UC and MP and increased for MN and MD. Overall L_2 seems to be the better choice.

We have also compared the results (data not shown) obtained by minimizing an unweighted versus a weighted sum of squares (Eqs. (13) and (14), respectively). The effect on the UC estimator was marginal. For MN, MD and MP there was a slight reduction in SD. For MN and MD the bias was reduced and for MP it was increased. As discussed in Section 2, weighting requires a priori knowledge of the variance. We do not have this information, see Eq. (7), but it might be possible to solve the problem by an iterative scheme. However, the benefits of weighting seem to be too small to justify such a time-consuming approach.

In Fig. 9, we show results for averaged data with $N_{AV} = 4$. The diffusion coefficient is $D = 1 \times 10^{-3}$ mm^2/s, the diffusion weighting factors are $\mathbf{b} = [0, 100, 200, \ldots, 1000]$ s/mm^2 and the SNR ρ/σ is in the range 2–10. The uncorrected estimators, LR an UC, are seen to perform poorly in terms of bias, even at high SNR. Among the bias-correcting estimators, MD performs best in terms of bias, whereas MN, MD, MP and ML are quite equal in terms of SD. The simple HS estimator has very large bias, in particular at low SNR.

As explained in Section 2, it is possible to apply a simplified two-step ML procedure for averaged data, where we correct the bias at each one of the N_b data points prior to least squares fitting. By keeping the N_{AV} individual datasets we can also employ the PI method or the scheme proposed by Gudbjartsson and Patz [5] (Eq. (20)). We have run simulations (data not shown) to compare the performance of these methods to MN, MD, MP and full-ML estimation. The simplified versions proved to be significantly inferior in terms of bias and in terms of SD.

Figs. 10 and 11 display the results for averaged data plotted as a function of the true diffusion coefficient with a fixed SNR value of $\rho/\sigma = 4$. The diffusion weighting factors are $\mathbf{b} = [0, 100, 200, \ldots, 1000]$ s/mm^2 and $\mathbf{b} =$

