Local estimation of the noise level in MRI using structural adaptation

Karsten Tabelow\textsuperscript{a,\textdagger}, Henning U. Voss\textsuperscript{b}, Jörg Polzehl\textsuperscript{a}

\textsuperscript{a}Weierstrass Institute for Applied Analysis and Stochastics, Berlin, Germany
\textsuperscript{b}Citigroup Biomedical Imaging Center, Weill Cornell Medical College, New York, USA

Abstract

We present a method for local estimation of the signal-dependent noise level in magnetic resonance images. The procedure uses a multi-scale approach to adaptively infer on local neighborhoods with similar data distribution. It exploits a maximum-likelihood estimator for the local noise level. The validity of the method was evaluated on repeated diffusion data of a phantom and simulated data using T1-data corrupted with artificial noise. Simulation results were compared with a recently proposed estimate. The method was also applied to a high-resolution diffusion dataset to obtain improved diffusion model estimation results and to demonstrate its usefulness in methods for enhancing diffusion data.

1. Introduction

Noise in Magnetic Resonance Imaging (MRI) affects data analysis in neuroscientific problems or clinical applications. For example, in functional MRI it is directly related to the sensitivity of the experiment (Worsley et al., 2002).

\textsuperscript{\textdagger}Corresponding author
Email address: karsten.tabelow@wias-berlin.de (Karsten Tabelow)

Preprint submitted to Elsevier October 27, 2014
In diffusion MRI it leads to variability and, even more important, to a bias of diffusion model parameter estimates (Pierpaoli and Basser, 1996; Basser and Pajevic, 2000; Jones and Basser, 2004). Consequently, quantification of the noise is required to access the quality of dMRI data and in modeling. Furthermore, estimates of the standard deviation of the MR signal are directly used in a number of data enhancing methods to discriminate noise variation from structural differences, see e.g. Aja-Fernández et al. (2008); Coupé et al. (2008); Becker et al. (2012); Rajan et al. (2012); Becker et al. (2014); Haldar et al. (2013) or the citations in the latter paper for a more comprehensive list.

MR magnitude image reconstruction from Fourier transformed $k$-space data of a single receiver coil leads to Rician distributed data (Gudbjartsson and Patz, 1995). For most parallel imaging methods the distribution depends on the reconstruction algorithm but is usually approximated by a (scaled) non-central $\chi$-distribution (Aja-Fernández et al., 2013) which includes the Rician distribution as a special case. The scale parameter $\sigma$ of these distributions is determined by the noise standard deviation of the complex valued noise in $k$-space, the local coil sensitivities and signal correlations between coils (Aja-Fernández and Tristán-Vega, 2012; Aja-Fernández et al., 2011). Accordingly, the noise level is generally not a global quantity over the MR image, but varies locally.

Almost all estimation methods for $\sigma$ rely on the properties of the noise in the image background, i.e., in the absence of a signal, see e.g. Aja-Fernández et al. (2009a) for a comprehensive list of procedures. Only a few methods have been developed that are, under different assumptions, suitable for the estima-
tion of the noise power also in the presence of an MR signal (Sijbers et al., 1998; Aja-Fernández et al., 2009b; Landman et al., 2009b; Aja-Fernández et al., 2013).

In this work we present a novel method for local adaptive noise estimation (LANE) in the presence of a signal, i.e., within tissue regions. The procedure is based on the propagation separation approach (Polzehl and Spokoiny, 2006) adapted for non-central $\chi$-distributed three-dimensional data. The method searches locally for maximal neighborhoods of a voxel with similar data distribution. This information can be used to infer on the noise parameter through maximum-likelihood estimation, cf. Sijbers et al. (1998) for Rician distributed data.

We demonstrate the effectiveness of the method in a) diffusion weighted data of a diffusion phantom measured repeatedly for a single diffusion gradient direction, b) simulated T1-data corrupted with artificial noise, and c) a diffusion MRI dataset. For the latter we additionally demonstrate how the result of the local noise estimation can be used to obtain an improved and unbiased estimate for the parameters of the diffusion tensor model (Basser et al., 1994b,a). Finally, we demonstrate how the local estimate of the noise power can be used to improve the results of a recently developed method for noise reduction in diffusion MRI data (msPOAS, Becker et al., 2014).

2. Theory

2.1. Noise distribution in multiple-coil MR acquisition

An MR image is acquired in frequency or $k$-space and has to be transferred to the common image domain via inverse Fourier transform (Callaghan, 1991). The noise present in $k$-space data of a single coil can be modeled as a
complex additive Gaussian variable with zero expectation and homogeneous standard deviation $\sigma$. For a single-coil acquisition the local signal of the magnitude image in the spatial domain then follows a Rician distribution (Gudbjartsson and Patz, 1995). For a multiple-channel RF coils (Roemer et al., 1990) the data from all $L$ coils is used in the reconstruction of the image data. In case of a sum-of-squares image reconstruction, the standardized signal $S_i/\sigma$ at the spatial position $x_i$ of voxel $i$ is usually considered to be non-central $\chi$ distributed with $2L$ degrees of freedom and non-centrality parameter $\theta_i$ (Constantinides et al., 1997). SENSE (Pruessmann et al., 1999) in general leads to the special case of a Rician distribution $L = 1$, with spatially scale parameter. With other parallel imaging methods like GRAPPA a non-central $\chi$ distribution with adjusted, location dependent distribution parameters serves as a valid approximation of the true data distribution (Aja-Fernández et al., 2011). For further reading we refer, e.g., to Aja-Fernandez et al. (2014). Thus, throughout this paper we will make use of the fact, that the data distribution can be at least well approximated by a non-central $\chi$ distribution.

