Master of Science Thesis

Denoising of Complex MRI Data by Wiener-like Filtering in the Wavelet Domain: Application to High b-value Diffusion Weighted Imaging

Adnan Bibic

Supervisor:
Ronnie Wirestam, PhD

Medical Radiation Physics
Clinical Sciences, Lund
Lund University, 2006

Abstract

Noise is a well-known problem in many imaging modalities. In magnitude magnetic resonance images, obtained by quadrature detection, the associated Rician distribution of noise constitutes a further complication. This type of noise is especially problematic at low signal-to-noise ratio (SNR) regions. The Rician noise distribution causes a non-zero minimum signal in the image, often referred to as the rectified noise floor. True low signal is likely to be concealed in the noise, and quantification is severely hampered in low-SNR regions.

To reduce this problem, real and imaginary Magnetic Resonance Imaging (MRI) data in the image domain were filtered, before construction of the magnitude image. The noise-reduction filtering (or denoising) was accomplished by Wiener-like filtering in the wavelet domain. The advantage of denoising the complex MRI data in the image domain is that the noise in the real and imaginary channels is Gaussian and most denoising tools in the wavelet domain are adapted to such a distribution. A further advantage, compared with filtering the complex k-space data, is that image artefacts caused by phase errors are minimized.

The proposed noise-removal scheme efficiently reduced the standard deviation and significantly lowered the rectified noise floor. The contrast of the images, especially in the low-SNR regions was increased accordingly. Denoising was tested on simulated diffusion-weighted (DW) images with slow and fast apparent diffusion coefficients (ADC) regions as well as on a dataset showing bi-exponential signal decay. Experimentally, the method was applied to diffusion-weighted images from a homogenous n-decane ($C_{10}H_{22}$) phantom and to data from a healthy volunteer.

The proposed de-noising algorithm is useful not only in DW-MRI but also for others kinds of MR images where SNR and/or image contrast is low.

Contents

1. Introduction

In magnetic resonance imaging (MRI), noise is a major disadvantage, especially in applications based on signal reduction, such as diffusion weighted imaging (DWI) and bolus-tracking perfusion MRI. Noise is problematic also in other applications where averaging is prohibited by temporal-resolution requirements, for example, in the dynamic time-series used for functional magnetic resonance imaging (fMRI). The choice of pulse-sequence type is also relevant for the resulting signal-to-noise ratio (SNR); gradient-echo (GE) pulse sequences, for example, typically show lower SNR than spin-echoes. Noise reduction is a delicate task, and the key to success is, obviously, to reduce or eliminate the noise with minimal destruction of true signal. The purpose of this study was to implement and evaluate noise reduction, or denoising, by filtering in the wavelet domain and to apply the technique to high b-value diffusion MRI where the SNR of the acquired data is extremely low. The method can, in principle, be used also in other MRI applications.

Several approaches to noise reduction by wavelet-domain filtering were tested before a viable concept was found: Filtering of (i) k-space data, (ii) data in the complex image space and (iii) in magnitude image space. The final denoising approach presented in this report was preceded by extensive tests of several wavelet-based noise reduction schemes, all similar to the Wiener-like filtering presented by Wirestam & Ståhlberg (2), the WienerChop model (3,5,6), the approximation to a Wiener-like filter by Nowak (12,13,14), denoising of magnitude images after bias-reduction, accomplished by a modification of the method proposed by Gudbjartsson (15) and of course combinations of the above-mentioned methods.

2. Theory

2.1 *Noise in MRI*

MR images are reconstructed from signal data measured by a quadrature detector and sampled in the spatial-frequency or k-space domain. The acquired k-space data (sometimes referred to as raw data) are collected in one real and one imaginary channel, and both these data components are affected by additive Gaussian noise. The major source of this noise is random thermal noise from the patient while some additional noise arises from the acquisition hardware.

The Gaussian distribution is symmetric and shows a bell-shaped density curve with a single peak (Fig.1):

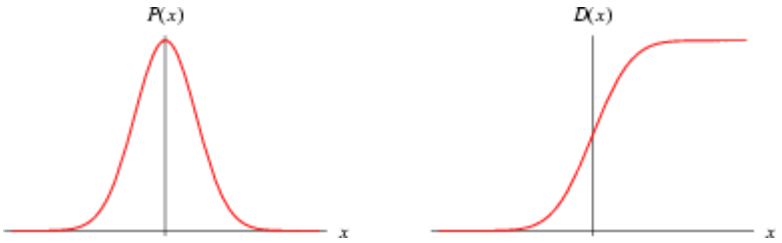


Fig. 1. The probability function of the Gaussian distribution (left) and its corresponding cumulative function (right).

