Hahn, Klaus R.; Prigarin, Sergej M.; Pütz, Benno; Hasan, Khader M.

Working Paper

DTI denoising for data with low signal to noise ratios

Discussion Paper, No. 409

Provided in Cooperation with:
Collaborative Research Center (SFB) 386: Statistical Analysis of discrete structures - Applications in Biometrics and Econometrics, University of Munich (LMU)

Suggested Citation: Hahn, Klaus R.; Prigarin, Sergej M.; Pütz, Benno; Hasan, Khader M. (2005) : DTI denoising for data with low signal to noise ratios, Discussion Paper, No. 409, Ludwig-Maximilians-Universität München, Sonderforschungsbereich 386 - Statistische Analyse diskreter Strukturen, München,
https://doi.org/10.5282/ubm/epub.1778

This Version is available at:
https://hdl.handle.net/10419/31023

DTI denoising for data with low signal to noise ratios

Klaus R. Hahn [a,e] Sergei Prigarin, [b] Benno Pütz, [c]
and Khader M. Hasan [d]

[a] *National Research Center for Environment and Health -gsf-, Neuherberg, Germany*

[b] *Institute of Computational Mathematics and Mathematical Geophysics, Novosibirsk, Russia*

[c] *NMR Research Group, Max-Planck-Institute of Psychiatry, Munich, Germany*

[d] *Department of Radiology, University of Texas Medical School, Houston, USA*

[e] *corresponding author*
Institute of Biomathematics and Biometrics
National Research Center for Environment and Health - gsf -
D-85764 Neuherberg
Ingolstaedter Landstr. 1
Germany
email: hahn@gsf.de
Tel (Fax) : +49-89-3187-4483 (-3369)

Preprint submitted to Elsevier Science *7 January 2005*

Abstract

Low signal to noise ratio (SNR) experiments in diffusion tensor imaging (DTI) give key information about tracking and anisotropy, e. g., by measurements with small voxel sizes or with high b values. However, due to the complicated and dominating impact of thermal noise such data are still seldom analysed. In this paper Monte Carlo simulations are presented which investigate the distributions of noise for different DTI variables in low SNR situations. Based on this study a strategy for the application of spatial smoothing is derived. Optimal prerequisites for spatial filters are unbiased, bell shaped distributions with uniform variance, but, only few variables have a statistics close to that. To construct a convenient filter a chain of nonlinear Gaussian filters is adapted to peculiarities of DTI and a bias correction is introduced. This edge preserving three dimensional filter is then validated via a quasi realistic model. Further, it is shown that for small sample sizes the filter is as effective as a maximum likelihood estimator and produces reliable results down to a local SNR of approximately 1. The filter is finally applied to very recent data with isotropic voxels of the size $1\times1\times1\,\mathrm{mm}^3$ which corresponds to a spatially mean SNR of 2.5. This application demonstrates the statistical robustness of the filter method. Though the Rician noise model is only approximately realized in the data, the gain of information by spatial smoothing is considerable.

1 Introduction

The basic experimental and theoretical tools for DTI are still in a state of rapid renewal. Nevertheless, there is general agreement that water diffusion magnetic resonance imaging can contribute substantially to the solution of central issues concerning the architecture of normal and diseased brain tissue. See Le Bihan [2003] for a recent review on the physical concepts and on potential applications to neuroscience and associated clinical fields, such as neurology, neurosurgery and psychiatry.

Thermal or Johnson noise in Diffusion Tensor data is one of the obstacles which obstruct the evaluation of experiments especially in low SNR situations. Within the DTI hierarchy of variables, the complex Gaussian noise of the signals is modified to the Rician family of distributions for the Diffusion Weighted magnitude Signals or DWIs, compare for details Henkelmann [1985] or Gudbjartsson and Patz [1995], then via the Stejskal Tanner equations to the noise distributions of the tensor coefficients and to those of further derived variables, like anisotropy or main diffusion directions describing anatomical details of nerve fiber bundles. Due to this chain of nonlinear mappings noise in derived variables has a very complicated structure for low SNR. Nevertheless, the analysis of low SNR experiments is of great interest. To reduce partial volume effects, which can cause "phantom connections", see Basser and Jones [2002], between anatomically separated fiber tracts, experiments with small voxels are desirable. To separate fast and slow diffusion fractions, low SNR experiments with high b-values, as described in Clark et al. [2002] or Yoshiura et al. [2003], are to be performed.

To increase the SNR, denoising is frequently performed voxelwise. To this end, experiments with the minimal number of necessary gradients and perhaps different b-values are repeated (Basser et al. [1994], Basser and Jones [2002]). In another experimental set up, multigradient arrangements with enhanced numbers of gradients are applied (Jones et al. [1999], Papadakis et al. [1999], Skare et al. [2000a]). In both cases, the multitude of measurements reduces noise effects. Scanning time or subject motion limit these procedures.

A complementary smoothing technique is offered by spatial smoothing, where, in a single data set, small samples of neighboring voxels are used to estimate, e.g., local mean values. Spatial smoothing is rather limited, care must be taken to avoid the introduction of blurring, which can be introduced by mixing information from different tissue units, like e.g., by averaging noisy diffusion directions from spatially close but different axon bundles. As will be shown, the application of convenient edge preserving filters can reduce blurring essentially. The analysis of the statistical properties of edge preserving filters is still in a preliminary state, see Winkler et al. [1999] and Winkler [2003] for reviews

of basic approaches. These comprise Bayesian methods which are conceptual transparent, but burdened with heavy computations and with model and (hyper)parameter ambiguities, or alternatively, chains of nonlinear filters, which are very fast and give excellent results, but lack presently a comprehensive theoretical foundation (Winkler and Liebscher [2002]).

At present, spatial smoothing or regularization in DTI is discussed for different random variables, like, e. g., the DWIs (Parker et al. [2000]), the tensor fields (Basser et al. [2000], Pajevic et al. [2002]), the Eigenvalues and anisotropy coefficients (Skare et al. [2000b], Bastin et al. [1998], Pierpaoli and Basser [1996]), or the principal diffusion directions (Poupon et al. [2000], Tench et al. [2002a,b]). Applying smoothing methods directly to the variable of interest has the advantages to restrict on low dimensional fields, minimizing computer time, and to deal with interpretable quantities, supporting special adapted regularization methods (Poupon et al. [2000], Tench et al. [2002a,b]). On the other hand, for some of these variables, like Eigenvalues, anisotropy coefficients, or main directions, already in the case of voxelwise smoothing, statistical bias is reported. These noise dependend deviations between true mean value and quantity of interest are investigated in model calculations for SNR's above 20 by perturbation theory (Anderson [2001]) or above approximately 5 by Monte Carlo simulations, see, e.g., (Skare et al. [2000b], Bastin et al. [1998], Basser and Pajevic [2000]). The results indicate that the voxelwise statistics of random variables involved in a DTI analysis is in general at least for medium (approximately 5-15) and low SNR (<5) no longer Gaussian and may be better described by skewed distributions (Skare et al. [2000b]). The fact that bias and variance in simulation studies increase with 1/SNR (Bastin et al. [1998], Anderson [2001], Basser and Pajevic [2000]) further indicates, that these distributions may change from voxel to voxel and form a specific random field for every variable in the brain. This is impressively exemplified by Monte Carlo simulations presented in Skare et al. [2000a], which show that bias and variance of anisotropy can have a strong dependence on the, in practice unknown, angles between local main diffusion and the diffusion gradients applied.

In the first part of this paper DTI data are analyzed for typical geometrical features which may constrain the filter construction. Discontinuities, e. g., in the anisotropy or in the direction field, favor edge preserving filters, like those applied in Parker et al. [2000]. On the other hand, as the fields of the variables have, inside those discontinuous boundaries, continuous regions with appreciable curvature, the linear approximation quality of the filter is also important. This is a non trivial observation, as most edge preserving filters, like the diffusion equation (Parker et al. [2000]), the local M-smoother (Chu et al. [1998]) or the nonlinear Gaussian filter chain from Aurich (Aurich and Weule [1995]), achieve optimal edge resolution only for underlying piecewise constant functions and can bias curvature. It is the main goal of this paper to establish a new spatial filtering method for DTI which combines edge detection with a

good generalization quality for piecewise curved fields.

As this filter shall be applicable in low SNR situations, it is necessary to find the DTI variables with the most convenient noise distributions for spatial smoothing. Therefore, a special Monte Carlo study is performed. Low and high b-value experiments with minimal and multigradient equipment are involved. Most attention is paid to low SNR situations, however, the smoothing strategies found are also convenient for higher SNR. In case of minimal gradient experiments with constant b-values the Rician distributed DWIs offer the most regular distributions. For minimal experiments with different b-values a "back" transformation of the mean tensor to a minimal set of virtual DWIs with approximately Gaussian statistics is convenient. For multigradient experiments a more ambiguous situation is apparent. In contrast to other variables, the angle of the main diffusion shows a rather regular distribution field close to Gaussian even for low SNR and may be convenient for spatial smoothing. For the Eigenvalues however, a residual bias cannot be reduced even by very high numbers of gradients. But, as will be shown, replications of the whole experiment with moderate numbers of gradients and voxelwise DWI averaging produce convenient prerequisites for further spatial smoothing, which, followed by a bias correction, can reduce bias in all variables already for realistic samples sizes close to zero.

Based on this analysis a new three dimensional nonlinear spatial filter for DWIs is presented. Its construction goes back to a chain of nonlinear Gaussian filters estimating local mean values (Aurich and Weule [1995]). This chain, designed for piecewise constant signals, corrupted by a bell shaped noise distribution, is modified to include the conditions of piecewise curved signals with spatially varying Rician noise. Finally, a bias correction is performed. The construction is applicable to individual spatial DWI fields, but with minor modifications also to scalar fields of other variables. The filter application is performed fully automatic, the only input parameter is the standard deviation of background noise. The filter is fast and both numerically and statistically robust. This filter is validated in low SNR situations by a quasi realistic "gold standard" model, based on measured data (Hahn et al. [2001]). A comparison with a maximum likelihood estimator of the DWI via the Rician distribution family (Sijbers et al. [1998]) demonstrates the convenient statistical effectiveness of the smoothing method. A new extension of the bias correction allows reliable analysis down to local SNR$\approx$ 1. Finally, the filter is applied to very new data which are measured on an isotropic grid with a voxel volume of $1\times1\times1\,\mathrm{mm}^3$ and a spatially mean SNR of about 2.5. Though noise of this experiment is only approximately Rician, the smoothing method results in an appreciable gain of information.

2 Materials and methods

2.1 MR data acquisition

Three experimental data sets are involved in the present study. The first experiment (Exp1) was performed on a General Electric (Milwaukee, WI) Signa 1.5 T Echospeed Horizon scanner using a spin echo echo-planar sequence with TR,TE=4200,120 ms. Three replications of a minimal experiment with the gradients $\{[1,0,1],\,[-1,0,1],\,[0,1,1],\,[0,-1,1],\,[1,1,0],\,[-1,1,0]\}/\sqrt{2}$, with a b-value of $b = 880\,\mathrm{s/mm^2}$, and with a 128x128 image matrix producing 1.875x1.875 $\mathrm{mm^2}$ in plane resolution were measured. The number of axial slices was 24 with a thickness of 3 mm and 1 mm gap. The corresponding $b = 0\,\mathrm{s/mm^2}$ reference experiment was repeated twice.

The second experiment (Exp2) was performed with the same scanner like Exp1 but 4 replications with the b-values $b = 400, 500, 700, 880\,\mathrm{s/mm^2}$ and only one measurement with $b = 0\,\mathrm{s/mm^2}$were acquired. The grid is like in Exp1, the gradients were $\left\{[\sqrt{2},0,0],\,[0,\sqrt{2},0],\,[0,0,\sqrt{2}],\,[1,1,0],\,[1,0,1],\,[0,1,1]\right\}/\sqrt{2}$.

A third experiment (Exp3) was performed to test the applicability of the presented theory to very small voxel sizes. The isotropic $1\times1\times1\,\mathrm{mm^3}$ diffusion weighted data were acquired from a consented normal volunteer (F40y) on a General Electric 1.5 MRI scanner using a dual spin echo prepared diffusion sequence that utilized ramp sampling and fat suppression. The tensor encoding scheme used is the principal icosahedral (Icosa6) as described elsewhere (Hasan et al. [2001], Hasan and Narayana [2003]). The field of view is 260 $\mathrm{mm^2}$ and the data matrix is 256x256 pixels, a total of 28 contiguous axial sections covering the corpus callosum were selected from a sagittal scout localizer. The b factor is $1000\,\mathrm{s/mm^2}$, TR=4.5 seconds and TE=82 ms, and the number of replications, NEX=4. The DWIs were averaged by the scanner.