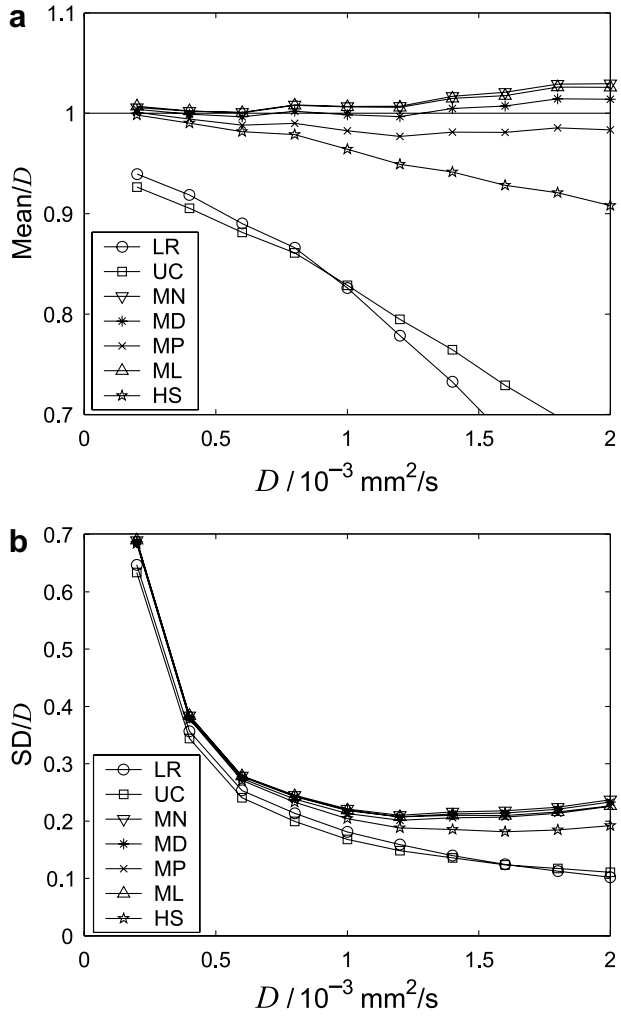

Fig. 10. Averaged data with $N_{AV} = 4$ plotted as a function of the true diffusion coefficient with a fixed SNR value of $\rho/\sigma = 4$. The figure shows the ratio of the sample mean to the true value of the diffusion coefficient (a) and the ratio of the sample standard deviation (SD) to the true value of the diffusion coefficient (i.e. the relative uncertainty) (b). The diffusion weighting factors are $\mathbf{b} = [0, 100, 200, \ldots, 1000]$ s/mm^2. The abbreviations in the legend are given in Table 1.

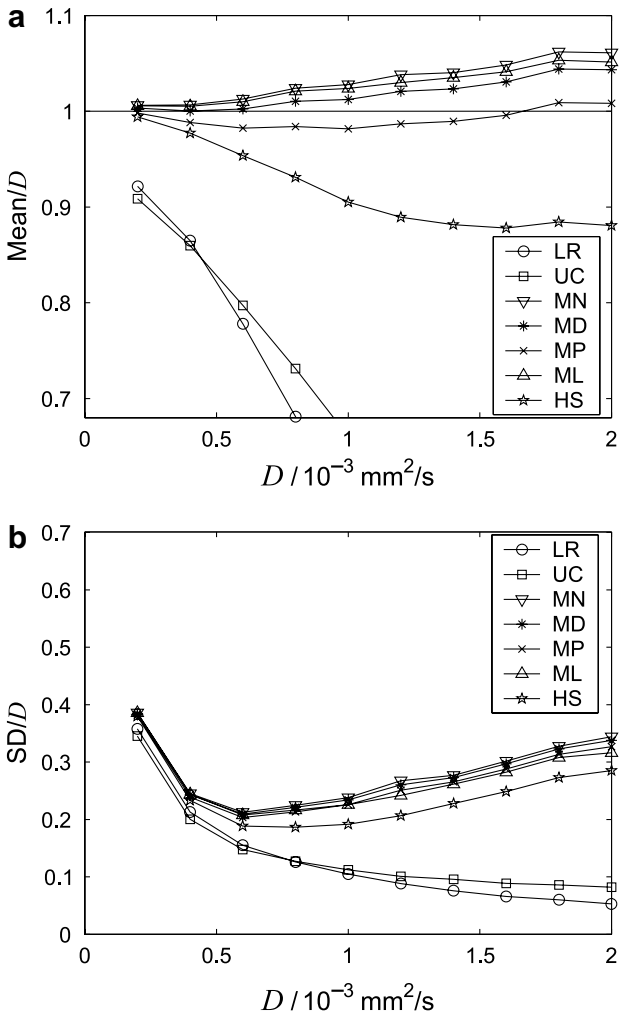

Fig. 11. Effect of altering the total range of the diffusion weighting factors. The figure shows averaged data with $N_{AV} = 4$ plotted as a function of the true diffusion coefficient. The parameters are the same as in Fig. 10, but the diffusion weighting factors are $\mathbf{b} = [0, 200, 400, \ldots, 2000]$ s/mm^2. The abbreviations in the legend are given in Table 1.

$[0, 100, 200, \ldots, 1000]$ s/mm^2, respectively. By using a maximum b-value of 2000 s/mm^2 we obtain a reduction in the overall SD (in particular for the slowest diffusion components), at the expense of a slight increase in bias. The relative uncertainty (SD/D) is at a minimum for $D \sim 1.2 \times 10^{-3}$ mm^2/s in Fig. 10 and $D \sim 0.6 \times 10^{-3}$ mm^2/s in Fig. 11. This expresses the aforementioned fact (see Fig. 7) that the estimators have a D-range in which the performance is optimal. In DTI we need to resolve slow components, so choosing the range $\mathbf{b} = [0, 200, 400, \ldots, 2000]$ s/mm^2 seems preferable for this purpose.