A Gaussian distribution of a random variable x with mean μ and variance σ² is given by the following equation:

$$P(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/(2\sigma^2)} \text{ on } x=(-\infty,\infty) \quad (1)$$

The next step in the reconstruction of MR images is a 2D Fourier transform of the raw data. The Fourier transform preserves the shape of the noise distribution which remains Gaussian. The MR images are normally converted to magnitude images (M) by calculating the absolute value pixel by pixel, i.e. |m[i,j]|, from the real and the imaginary images (see Eq. 2).

$$\left|m[i,j]\right|=\left[\left(s[i,j]\cos(\varphi)+n_{\mathrm{Re}}[i,j]\right)^2+\left(s[i,j]\sin(\varphi)+n_{\mathrm{Im}}[i,j]\right)^2\right]^{\frac{1}{2}} \quad (2)$$

where m[i,j] is the magnitude signal in pixel [i,j] , s[i,j] is the true signal, φ is the phase angle and n is the noise contribution to the real (Re) and imaginary (Im) part in the given pixel [i,j].

From Eq 2 it is obvious that the construction of M involves both the true signal and the noise, leading to a modified noise distribution and of course loss of all information about the phase angle φ.

The noise in M becomes signal dependent (following a Rician distribution) and its probability function is a follows:

$$P(m)=\frac{m}{\sigma^2}e^{-(m^2+s^2)/2\sigma^2}\hat{I}_0\left(\frac{s\cdot m}{\sigma^2}\right) \text{ on } m=[0,\infty) \quad (3)$$

where $\hat{I}_o$ is the modified zeroth order Bessel function of the first kind (20).

If the true signal $|s|=0$, Eq. 3 reduces to a Rayleigh distribution (see Fig. 2) given by:

$$P(m) = \frac{m}{\sigma^2} e^{-m^2/2\sigma^2} \text{ on } m = [0, \infty) \quad (4)$$

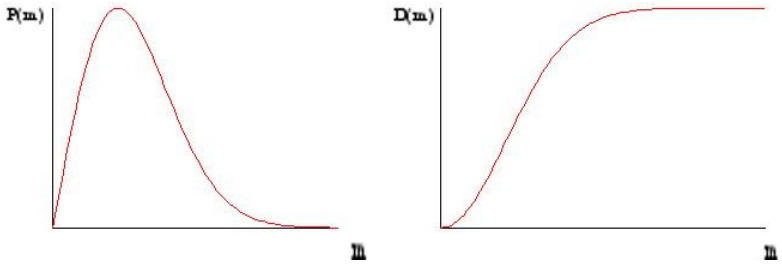


Fig. 2. The probability function of the Rayleigh distribution (left) and its cumulative function (right).

The probability and cumulative functions of the Rician are intermediate to the Rayleigh distribution (Fig.2) and the Gaussian distribution (Fig.1.) If the true signal $s \to 0$ then (Rician distribution) $\to$ (Rayleigh distribution) while if $s \to \infty$ then (Rician distribution) $\to$ (Gaussian distribution). If the SNR is below 10 dB, which infers that the mean pixel value is less than three times the standard deviation, the noise distribution can not be approximated by a Gaussian function.

In the absence of any true signal, the mean (η) and standard deviation (σ_η) of the magnitude signal are given by Eq. 5:

$$\eta = \sigma\sqrt{\frac{\pi}{2}} \text{ and } \sigma_\eta = \sigma\sqrt{2 - \frac{\pi}{2}} \quad (5)$$

where σ is the standard deviation of the MR signal in the real and imaginary channels (it is assumed that the standard deviations are equal in both channels) (1, 7). Consequently, this phenomenon gives rise to a minimum mean signal in the magnitude image, sometimes referred to as the rectified **noise floor** (1).

2.2 *The Discrete Wavelet Transform*

Analogous to Fourier series analysis, where a signal is decomposed into its independent Fourier modes with sinusoids as the basis function, orthogonal wavelet analysis is the decomposition of the signal into its independent wavelet modes using an orthonormal family of basis functions. This is accomplished by convolving the given function (input image) *f* with translated and dilated wavelet functions. Orthogonal wavelet basis functions are

recommended in the present context, because they preserve a Gaussian noise distribution.

The formula to compute the scaling ($s_{j_0,k}$) and the wavelet coefficients ($w_{j,k}$) is given by Eq. 6:

$$f_J(t) = \sum_k s_{j_0,k}\varphi_{j_0,k}(t) + \sum_{j=j_0}^{J-1}\sum_k w_{j,k}\psi_{j,k}(t) \qquad (6)$$

where $\varphi_{j_0,k}(t)$ is a low-pass scaling function called scaling or father wavelet which is given by:

$$\varphi_{j_0,k}(t) = 2^{-J/2}\varphi(2^{-J}t - k) \qquad (7)$$

and $\psi_{j,k}(t)$ is a band-pass wavelet function called the generating or mother wavelet which is given by:

$$\psi_{j,k}(t) = 2^{-J/2}\psi(2^{-J}t - k) \qquad (8)$$

where *k* is the translation, *t* is the dilation or compression parameter and j -s are different wavelet scales or frequency bands $j=j_0, j_0+1, j_0+2,\ldots,J$.