2.2 Noise, edges and curvature in DTI data

To illustrate typical features in DTI data, we present in Fig. 1 some data of Exp1 for a region around the corpus callosum within an axial slice. To reduce noise the three sets of DWIs and the two references were averaged voxelwise. The mean tensor was derived via the Stejskal Tanner equations, Eq. (2). In the left panel of Fig. 1 the tensor field component $d_{11}(\vec{x})$ is presented, see Eq. (1), in the right panel the mean DWI, $|S_4(\vec{x})|_{\mathrm{mean}}$, for $|\vec{g}_4\rangle = [0,-1,1]/\sqrt{2}$. A closer inspection of the data still shows appreciable noise with spatially varying variance, compare, e. g., in d_{11} the yellow valleys of the pyramidal tracts and

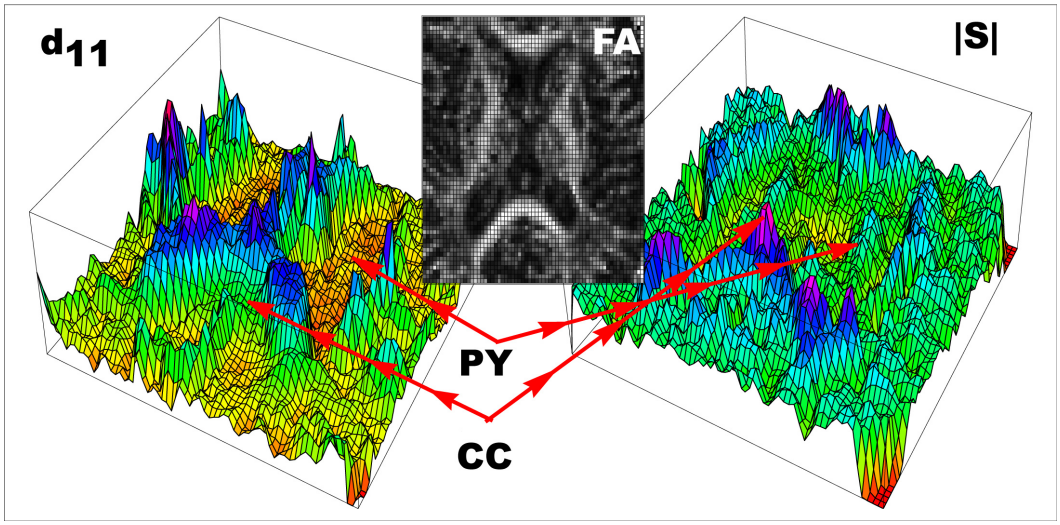

Fig. 1. The left panel shows d_{11} on an axial slice around corpus callosum (CC). Red-blue maxima are mainly due to high diffusion within the ventricles. The right panel shows the mean DWI for the gradient $[0, -1, 1]/\sqrt{2}$ in the same region, here the maxima are mainly due to the signal from diffusion in the corpus callosum. The data are denoised voxel wise via DWI averaging based on two replications for the reference experiment and three replications for $b = 880\,\mathrm{s/mm^2}$, red arrows indicate parts of corpus callosum and of pyramidal tract (PY), the few extremal voxels in the corners are outside the brain.

the green bow of the corpus callosum region in front. Especially the boundaries of the corpus callosum show steep intensity variations within very few voxels due to the ventricles, see the red-blue maxima of water diffusion in d_{11}. Such steep variations, whose width is mainly due to partial volume effects, are hereafter called edges. Similarly, in $|S|$ the water diffusion is separated from that in white matter by edges. As in the $|S|$-image large intensities correspond to small diffusion, also the transition from corpus callosum to neighboring white matter tissue is frequently stepwise, see, e. g., the steep red-blue bow of the corpus callosum DWI in the front. This is related to a steep variation of the local anisotropy or Eigenvalues when the corpus callosum is left within white matter. Nonetheless, also regions with smoothly curved intensities are apparent in both panels. This feature may be more obvious in Fig. 5, where a denoised DWI is presented.

The presence of edges is further supported by the following observation: According to empirical knowledge about the streams of axon tracks (Nieuwenhuys et al. [1983], Wiegell et al. [2000]) it is evident that neighboring nerve bundles can have strongly different main directions in white matter. This has the consequence that the Eigenvectors in regions with different fiber streams change in a discontinuous way. As the tensor is real and symmetric or Hermitean, the tensor field, $d(\vec{x})$, can be formulated as a function of the Eigenvalue and normalized Eigenvector fields $\lambda_i(\vec{x})$ and $|i(\vec{x})\rangle$ in braket notation (Messiah [1972]); where the ket, $|\ldots\rangle$, is a three dimensional vector, the bra, $\langle\ldots|$, its

transpose and $\langle \ldots | \ldots \rangle$ a scalar product.

$$d(\vec{x}) = \begin{pmatrix} d_{11}(\vec{x})\ d_{12}(\vec{x})\ d_{13}(\vec{x}) \\ d_{21}(\vec{x})\ d_{22}(\vec{x})\ d_{23}(\vec{x}) \\ d_{31}(\vec{x})\ d_{32}(\vec{x})\ d_{33}(\vec{x}) \end{pmatrix} = \sum_{i=1}^{3} \lambda_i(\vec{x}) |i(\vec{x})\rangle\langle i(\vec{x})| \tag{1}$$

This representation implies that spatial discontinuities in the Eigenvalues or Eigenvectors map to the tensor field and consequently to other derived quantities like the anisotropy coefficients. They will in general also transform to the DWIs which are connected to $d(\vec{x})$ by the Stejskal Tanner equations.

$$|S_j(\vec{x})| = f|S_0(\vec{x})| \exp\left(-b \sum_{i=1}^{3} \lambda_i(\vec{x}) \langle i(\vec{x})|\vec{g}_j\rangle^2\right) \tag{2}$$

Where b is the b-value of the experiment, f a volume fraction, which is only for slow diffusion components different from unity, $|\vec{g}_j\rangle$ a normalized diffusion gradient and $|S_j(\vec{x})|$ and $|S_0(\vec{x})|$ are the DWIs and reference .

2.3 The random fields of the DTI variables

According to Jensen's inequality for concave or convex transformations of random variables (Feller [1971]), statistical bias effects like those discussed in the Introduction are to be expected for the nonlinearly transformed random variables of DTI. The mentioned results of Monte Carlo simulations, where noise propagation in DTI variables is studied for medium and high SNRs further indicate that the random properties of DTI variables should be described by locally varying random fields in the brain. This is evident from the structure of the Stejskal Tanner equations, see Eq. (2). The exact magnitude signals without noise, $\mathrm{DWI}_{\mathrm{exact}}$, depend on the local angles between diffusion and gradients and on the local Eigenvalues or diffusion strengths. When noise is introduced, this dependence is mapped to the specific DWI distributions within the Rician family determined by the SNR= $|S_{\mathrm{exact}}|$ / σ_0, where σ_0 equals the Rayleigh corrected standard deviation of background noise. Superpositions of the logarithm of noisy DWIs determine the tensor coefficients and so carry over their local dependence on diffusion and gradients to the statistical properties of derived random variables, like anisotropy.

2.4 A Monte Carlo simulation of the random fields

An unbiased, and spatially uniform bell shaped distribution for a noisy variable would be ideal for any attempt to apply spatial smoothing. In this context Monte Carlo simulations are performed, to quantify the random field properties. The tensor model describes a three dimensional cigar-shaped diffusion ellipsoid, where the two larger axes rotate within the x-y plane of the laboratory system around the z axis. The tensor has the form:

$$d = \begin{pmatrix} d_{11} & d_{12} & 0 \\ d_{21} & d_{22} & 0 \\ 0 & 0 & \frac{d_{11}+d_{22}-\sqrt{(d_{11}-d_{22})^2+4d_{12}^2}}{2} \end{pmatrix}, \tag{3}$$

the element d_{33} is the smaller Eigenvalue of the upper submatrix. As all unknown parameters are contained in this submatrix, the simulated measurements are performed only in the x-y plane, i.e. i=1,2 in Eq. (2); for a minimal experiment only 3 plane gradients are necessary. The diffusion model, Eq. (3), is probed in low SNR situations.

For fixed diagonal tensor coefficients d_{11} and d_{22} the off diagonal d_{12} is changed in 50 equidistant steps within the interval $d_{12} \in \left[-0.8\sqrt{d_{11}d_{22}},\, 0.8\sqrt{d_{11}d_{22}}\right]$. This variation serves as a simple model for spatially varying diffusions within the brain. The interval is defined by the positive definiteness condition of the tensor and covers, without the factor 0.8, all possible off diagonal diffusions. The factor is introduced to limit the ratio of it's sorted Eigenvalues $\lambda_1 > \lambda_2$ to $\lambda_1/\lambda_2 < 10$, preventing unrealistic high anisotropy. Low and high b-value experiments are simulated via Eq. (2). For the first group of experiments b=900 s/mm^2 and f=1 are applied with the diagonal components d_{11}=.00156 mm^2/s and d_{22}=.00084 mm^2/s to achieve a realistic mean diffusity of $(d_{11}+d_{22}+d_{33})/3 \approx$.0009 mm^2/s (Bastin et al. [1998]). For b=3500 s/mm^2 the procedure of Clark et al. [2002] was used. There, the slow tensor component is fitted by a monoexponential model to the signal decay within a high b-value interval. The fraction, f, of the slow component is achieved by an extrapolation of high b-value data. The slow diagonal diffusion for the Monte Carlo simulation is defined to be d_{11}=.0007 mm^2/s and d_{22}=.00035 mm^2/s with f= .31. This is in agreement with the fraction and approximately with the slow three dimensional mean diffusity of .00035 mm^2/s found in corpus callosum by Clark et al. [2002]. For both models, the fractional anisotropy FA is approximately in the range FA $\in$ [.4, .9]. The reference signal was assumed to be $|S_0| = 1000$.

In recent experimental arrangements the frame of minimal experiments is extended to more gradients. In this context Jones, Papadakis and Skare discuss the propagation of noise from the DWIs to the tensor coefficients for higher numbers and varying directions of gradients (Jones et al. [1999], Papadakis et al. [1999], Skare et al. [2000a]). Skare proposes a condition number, *cond*, depending only on the gradient directions, which controls an error in the tensor components due to noise in the signals. In our simulations uniformly distributed gradient directions are applied which were proposed in Jones et al. [1999] with a minimal condition number, $cond = 1.4$, for a plane measurement. All applied gradient sets include the direction $\vec{g}_{12} = [1, 1]/\sqrt{2}$ which is used as reference projection in the graphical presentations.

To introduce thermal noise, noise simulated by a complex Gaussian distribution with standard deviation σ_0 is added to the signals independently for every gradient direction, see, e. g., Skare et al. [2000b] or Pierpaoli and Basser [1996] for a precise algorithmic definition of the procedure applied. Different noise levels with standard deviations σ_0 are involved in the simulations. Experiments for b= 900 s/mm^2 are performed with σ_0 = 30n, with n=2...6. The slow diffusion component was investigated with σ_0= 15, 20, and 30. As a SNR=$|S_0|/\sigma_0 = 33$ corresponds for a 1.5 T DTI scanner approximately to a voxel size of 2×2×3 mm^3, the voxel sizes investigated vary for b=900 s/mm^2 approximately from 12 to 2 mm^3 and for b=3500 s/mm^2 from 24 to 12 mm^3.

A large number of Monte Carlo simulations was performed to explore the systematics in the distributions and to find the most convenient variables for spatial smoothing. Typical results and the consequences for filtering are presented and discussed in the sections Results and Discussion.

2.5 A nonlinear spatial filter

For spatial smoothing, we propose to apply a chain of nonlinear Gaussian filters, which goes back to Aurich and Weule [1995], and estimates the mean of a noisy scalar intensity function $f(x), x \in R^n, n$= dimension of space. The chain iterates nonlinear filters which combine a spatial window, Φ, and an intensity window, Ψ. The corresponding parameters vary according to special rules by which edge detection and generalization or smoothing quality are balanced. After the presentation of the standard version, several new modifications which are convenient for spatial DTI filtering will be introduced.