In Table 2, we show numerical results for non-averaged and averaged data that were obtained with SNR $\rho/\sigma = 4$ and true diffusion coefficient $D = 1 \times 10^{-3}$ mm^2/s. For non-averaged data the optimal choice of estimator is MP or HS, whereas LR, UC and PI should not be used. For averaged data MD or ML seems preferable, whereas LR, UC and HS should be avoided.

In the simulations, we assumed the noise level, σ, to be known. In a real diffusion imaging experiment σ is estimated from regions of interest that contain only noise pixels [5,13,25]. Our experience shows that this can be done with a precision better than 5%. We have tested the sensitivity of the algorithms with respect to the σ estimate. With an error of 5% and parameters $D = 1 \times 10^{-3}$ mm^2/s, $\mathbf{b} = [0, 100, 200, \ldots, 1000]$ s/mm^2 and $\rho/\sigma = 4$, we find the following effect for non-averaged data: MN, 3.0%; MD, 2.4%; MP, 1.2%; PI, 3.0%; ML, 3.1% and HS, 1.8%. Hence MP and HS are the most robust algorithms with respect to errors in the noise estimate. For averaged data with $N_{AV} = 4$ we find the following effect: MN, 2.5%; MD, 2.4%; MP, 2.1%; ML, 2.6% and HS, 1.6%.

Our simulations indicate that the MD and MP estimators perform better than ML and MN in terms of bias. This may seem surprising. The ML estimator is known to be

Table 2
Estimates and standard deviations (SD) obtained at signal-to-noise ratio $\rho/\sigma = 4$

Estimator	Non-averaged				Averaged			
	Dense sampling		Sparse sampling		Narrow range		Wide range	
	Estimate	SD	Estimate	SD	Estimate	SD	Estimate	SD
LR	0.92	0.43	0.92	0.54	0.83	0.18	0.58	0.10
UC	0.84	0.35	0.85	0.45	0.83	0.17	0.66	0.11
MN	1.07	0.53	1.15	0.91	1.01	0.22	1.03	0.24
MD	1.03	0.49	1.10	0.83	1.00	0.22	1.01	0.23
MP	0.97	0.43	1.01	0.70	0.98	0.21	0.98	0.23
PI	1.20	1.13	1.27	1.21	n/a	n/a	n/a	n/a
ML	1.05	0.49	1.11	0.77	1.01	0.22	1.02	0.23
HS	1.00	0.45	1.05	0.73	0.96	0.20	0.91	0.19

The results are given in units of 10^{-3} mm^2/s. The true diffusion coefficient is $D = 1 \times 10^{-3}$ mm^2/s. For non-averaged data, the diffusion weighting factors were $\mathbf{b} = [0, 100, 200, \ldots, 1000]$ s/mm^2 (dense sampling) and $\mathbf{b} = [0, 200, 400, \ldots, 1000]$ s/mm^2 (sparse sampling). For averaged data, the diffusion weighting factors were $\mathbf{b} = [0, 100, 200, \ldots, 1000]$ s/mm^2 (narrow range) and $\mathbf{b} = [0, 200, 400, \ldots, 2000]$ s/mm^2 (wide range). For non-averaged data, the optimal choice of estimator is MP or HS. LR, UC and PI should not be used. For averaged data, MD or ML seems to be a good choice, whereas LR, UC and HS should be avoided.

asymptotically unbiased, i.e. unbiased as the total number of measurements, $N_{\mathrm{AV}} \times N_b$, approaches infinity. In this limit ML can also be shown to have the lowest SD (the Cramér-Rao lower bound). The MN estimator is also asymptotically unbiased. This is readily understood for averaged data in the limit $N_{\mathrm{AV}} \to \infty$, where the averaged intensity at each b-value approaches m_{MN} (Eq. (3)). Furthermore, we also have $m_{\mathrm{MD}} \to m_{\mathrm{MN}}$ (Eq. (4)) and $m_{\mathrm{MP}} \to m_{\mathrm{MN}}$ (Eq. (5)) in this limit. Hence the MD and MP estimators are also asymptotically unbiased for $N_{\mathrm{AV}} \to \infty$.