A combination of $\varphi_{j_0,k}(t)$ and $\psi_{j,k}(t)$ functions gives a family of functions, called “wavelet packets”. These functions give rise to the orthonormal basis (10). In addition to the orthogonal wavelets there are also biorthogonal, complex, infinite, multi- and other wavelets.

The very first wavelet was introduced by Alfred Haar in 1909, and the wavelet packet named after him is the simplest finite supported wavelet which does not demand any powerful computational resources. The Haar wavelet is the only orthogonal wavelet that has symmetric analyses and synthesis filters, and the advantage of using a symmetric wavelet in image processing applications could be that the human vision is more tolerant to symmetric errors. A number of wavelet families have been presented, such as Morlet, Mayer, Gaussian wavelet and Mexican Hat, and many scientists have become interested in wavelets. However, all of the wavelets mentioned above are of infinite support and not suitable for the discrete wavelet transform (DWT). The introduction of the Daubechies wavelet packets in the early 1990s, which made DWT practicable, resulted in an increase of image-processing applications. Daubechies developed wavelets with maximum regularity, i.e. wavelets with good compression properties although only for wavelet coefficients ($w_{j,k}$) and

not for scaling coefficients ($s_{j_0,k}$). In order to resolve this, Coifman modified the Daubechies wavelet packets to show similar compression properties in both types of coefficients, and the resulting packets are called Coiflets (10, 11).

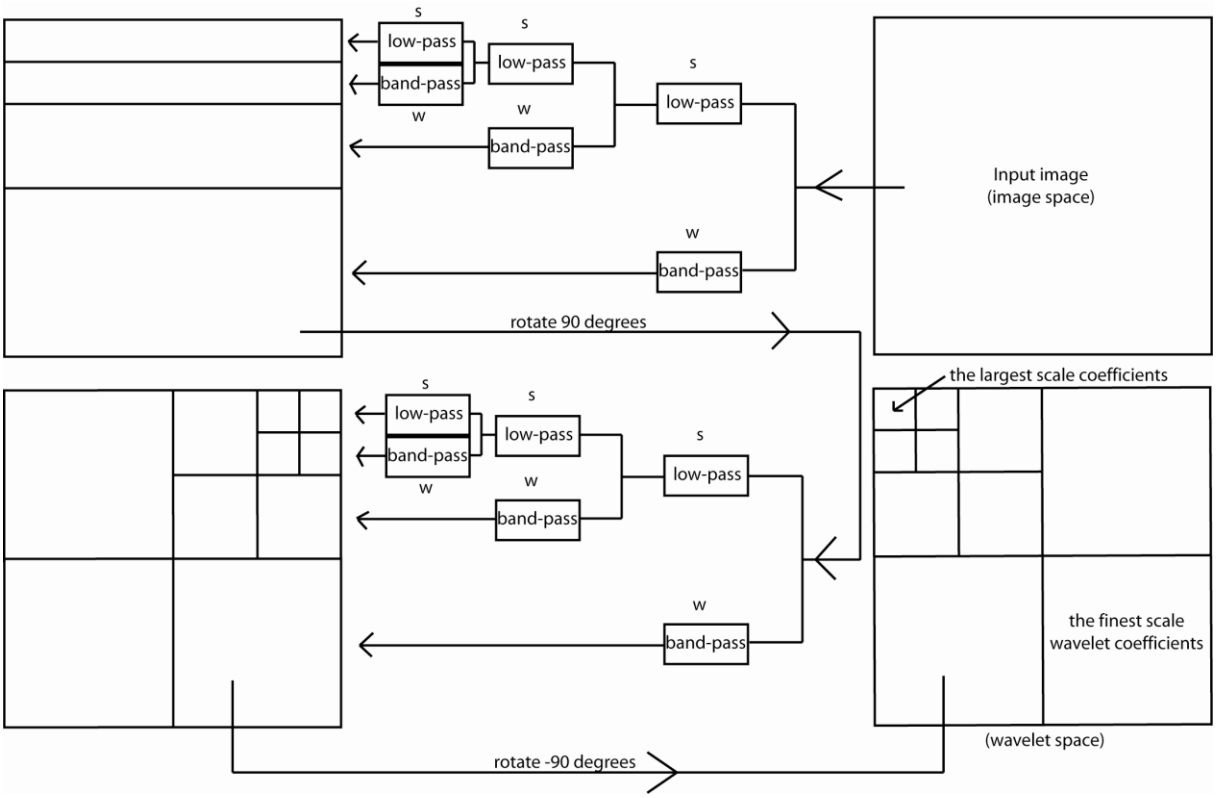


Fig. 3a. Block diagram of DWT for a two-dimensional image. The wavelet transform is computed using the pyramidal algorithm along each dimension.