The probability density $p_S$ for the distribution $P_S$ of the signal $S_i$ generally depends on three parameters $\theta_i, \sigma_i, L_i$ and is given by

$$p_S(S_i; \theta_i, \sigma_i, L_i) = \frac{S_i^{L_i} \theta_i^{(1-L_i)}}{\sigma_i^{(L_i+1)}} e^{-\frac{1}{2} \left( \frac{S_i^2}{\sigma_i^2} + \theta_i^2 \right)} I_{L_i-1} \left( \frac{\theta_i S_i}{\sigma_i} \right),$$

(2.1)

where $I_{L_i-1}$ denotes the $(L_i - 1)$-th order modified Bessel function of the first kind. The mean $\mu$ and variance $v$ of the distribution are given by

$$\mu(\theta_i, \sigma_i, L_i) = \sigma_i \sqrt{\frac{\pi}{2}} I_{\frac{L_i}{2}} \left( \frac{\theta_i^2}{2} \right),$$

(2.2)

$$v(\theta_i, \sigma_i, L_i) = \sigma_i^2 \left( 2L_i + \theta_i^2 \right) - \mu^2(\theta_i, \sigma_i, L_i),$$

(2.3)
where $L_{1/2}^{(L-1)}$ is a generalized Laguerre polynomial. In general, $\sigma_i$ and $L_i$ vary smoothly with location.

2.2. Estimating a local smooth noise standard deviation using adaptive weights

We propose a method for estimation of the local scale parameter $\sigma_i$ in two- or three dimensional MR images. The method does not require multiple or replicated measurements of the volumes. We assume the effective number of coils $L_i$ to be known. The parameter $\sigma_i = \sigma(x_i)$ generally depends on the reconstruction algorithm and is assumed to be a smooth and slowly varying function of location.

The parameter function $\theta_i$ is supposed to be locally constant with $x_i$ (Becker et al., 2012, 2014). This assumption is motivated by the observation that the expected signal intensity or equivalently the non-centrality parameter $\theta_i$ relates to properties of the biological tissue. Then, $\theta_i$ is approximately constant in regions with one type of tissue, while values for different tissue may considerably differ. In the following the term “homogeneity region” refers to this property of the non-centrality parameter.

We use the propagation-separation approach (Polzehl and Spokoiny, 2006) to iteratively infer on the non-centrality parameter $\theta$ and its homogeneity regions. Within this approach estimation of the non-centrality parameter $\theta_i$ and the local scale parameter $\sigma_i$ by weighted maximum likelihood is restricted to only use measurements from the adaptively refined homogeneity regions. The local estimates of $\sigma_i$ are stabilized by a spatial median filter.

Let $K_\text{loc}$ and $K_\text{ad}$ be two non-decreasing kernel function supported on the interval $[0, 1)$, $k^*$ a pre-specified number of iteration steps and $\{h(k)^k\}_{k=0}^{k^*}$ a monotone sequence of bandwidths. Let $\lambda > 0$, $h_{\text{med}} > 0$, $N_0$ denote further
parameters of the procedure, which are explained in detail below.

We propose the following iterative algorithm (LANE):

- \( k = 0 \): Initialize \( \tilde{\sigma}^{(0)}_i = \sigma \) using a global initial estimate (or guess) for the noise standard deviation. Set the initial estimate for the non-centrality parameter constant, i.e., \( \hat{\theta}_i^{(0)} = 1 \), to enforce a non-adaptive weighting scheme \( w_{ij}^{(1)} \) for the first iteration step, see below. Set \( N_i^{(0)} = 1 \) for all voxel \( i \), cf. also below. Set \( k = 1 \).

- For each voxel \( i \) compute adaptive weights

\[
    w_{ij}^{(k)} = K_{\text{loc}} \left( \frac{||x_i - x_j||}{h^{(k)}_{ij}} \right) K_{\text{ad}} \left( \frac{N_i^{(k-1)} \Delta_{ij}}{\lambda} \right),
\]

and their sum \( N_i^{(k)} = \sum_j w_{ij}^{(k)} \) over all voxel \( j \). Here, \( ||\cdot|| \) denotes the Euclidean norm to calculate the spatial distance between two voxel \( i \) and \( j \) and \( \Delta_{ij} = \mathcal{K}_{\mathcal{L}} \left( P_S \left( \hat{\theta}_i^{(k-1)}, \tilde{\sigma}_i^{(k-1)} \right), P_S \left( \hat{\theta}_j^{(k-1)}, \tilde{\sigma}_i^{(k-1)} \right) \right) \) is the Kullback-Leibler \( \mathcal{K}_{\mathcal{L}} \) divergence between the probability distributions \( P_S \left( \hat{\theta}_i^{(k-1)}, \tilde{\sigma}_i^{(k-1)}, L_i \right) \) and \( P_S \left( \hat{\theta}_j^{(k-1)}, \tilde{\sigma}_i^{(k-1)}, L_i \right) \). We dropped the dependence on \( L_i \) for brevity of the notation. In distributions we use \( L_i \) and the same estimated scale parameter from the previous iteration step \( \tilde{\sigma}_i^{(k-1)} \) which reflects our assumptions on the spatial smoothness of \( \sigma_i = \sigma(x_i) \) and \( L_i \). The first term of \( w_{ij}^{(k)} \) defines non-adaptive weights depending on the location kernel \( K_{\text{loc}} \) and the spatial distance of voxel \( i \) and \( j \). The second term evaluates the statistical difference between the estimated non-centrality parameters \( \hat{\theta}_i^{(k-1)} \) and \( \hat{\theta}_j^{(k-1)} \) from the previous iteration step.

\( \lambda \) is the adaptation bandwidth of the procedure. It controls the amount of spatial adaptation in the definition of the weights. In the extreme
case of \( \lambda \to \infty \) the second term equals to \( K_{ad}(0) \), such that the weighting schemes \( w_{ij}^{(k)} \) are non-adaptive. For \( \lambda = 0 \) the weights \( w_{ij}^{(k)} \) vanish for \( i \neq j \) due to the bounded support of the kernel function \( K_{ad} \). For a proper choice of \( \lambda \) (see below), \( w_{ij}^{(k)} \) will, with high probability and increasing \( k \), approach 1 or 0, depending on voxel \( i \) and \( j \) belonging to the same homogeneity region or not, respectively. Thus, for large \( k \), the weighting scheme \( W_i^{(k)} = (w_{i1}^{(k)}, \ldots, w_{in}^{(k)}) \) describes the homogeneity region of voxel \( i \).