One filter step is defined by :

$$F_\eta^\mu \circ f(x) = \frac{\sum\limits_{y \in \text{neighborhood of } x} \Phi(x,y)\Psi(f(x),f(y))f(y)}{\sum\limits_{y \in \text{neighborhood of } x} \Phi(x,y)\Psi(f(x),f(y))} \tag{4}$$

with $\Phi(x,y) = \exp(-\frac{(x-y)^2}{2\eta^2})$ and $\Psi(f(x),f(y)) = \exp(-\frac{(f(x)-f(y))^2}{2\mu^2})$

it's k iterations are defined by :

$$f_k^{\text{smooth}}(x) = F_{\eta c^{k-1}}^{\mu d^{k-1}} \circ \ldots \circ F_{\eta c}^{\mu d} \circ F_\eta^\mu \circ f(x) \tag{5}$$

The number of iterations, k, the spreading factors of the windows, c and d, and the initial width of the spatial window, η, are regarded as parameters. The initial intensity window width is $\mu = 3\sigma$, where σ =standard deviation of noise on f.

To enable an adaptive application of the filter for the reader, a short presentation of the basic parametrization strategy will be given. From the construction of the nonlinear window, Ψ, it is clear that for vanishing noise, the filter result converges to the input data. Regard now an intensity step H in an elsewhere constant spatial function which is corrupted by Gaussian noise with standard deviation $\sigma > 0$. The edge to noise ratio, ENR, is given by

$$ENR = H/2\sigma \tag{6}$$

Iterating the chain k times reduces σ by $1/\alpha^k$ and allows for large samples an edge resolution down to

$$ENR = 1/\alpha^k \tag{7}$$

where $\alpha > 1$ is a speed factor determining the parameters c, d, and η via: $c = \sqrt[n]{\alpha^2}$, $d = 1/\alpha$, and $\eta = \frac{\sqrt[n]{4\alpha^2}}{2\sqrt{\pi}} \cdot \Delta_g$, where Δ_g is the smallest width in the n-dimensional anisotropic or isotropic voxels. Frequently k is chosen so high that the filter reaches its fix point, i. e., that further iterations do not change the results. Due to the nonlinearity of the chain, this parametrization could be theoretically justified up to now only for stepwise constant signals with Gaussian noise, see Mühlhaus [1997] for more details.

It will be shown in the section Results that in the present context the filter is mainly applied to noisy DWI fields. As magnitude signals are piecewise curved,

and as their noise is Rician, several modifications of the basic construction are introduced. In Rician distributions, for low SNR, $\sigma = \sigma(\mathrm{MNR})$, where MNR= $|S|_{\mathrm{mean}}$ / σ_0. Therefore a heteroscedastic $\mu = 3\sigma(\mathrm{MNR})$ is introduced into Ψ. Increasing k and α increases the edge resolution capacity, but, favoring estimates with curvature = 0, it reduces the ability to approximate curved functions. Therefore a convenient balance suited for DTI data must be found by data or model validation. A further improvement of the generalization properties can be achieved by the use of $f(x)$ instead of $f_{k-1}^{\mathrm{smooth}}(x)$ in the last iteration, see Winkler et al. [1999]. As the present filter produces finally an estimate of the mean value of the magnitude signal, the Rician bias correction should be applied for low SNR's to the resulting $f_k^{\mathrm{smooth}}(x)$.

In Fig. 4 applications of this modified filter chain to simulated one dimensional noisy DWIs are presented in panels a), b) and c). Denoising of functions with two different shapes is performed with the same parameters k=4 and $\alpha = \sqrt{2}$ to demonstrate robustness of the method. The intermediate and final filter results of the iterations are shown. The final mean value estimate of the chain (yellow) is bias corrected (violet), the corresponding true model is indicated by a black line. Different sample sizes=$n_R \times 500$ are involved, Δ_g is constant for all calculations a) - d). The mean value estimates show a good relative precision also for skewed noise in low SNR situations, as the consecutive bias corrected magnitude signals agree in general well with the model. For SNR$<$ 1, due to decreasing statistical stability of the bias correction, see *section* 3.1.3, the estimation errors increase. Note the improvement of generalization in panel b) for n_R=4 compared to panel a). Discontinuities are already well reproduced in panel a). Panel c) shows an application to stepwise constant signals with partly very low SNRs. The transition to n_R=4, not shown, improves the generalization again but not the resolution of the step at ENR=1/2. To achieve that, an essentially higher n_R would be necessary. In panel d) a linear Gaussian filter is applied to the n_R=4 situation. The blurring effects at the edges over a scale much larger than Δ_g are inherent to linear spatial filters.

To cover the whole brain, in DTI experiments frequently anisotropic voxels with enlarged grid in axial direction are applied. It is more convenient for spatial smoothing to use isotropic voxels which increase the local sample sizes and thus improve the mean value estimates. This is due to the more symmetric grouping of the data enabling a more effective localization of the spatial windows. Of similar importance is the dimension n. With n, the local sample sizes grow and therefore the widths, ηc^k, of the spatial windows can be reduced leading to a better localization of the data necessary for averaging.

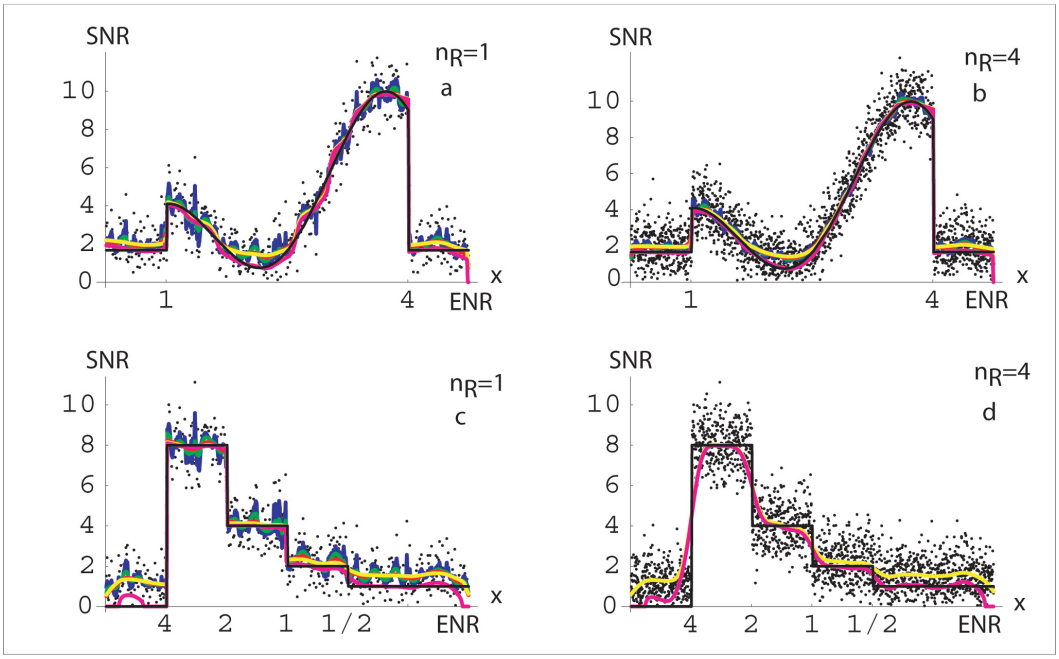

Fig. 2. Denoising of magnitude signals for dimension $n = 1$ is illustrated. On the data points of panels (a-c) the nonlinear filter chain is applied. The results of the consecutive iterations (k=4) are given by blue, green, red and yellow curves, the final bias corrected approximation is violet. True signals are indicated by black lines. The individual graphs give signal to noise ratios (SNR) versus space coordinate (x) and edge to noise ratios (ENR) at discontinuities. In panel (d) a linear Gaussian filter followed by the bias correction is applied. The sample sizes are $n_R \times 500$.

3 Results

3.1 Results of the Monte Carlo simulations

3.1.1 The investigated distribution properties and spatial smoothing

In the simulations the distributions of the following DTI variables are investigated : $|S_{12}|$, d_{12}, the Eigenvalues $\lambda_1 > \lambda_2 = \lambda_3$, the Trace= $\frac{\lambda_1+\lambda_2+\lambda_3}{3}$, the fractional anisotropy FA (Basser [1995]) and the angles α, of the main Eigenvectors. The direction of a noisy main Eigenvector is defined only modulo π, therefore it is aligned to the exact main Eigenvector, before its angle with [1,0] is calculated. The distribution of any variable q is characterized by a measure of the bias effect and by higher moments μ_n about the mean $mean_q$. In detail, these are the relative bias $(mean_q - q_{\text{exact}})/q_{\text{exact}}$, the bias uncertainty $\sqrt{\mu_2}/|q_{\text{exact}}| = \sigma_q/|q_{\text{exact}}|$, the skewness μ_3/σ_q^3, the normalized standard deviation of noise quantifying heteroscedasticity $\sigma_q/\max(\sigma_q)$, and finally the kurtosis $(\mu_4/\sigma_q^4 - 3)$ measuring the peakedness of the distribution or the degree of tail heaviness. To avoid singularities, the angular bias is quantified

by $(mean_\alpha - \alpha_{\text{exact}})/10^o$, the uncertainty by $\sigma_\alpha/30^o$, the other moments for $q = \alpha$ are calculated like above. All moments are plotted in the Figures versus the SNR $= |S_{12}^{\text{exact}}|/\sigma_0$, varying with the diffusion coefficient d_{12}. The quantity $|S_{12}^{\text{exact}}|$ is derived by the gradient $\vec{g}_{12} = [1, 1]/\sqrt{2}$ from the model diffusion not distorted by noise, see Eq.(3). The distribution properties are chosen with respect to their importance for spatial smoothing. As the discussed filters estimate the mean values, the bias and the possibility of its correction by a convenient parametrization is important. However, as bias is an asymptotic quantity and as spatial or voxel wise smoothing is performed on rather small samples, the bias uncertainty is also discussed. Reduced uncertainty of a variable is also essential for nonlinear filters, as their ability to detect edges increases with reduced noise. Skewness measures deviations from symmetry. Enlarged skewness combined with kurtosis>0 increases the risk of outliers which may corrupt the estimated mean. Such situations demand large sample sizes which may not be available. Kurtosis is zero for Gaussian and 6 for exponential, skewness is zero for Gaussian, 0.6 for Rayleigh, and 2 for one sided exponential distributions. As spatial filters need information about the noise level, spatially uniform noise, which can easily be estimated from the data, is convenient. Therefore, heteroscedasticity and the possibility of its parametrization which can be used as filter input is also discussed.

3.1.2 Results for minimal experiments

In Fig. 3 A, B, C results of experiments with three uniformly distributed gradients in the x-y plane (*cond*=1.4) and b=900 s/mm^2 are presented, Fig. 3 D presents the same minimal experiment for b=3500 s/mm^2. The moments are derived via Monte Carlo sampling with 50 000 iterations per parameter combination or model diffusion. Results for the voxel volume 4 mm^3 are presented in Fig. 3A, C and for 12 mm^3 in Fig. 3B, D. The vertical bars at SNR=1.8, 3.8, and 11.3 indicate the transition d_{12}=0 mm^2/s. On the left sides of these transitions the exact main diffusion directions increase their angles with [1,0] from 0° to 35°, on the right sides these angles decrease from 0° to –35°. Anisotropy, FA $\in$ [.4, .9], is roughly proportional to the magnitude of the angles.

Due to the convenient condition number for the gradient set the relative bias effects in Fig. 3 A/a are quite small. For a different gradient set, e. g.,$\{[1, 0]$, $[0, 1]$, $[1, 1]/\sqrt{2}\}$, the bias in all variables is nearly doubled. However, the bias uncertainty, see Fig. 3 A/b, can deteriorate single measurements appreciably. The maximum uncertainty for FA is close to 60% and that of the angle close to 25°, which can produce severe errors in special tracking algorithms, see Anderson [2001]. In Fig. 3 B/a,b both quantities are reduced due to the enlarged SNR. In all variables of Fig. 3 A, except $|S_{12}|$, skewness is appreciable, the absolute values of skewness are at least close to that of a Rayleigh distribution,

see Fig. 3 A/c. In Fig. 3 B/c the skewness of d_{12}, λ_1 and α comes close to 0.6. Kurtosis for both experiments is essentially positive for all variables. The corresponding patterns in the plots, not shown, are similar to those of skewness, maximum kurtosis for λ_1 is ≈ 6 (A) and ≈ 2.5 (B). In both experiments a kurtosis peak for $\alpha \approx 0°$ is striking, the heights of the peaks are ≈ 4 (A) and ≈ 1 (B). Heteroscedasticity in Fig. 3 B/d can become even larger than that in Fig. 3 A/d, the exception is again $|S_{12}|$. A minimal experiment for b=3500 s/mm^2 with a voxel volume of 12 mm^3 is shown in Fig. 3 D. Due to the enhanced b-value and the weighting factor the SNR scale is reduced compared to that of Fig. 3 B and is close to that of Fig. 3 A. The results in Fig. 3 D and Fig. 3 A are similar in the uncertainty, skewness and heteroscedasticiy, the relative bias is however appreciably enhanced, kurtosis is for λ_1 between 2 and 3.