The DWT transforms a matrix to another matrix of the same dimensions by convolving input function with translated and dilated wavelet-base functions $\varphi_{j_0,k}(t)$ and $\psi_{j,k}(t)$ and sequentially calculating the higher level coefficients from the lower level ones for different *j*-scales, these recursions are called cascade algorithm or pyramidal algorithm. However, the calculated coefficients are stored with the largest scale coefficient is stored in the [0,0] position, while the other matrix elements contain finer and finer coefficients and the 4th quadrant contains only the finest scale wavelet coefficients (see Fig. 3a).

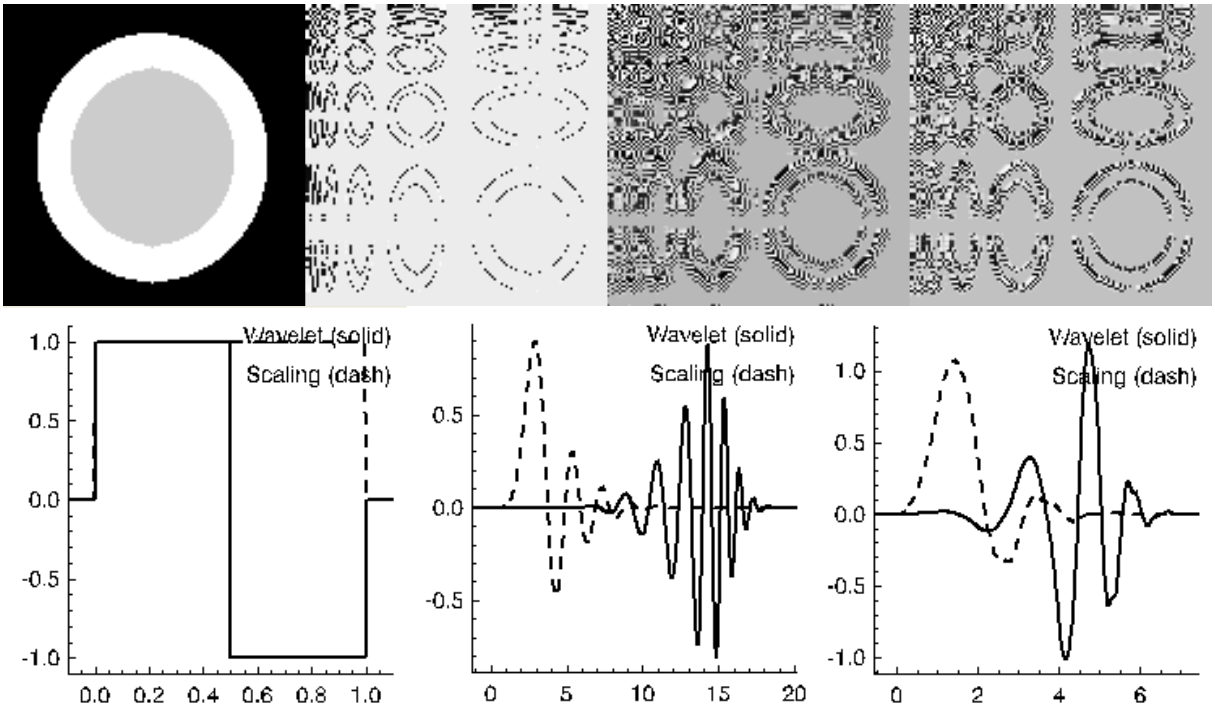


Fig. 3b. An example of a simulated image transformed to wavelet domain with different wavelet families and its corresponding wavelet-base functions (images in wavelet domain are histogram equalized). From left to right: Haar wavelet, Daubechies of order 12 and Daubechies of order 5

As we can see in Fig. 3b, the wavelets are localized in both frequency (via dilations) and in space (via translations), this localization in space is an advantage compared with Fourier transform where a small change in frequency domain will produce changes everywhere in the space domain.

2.3 *Diffusion weighted MR imaging (DWI)*

So-called diffusion- weighted MR imaging (DWI) is a method for analysis of thermal movements of water within the brain tissue at the molecular level. Water molecules are constantly in random movement driven by internal kinetic (thermal) energy, a phenomenon known as “Brownian motion”. The movement, according to the Maxwell-Boltzmann distribution law, depends only on the temperature of the water (see Eq. 9).

$$f(v) = 4\pi N\left(\frac{m}{2\pi kT}\right)^{3/2} v^2 e^{-mv^2/2kT} \qquad (9)$$

Equation 9 is the probability density function of the velocities v of the N molecules of mass *m* contained in a substance, and *k* is the Boltzmann constant (9).