- If \( N_i^{(k)} := \sum_j w_{ij}^{(k)} > N_0 \) we obtain estimates for \( \theta(x_i) \) and \( \sigma(x_i) \) by weighted log-likelihood

\[
\left( \hat{\sigma}_i^{(k)}, \hat{\theta}_i^{(k)} \right) = \operatorname{argmax}_{(\theta,\sigma)} \sum_j w_{ij}^{(k)} \log p_S(S_j; \theta, \sigma). \tag{2.5}
\]

Otherwise we set \( \hat{\sigma}_i^{(k)} := \hat{\sigma}_i^{(k-1)} \) and

\[
\hat{\theta}_i^{(k)} := \sqrt{\left( \frac{\sum_j w_{ij}^{(k)} S_j^2}{\sum_j w_{ij}^{(k)}} - 2L_i(\hat{\sigma}_i^{(k-1)})^2 \right)} +
\]

employing the moment equation (2.3). The adaptive weights enforce that observations \( S_j \) following a distribution with significantly different non-centrality parameter \( \theta \) are not utilized in the estimator above. Within the iteration process (for increasing \( k \)), the sum of all weights \( N_i^{(k)} \) increases, the inference on the homogeneity regions will be more informed, and the estimates above stabilize.

- We further stabilize the estimate of \( \sigma(x_i) \) using the median filter

\[
\hat{\sigma}_i^{(k)} = \text{median}_{j : \|x_i - x_j\| < h_{med}} \hat{\sigma}_j^{(k)}
\]
over a ball with radius $h_{med}$ centered in $i$. The median filter reduces the variability of the estimate. The robustness of the median filter avoids a bias that may be caused by insufficient adaptation in some voxel $j$.

- If $k = k^*$ stop, else increase $k$ by 1 and continue with the second step, i.e., with the definition of the weights.

The parameter $N_0$ defines a minimal sum of weights $N_i^{(k)}$ before the initial guess for $\sigma_i$ is updated. $N_0 \gg 1$ guarantees the identifiability in (2.5) using observations from a sufficiently large homogeneous vicinity of voxel $i$.

2.3. Maximum Likelihood estimator for non-central $\chi$-distribution

Given a sample $S = (S_1, \ldots, S_n)$ the local weighted log-likelihood (2.5), up to terms that do not depend on the parameters, takes the form, see Eq. (2.1),

$$l(S; W_i; \theta, \sigma) = \sum_j w_{ij} \log p_s(S_j; \theta, \sigma)$$

$$= -N_i \left((L + 1) \log \sigma + (L - 1) \log \theta + \frac{\theta^2}{2} + \frac{\xi^2}{2\sigma^2}\right)$$

$$+ \sum_j w_{ij} \log I_{L-1} \left(\frac{\theta S_j}{\sigma}\right)$$

(2.6)

with $N_i = \sum_j w_{ij}$ and $\xi = \frac{1}{N_i} \sum_j w_{ij} S_j^2$.

Here, we assumed the probability density $p_s$ to be non-central $\chi$ with non-centrality parameter $\theta$ and $2L$ degrees of freedom. The degrees of freedom $2L$ are assumed to be known and may depend on $i$. The locality of the estimates is then realized by employing the weighting scheme $W_i$ for voxel $i$. 

8
In order to solve the optimization problem in Eq. (2.5) we take the derivatives with respect to $\theta^2$ and $\sigma$ and set them to zero. This yields

$$\frac{\sigma}{\theta} \frac{\partial l}{\partial \sigma} (S; W_i; \theta, \sigma) = -N_i \left( \frac{L + 1}{\theta} - \frac{\xi}{\theta \sigma^2} \right) - \sum_j w_{ij} \frac{I_L \left( \frac{\theta S_j}{\sigma} \right)}{2I_{L-1} \left( \frac{\theta S_j}{\sigma} \right)} \frac{S_j}{\sigma} = 0$$

$$\frac{\partial l}{\partial \theta} (S; W_i; \theta, \sigma) = -N_i \left( \frac{L - 1}{\theta} + \theta \right) + \sum_j w_{ij} \frac{I_L \left( \frac{\theta S_j}{\sigma} \right) + I_{L-2} \left( \frac{\theta S_j}{\sigma} \right) S_j}{2I_{L-1} \left( \frac{\theta S_j}{\sigma} \right)} \frac{1}{\sigma} = 0$$

By adding both equations and re-arranging the result we get

$$\hat{\theta}^2 = \frac{\xi}{\sigma^2} - 2L = \frac{1}{N_i \sigma^2} \sum_j w_{ij} S_j^2 - 2L \quad (2.7)$$

Substituting this into the local likelihood Eq. (2.6) and again removing the constant term that does not depend on $\sigma$ yields

$$\tilde{l}(S; W_i; \sigma) = -N_i \left( \xi/\sigma^2 + 2 \log \sigma + \frac{L - 1}{2} \log (\xi - 2L\sigma^2) \right)$$

$$+ \sum_j w_{ij} \log I_{L-1} \left( \frac{S_j}{\sigma^2} \sqrt{\xi - 2L\sigma^2} \right). \quad (2.8)$$

Finally we get an estimator for $\sigma_i$ as a solution of an optimization problem in the univariate parameter $\sigma$:

$$\hat{\sigma}_i = \sqrt{\frac{N_i}{N_i - 1}} \arg\max_{\sigma^2 \geq \frac{\xi}{2L}} \tilde{l}(S; W_i; \sigma). \quad (2.9)$$

If the solution of the optimization problem Eq. (2.7) is at the left border the parameter domain, i.e., $\sigma^2 = \frac{\xi}{2L}$, then $\hat{\theta}_i = 0$, i.e., the $\chi$-distribution is central. The estimator in Eq. (2.9) is biased for small values of $\theta/\sigma$ (Sijbers
et al., 1998). In particular, for extremely low SNR, its density has two modes, one arising from the estimates from the central \( \chi \)-distribution case, and one for the case where \( \xi_i / \sigma^2 - 2L > 0 \). This effect is demonstrated in Fig. 1. The effect may lead to underestimating \( \sigma \) in case of very low SNR (\(< 1.5\))..