For non-averaged data, we can give a heuristic argument to substantiate the fact that the MN estimator must be asymptotically unbiased. If we consider averaged data where N_{AV} as well as N_b are large, then (as argued above) the MN estimator is unbiased. If we distribute the $N_b \times N_{\mathrm{AV}}$ measurements evenly along the b-axis, the estimated diffusion coefficient should not be altered too much. Hence the MN estimator must be unbiased for non-averaged data as well. From the same line of reasoning the MD and MP estimators must now be asymptotically *biased*.

In order to verify the heuristic argument we ran a simulation with very dense sampling, $\mathbf{b} = [0, 1, 2, \ldots, 1000]$ s/mm^2, with SNR $\rho/\sigma = 3$ and $D = 1 \times 10^{-3}$ mm^2/s. The results were (mean/SD in units 10^{-3} mm^2/s) MN, 1.004/0.072; ML, 1.004/0.072; MD, 0.94/0.066 and MP, 0.84/0.056. Indeed we confirm that MN and ML are asymptotically unbiased, whereas MD and MP are asymptotically biased. MD is a more "conservative" choice than MP in the sense that the asymptotic bias is lower. The sample size required for MN and ML to be unbiased is too large to be encountered in clinical situations. Our simulations are in favor of MD or MP for limited samples.

The question of the optimal estimation of the diffusion tensor is interesting and has recently been discussed by Koay et al. [26]. An alternative procedure for estimating the diffusion tensor might be to maximize the joint likelihood function for the entire tensor, $L(\boldsymbol{\theta};\mathbf{y})$, with $\boldsymbol{\theta} = [\rho/\sigma, D_{11}, D_{22}, D_{33}, D_{12}, D_{13}, D_{23}]$, and a measurement vector $\mathbf{y}$ that includes the observed magnitudes for all directions and all N_b values of the diffusion weighting factor. We have concluded earlier that one-step ML estimation of (ρ, D) yielded better results than the two-step procedure by which we used ML principles to correct the bias in the measurements at each value of the diffusion-weighting factor prior to least squares fitting. This suggests that simultaneous ML estimation of the entire diffusion tensor might be the optimal procedure. However, this requires optimization in a seven-dimensional parameter space, so the computational cost will be formidable.

5. Conclusion

We have discussed various estimators of the diffusion coefficient in the presence of noise for non-averaged as well as averaged data. We found that the MP and MD estimators work well when the number of measurements is limited, which is the case of interest for clinical applications. The simple HS estimator works well for non-averaged data, but it is not adequate for averaged data. The L_2 norm works better than L_1. There is not much to gain by minimizing a weighted, relative to an unweighted, sum of squares. ML estimation is possible for averaged data, but does not perform better than MD with the number of measurements typical for clinical situations. Proper estimation of the diffusion coefficient allows for high resolution DTI. Smaller voxels reduce the ambiguities associated with crossing fibers in tractography. The results given in this paper apply equally well to any situation where the task is to extract data from a noisy magnitude dataset where the true magnitude is mono-exponential. Other examples are measurements of the transverse relaxation rate, R_2 and effective transverse relaxation rate, R_2^*.