In brain tissue, the thermal movement (i.e. the self-diffusion) of the water is not truly random since the biological tissue contains structures and complex geometry, which limit the degree of free diffusion, especially in certain directions. It is thus possible to visualize different brain-tissue structures by studying the effects of diffusion on the MR signal and by reconstructing the preferred paths of the water from the different diffusion movements in different cell structures and along different directions. The diffusion coefficients measured by MRI is normally termed apparent diffusion coefficients (ADCs).

The diffusion-weighted MRI pulse sequences are basically obtained by adding two strong gradient pulses to the pulse sequence (see Fig. 3). The first gradient labels a spin with a phase angle and the second pulse re-phases the stationary spin, whereas spins that have moved obtain a resulting net phase angle at the time of the echo. The random motion of a large population of spins results in voxel phase dispersion and an associated signal loss. The degree of signal attenuation caused by diffusion is exponentially related to the magnitude of the molecular translation (i.e. the apparent diffusion coefficient) and the employed sensitivity of the pulse sequence (the so-called b-value).

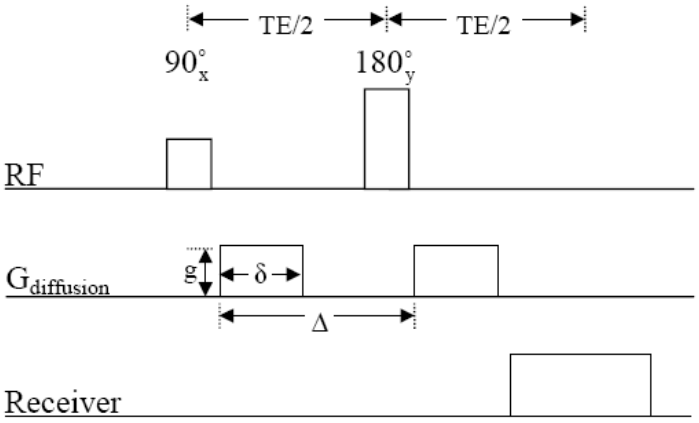


Fig.4. Schematic illustration of a basic diffusion-sensitive pulse sequence [8].

The b-value depends on the amplitude of the diffusion gradients *g*, the duration of the gradients δ and the time between the starting points of the two diffusion gradient pulses Δ. The definition of b is given by Eq.10.

$$b = (\gamma \delta g)^2 (\Delta - \delta / 3) \qquad [\mathrm{s/m^2}] \qquad (10)$$

where γ is the gyromagnetic ratio [8]. By acquiring several images with different b-values and different directions, given by the unit vector, **g**, the ADC along each direction can be calculated by using Eq 11.

$$ADC_g = \frac{1}{b}\ln\left(\frac{I_0}{DWI_g}\right) \tag{11}$$

where I_0 is the non-weighted image intensity or signal and DWI_g is the diffusion weighted intensity along direction **g**. To characterize the directional dependence of diffusion, ADC must be expressed in tensor form. By calculating an ADC map for each direction an ADC vector can be constructed (see Eq. 12). A minimum of six directions is necessary to completely describe the 3×3 diffusion tensor matrix and the commonly used gradient scheme is $(G_x, G_y, G_z) = \{(1,1,0), (1,0,1), (0,1,1), (-1,1,0), (-1,0,1), (0,-1,1)\}$.

$$ADC = M^{-1}\begin{bmatrix} ADC_{110} \\ ADC_{101} \\ ADC_{011} \\ ADC_{-110} \\ ADC_{-101} \\ ADC_{0-11} \end{bmatrix} = \begin{bmatrix} ADC_{xx} ADC_{xy} ADC_{xz} \\ ADC_{yx} ADC_{yy} ADC_{yz} \\ ADC_{zx} ADC_{zy} ADC_{zz} \end{bmatrix} \tag{12}$$

where **M** is a transformation matrix, depending only on the diffusion directions that are applied. The three upper and lower off-diagonal elements in the ADC matrix are identical, i.e. $ADC_{xy} = ADC_{yx}$, $ADC_{xz} = ADC_{zx}$ and $ADC_{yz} = ADC_{zy}$. In some cases the average diffusivity <ADC> is required, and it can be calculated by averaging the diagonal elements of the diffusion tensor (see Eq. 13).

$$< ADC >= \frac{ADC_{xx} + ADC_{yy} + ADC_{zz}}{3} \tag{13}$$

Images constructed from $< ADC >$ are often called trace maps. Finally, the diffusivity can be expressed as the fractional anisotropy (FA). This is a measure of deviation from isotropy, i.e. how much the diffusion ellipsoid deviates from a sphere, and it is mathematically defined as

$$FA = \sqrt{\frac{3}{2}}\frac{\sqrt{(\lambda_1 - \bar{\lambda})^2 + (\lambda_2 - \bar{\lambda})^2 + (\lambda_3 - \bar{\lambda})^2}}{\sqrt{\lambda_1^{\,2} + \lambda_2^{\,2} + \lambda_3^{\,2}}} \tag{14}$$

where λ_1, λ_2 and λ_3 are the principal eigenvalues of the diffusion tensor that can be calculated by solving the determinant of the tensor in Eq. 12.