2.4. Choice of parameters of the procedure

The specific choice for the kernel functions \( K_{\text{loc}} \) and \( K_{\text{ad}} \) has only minor influence on the estimation results, see, e.g., Section 6.2.3 in Scott (1992). We choose them as:

\[
K_{\text{loc}}(x) = \begin{cases} 
1 - x^2 & x < 1 \\
0 & x \geq 1 
\end{cases} \quad \text{and} \quad K_{\text{ad}}(x) = \begin{cases} 
1 & x < 0.5 \\
2 - 2x & 0.5 \leq x < 1 \\
0 & x \geq 1 
\end{cases}
\]
for computational efficiency. We choose $h^{(0)} = 1$. The monotone sequence $\{h^{(k)}\}_{k=0}^{k^*}$ of bandwidths is chosen such that for non-adaptive weights

$$\tilde{\omega}^{(k)}_{ij} = K_{\text{loc}} \left( \frac{\|x_i - x_j\|}{h^{(k)}} \right)$$

such that

$$\frac{\sum_j (\tilde{\omega}^{(k)}_{ij})^2}{(\sum_j \tilde{\omega}^{(k)}_{ij})^2} = c_h^{-1} \frac{\sum_j (\tilde{\omega}^{(k-1)}_{ij})^2}{(\sum_j \tilde{\omega}^{(k-1)}_{ij})^2},$$

i.e. the variance of a non-adaptive estimate would be reduced reduced by the factor $c_h$ going from step $k - 1$ to step $k$. We choose $c_h = 1.25$.

$\lambda$ is the main parameter of the procedure as it controls the amount of adaptation of the method. A reasonable value is $\lambda = 5$, see Figure 5.

We recommend $k^* = 20$, $N_0 = 2$ and $h_{\text{med}} = 5$. The mean number of effective coils $L$ (or $L(x_i)$) needs to be specified by the user.

2.5. Application to an unbiased estimation of the diffusion tensor for DTI data

Diffusion weighted imaging (dMRI; Jones, 2010) has become a widely used standard tool for structural in-vivo examination of the brain in recent years. There, the application of an additional diffusion weighting gradient in the magnetic field leads to a signal attenuation that is directly related to the diffusion properties of the water in tissue along the considered gradient direction. Additionally, at least one non-diffusion weighted image volume is acquired for comparison. Very often, in the Gaussian diffusion approximation, the directional dependence of the diffusion properties is described within the diffusion tensor model exploited in diffusion tensor imaging (DTI; Basser et al., 1994b,a). The physical model for DTI is formulated in terms of a noiseless situation. Let $\zeta_{p,i}$ denote the noiseless image value at voxel $i$ for
the $p$-th volume of the diffusion weighted MRI dataset, corresponding to the diffusion weighting gradient in direction $\vec{g}_p$ and the $b$-value $b_p$ (Basser et al., 1994a). This series also includes the non-diffusion weighted image volumes. The diffusion tensor model then describes the data by a symmetric, positive definite $3 \times 3$ matrix $D_i$ and a parameter $\zeta^0_i$ with

$$
\zeta_{p,i} = \zeta^0_i \exp(-b_p \vec{g}_p^T D_i \vec{g}_p) \quad \forall p.
$$

(2.10)

Positive definiteness of the diffusion tensor $D_i$ can be enforced by re-parametrization using $D_i = R_i^T R_i$ with an upper triangular matrix $R_i$, see Koay et al. (2006) or Ghosh et al. (2013).

Traditionally the diffusion tensor is estimated using non-linear regression

$$(\hat{\zeta}_i^0, \hat{R}_i) = \arg\min \sum_p (S_{i,p} - \zeta_{i,p})^2
$$

(2.11)

thereby ignoring the difference between non-centrality parameter and expectation of a non-central chi-distribution. For small SNR this causes a bias in the tensor estimate. By use of the (local) noise variance as determined by the method from the previous sections, one can now develop an unbiased estimator for the diffusion tensor.

The log-likelihood function for the diffusion tensor model assuming a non-central $\chi$ distribution with $2L_i$ degrees of freedom and non-centrality parameter $\zeta_{p,i}/\sigma_{p,i}$ for the standardized observed image intensities $S_{p,i}/\sigma_{p,i}$ is given by

$$
l(\{S_{p,i}\}_p; \{\sigma_{p,i}\}_p, L_i; \zeta^0_i, R_i) =
\sum_p \left[ \log \left( \frac{S_{p,i}^{L_i} (1-L_i)}{\sigma_{p,i}^2} \right) - \frac{1}{2} \left( \frac{S_{p,i}^2}{\sigma_{p,i}^2} + \zeta_{p,i}^2 \sigma_{p,i}^2 \right) + \log \left( I_{L_i} - \left( \frac{\zeta_{p,i} S_{p,i}}{\sigma_{p,i}^2} \right) \right) \right],
$$

(2.12)
where \( \zeta_{p,i} \) is given by Eq. (2.10) and the re-parametrization in terms of the model parameters \( \zeta^0_i, R_i \). Estimates are then obtained by maximizing the log-likelihood.

This approach has been already considered in the literature for the estimation of the diffusion tensor (Landman et al., 2009a) or for the extended diffusion kurtosis model (Veraart et al., 2011a,b; Ghosh et al., 2013; André et al., 2014). The procedure avoids the bias in the parameter estimation that is caused by the skewness of the \( \chi \)-distribution causing the deviation of its expectation value from the non-centrality parameter in a noisy situation, see Eq. (2.2).

An alternative is provided by quasi-maximum likelihood, i.e., by minimizing the negative Gaussian log-likelihood

\[
R(\{S_{p,i}\}_p; \{\sigma_{p,i}\}_p, L_i; \zeta^0_i, R_i) = \sum_p \left[ \frac{(S_{p,i} - \mu(\zeta_{p,i}/\sigma_{p,i}; \sigma_{p,i}, L_i))^2}{\nu(\zeta_{p,i}/\sigma_{p,i}, \sigma_{p,i}, L_i)} \right]^{(2.13)}
\]

using (2.2), (2.3), (2.10) and \( D_i = R_i^T R_i \), and effectively approximating the non-central \( \chi \)-distribution by a Gaussian distribution with appropriate moments.