Appendix A

In this appendix we provide details about ML estimation for averaged data. We need to calculate the probability $p_{\underline{m}_{\mathrm{ave}}}(m; m_0)$ of observing the averaged magnitude m given

the true magnitude m_0 numerically (Eq. (10)). Then, assuming independent measurements, the logarithm of the likelihood of the model parameters (D,ρ) given the averaged measurement vector $\mathbf{m}$ is defined by Eq. (21) as

$$\ln[L((D,\rho);\mathbf{m})] \propto \sum_{k=1}^{N_b} \ln[p_{\underline{m}_{\text{ave}}}(m_k;\rho\exp(-Db_k))] \tag{A.1}$$

We tabulate $\ln[p_{\underline{m}_{\text{ave}}}(m;m_0)]$ in the range $m_0/\sigma=[0,20]$ and $m/\sigma=[0,30]$ and use a simple bilinear scheme for interpolation. If $m_0/\sigma>20$, we use the Gaussian approximation

$$p_{\underline{m}_{\text{ave}}}(m;m_0) \approx N(m_{\text{MN}},\sigma^2_{\underline{m}}/N_{\text{AV}}) \tag{A.2}$$

where m_{MN} and $\sigma_{\underline{m}}$ are given in Eqs. (3) and (7), respectively. The procedure for evaluating Eq. (A.1) is illustrated in Fig. 12. The ML estimates of (D,ρ) for a given realization of $\mathbf{m}$ are the values that maximize Eq. (A.1).

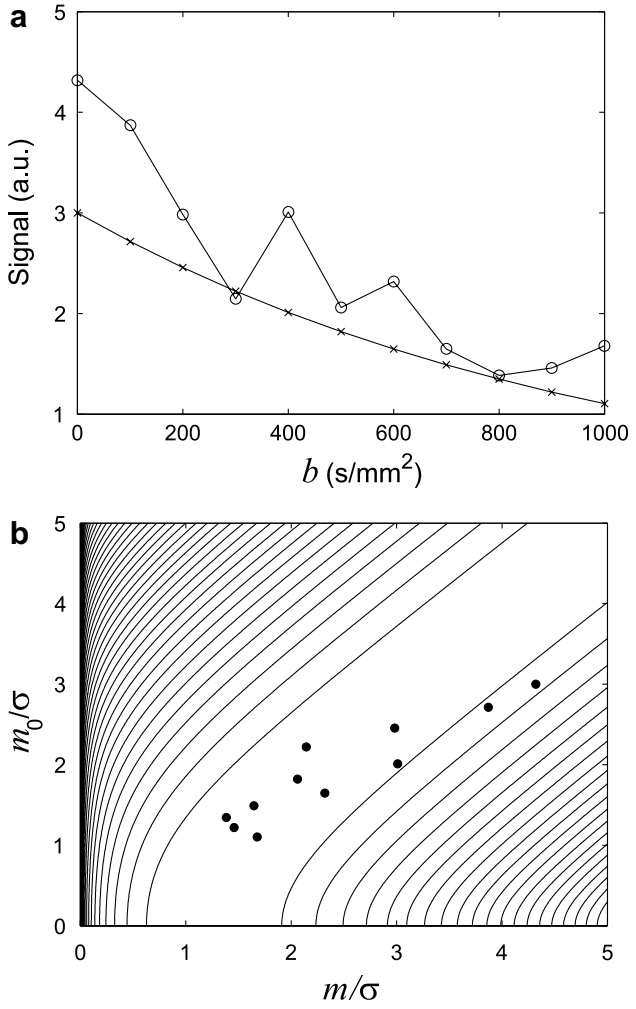

Fig. 12. Calculation of the logarithm of the likelihood for averaged magnitude data. (a) The circles show a possible outcome of a measurement of the averaged signal magnitude, $\mathbf{m}$, in the presence of noise. The number of averages, N_{AV}, is four and the noise level, σ, equals unity. The crosses show values of the true magnitude, $\mathbf{m}_0$, that correspond to a specific realization of the model parameters (which we want to estimate). In this example $D=1\times10^{-3}\,\text{mm}^2/\text{s}$ and $\rho/\sigma=3$. (b) A contour plot of the logarithm of the likelihood as a function of the normalized true and measured magnitudes (with $N_{\text{AV}}=4$) calculated numerically using Eq. (11). The dots correspond to the data points in (a).