$$\det\begin{bmatrix} ADC_{xx}-\lambda & ADC_{xy} & ADC_{xz} \\ ADC_{yx} & ADC_{yy}-\lambda & ADC_{yz} \\ ADC_{zx} & ADC_{zy} & ADC_{zz}-\lambda \end{bmatrix} = 0 \qquad (15)$$

The mean eigenvalue is given by Eq. 16:

$$\bar{\lambda} = \frac{\lambda_1 + \lambda_2 + \lambda_3}{3} \qquad (16)$$

3. Methods

The complex image data obtained after 2D Fourier transform of the raw data (k-space data) was used for denoising, i.e. the real and imaginary parts in the image domain. The real and imaginary parts of the image in the complex image space still show the Gaussian distributed noise, required by the wavelet shrinkage method.

The complex image signal is given by the following:

$$x = s + n \qquad (17)$$

where ***s*** is the noise-free signal and ***n*** is Gaussian noise of variance σ^2.
After the discrete wavelet transform of noisy data the corresponding wavelet coefficients were obtained:

$$y = \Theta + z \qquad (18)$$

where $y = W_n x$, $\Theta = W_n s$ and $z = W_n n$ (W_n denotes a wavelet transform).

The noise in wavelet space remains unchanged since an orthogonal wavelet transforms white noise into white noise and the noise is uniformly spread over all wavelet coefficients [4].

To obtain $\hat{\Theta}$ of the *i, j*th pixel (i.e. the filtered signal in the wavelet space) each noisy wavelet coefficients was thresholded by multiplication of the wavelet coefficients *y(i,j)* by the hard-threshold filter $h_h(i,j)$:

$$\hat{\theta}(i,j) = y(i,j) \cdot h_h(i,j) \qquad (19)$$

The hard threshold filter H_h sets any $|y(i,j)|$ below $\rho\sigma$ to zero and keeps the remaining data unchanged, i.e. the filter is given by the following expression [2]:

$$h_h(i,j) = \begin{cases} 1, & if\ |y(i,j)| > \rho\sigma \\ 0, & otherwise \end{cases} \quad (20)$$

where ρ is an empiric threshold factor and σ is standard deviation of the noise. The threshold factor ρ is a user-defined parameter. Larger values of ρ typically result in a larger degree of noise reduction, but may, on the other hand, lead to image smoothing and artifacts. Hence, optimization of this threshold factor is required. The standard deviation σ in the real and imaginary part of the complex image was obtained from the finest-scale wavelet coefficients. In 2D wavelet space the finest wavelet coefficients can be found in 4th quadrant of the matrix. Filtered signals was then estimated according to Eq. 21

$$\hat{s}_1 = W_1^{-1} H_h W_1 x \quad (21)$$

An overview of this part of the filtering procedure is given by the block diagram in Fig. 5:

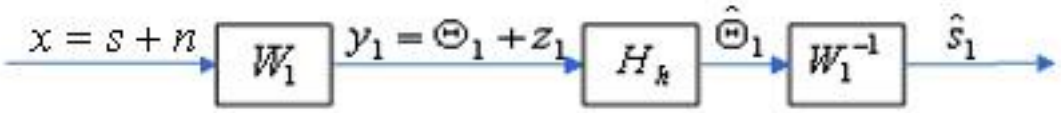


Fig. 5. Block diagram of the initial wavelet-domain filtering. The wavelet transform $\mathbf{W_1}$ was used to transform a signal into the wavelet space and the hard threshold filter $\mathbf{H_h}$ was applied to signal y_1. The inverse wavelet transform $\mathbf{W_1^{-1}}$ was used to reconstruct the filtered signal $\hat{\Theta}_1$ to the signal $\hat{s}_1$ in complex image space.

To further improve the estimated data $\hat{s}_1$ [3], a Wiener-like filter H_W was employed to shrink each wavelet coefficient $\hat{\theta}(i,j)$ by a factor $h_w(i,j)$ given by Eq. 22:

$$h_w(i,j) = \frac{|\theta(i,j)|^2}{|\theta(i,j)|^2 + \sigma^2} \quad (22)$$

The Wiener-like filter requires knowledge of the signal and its noise properties. The variance σ² can be estimated from the finest-scale wavelet coefficients, according to the description above. However, $\theta(i,j)$ is the noise-free signal in the pixel *(i,j)*, i.e. the unknown signal to be retrieved. At this point, the closest approximation is to use the hard- threshold filtered data. Eq. 22 is then modified to the following:

$$h_w(i,j) = \frac{|\hat{\theta}(i,j)|^2}{|\hat{\theta}(i,j)|^2 + \sigma^2} \quad (23)$$

From the equation above it is seen that the filter has the highest effect at small

$\theta^2(i,j)$ compared with σ^2 and its effect decreases monotonically when $\theta^2(i,j) \to \infty$ and $h_w(i,j) \to 1$. Consequently, the Wiener-like filter does not primarily affect the large scaling or wavelets coefficients but it does influence the small ones.