These results, presented in Fig. 3 A, B, D, together with those mentioned in *section* 2.4, suggest the following: At mean and low SNR all variables, except, perhaps, the tensor Trace, show the features of a locally varying random field, when variation in SNR is identified with local variation of diffusion in the brain. Their deviations from spatially uniform bell shaped distributions increase with 1/SNR, the quantitative ranges of the central moments cause doubts on the possibility to apply spatial smoothing in a straightforward manner. Only the Trace, at least for b=900 s/mm^2 , shows convenient features for direct spatial smoothing, though skewness may be problematic for too low SNR. However, the information contained in the Trace is rather limited.

The DWIs are the only variables where the distributions are explicitly known and can be simply parametrized via SNR or MNR, the statistical bias in any DWI can be corrected, when convenient mean value estimates are available. Concerning skewness, above SNR$\approx$4 the DWIs are symmetric, below, the maximal skewness of $|S_{12}|$ in Fig. 3 A is well below the corresponding maxima of the residual variables. In Fig. 3 D this maximum is at least below those of λ_1, Trace, and d_{12}. Kurtosis is ≈ 0 in all cases. Heteroscedasticity is minimal for $|S_{12}|$ in Fig. 3 A, in Fig. 3 D λ_1 and Trace are more uniform. However, as in DWIs the variance scales with MNR, varying variance can be incorporated approximately into spatial filters for DWIs as will be shown below. Concluding, in low SNR situations of minimal experiments the DWIs may be convenient for spatial smoothing. Also in situations with medium SNRs, see Fig. 3 B, the spatially uniform Gaussian distribution with neglectable bias shifts make them to the preferable candidates. Finally, as DWIs are basic in the DTI chain of variables, "good" approximations to the true DWIs should be sufficient to derive any residual variable in "good" approximation, due to continuity of the mappings between the variables.

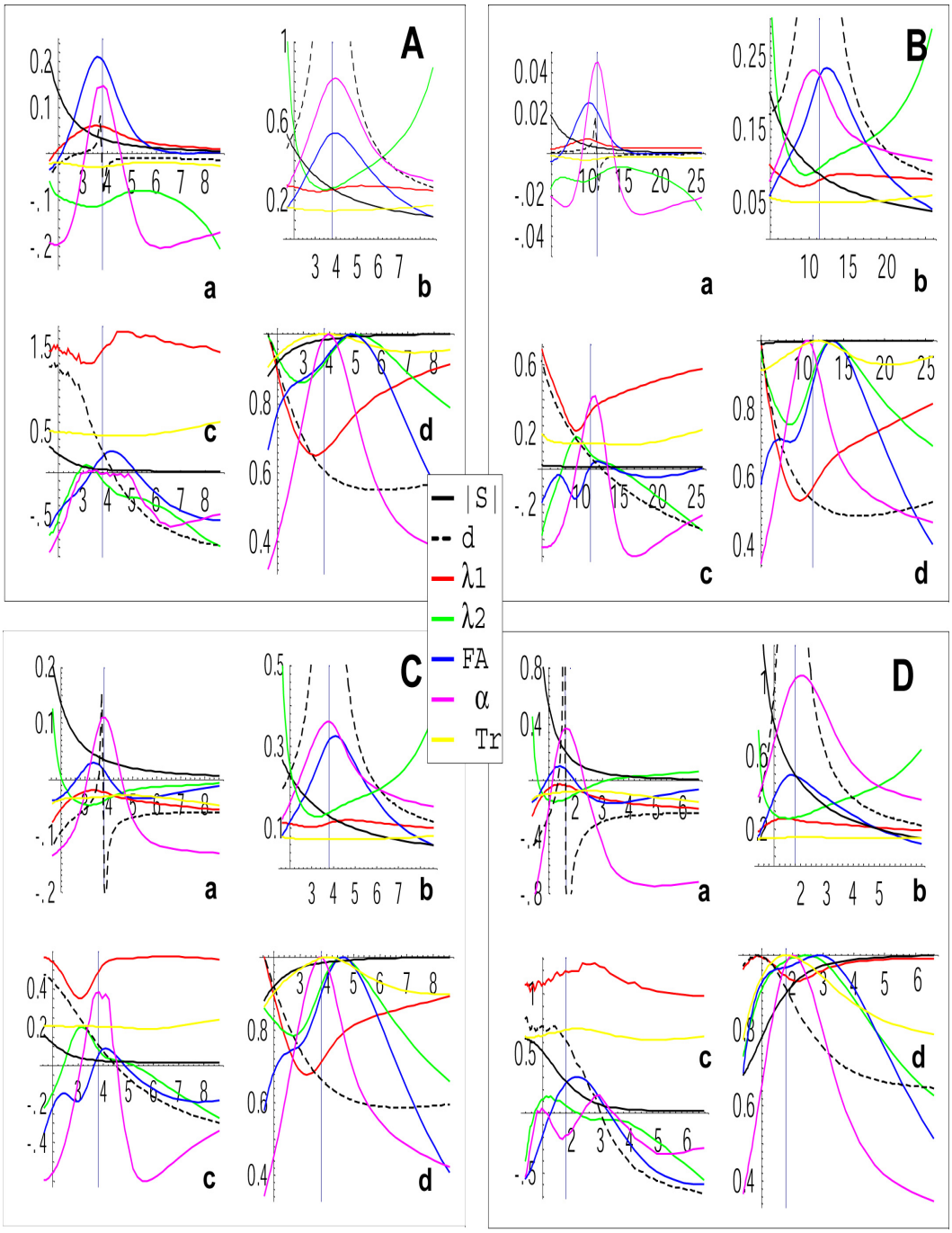

Fig. 3. Monte Carlo results for minimal experiments with $b = 900\,\mathrm{s/mm^2}$ in panels A/B/C and with $b = 3500\,\mathrm{s/mm^2}$ in D are presented. Relative bias (a), uncertainty of relative bias (b), skewness (c) and heteroscedasticity (d) of the variables $|S_{12}|$, d_{12}, λ_1, λ_2, FA, α, and Trace (see color legend) are plotted versus SNR. In experiments A/C a voxel volume of $4\,\mathrm{mm^3}$, in B/D of $12\,\mathrm{mm^3}$ is involved. For panel C the experiment A is repeated $n_R = 4$ times, the variables are derived from the averaged DWIs.

3.1.3 Results for experiments with voxelwise denoising

To increase the SNR, which would improve the conditions for spatial smoothing, essentially two variants of voxelwise denoising are in use. First, experiments with low numbers of gradients are repeated to reduce noise by voxelwise averaging (Anderson [2001]). In the second variant, multigradient experiments are applied (Jones et al. [1999], Papadakis et al. [1999], Skare et al. [2000a]).

Concerning the first variant for constant b-values, we assume that the minimal experiments discussed above are repeated n_R times. The naive transition to the averaged random variable of interest (e. g., $\widetilde{\mathrm{FA}} = \sum_{i=1}^{n_R} \mathrm{FA}_i/n_R$) reduces its uncertainty and skewness by a reduction factor $1/\sqrt{n_R}$ and also kurtosis by $1/n_R$, but not bias and heteroscedasticity. A convenient bias and variance parametrization is only known for DWIs. Therefore, the conditions for further spatial smoothing are in general not improved in an optimal way.

When averaging is performed on the DWIs, and when all residual variables are derived from these estimates of $|S|_{\mathrm{mean}}$, we find reduction of uncertainty, skewness and kurtosis in nearly all variables. The effect on bias is ambiguous. At least partially it is reduced in derived variables, but, due to a more peaked distribution around the biased $|S|_{\mathrm{mean}}$, bias for derived variables can even be increased by denoising of DWIs. See Fig. 3 C for such results, they are derived from experiments presented in Fig. 3 A, $n_R = 4$. Increased bias can be found e.g. for the Eigenvalues λ_1 and λ_2 for high FA and low SNR or in the tensor d_{12} in Fig. 3 C/a. This indicates that in addition to denoising a bias correction (*bc*) is important. Due to the reduction of uncertainty and skewness, and the parametrization of heteroscedasticity the statistics of the DWIs in Fig. 3 C complies well with further spatial smoothing increasing an effective n_R. To estimate the improvement, assume for simplicity strict homogenuos diffusion within the $n_F - 1$ nearest voxels surrounding the voxel of interest and perform unweighted spatial averaging on the mean DWIs based on n_R experimental replications. The result would be equivalent to the result for $n_{\mathrm{effective}} = n_R \cdot n_F$ experimental replications. Relative bias and bias uncertainty for an effective $n_R = 16$ applied to the experiment of Fig. 3 A and DWI averaging are presented in Fig. 4 A/a,b. To the estimates of $|S|_{\mathrm{mean}}$ in Fig. 4 A/a the Rician *bc* relating MNR to SNR, see Fig. 4 A/d, is then applied. Note, that in this correction the noise level before denoising, that of Fig. 3 A, must be applied. All variables which are derived from these estimates of $|S|_{\mathrm{exact}}$ are then approximately unbiased, see Fig. 4 A/c. The precision of this *bc* is limited by the standard deviation (std) of $|S|_{\mathrm{mean}}$, $\sigma(|S|_{\mathrm{mean}}) = \sigma/\sqrt{n_R}$, where σ equals the std of the DWI before denoising, see Fig. 4 A/d. For $\sigma(|S|_{\mathrm{mean}}) \to 0$ also for the variables the bias $\to 0$. In case of low SNR's and a too small n_R the uncertainty can create poor mean value estimates mapping to $|S|_{\mathrm{exact}}$ ≈ 0 with singular tensor components, see Eq. (2), and heavy tail statistics, in other cases they may be even outside the domain of the mapping. Statistics,

limitations and extensions of *bc* are discussed in *section* 4.1 of Discussion.

In another type of data acquisition, minimal experiments are repeated with different b-values. This arrangement goes back to Basser et al. [1994]. In numerous simulations the system of Stejskal Tanner equations was solved for the mean tensor by multivariate linear regression without the weighting proposed in Basser et al. [1994], as bias and skewness in the tensor distributions were then considerably smaller. A typical result for a voxel size of $5.3\,\mathrm{mm}^3$ and b-values b_1=500, b_2=700, and b_3=$900\,\mathrm{s/mm}^2$ is presented in Fig. 4 B, the tensor components d_{11}, d_{12} and d_{22} are shown. To perform additional spatial smoothing a back transformation proposed in Hahn et al. [2001] is convenient to reduce the noise distortions caused by the Stejskal Tanner equations. A transformation from the mean tensor to virtual DWIs via Eq. (2) with $b = (b_1 + b_2 + b_3)/3$ and the same gradients as in the "real" experiment produced in all cases which were studied DWI distributions with clearly better prerequisites for spatial smoothing than those of the tensor components, see Fig. 4 B for a comparison between tensor and virtual DWIs.

In case of multigradient experiments with constant b-value the multitude of DWIs reduces noise voxelwise in the tensor calculated via singular value decomposition. In Fig. 4 C an experiment like in Fig. 3 A but with 12 uniformly distributed gradients is presented. In this experiment any diffusion direction is projected to (practically) every orientation in space via the 12 gradients. This prevents a limitation to projections with decreased SNR only, see Eq. (2), which can enhance bias in minimal experiments, see Fig. 3 C, for a minimal experiment with the same number of DWIs. The relative bias in Fig. 4 C/a is well reduced. Skewness is in general similar to that of the replicated experiment, see Fig. 3 C/c and Fig. 4 C/c, however, the skewness of the angle is essentially reduced in the multigradient experiment. Less perfect is kurtosis, again a peak (≤ 2) for $\alpha \approx 0^o$ is apparent. These moderate tails have their origin in the rather steep slopes of α, when d_{12} crosses zero. The second moments, see Fig. 4 C/b,d are close to those of the experiment presented in Fig. 3 C, they are mainly determined by the number of DWIs. With respect to further spatial smoothing, the situation is ambiguous. Noise is not reduced in the DWIs and back transformations to virtual DWIs did not bring essential improvements. Though the bias is close to zero for the derived variables, the other conditions are at least not optimal. Skewness for d_{12} and λ_1 can become effective, heteroscedasticity for α, FA, λ_1 and d_{12}, kurtosis for α, see Fig. 4 C/c,d. On the other hand, the relative noise level of the variables, see Fig. 4 C/b, is low, so that the effect of heteroscedasticity on spatial filtering may be of second order. Concluding, we find that α may be convenient for spatial smoothing, followed by the Trace which again reveals a very uniform behavior.