Appendix B

In this appendix, we describe the algorithm that we used for calculating the Hessian matrix.

The model parameters must be scaled to the same order of magnitude. Measuring ρ in units of σ and D in units of $10^{-3}\,\text{mm}^2/\text{s}$, we have both parameters on the order of unity. In order to calculate the second derivatives that constitute the Hessian we use low order finite difference expressions. This amounts to evaluating the cost function on a 3 by 3 grid [19]. We use a step size of 0.003 in each direction. In most cases this provides the necessary precision for classifying the point to which the algorithm converged. However, there are some cases where the larger eigenvalue exceeds the smaller one by several orders of magnitude. Then the algorithm with fixed step length does not yield sufficient precision. In order to cover all cases we therefore use an iterative scheme to calculate the Hessian. The steps are as follows

- *First iteration.* Use a step length of 0.003 in each direction. Calculate the first approximation to the Hessian. Calculate eigenvalues and eigenvectors.
- *Iteration n.* Use the eigenvectors found in iteration $n-1$ to construct a rotation matrix. Use this rotation matrix to align approximately the finite difference grid with the main axes of the true Hessian. Adapt the step lengths by choosing them to be proportional to the inverse square root of the eigenvalues found in iteration $n-1$.

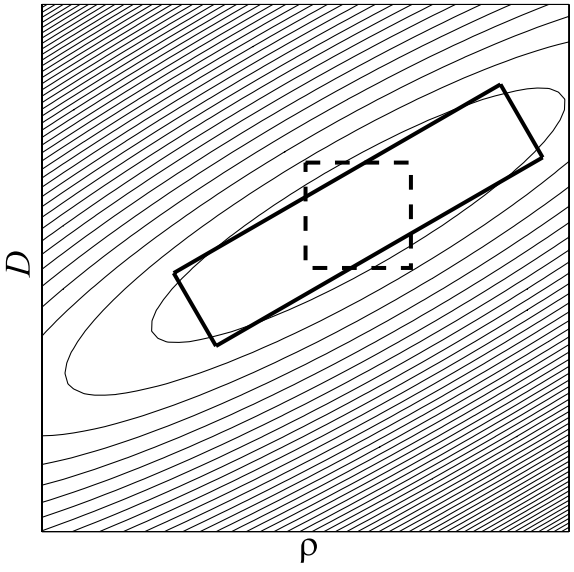

Fig. 13. Principle drawing of the iterative algorithm for calculating the Hessian. We start with a 3 by 3 finite difference grid with equal step lengths in the ρ and D directions (illustrated by the square drawn with dashed lines). We then rotate the grid and adapt the step lengths (illustrated by the rectangle drawn with full lines).

- Repeat until the smaller eigenvalue has converged with the desired precision.

The procedure is illustrated in Fig. 13.

References

[1] S.O. Rice, Mathematical analysis of random noise, Bell Syst. Tech. J. 23 (1944) 282.

[2] R.M. Henkelman, Measurement of signal intensities in the presence of noise in MR images, Med. Phys. 12 (1985) 232–233, Erratum in 13 (1986) 544.

[3] A.J. Miller, P.M. Joseph, The use of power images to perform quantitative analysis on low SNR MR images, Magn. Reson. Imaging 11 (1993) 1051–1056.

[4] G. McGibney, M.R. Smith, An unbiased signal-to-noise ratio measure for magnetic resonance images, Med. Phys. 20 (1993) 1077–1078.

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

[6] J.M. Bonny, M. Zanca, J.Y. Boire, A. Veyre, T_2 maximum likelihood estimation from multiple spin-echo magnitude images, Magn. Reson. Med. 36 (1996) 287–293.

[7] A.H. Andersen, On the Rician distribution of noisy MRI data, Magn. Reson. Med. 36 (1996) 331–332.