Furthermore, it has been proposed that the hard-threshold filtered data should be transformed from image space to wavelet space with a different wavelet family before the second filtering[3],[5].

$$\hat{\Theta}_{21} = W_2 \hat{s}_1 \qquad (24)$$

With the estimated data $\hat{\Theta}_{21}$, the Wiener-like filter was constructed and applied to the hard-threshold filtered data (see the block diagram in Fig. 6):

Fig.6. Block diagram of wavelet-domain filtering. The wavelet transform $\mathbf{W}_1$ was used to transform the signal x into wavelet space and the hard threshold filter $\mathbf{H_h}$ was applied to signal y_1 to produce the pilot signal $\hat{s}_1$. This estimate was then used to construct the Wiener-like filter $\mathbf{H}_W$ in the $\mathbf{W}_2$ domain, which is applied to the pilot signal $\hat{s}_1$.

To achieve additional reduction of the standard deviation and the mean signal of the background, the estimated value $\hat{s}_2$ was used to construct a new Wiener-like filter that was applied to the original noisy signal x according to Fig. 7. This approach is an extension of the so-called *WienerChop* technique [3], [5], [6].

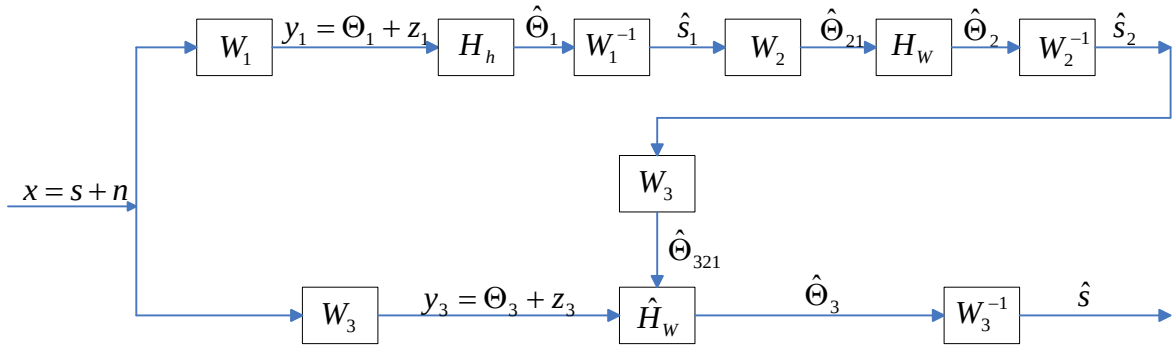


Fig.7. Block diagram of the complete wavelet-domain filtering procedure. The top row is the same as in Fig.6, The addition here is that the estimated signal is used to construct another wiener-like filter applied to the original noisy signal x in the $\mathbf{W}_3$ domain.

After filtering of the real as well as the imaginary part of the image, a magnitude image was calculated.

3.1 Simulations

The simulated synthetic image data were designed as two homogenous regions with different apparent diffusion coefficients, i.e. $ADC_1 = 0.1 \times 10^{-3}\, mm^2 s^{-1}$ and $ADC_2 = 1.0 \times 10^{-3}\, mm^2 s^{-1}$. The diffusion-weighted signal intensities along the direction *g* were computed as follows:

$$DWI_g = I_0 e^{-b \cdot ADC_g} \tag{25}$$

The b-value is the scalar magnitude of the diffusion weighting (see Eq. 10). The signal-to-noise ratio at $b = 0$ was set to 15:1 for both regions, with the initial signal $I_0 = 1000$ arbitrary units. Gaussian noise was added to the signal in k-space for each channel (see Fig.8). In total, thirty images were constructed with b-values ranging from 0 to 5800 s·mm^{-2}.