Application of more gradients reduces uncertainty, skewness and also kurtosis, but the relative bias remains essentially stable due to the remaining tensor

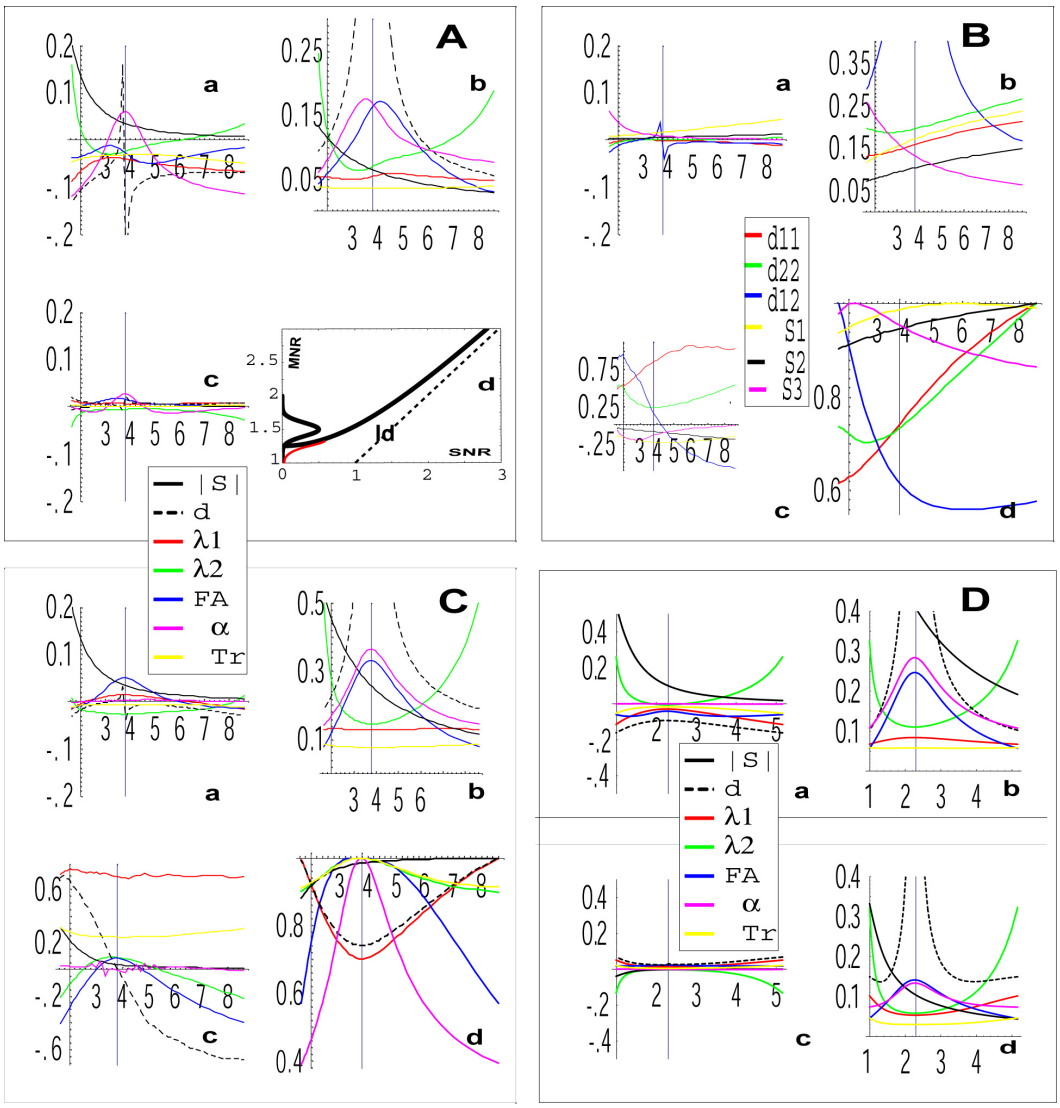

Fig. 4. Replications and multigradient experiments with voxel volumes 4 and 2.5 mm^3 are presented in A-C and D. Experiment for A like in Fig. 3 C, but n_R=16. Bias (a), bias uncertainty (b) and corrected bias (c) are plotted, see color legend for variables. In (d) the bias correction, mapping MNR to SNR, is illustrated. The distribution, centered at MNR=1.5, indicates the uncertainty in MNR, the red curve segment an extension of the bias correction. In B the averaged tensor components and virtual DWIs, see color legend, for minimal experiments with b =500, 700 and 900 s/mm^2 are shown. The moments in (a-d) are like in Fig.3. In C a 12-multigradient experiment is presented. For the variables see color legend, the moments are like in B. In D bias and uncertainty for experiments with very low SNR (1-5) are shown. A 60-multigradient experiment in a,b; in c,d a 12-multigradient experiment with 5_Rx4_F replications and *bc*.

bias. The reduction of kurtosis in α is however quite slow, as residual noise in d_{12} is very effective. An example for very low SNRs is given partly in Fig. 4 D/a,b for 60 multigradients. The relative bias with uncertainty for experiments with b=900 s/mm^2 and a voxel volume of 2.5 mm^3 is presented, the corresponding SNR range is 1–5. The relative bias remains when we enlarge the number of gradients, up to 200 were applied numerically. As DWI denoising permits a general way to a *bc*, a mixed procedure is plausible. Such a multigradient arrangement with only a moderate number of gradients and several replications of the whole experiment with subsequent spatial DWI smoothing plus *bc* is exemplified in Fig. 4 D/c,d, where for a b value and a voxel volume like in Fig. 4 D/a,b 12 multigradients, an effective $n = 5_R$ x 4_F, and *bc* are applied. When n is increased, this bias converges to zero for all variables. For comparing analysis of higher moments and estimation of realistic numbers n, see section Discussion.

3.1.4 Negative Eigenvalues

In experiments with low SNR noise can produce large fractions of voxels where the tensor violates positive definiteness and can no more be interpreted as a quantity describing diffusion. Especially in regions with large ansiotropy the smaller Eigenvalues are close to zero and frequently become negative by noise distortions. It was proposed in Ahrens et al. [1998] to constrain the Eigenvalues positive within the least square algorithm which derives voxelwise a mean tensor. The presented combination of voxelwise and spatial smoothing with *bc* should offer a basic solution to this problem, which is within our model related to the bias of λ_2. In the simulations of minimal experiments with b=900 s/mm^2 and voxel sizes 2.5, 3, 4, and 6 mm^3, the fractions of negative definite tensors are .19, .15, .11, and .05 when voxels with $FA_{\text{exact}} \geq 0.8$ are considered. When we apply 12 multigradients the fractions are below .03. This is in line with the mainly negative or positive bias of λ_2 for high FA in the corresponding minimal or multigradient experiments before additional denoising. After voxelwise denoising by DWI averaging without any discrimination between "negative" or "positive" voxels via 20 effective replications for the smallest voxel and 16 else, all fractions are below $< 10^{-3}$, as denoising alone produces positive bias in all experiments. Note, that due to the hermitean properties of the tensor, the main directions are not corrupted by the bias in λ_2, see panels a,b in Fig. 3 A,C and in Fig. 4 A. Applying *bc*, the fractions, ordered like above, are .04, .02, .002 and 10^{-5} for the minimal experiments and $<$ .003 for 12 multigradients. When the number of effective replications involved is roughly realistic and when we can trust the bias correction, it seems that the problem of low diffusions can be approached in an essentially data driven way. In this context the possibility of a sorting bias for noisy Eigenvalues is frequently discussed in model simulations, see, e. g., Skare et al. [2000b] or Basser and Jones [2002]. In our simplified model this effect could not be studied as the diffusion is

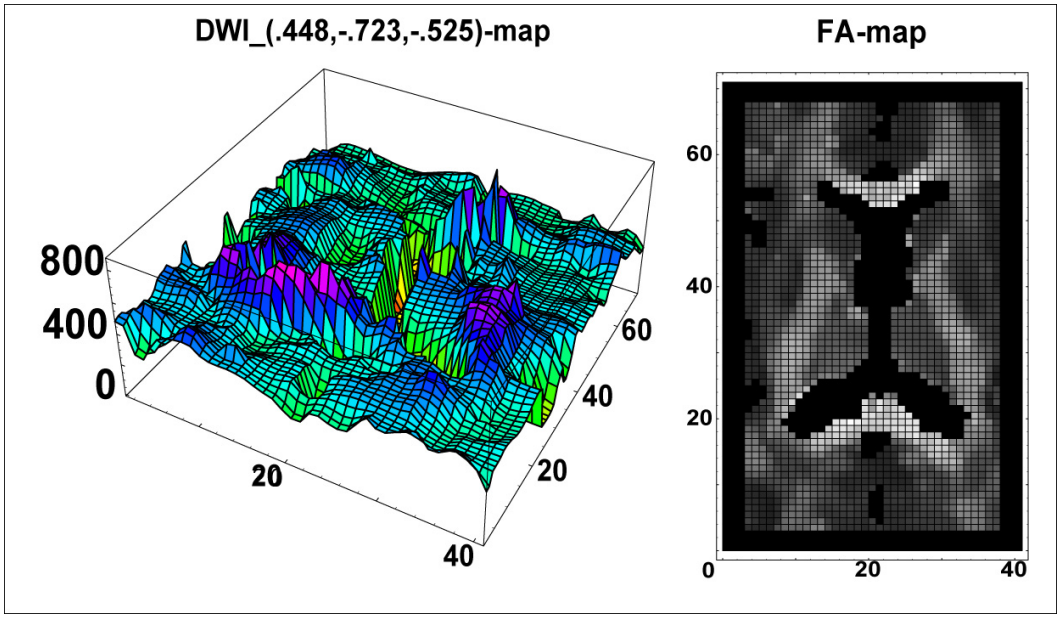

Fig. 5. Left panel: An axial DWI-map of the gold standard tensor for the gradient indicated, else see text; right panel shows the FA-map of this slice. Voxels which are excluded from the error analysis are colored black.

effectively two dimensional and therefore the Eigenvalues are always sorted correctly. The mentioned bias reduction should also reduce these effects in arbitrary three dimensional situations.

3.2 Validation of the spatial filtering method

The Monte Carlo simulations give indications for convenient strategies how to apply spatial filters in low and medium SNR situations. To validate the DWI based method a quasirealistic anatomical model is introduced. It is derived from the data of Exp2 with a voxel size $1.9\times1.9\times4\,\mathrm{mm}^3$, see *section* 2.1 for more details. A smoothed tensor is calculated from the data as described in Hahn et al. [2001] and is used as "gold standard". In a region around the corpus callosum with a size of $71\times41\times12$ voxels this tensor is transformed to the DWIs by Eq. (2) for b=900 s/mm^2, $|S_0|$=1000, and for six gradients proposed by Jones (Jones et al. [1999]), *cond* ≈1.5. The DWIs are then distorted by noise as described in *section* 2.4 and three dimensional versions of the filter are applied to the whole region. To achieve an optimal treatment of curvature and edges in the DWIs a series of tests with the gold standard diffusion, not presented in detail, lead to the following optimal filter parametrization: $k = 3$, $f_{k-1}^{\mathrm{smooth}}(x) = f(x)$, $\mu \in \{3\sigma_0, \sigma_0, \sigma_0/2\}$, and η and c for α=2; notation like in *section* 2.5.