2.6 Misspecification of \( L_i \)

The effective number of coils \( L_i \) or equivalently the degrees of freedom \( 2L_i \) depends on the reconstruction algorithm used. If information from \( L \) independent receiver coils is combined as a root sum of squares (RSoS) we get \( L_i \equiv L \). In case of SENSE (Pruessmann et al. (1999)) the image is obtained from a linear combination of complex Gaussian images and hence \( L_i \equiv 1 \). Reconstruction methods like GRAPPA (Griswold et al. (2002)) or ZOOPPA (Heidemann et al. (2012)) in general lead to a spatially varying \( L_i \),
see e.g. Aja-Fernández et al. (2011). In this case we have $1 < L_i << L$ with $L_i$ tending to be small in case of high acceleration factors.

If $L_i$ can be computed from parameters of the reconstruction method it should definitively be used. Unfortunately $L_i$ and $\sigma$ cannot be estimated simultaneously due to severe identifiability problems. Still a non-central $\chi$-distribution with $2L$ degrees of freedom and non-centrality parameter $\theta$ can be well approximated by a Rician distribution in terms of the Kullback-Leibler-divergence between these distributions, see Table 1. This approximation, in general, improves on a Gaussian approximation.

<table>
<thead>
<tr>
<th>True $L_i$</th>
<th>$\theta$</th>
<th>$\text{eff. } \theta \text{ (L=1)}$</th>
<th>Expected value</th>
<th>KL-div.(Rice)</th>
<th>KL-div.(Gauss)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.5</td>
<td>1</td>
<td>1.60</td>
<td>1.85</td>
<td>0.00609</td>
<td>0.02250</td>
</tr>
<tr>
<td>1.5</td>
<td>2</td>
<td>2.29</td>
<td>2.49</td>
<td>0.00174</td>
<td>0.00683</td>
</tr>
<tr>
<td>1.5</td>
<td>4</td>
<td>4.13</td>
<td>4.25</td>
<td>0.00003</td>
<td>0.00013</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1.93</td>
<td>2.11</td>
<td>0.00817</td>
<td>0.01511</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2.53</td>
<td>2.69</td>
<td>0.00294</td>
<td>0.00571</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>4.25</td>
<td>4.37</td>
<td>0.00010</td>
<td>0.00021</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>2.41</td>
<td>2.54</td>
<td>0.00690</td>
<td>0.00900</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2.92</td>
<td>3.05</td>
<td>0.00338</td>
<td>0.00455</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>4.49</td>
<td>4.60</td>
<td>0.00025</td>
<td>0.00037</td>
</tr>
</tbody>
</table>

Table 1: Approximation of non-linear $\chi$-distributions by Rician- and Gaussian-distributions.

Using an incorrect value $L$ instead of $L_i$ results in a model misspecification. The consequence is that instead of estimating $\sigma_i$ in $p_S(., \sigma_i, L_i)$ we aim for a projection parameter $\bar{\sigma}_i$ in $p_S(., \bar{\sigma}_i, L)$ minimizing the Kullback-Leibler divergence $\mathcal{KL}(p_S(., \sigma, L), p_S(., \sigma_i, L_i))$. Note that $\bar{\sigma}_i$ converges to $\sigma_i$ with increasing SNR and $\bar{\sigma}_i < \sigma_i$ if $L < L_i$.

If $L_i$ is unknown we propose to use $L = 1$. The consequences depend on the intended use of the estimated scale parameter $\sigma_i$. In msPOAS, Becker et al. (2014), adaptation is driven by the Kullback-Leibler-divergence of distributions in neighboring points in orientation space $R^3 \times S^2$. Using $L = 1$
is therefore appropriate.

In modeling, e.g., for estimating parameters in a diffusion tensor model, a correct assessment requires knowledge of $L_i$ to completely remove the bias using Eq. (2.12) or (2.13). Still specifying $L = 1$ or some $L < L_i$ partially removes the bias and leads to improved estimates.

3. Materials and methods

The proposed method is suitable for any MR data that comply with the assumptions outline above. We use T1-weighted data for the simulation experiment. Further, we focus on diffusion weighted data, as there SNR is inherently low and the knowledge of the local noise variance can be used for improved diffusion models and data enhancement.

3.1. Experimental data (Diffusion weighted images of a diffusion phantom)

Data were obtained from a DTI phantom with straight and crossing fibers assembled from parts of the phantom published by Pullens et al. (2010). Diffusion weighted images were acquired on a 7.0 Tesla 70/30 Bruker Biospec small animal MRI system with 450 mT/m maximum gradient amplitude and 4500 T/m/s maximum slew rate. Radio-frequency power was transmitted by a 72 mm diameter linear coil and picked up by a quadrature rat brain coil. We used a single-shot spin-echo EPI sequence with echo and repetition times of $TE = 46\,\text{ms}$ and $TR = 2500\,\text{ms}$, respectively. Four slices in the x-y plane with matrix size $128 \times 128$ (resolution $0.2344 \times 0.2344\,\text{mm}^2$) and 2 mm thickness were scanned without skip. The acquisition matrix size was $128 \times 91$ (partial Fourier overscan factor of 1.4). We acquired 10 images
without diffusion weighting and 1000 diffusion weighted images with diffusion gradient direction along the x-direction. The diffusion gradient width was \( \delta = 4.5 \text{ ms} \) and the spacing \( \Delta = 9.2 \text{ ms} \), yielding a \( b \)-value of 1000 \( \text{s/mm}^2 \). Effective \( b \)-values for the non-diffusion weighted and the diffusion weighted scans were 2 \( \text{s/mm}^2 \) and 1031 \( \text{s/mm}^2 \), respectively. The total scan time for this scan was 42 min.