[8] C.D. Constantinides, E. Atalar, E.R. McVeigh, Signal-to-noise measurements in magnitude images from NMR phased arrays, Magn. Reson. Med. 38 (1997) 852–857, Erratum in 52 (2004) 219.

[9] J. Sijbers, A.J. den Dekker, P. Scheunders, D. Van Dyck, Maximum-Likelihood estimation of Rician distribution parameters, IEEE Trans. Med. Imaging 17 (1998) 357–361.

[10] J. Sijbers, A.J. den Dekker, E. Raman, D. Van Dyck, Parameter estimation from magnitude MR images, Int. J. Imag. Syst. Tech. 10 (1999) 109–114.

[11] O.T. Karlsen, R. Verhagen, W.M.M.J. Bovée, Parameter estimation from Rician-distributed data sets using a maximum likelihood estimator: application to T_1 and perfusion measurements, Magn. Reson. Med. 41 (1999) 614–623.

[12] O. Dietrich, S. Heiland, K. Sartor, Noise correction for the exact determination of apparent diffusion coefficients at low SNR, Magn. Reson. Med. 45 (2001) 448–453.

[13] J. Sijbers, A.J. den Dekker, Maximum likelihood estimation of signal amplitude and noise variance from MR data, Magn. Reson. Med. 51 (2004) 586–594.

[14] A. Papoulis, Probability, Random Variables and Stochastic Processes, second ed., McGraw-Hill, Tokyo, 1984.

[15] M. Abramowitz, I. Stegun, Handbook of Mathematical Functions, National Bureau of Standards, US Government Printing Office, Washington, DC, 1964, p. 362.

[16] C. Helstrom, Computing the distribution of sums of random sine waves and of Rayleigh-distributed random variables by saddle-point integration, IEEE Trans. Commun. 45 (1997) 1487–1494.

[17] N. Beaulieu, An infinite series for the computation of the complementary probability distribution function of a sum of independent random variables and its application to the sum of Rayleigh random variables, IEEE Trans. Commun. 37 (1990) 1463–1474.

[18] G.K. Karagiannidis, S.A. Kotsopoulos, On the distribution of the weighted sum of L independent Rician and Nakagami envelopes in the presence of AWGN, KICS J. Commun. Networks 3 (2001) 112–119.

[19] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, Numerical Recipes in C, Cambridge, New York, 1994.

[20] J.F.L. Simmons, B.G. Stewart, Point and interval estimation of the true unbiased degree of linear polarization in the presence of low signal-to-noise ratios, A&A 142 (1985) 100–106.

[21] D.K. Jones, P.J. Basser, "Squashing peanuts and smashing pumpkins: How noise distorts diffusion-weighted MR data, Magn. Reson. Med. 52 (2004) 979–993.

[22] L.J. Larsen, M.L. Marx, An Introduction to Mathematical Statistics and Its Applications, second ed., Prentice-Hall, New Jersey, 1986.

[23] J.C. Lagarias, J.A. Reeds, M.H. Wright, P.E. Wright, Convergence properties of the Nelder-Mead simplex method in low dimensions, SIAM J. Optimization 9 (1) (1998) 112–147.

[24] J.A. Jones, P. Hodgkinson, A.L. Barker, P.J. Hore, Optimal sampling strategies for the measurement of spin-spin relaxations times, J. Magn. Reson. B 113 (1996) 25–34.

[25] J. Sijbers, A.J. den Dekker, M. Verhoye, J. Van Audekerke, D. Van Dyck, Estimation of noise from magnitude MR images, Magn. Reson. Imaging 16 (1998) 87–90.

[26] C.G. Koay, L.C. Chang, J.D. Carew, C. Pierpaoli, P.J. Basser, A unifying theoretical and algorithmic framework for least squares methods of estimation in diffusion tensor imaging, J. Magn. Reson. 182 (2006) 115–125.