Five different noise levels are investigated, σ_0=60, 90, 120 and 5 replications

σ_0	60	90	120	5×120	5×150
filter-mean [%]	20,71,90	11,58,83	6,48,78	28,79,94	23,74,92
noisy-exact [%]	0,3,18	0,0,3	0,0,1	0,5,24	0,1,10
filter-exact [%]	18,68,89	7,50,78	2,30,65	8,58,86	1,31,71
(filter+bc)-exact [%]	19,70,90	9,54,80	5,42,73	23,74,91	16,65,87
$n_{effective}$	5.3	9	10.5	30	30
SNR	18 (7.8)	16 (5.2)	13 (3.9)	21 (3.9)	17 (3.1)

Table 1
For five noise levels applied to the "gold standard" signal, numbers of voxels in percent are given in rows 2-5. Row 2: voxels with relative errors 0.05, 0.1 and 0.15 between filter estimates and true mean values. Row 3-5: same, for errors between quantities indicated. Row 6: numbers of equivalent experimental replications. Row 7: SNR after spatial smoothing (and of the noisy DWIs).

for 120 and 150 with DWI averaging. For the error analysis voxels inside CSF are excluded by a tensor Trace condition, Trace<.0012 mm^2/s. In addidtion, to exclude boundary effects of the filter only 8 axial slices, marginated by three voxels, are involved in the error analysis. The total number of analysed voxels is 14660, the FA map with excluded voxels and a DWI are presented for a middle slice of the model in Fig. 5.

The spatially mean SNRs of these experiments are given in brackets in the last row of Tab. 1. For the rows 2-5 the relative errors $||S|_\alpha - |S|_\beta|/|S|_\beta \leq ratio$ for $\alpha \in$ {noisy, filtered, filtered and bias corrected} and $\beta \in$ {exact, exact mean}, and $ratio \in \{0.05, 0.1, 0.15\}$ are investigated. The relative deviations are calculated voxel wise for all gradients. When in a voxel all six DWIs fulfill the error condition, the voxel is counted, to add to the relative numbers of voxels (in percent) of Tab. 1. Row 2 quantifies the ability of the filter to approximate the exact DWI mean values, e.g. for σ_0=60 in 71% of the voxels the relative deviation is below 0.1. In rows 3-5 the gain of information by spatial smoothing is illustrated. Row 3 demonstrates the large gap between exact and noisy DWIs. Application of the filter, row 4, improves the situation considerably. For the replicated experiments even very low SNR situations can be approximated well with high probability. When the bias correction is added (M2 of *section* 4.1), the information quality systematically improves, see row 5. To compare the effect of spatial smoothing with denoising by voxelwise averaging, the standard deviations after smoothing for all voxels with MNRs $\geq$ 3.5 are calculated, the equivalent effective numbers of replications are given in row 6. In row 7 the corresponding SNR improvement is presented. As is evident from the results of Tab. 1, spatial filtering is quite effective and can replace on average 5–10 replications of experiments in the SNR region investigated. In measures which compare filter results with the model "truth" we find that

with reduced initial noise level in the DWIs the results improve.

4 Discussion

The aim of the presented study is to explore the possibility of spatial filtering on DTI variables in low and medium SNR situations. It was exemplified, that the distributions of the DTI variables are those of spatially varying non Gaussian random fields which in general conflicts the basic demand of spatial smoothing for uniform noise close to Gaussian. The local variability of the distributions depends in a complex way on basic system parameters like, e. g., the angles between local diffusion and the measuring gradients of the individual experiment. According to the presented analysis and in agreement with earlier case studies by Hahn et al. [2001, 2002a,b], it seems that even for low SNR the DWIs offer a possibility to bring the advantages of spatial smoothing into the game. As the proposed spatial filter is a mean value estimator, bias correction of the DWIs is important in low SNR situations to eliminate, via the continuous mappings of DTI, bias in derived variables. The statistical properties and limitations of the bias correction will be analyzed in the following.

4.1 How low can we go with SNR ?

The bias correction is based in our approach on the map $bc : MNR_{exact} \rightarrow SNR_{exact}$, where the domain of bc is identified with MNR, the set of estimated $|S|_{mean}/\sigma_0$. The precision of $|S|_{exact}/\sigma_0 \approx bc(|S|_{mean}/\sigma_0)$ depends on the uncertainty of the MNR distribution, and on the MNR level, as stability in the map decreases with MNR due to the steep slope of bc in this region, see Fig.4 A/d. For MNRs below the Rayleigh minimum, $\sqrt{\pi/2}$, bc is not defined. The extension $bc = zero$ is natural, but introduces a second mode in the distribution of estimated SNRs. To avoid this mode a smooth C^1-extension was introduced for MNR<1.33, where $bc = (\cdot/a)^b$, with (a,b)=(1.44,8.76), and $bc(1.33) = 0.5$, see Fig.4 A/d. Averaging of the DWIs to estimate $|S|_{mean}$ is henceforth shortened to method M1, M1 followed by bc with extension $(\cdot/a)^b$ to M2, and M1 plus bc with extension $zero$ to M3.

By a numerical comparison, not presented in detail, we found that M3 is for all sample sizes equivalent to a direct Maximum Likelihood estimation ML of $|S|_{exact}$ via the complete Rician distribution. This means that M3 produces for large sample sizes unbiased Gaussian distributed SNRs with the mean $|S|_{exact}$ and with optimal variance equal to the minimum variance bound, which is a lower bound on the variance of any unbiased estimator, see Sijbers et al.

SNR	$\frac{\text{Bias}}{\lvert S\rvert}$ M1	$\frac{\text{Bias}}{\lvert S\rvert}$ M2	$\frac{\text{Bias}}{\lvert S\rvert}$ M3=ML	$\frac{\sqrt{\text{MSE}}}{\lvert S\rvert}, \frac{c_{95}}{\lvert S\rvert}$ M1	$\frac{\sqrt{\text{MSE}}}{\lvert S\rvert}, \frac{c_{95}}{\lvert S\rvert}$ M2	$\frac{\sqrt{\text{MSE}}}{\lvert S\rvert}, \frac{c_{95}}{\lvert S\rvert}$ M3=ML	sample size
1	0.55	-0.07	-0.12	0.65,1.15	0.57,0.87	0.63,1.0	5
	0.55	-0.06	-0.09	0.60,0.97	0.44,0.71	0.50,1.0	10
	0.55	-0.04	-0.05	0.58,0.84	0.33,0.59	0.36,0.84	20
	0.55	-0.03	-0.04	0.57,0.79	0.26,0.52	0.29,0.57	30
2	0.14	-0.015	-0.02	0.25,0.48	0.25,0.49	0.26,0.51	5
	0.14	-0.006	-0.006	0.20,0.38	0.18,0.35	0.18,0.35	10
	0.14	-0.003	-0.003	0.17,0.3	0.12,0.24	0.12,0.24	20
	0.14	-0.002	-0.002	0.16,0.28	0.1,0.2	0.1,0.2	30
3	0.06	-0.002	-0.002	0.15,0.3	0.15,0.3	0.15,0.3	5
	0.06	-0.001	-0.001	0.12,0.22	0.11,0.21	0.11,0.21	10
	0.06	-0.0002	-0.0002	0.09,0.18	0.08,0.15	0.08,0.15	20
	0.06	-0.0003	-0.0003	0.08,0.16	0.06,0.12	0.06,0.12	30

Table 2
Simulated relative bias, rooted mean squared errors and deviation thresholds ($P = 0.95$) for methods M1,M2 and M3=ML to estimate $|S|$ are given; 10^5 iterations are applied to calculate the expectation values. The effective sample sizes vary from 5 to 30, the SNRs from 1 to 3, M1 is the mean value estimator, M2 is identical to M1 but bias corrected with extension $(\cdot/a)^b$, and M3 like M2 but with extension *zero*, ML is the Maximum Likelihood estimator based on the Rician distribution.

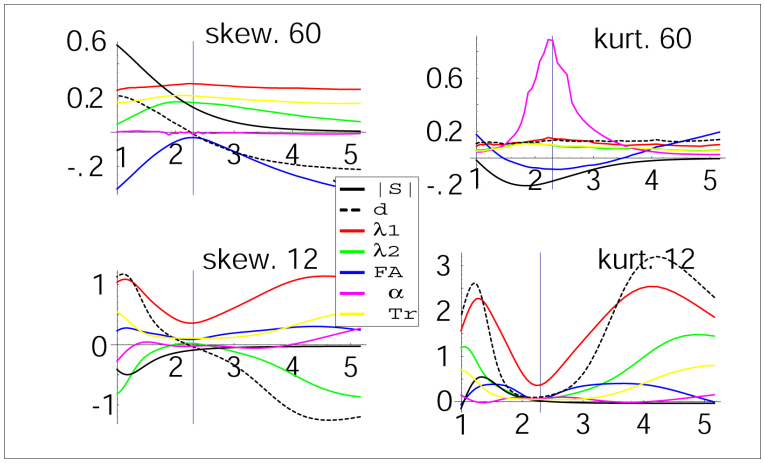

Fig. 6. Skewness and kurtosis of the experiments presented partly in Fig. 4/D; upper panels 60 multigradients, lower panels 12 multigradients and consecutive DWI averaging over $5_R \times 4_F$ replications via M2.

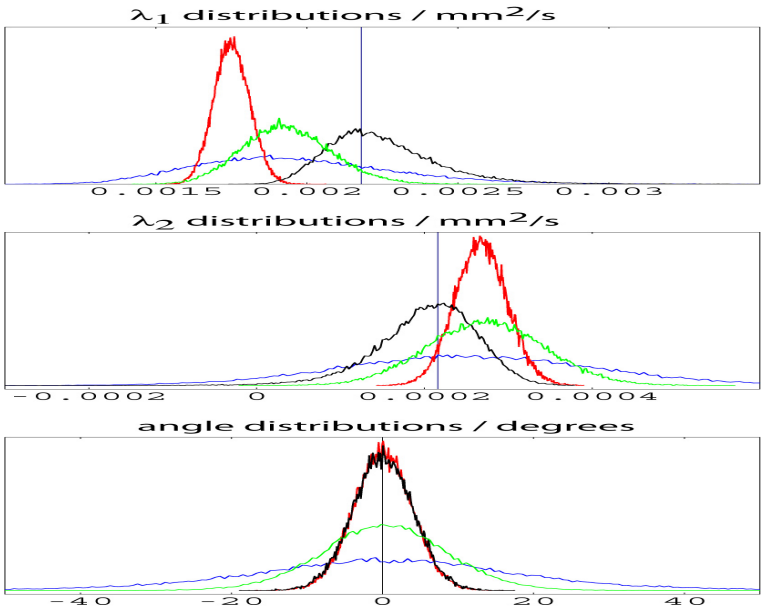

Fig. 7. Worst case distributions of λ_1, λ_2 and α. Experiments with 60 (green) and 12 multigradients (blue: n_R=1, red: M1 for n_R=20, black: M2 for n_R=20) are presented. Residual parameters are like in Fig. 6. Black bars equal the exact values of the variables.

[1998] for further properties of this ML estimator. When the sample sizes and SNR are reduced, it is clear from Fig.4 A/d that *bc* skews the distributions of $|S|$ estimated. When the Rayleigh minimum is approached, for M3 tensor components or Eigenvalues increase rapidly, and high kurtosis is produced in their distributions corrupting the estimation of individual bias corrected DTI variables. The Rayleigh minimum itself produces infinite diffusion indicating the limit of the noise model, based essentially on Eq. (2). Those outliers are reduced by M2, which maps MNR<1.33 smoothly to SNR$\in [0, 0.5]$.

For quantification, in Table 2, for M1, M2 and M3 bias, mean squared error $MSE = \mathbf{Exp}[(|S|_{exact} - |S|)^2] = bias^2 + variance$, and confidence thresholds c_{95} for $\mathbf{P}[(||S|_{exact} - |S|| < c_{95})] = .95$, where **Exp** denotes expectation value and **P** probability, are presented. The bias in M1 expresses the Rician shift and is corrected somewhat better in M2 than in M3. The relative rooted MSE is considerably improved by *bc*, again more in M2 than in M3. For SNR=1 about 30 effective replications are necessary to remain for M2 approximately within a 50 % relative deviation from the exact value for $\mathbf{P} = 0.95$; for SNR=2 we find a 20 % deviation. This indicates that for very low SNR a coupling of experimental replications with spatial smoothing is probably unavoidable to increase effective sample sizes sufficiently.

Weak approximation of few DWIs with very low SNR does not necessarily imply weak approximation of the tensor, as for multigradients in the derivation of the tensor also higher SNR information is involved, see for an example Fig.