**Analysis**

To describe possible trends over time we fitted an adaptive local polynomial of order 2 to the series of 1000 replicates in each voxel \( i \). Following the proposal in Landman et al. (2009b) (Eq. (4),(5) and (7)) we then obtained a robust estimate \( \tilde{\sigma}_i \) as a characterization of the “ground truth”. We then, for all 1000 images, estimated \( \sigma_i \) using the proposed method (LANE) leading to estimates \( \hat{\sigma}_{i,j} \) in location \( x_i \) and image \( j \). The effective number of coils was specified as \( L = 1 \) which is correct in this case. The other parameters of the procedure were \( \lambda = 5 \), \( k^* = 20 \), and \( h_{med} = 5 \).

For comparison we also determined an estimate of the effective parameter \( \sigma_{eff,i} \) by the method proposed in Aja-Fernández et al. (2013)(Eq. (24)) using a cube of \( 5 \times 5 \times 5 \) voxel centered in \( x_i \) as a local neighborhood. We refer to this method by AF13.

### 3.2. Simulated data (T1 image)

**Data**

For the simulation experiments we used a 3D synthetic BrainWeb MR volume (Collins et al., 1998). Specifically, we generated a noise-free T1 image with 1mm slice thickness and 0% intensity non-uniformity in the provided 12bit short raw format. The image intensity had a range from 0 to 16.
4096 with a median value of approximately 1800 for gray matter areas and 2400 for white matter regions. To simulate a parallel imaging process we defined 8 artificial linear sensitivity maps along the $x$- and $y$-axis and the diagonals, and created pseudo image acquisition for each of the 8 simulated coils, see Fig. 2. We added complex Gaussian noise with standard deviation

$$\sigma_K = 50, 100, 200, 400, 800$$

in $k$-space, specifying a correlation between coils $k$ and $l$ of $\rho = 0.15^{(d_{kl})}$ where $d_{kl}$ refers to a distance between the spherically arranged coils. The final magnitude image was obtained using a SENSE1 reconstruction (Sotiropoulos et al., 2013b), which is the standard imaging protocol used in the Human Connectome Project (Sotiropoulos et al., 2013a). The parameter $\sigma_i$ in the resulting magnitude image ranges from $0.66\sigma_K$ to $1.98\sigma_K$ with maximum values in the center of the image, see Fig. 4g.
Analysis

We applied our method (LANE) described in the Theory part of this paper to determine the noise parameter over the image within a brain mask covering most of the white and gray matter. The parameters of the procedure were $\lambda = 5$, $k^* = 20$, $h_{med} = 5$ and $L = 1$. The analysis was restricted to a white/gray matter mask.

For comparison we again obtained estimates of $\sigma_i$ using the method AF13.

Additionally we investigated the quality of results depending on parameters $\lambda$, $h_{med}$ and $k^*$.

3.3. Experimental data (diffusion weighted imaging)

MRI

We re-analyzed a dataset described already in Becker et al. (2012) and Becker et al. (2014). Data were acquired from a whole body 7T MAGNETOM scanner (Siemens Healthcare) with a maximum gradient amplitude of 70 mT/m and a maximum slew rate of 200 T/m/s (SC72, Siemens Healthcare, Erlangen, Germany). The scan was performed using a single channel transmit, 24-channel receive phased array head coil (Nova Medical, Wilmington, MA, USA). An optimized monopolar Stejskal-Tanner sequence according to Morelli et al. (2010) together with the ZOOPPA approach described in Heidemann et al. (2012) was used with TR 14.1 s, TE 65 ms, BW 1132 Hz/pixel, and ZOOPPA acceleration factor of 4.6. A total of 91 slices with 10% overlap were acquired at a field-of-view (FoV) of $143 \times 147 \text{mm}^2$ resulting in an isotropic high resolution of 800 $\mu$m. Diffusion weighting gradients were applied along 60 different directions at a b-value of 1000 s/mm$^2$. 7 interspersed non-diffusion weighted images were acquired. The scan was repeated
4 times. The subject was a healthy adult volunteer after obtaining written informed consent in accordance with the ethical approval from the University of Leipzig. Total acquisition time was 65 min.

Analysis

We applied the procedure LANE introduced in the Theory part to determine the local noise parameter $\sigma_{p,i}$ over all $p = 1, \ldots, 268$ images (with and without diffusion weighting) within a brain mask. The parameters of the procedure were $\lambda = 5$, $k^* = 16$ (for reduced computational costs), $h_{med} = 5$ and $L = 1$.

For a comparison we estimated $\sigma_{p,i}$ using method AF13. Additionally we computed the spatially robust voxel-wise method with one repetition (SRV1) estimator introduced in Landman et al. (2009b) that employs an robust estimate of the standard deviation of a truncated set of leave-one out residuals from a tensor model.

The data were then smoothed using a version of the msPOAS algorithm (Becker et al., 2014) that has been adapted to use local estimates of $\sigma$ as indicated there as well as using a global estimate for $\sigma$.

Diffusion tensors where estimated using non-linear regression (2.11) and the quasi-likelihood (2.13) for both the original and smoothed data. Quasi-likelihood was used instead of the likelihood (2.12) since it is suitable for both the original and the smoothed data.

3.4. Software

The new method LANE for estimation the local noise parameter proposed in this paper is implemented within our R-package dti (Tabelow and Polzehl, 2014) (version 1.2-0). This package is freely available on CRAN
The implementation uses FORTRAN and native R-code. The package also provides an implementation of the maximum-likelihood and quasi-maximum-likelihood estimates of the diffusion tensor model parameters. The package dti and R were used for all calculations, including AF13 and SRV1, in this paper.

On a single diffusion weighted volume from the dMRI dataset the method takes approximately 140 min for 16 iterations on a HP SL390s G7 compute server with two Xeon, Six-Core 3467@MHz and 96 GB RAM running SuSE Linux. Our implementation uses OpenMP parallelization for speed-up. With use of 12 cores the computation time reduces to 13 minutes for the same data.