4/D, for 60 and 12 gradients. For both experiments the mean SNR without noise ≈ 2.7 when averaging is performed over all DWIs of a measurement. See Fig. 6 for skewness and kurtosis complementing the bias information of Fig. 4 D. For 60 multigradients only kurtosis of $\alpha, d_{12} \approx 0$ is remarkable. This is due to a steep, but not singular slope in the aligned Eigenvector components. For 12 multigradients and 5_Rx4_F replications plus *bc* however, skewness and kurtosis of λ_1 and d_{12} increase up to 1 and 2.5, essentially due to effects of the Rayleigh minimum in *bc* of M2. In the corresponding minimal experiment, not shown, the patterns of skewness and kurtosis are similar. The angular peak in kurtosis goes up to ≈ 4, the skewness and kurtosis maxima of λ_1 are enhanced up to ≈ 2 and ≈ 15. In Fig. 7 the worst case distributions of λ_1, λ_2 and the angle between noisy and exact Eigenvector are presented, for the Eigenvalues the experiments at SNR=1, for the angle at SNR=2.3 were chosen, see Fig. 6. The bias corrected variables (black) peak close to the true values (bars), indicating maximum probability for good approximations and reasonable confidence intervals. For $M1_{12}$ (red) variance is lower, but the modes for the Eigenvalues are biased. For 60 multigradients (green) the distributions of Eigenvalues are intermediate, biased and with mean variance. For the angle 60 multigradients produce an appreciable tail up to about $\pm 20^o$. Summarizing, even in the worst cases, $M2_{12}$ (black) seems to give reasonable approximations; in contrast, the distributions for 12 multigradients alone (blue) produce a rather undetermined situation.

The number of experimental replications, n_R, can be related to the total scanning time, T, via $T = TR * (n_{gradients} + 1) * n_R$. For a full three dimensional measurement with 21 multigradients, $TR = 7$ seconds, and $n_R = 4$ this results in $T \approx 10$ minutes, indicating that the numbers of gradients and experimental replications chosen in the simulations are within a "realistic" frame. Thus our study indicates, that experiments with local SNRs down to ≈ 1 can be analyzed with good precision, when convenient gradient arrangements are used and when M2 is applied. Prerequisite is however, that the Rician noise model is approximately realized. Dietrich et al. [2001] investigated the noise model of two experiments with very low SNRs > 0.5 . The bias of diffusion coefficients (ADCs) from a clinical scanner could be corrected by *bc* of M3, concluding that noise was essentially Rician. The second equipment produced no Rician noise and bias could only be corrected by an empirical function modelling *bc*. To determine the parameters of this function additional phantom measurements were necessary.

4.2 Analysis of real data with voxel size $1 \times 1 \times 1$ *mm* ^{3}

To test the applicability of the presented theory, a very recent high resolution experiment with isotropic voxels of size $1 \times 1 \times 1$ mm^3 was analyzed, see *section*

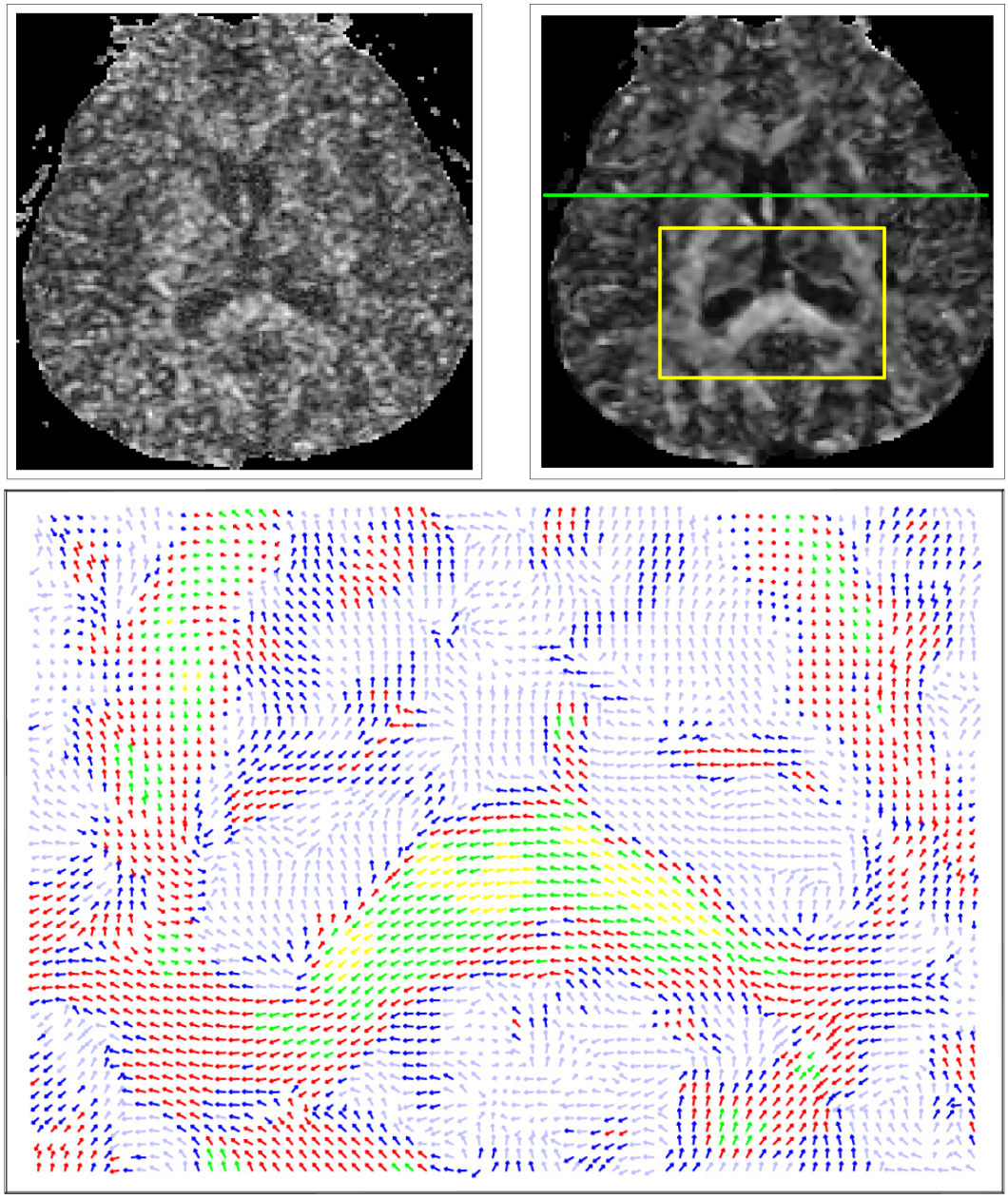

Fig. 8. Upper panels : FA maps before (left) and after spatial smoothing (M1). Yellow box indicates region of main diffusion directions plotted below. Coloring of the vectors reflects FA. Light blue <0.1, blue 0.1-0.2, red 0.2-0.4 green 0.4-0.6, yellow 0.6-1. Green horizontal bar indicates position of coronal slice presented in Fig. 9/10.

2.1 for experimental details (Exp3). Isotropic voxels are best suited for filter applications, see *section* 2.5. The measurements were repeated, NEX=n_R=4, and the DWIs were averaged inside the scanner. Determination of the noise level outside the brain region revealed a deviation from Rician noise. The Rayleigh identity $mean_{exp} \approx 1.9 \times std_{exp}$ was not fulfilled well, std_{exp} was about 30 % too high. This may be due to temporal instabilities of the scanner

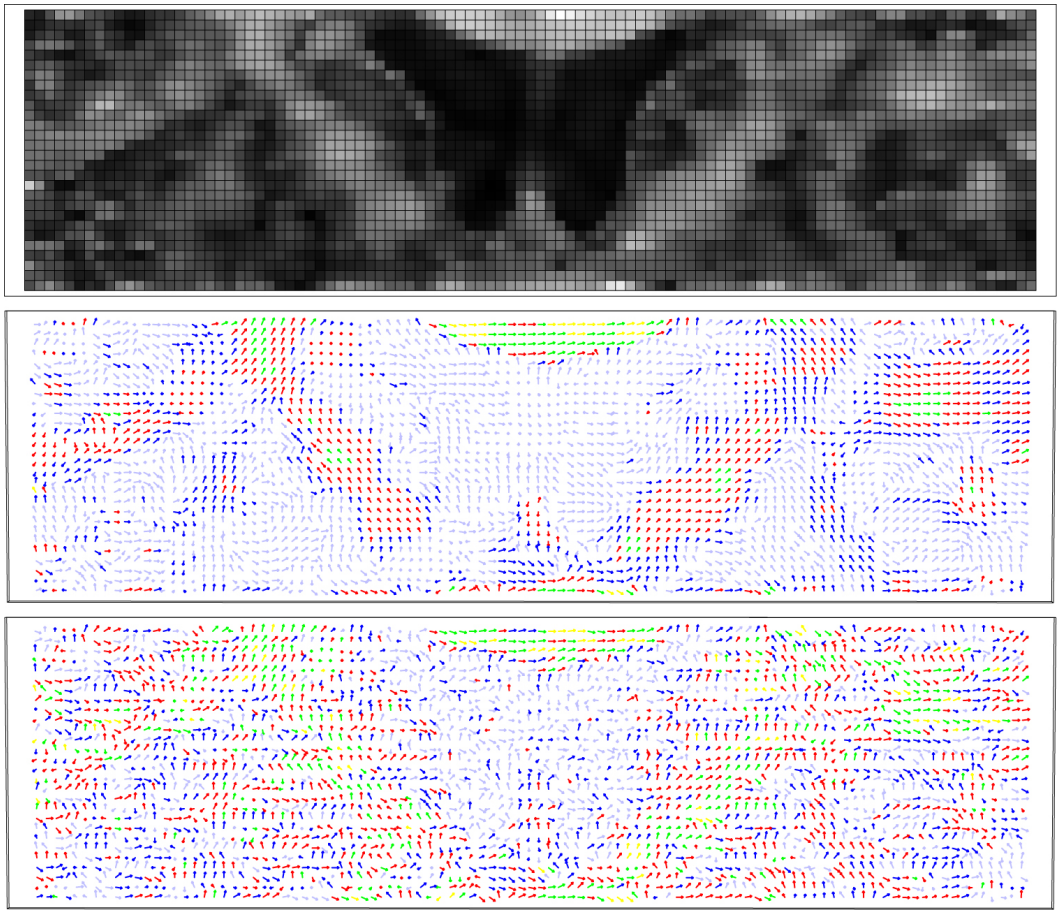

Fig. 9. Upper panel, FA map of a coronal slice after spatial smoothing (M1), middle panel : the corresponding main diffusion directions, low panel : before smoothing; colors like in Fig. 8.

or to "loading" effects created by the subject inside the coil. Inside the brain, tissue pulsations and motion may cause similar deviations. A discussion on the issue of scanner stability and its impact in fMRI can be found in Weisskopf [1996].