4. Results

4.1. Experimental data (Diffusion weighted images of a diffusion phantom)

In Fig. 3 we summarize the results obtained for the repeated diffusion weighted images of the phantom. The relative mean absolute error

\[ rMAE(x_i) = \frac{1}{1000} \sum_j |\hat{\sigma}_{ij} - \tilde{\sigma}_i|/\tilde{\sigma}_i, \]

the relative bias

\[ rBias(x_i) = \frac{1}{1000} \sum_j \frac{\hat{\sigma}_{ij} - \tilde{\sigma}_i}{\tilde{\sigma}_i} \]

and the relative standard deviation

\[ rSD(x_i) = \frac{sd_j \hat{\sigma}_{ij}}{\tilde{\sigma}_i} \]

are defined in comparison to estimates \( \tilde{\sigma}_i \) considered as ground truth. The estimates \( \hat{\sigma}_i \) were generated applying the robust variance estimate proposed
in Landman et al. (2009b) on residuals of an adaptive local polynomial fit to the series of 1000 replicates. Results refer to the region inside the phantom and are presented in terms of rMAE, e) and f), rBias, g) and h), and rSD, i) and j). The density plots in k)-m) illustrate that LANE provides a smaller rMAE and rBias over a wide range of voxel, while AF13 provides a considerably smaller rSD. LANE fails in regions where the assumption of a spatially
Figure 4: Simulation results for a slice of the BrainWeb MR volume. a) Original slice. b)-e) Slice after adding complex Gaussian noise with varying standard deviation using a SENSE1 Sotiropoulos et al. (2013b) reconstruction. The four situations are characterized by a median SNR of approximately 14.9, 7.4, 3.7, 1.9 within the grey/white matter area and of 9.2, 4.6, 2.3 and 1.15 in the center of the images, respectively. f) image of locally varying effective $\sigma_i$ parameters, g)-j) relative error of local estimates $\hat{\sigma}_i$ using the method LANE proposed in this paper. k)-n) relative error of local estimates $\hat{\sigma}_i$ using the method AF13 described in Aja-Fernández et al. (2013). The relative errors are given on a log-scale, with green indicating a perfect estimate.

smooth scale parameter $\sigma$ is violated, see b) and c).

4.2. Simulated data (T1 image)

In Fig. 4 summarizes the results of the noise estimation for an arbitrarily selected slice (Fig. 4a) of the simulated T1 data. Figs. 4b)-e) show the same slice corrupted with complex Gaussian noise with standard deviation $\sigma_K =$
100, 200, 400, 800 using artificial sensitivity maps, correlation between coils and a SENSE1 reconstruction (Sotiropoulos et al., 2013a), cf. Fig. 2. This leads to a location dependent parameter σ as shown in Fig. 4f. In Figs. 4g)-j) we show the logarithmic ratio of the local estimated noise parameter and its theoretical value (Fig. 4g) for all four noise levels. The color scale is defined such that green refers to the ideal log-ratio of 0. The range of the log-ratio varies with the noise level and is given below the color scale. In Figs. 4k)-n) we show the corresponding results for the method proposed in Aja-Fernández et al. (2013).

Fig. 5 provides the mean relative absolute error as a function of the parameters λ, h_{med} and k^{*} and the five noise levels. The mean relative absolute error for the method proposed in Aja-Fernández et al. (2013) is given as a comparison. The profiles indicate a stable behavior of the procedure with respect to the chosen parameters. Large values of h_{med} and k^{*} provide slight improvements at the cost of significantly higher numerical effort. Improvements are mainly due to a reduced variability of the estimates. The parameter λ steers the adaptation to the structure in the non-centrality parameter of
the \( \chi \)-distribution leading to a balance between excluding observations from the tails of the non-central \( \chi \)-distribution when the non-centrality parameters coincide (same homogeneity region), and including observations with different non-centrality parameters (distinct regions). The optimal balance depends on the SNR and the homogeneity structure, with inadequate choice resulting in a bias of the estimates. Our recommendation \( \lambda = 5 \) serves as a good compromise, with slightly smaller / larger values being preferable in high / low SNR situations.

4.3. Experimental data (diffusion weighted imaging)

In Fig. 6 we illustrate the results for the diffusion weighted dataset. In a) and d) we show the mean, over all gradients, estimated \( \sigma_i \) for one slice of the non-diffusion weighted images and the diffusion weighted images, respectively. Figs. 6 b) and e) provide the ratio of the standard deviation \( sd_j \hat{\sigma}_{ij} \) over all gradients, and the mean estimated \( \sigma_i \) while c) and f) show densities of the values in b) and e). Note that the standard deviation \( sd_j \hat{\sigma}_{ij} \) reflects both the variability of the true \( \sigma_i \) over the consecutively acquired images and the variability of the individual estimates. Results suggest that estimates from individual images are close to the mean estimate. Therefore it suffices to obtain estimates from a one or a few diffusion weighted images per b-value.

Figure 7 provides an comparison between the three considered methods in terms of a mean estimated scale parameter \( \sigma_i \). Methods LANE and SRV1 produce similar spatial patterns while the estimate provided by AF13 is almost homogeneous. SRV1 in general produces slightly larger estimates than LANE which coincides with the findings in Landman et al. (2009b)(Table 1) that “estimated value is artificially increased by model mismatch, and
Figure 6: Results for the diffusion weighted dataset. Mean estimated $\sigma$ over all a) non-diffusion weighted and d) diffusion weighted images. Ratio of the standard deviation of the estimated $\sigma$ over all b) non-diffusion weighted and e) diffusion weighted images to the corresponding mean estimate. c) and f) show the corresponding densities of this relative standard deviation.

Figure 7: Comparison with alternative methods: a) Mean estimated $\sigma$ of diffusion weighted images for method LANE, b) estimated $\sigma$ for SRV1, c) Mean estimated $\sigma$ for AF13.

suitability depends on accuracy of the model."