Nevertheless, methods M1 and M2 for the filter variant of *section* 3.2 were applied to the data on the basis of std_{exp}, as the deviation from a Rayleigh distribution was not too severe, and the methods should be robust. To be on the save side, also several calculations with up to 30 % enlarged and reduced noise levels of std_{exp} were performed, the results differed slightly from the presented ones and showed the expected over- or undersmoothing, the (subjectively) best results were achieved with std_{exp}. In Fig. 8, upper panels, representative FA maps of an axial slice after DWI averaging, left panel, and after consecutive spatial smoothing (M1), right panel, are presented. M1 improves the information quality essentially, no apparent filter caused artifacts in the smoothed image could be detected. In the lower panel the main diffusion directions, for the region indicated, is presented, the arrows are colored according to their

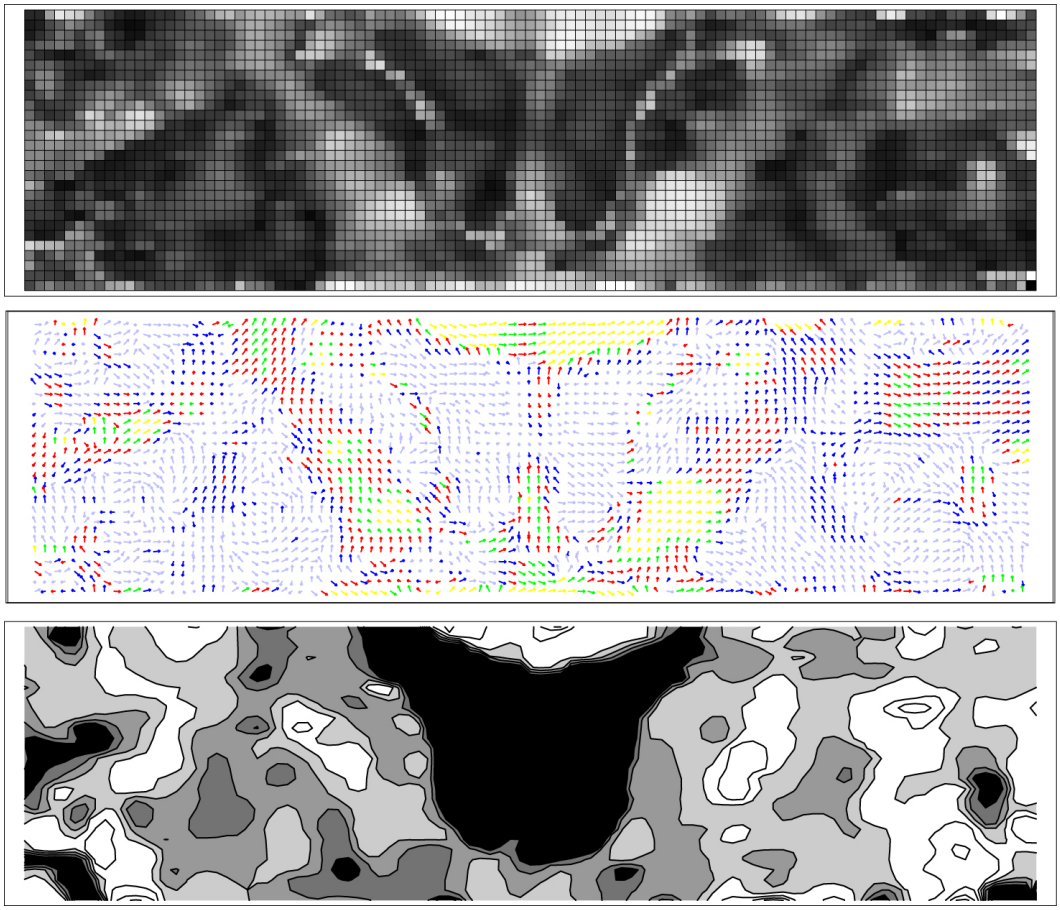

Fig. 10. Coronal slice of Fig. 9 , but with bias correction (M2), colors like in Fig. 8. Low panel, SNR map of DWI for gradient (0,0.53,-0.85). SNR ranges, black : $0-1$, dark gray : $1-1.5$, gray : $1.5-2$, light gray : $2-2.5$, white : $2.5-4.3$.

FA value. The direction field shows the anatomically expected fiber streams and a good coherence, which may make tracking feasible. In regions with high FA in several voxels the lowest Eigenvalues become negative. They are corrected by shifting them to a small positive value. The coherence of the main directions is not distorted by these negative Eigenvalues. In Fig. 9 a coronal slice is presented. Again FA map and direction field seem to be realistic, the information gain for directions by application of M1 is illustrated in the lowest panel showing main directions before spatial smoothing. In Fig. 10 the bias correction is added (M2), in the FA map (upper panel) the contrast is enhanced, fine structures with enlarged anisotropy around CSF become clearer. The anatomical origin of these structures could not be identified. The effect of *bc* on the directions is reduced compared to the Eigenvalues, as can be seen in the middle panel. To give an idea of the SNR range inherent in the data, the lowest panel shows an approximative SNR map calculated by *bc* for one DWI, we find for the averaged SNR$\approx$ 2.5, liquor was excluded. The presented analysis exemplifies that the methods M1 and M2 are robust denoising tools even for DTI data with very low SNR and with noise which is probably only

approximately Rician . Their application improves the information quality of such data considerably. Parallel imaging, or the application of higher field strengths may further improve the statistics for minimal experiments to apply spatial smoothing.

5 Conclusion

A robust denoising method for DTI data with low SNR is proposed and validated. A convenient coupling between voxelwise and spatial smoothing seems to enable the analysis of experiments with a local SNR down to ≈ 1. The presented method of DWI averaging is of general applicability and also useful in higher SNR situations. This is supported by a different investigation from Anderson [2001] who studied noise effects on bias and variance in Eigenvalues and Eigenvectors for SNRs from approximately 20 to 100. For SNRs around 20, averaging of DWIs by experimental replications is recommended. For higher SNRs Anderson's results imply an equivalence between DWI and tensor averaging. This is in line with a proposal by Pajevic et al. [2002] who apply B-splines to obtain a continuous representation of the tensor coefficient fields. This method can be regarded as a special variant of spatial smoothing applied to the tensor fields.

6 Acknowledgement

We thank Prof. V. Liebscher and Prof. G. Winkler for helpful discussions on several statistical issues, Prof. P.A. Narayana enabled the small voxel experiment in his laboratory, and Prof. D.P. Auer supported us with her profound DTI competence.

References

E.T. Ahrens, D.H. Laidlaw, C. Readhead, et al. MR microscopy of transgenic mice that spontaneously aquire experimental allergic Encephalomyelitis. *Magn. Reson. Med.*, 40:119–132, 1998.

A. W. Anderson. Theoretical analysis of the effects of noise on diffusion tensor imaging. *Magn. Reson. Med.*, 46:1174–1188, 2001.

V. Aurich and J. Weule. Non-linear Gaussian filters performing edge preserving diffusion. In *Proc. of 17th DAGM Symposium*, pages 538–545, Bielefeld, 1995. Springer.

P.J. Basser. Inferring microstructural features and the physiological state of tissues from diffusion-weighted images. *NMR in Biomedicine*, 8:333–344, 1995.

P.J. Basser and D.K. Jones. Diffusion-tensor MRI: theory, experimental design and data analysis - a technical review. *NMR in Biomedicine*, 15:456–467, 2002.

P.J. Basser, J. Mattiello, and D. Le Bihan. Estimation of the effective self-diffusion tensor from the NMR spin echo. *J. Magn. Reson.*, 103:247–254, 1994.

P.J. Basser and S. Pajevic. Statistical artifacts in diffusion tensor MRI (DT-MRI) caused by background noise. *Magn. Reson. Med.*, 44(1):41–50, 2000.

P.J. Basser, S. Pajevic, C. Pierpaoli, et al. In vivo fiber tractography using DT-MRI data. *Magn. Reson. Med.*, 44(4):625–632, 2000.

M.E. Bastin, P.A. Armitage, and I. Marshall. A theoretical study of the effect of experimental noise on the measurement of anisotropy in diffusion imaging. *Magn. Reson. Imag.*, 16(7):773–785, 1998.

D. Le Bihan. Looking into the functional architecture of the brain with diffusion MRI. *Nature Reviews/Neuroscience*, 4(6):469–480, 2003.

C.K. Chu, I. Glad, F. Godtliebsen, et al. Edge-preserving smoothers for image processing. *JASA*, 93:526–541, 1998.

C.A. Clark, M. Hedehus, and E. Moseley. In vivo mapping of the fast and slow diffusion tensors in human brain. *Magn. Reson. Med.*, 47:623–628, 2002.

O. Dietrich, S. Heiland, and K. Sartor. Noise correction for the exact determination of apparent diffusion coefficients at low SNR. *Magn. Reson. Med.*, 45:448–453, 2001.

W. Feller. *An Introduction to Probability Theory and its Application*, volume II. John Wiley & Sons, 1971.

H. Gudbjartsson and S. Patz. The Rician distribution of noisy MRI data. *Magn. Reson. Med.*, 34:910–914, 1995.

K. Hahn, S. Prigarin, and B. Pütz. Edge preserving regularization and tracking for diffusion tensor imaging. In *Proc. MICCAI 2001*, volume 2208, pages 195–203. Springer LNCS, 2001.

K. Hahn, S. Prigarin, and B. Pütz. Glättung von Tensorfeldern und Modellierung von Diffusionspfaden für die Sichtbarmachung von Nervenbahnen im menschlichen Gehirn. In *Proc. Bildverarbeitung für die Medizin*, pages 55–58. Springer, 2002a.

K. Hahn, S. Prigarin, and B. Pütz. The hierarchy of random variables in diffusion tensor imaging and the problems: Spatial smoothing, resolution and tracking. In *Proc. NeuroImage: Human Brain Mapping*, page (order to appear) 360. OHBM, 2002b.

K.M. Hasan and P.A. Narayana. Computation of the fractional anisotropy and mean diffusivity maps without tensor decoding and diagonalization: Theoretical analysis and validation. *Magn. Reson. Med.*, 50(3):589–598, 2003.

K.M. Hasan, D.L. Parker, and A.L. Alexander. Comparison of gradient en-

coding schemes for diffusion-tensor MRI. *J. Magn. Reson. Imag.*, 13(5): 769–780, 2001.

M. Henkelmann. Measurement of signal intensities in the presence of noise in MR images. *Med. Phys.*, 12(2):232–233, 1985.

D.K. Jones, M.A. Horsfield, and A. Simmons. Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging. *Magn. Reson. Med.*, 42(3):515–525, 1999.

A. Messiah. *Quantum Mechanics I.* North-Holland Publishing, 1972.

E. Mühlhaus. *Die Sprungerhaltende Glättung verrauschter harmonischer Schwingungen.* PhD thesis, Heinrich-Heine Universität, Düsseldorf, 1997.

R. Nieuwenhuys, J. Voogd, and V. Huijzen. *The Human Central Nervous System.* Springer, Berlin, 1983.

S. Pajevic, A. Aldroubi, and P.J. Basser. A continuous tensor field approximation of discrete DT-MRI data for extracting microstructural and architectural features of tissue. *J. Magn. Reson.*, 154:85–100, 2002.

N.G. Papadakis, Da Xing, C.L.-H. Huang, et al. A comparative study of acquisition schemes for diffusion tensor imaging using MRI. *J. Magn. Reson.*, 137(1):67–82, 1999.

G.J.M. Parker, J.A. Schnabel, M.R. Symms, et al. Nonlinear smoothing for reduction of systematic and random errors in diffusion tensor imaging. *J. Magn. Reson. Imag.*, 11(6):702–710, June 2000.

C. Pierpaoli and P.J. Basser. Toward a quantitative assessment of diffusion anisotropy. *Magn. Reson. Med.*, 36(6):893–906, 1996.

C. Poupon, C.A. Clark, V. Frouin, et al. Regularization of diffusion-based direction maps for the tracking of brain white matter fascicles. *NeuroImage*, 12(2):184–195, August 2000.

J. Sijbers, A.J. den Dekker, P. Scheunders, et al. Maximum-likelihood estimation of Rician distribution parameters. *IEEE Trans. Med. Im.*, 17(3): 357–361, 1998.

S. Skare, M. Hedehus, M.E. Moseley, et al. Condition number as a measure of noise performance of diffusion tensor data acquisition schemes with mri. *J. Magn. Reson.*, 137:67–82, 2000a.

S. Skare, T.-Q. Li, B. Nordell, et al. Noise considerations in the determination of diffusion tensor anisotropy. *Magn. Reson. Imag.*, 18(6):659–669, 2000b.

C.R. Tench, P.S. Morgan, L.D. Blumhardt, et al. Improved white matter fiber tracking using stochastic labeling. *Magn. Reson. Med.*, 48:677–683, 2002a.

C.R. Tench, P.S. Morgan, M. Wilson, et al. White matter mapping using diffusion tensor MRI. *Magn. Reson. Med.*, 47:967–972, 2002b.

R.M. Weisskopf. Simple measurement of scanner stability for functional NMR imaging of activation in the brain. *Magn. Reson. Med.*, 36:643–645, 1996.

M.R. Wiegell, H.B.W. Larsson, and Van J. Wedeen. Fiber crossing in human brain depicted with diffusion tensor MR imaging. *Radiology*, 217(3):897–903, 2000.

G. Winkler. *Image Analysis, Random Fields and Markov Chain Monte Carlo Methods.* Applications of Mathematics, Stochastic Modelling and Applied

Probability. Springer, 2003.

G. Winkler, V. Aurich, K. Hahn, et al. Noise reduction in images: Some recent edge-preserving methods. *Jour. of Pattern Recognition and Image Analysis*, 9(4):769–766, 1999.

G. Winkler and V. Liebscher. Smoothers for discontinuous signals. *Jour. of Nonparam. Statist.*, 14(1):202–222, 2002.

T. Yoshiura, F. Mihara, A. Tanaka, et al. High b value diffusion-weighted imaging is more sensitve to white matter degeneration in Alzheimer‘s disease. *NeuroImage*, 20(1):413–419, 2003.