Our last result concerns the effect using the locally estimated $\sigma_i$ on estimates of the Fractional Anisotropy (FA) in a tensor model. The locally estimated $\sigma_i$ is employed in two different occasions: for suitable standardiza-
tion of the signal in our msPOAS smoothing algorithm (Becker et al., 2014), and as a parameter in the quasi-likelihood function employed for estimating the tensor parameters, see Eq. (2.13). Fig. 8 provides results in terms of estimated FA and estimated main diffusion direction based on different estimates of the diffusion tensor, $D_{\text{orig}}^{\text{nlreg}}$ and $D_{\text{locPOAS}}^{\text{nlreg}}$ employing Eq. (2.11) on the original data and data smoothed by msPOAS (Becker et al., 2014) (16 steps) using local estimates $\sigma_i$, respectively, and $D_{\text{orig}}^{\text{qlike}}$ and $D_{\text{locPOAS}}^{\text{qlike}}$ where the quasi-likelihood Eq. (2.13) was used instead of non-linear regression. For comparison we also computed an estimate $D_{\text{gPOAS}}^{\text{qlike}}$ using Eq. (2.13) and data smoothed by msPOAS (16 steps) employing a global estimate of $\sigma$.

Fig. 8a) provides the estimated FA obtained from $D_{\text{locPOAS}}^{\text{qlike}}$. Fig. 8 b), d) and e) show color coded FA obtained from the quasi-likelihood estimates $D_{\text{orig}}^{\text{qlike}}$, $D_{\text{gPOAS}}^{\text{qlike}}$ and $D_{\text{locPOAS}}^{\text{qlike}}$, respectively. We observe a clearly improved color coded FA in e) compared to b) (original data) and d) where the inadequate global $\sigma$ in msPOAS leads to a spatially varying quality of the result. Fig. 8 c) and f) reflect the change in FA when employing quasi-likelihood instead of non-linear regression for the original and the smoothed (local $\sigma$) data, respectively. The FA-differences are provided on a color scale with green referring to a zero difference. Both plots indicate a FA-dependent bias of the regression estimate, with the effect being much more prominent and stable for the smoothed data.

5. Discussion and Conclusion

We developed a novel method LANE for the local estimation of the noise parameter in magnetic resonance imaging in the presence of an MR signal. The method can thus be applied to estimate the noise level in regions with
Figure 8: Fractional anisotropy (FA) maps for the first repetition of the diffusion weighted dataset. a) FA gray-scale image, and e) corresponding color-coded FA image from data smoothed by msPOAS with local variance estimates. d) Color-coded FA estimates from data smoothed by msPOAS with a single global estimate of $\sigma$. b) Result for original (unsmoothed) data. a), b), d), and e) provide results obtained using the quasi-likelihood (2.13). c) and f) refer to the local differences for the FA estimates when the “correct” quasi-likelihood is used compared to the ”biased” non-linear regression estimate Eq. (2.11).

tissue, which is not accessible by methods that rely on the background distribution, see e.g. the comprehensive list of methods in Aja-Fernández et al. (2009a).

The method assumes the standardized MR signal to follow a non-central $\chi$ distribution with a locally constant non-centrality parameter, and spatially slowly varying noise parameter and number of degrees of freedom. The procedure is based on the propagation-separation approach (Polzehl and Spokoiny, 2006) to infer on local homogeneity regions of the non-centrality parameter of the signal distribution. Then a local weighted Maximum Likelihood estima-
tor (over these regions) is used to obtain local estimates for the noise parameter. The parameters of the method influence the smoothness of the resulting estimate (e.g. $h_{med}$) or reduce the potential bias due to mis-specification of the homogeneity regions (e.g. $\lambda$). The assumption of a scaled non-central chi-distribution is exclusively employed in the weighted Maximum Likelihood estimator. The method can therefore be easily adapted to alternative scaled distributional models. Our method is basically a 3D method and tries to identify the homogeneity regions of the non-centrality parameter of the data distribution. If inhomogeneity occurs across slice due to the acquisition, the method will detect it and act as if it were defined in 2D. As a consequence the estimates for the non-centrality parameter will only be slightly more variable, as the estimator then pools only from a fewer number of voxel.

In this paper we demonstrated for real and simulated data the robustness and validity of the method comparing it to the ground truth. Furthermore, we compared its performance with the intriguingly simple and fast method proposed by Aja-Fernández et al. (2013).

Noise estimation for magnetic resonance imaging data is often essential to evaluate image quality and to successfully run image enhancing methods like noise reduction. We showed in this paper, that local estimation of the noise estimation outperforms the use of a global parameter in the noise reduction method msPOAS (Becker et al., 2014) as an example. However, this improvement will also apply to other denoising methods, see e.g., Aja-Fernández et al. (2013).

An often underestimated fact is the bias introduced in parameter estimates of diffusion models for dMRI data like DTI or DKI at low SNR, see
e.g. Jones and Basser (2004). We also demonstrated, that the estimation of
the model parameters (for DTI) will benefit from the local determination of
the noise as proposed, see e.g. also Landman et al. (2009a), Veraart et al.
(2011a), Veraart et al. (2011b), or Ghosh et al. (2013) for related ideas. An
efficient solution of this problem still requires to specify reasonable values for
the (local) effective number of coils $L_i$. The use of $L_i \equiv 1$ improves the esti-
mates but only partially solves the bias problem. While the diffusion model
parameter estimation using maximum likelihood based on the non-central
$\chi$ distribution is valid for the original unsmoothed data, data processed by
non-local noise reduction methods like msPOAS follows a different signal dis-
tribution, such that a quasi-likelihood formulation of the estimation problem
is preferable.

The package primarily used for the analysis in this paper is freely avail-
able: dti (Tabelow and Polzehl, 2014).

6. Acknowledgment

We thank A. Anwander and R. Heidemannn for the permission to re-use
the diffusion weighted dataset for comparison. This work is supported by
the DFG Research Center MATHEON. We thank two anonymous reviewers
for their helpful comments.

References

Aja-Fernández, S., Brion, V., Tristán-Vega, A.. Effective noise estimation
and filtering from correlated multiple-coil MR data. Magnetic Resonance


Tabelow, K., Polzehl, J.. dti: DTI/DWI Analysis; 2014. R package version 1.2-0.