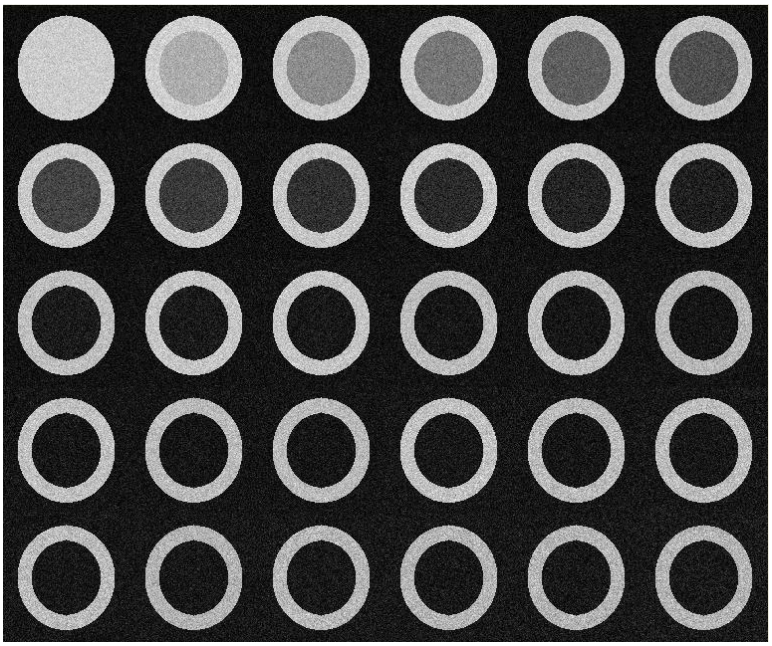


Fig.8. Simulated image data set, displayed as magnitude images at b values ranging from 0 to 5800 s·mm^{-2}.

Additionally, a similar dataset showing bi-exponential signal decay was simulated, in order to assess the noise effects on bi-exponential (i.e. two-component) diffusion data. The bi-exponential function, before addition of noise, was given by Eq. 26:

$$DWI(b) = A\, e^{-bADC_A} + B\, e^{-bADC_B} \tag{26}$$

The employed parameters were A=680, B=1000-A=320, ADC_A=1.25·10^{-9} m^2/s, ADC_B=0.18·10^{-9} m^2/s) [19], and the SNR at b=0 was 15.

In these simulation experiments, the resulting average signal and standard deviation (SD) were investigated, before and after filtering. To avoid possible image smoothing effects, the signal and SD were not determined from a region-of-interest (ROI), but from one representative pixel in the image (within the area of interest) after 1000 simulations for each b-value. The employed filtering scheme is shown in Fig. 7, where the selected wavelet families were as follows: $\mathbf{W_1}$ was the Haar wavelet, $\mathbf{W_2}$ was Daubechies of order 12 and $\mathbf{W_3}$ was Daubechies of order 5. The threshold factor in the hard-threshold filtering part was $\rho = 2$ (see Eq. 20).

Finally, a homogeneous object was simulated in order to assess improvement in contrast at different SNRs after noise reduction. The contrast is the difference in visual characteristics that makes it possible to distinguish an object from other objects or from the background in an image. When comparing contrast in original and filtered images, the following definition of contrast was employed [13, 14]:

$$c = \frac{S_{mean,1} - S_{mean,2}}{S_{mean,1} + S_{mean,2}} \tag{27}$$

where $S_{mean,1}$ and $S_{mean,2}$ are obtained as the mean values in two different regions over a homogeneous region-of-interest (ROI).

3.2 Experiments

The performance of the proposed noise-reduction algorithm was illustrated by application to phantom and in vivo data. Relatively high b-values and long echo times, relevant to, for example, q-space imaging, were employed in order to test the denoising technique under realistic low-SNR conditions. Diffusion imaging was performed at 3 T, using a locally modified double refocused spin-echo EPI pulse sequence, implemented on an MR unit with a maximum gradient amplitude of 40 mT/m (Magnetom Allegra, Siemens AG, Erlangen, Germany). Experiments on volunteers were approved by the local ethics committee.

The same filtering scheme and the same combination of wavelet families and orders as for the simulated images were employed: $\mathbf{W_1}$ was the Haar wavelet, $\mathbf{W_2}$ was Daubechies of order 12 and $\mathbf{W_3}$ was Daubechies of order 5. The threshold factor was $\rho = 2$.

Phantom experiment

A special spherical homogenous phantom was used, containing n-decane, a liquid linear hydrocarbon (i.e. a carbon chain without branched carbon backbone) with the chemical formula $C_{10}H_{22}$.

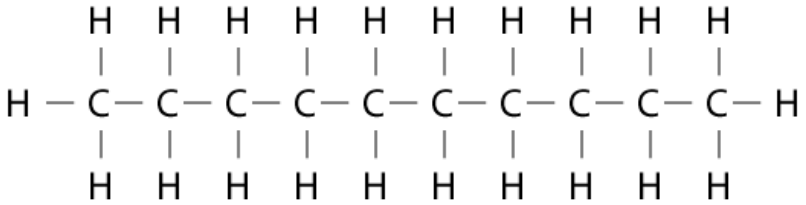


Fig. 9. The chemical structure formula of n-decane $C_{10}H_{22}$

The n-decane phantom was imaged using 30 different b-values ranging from 0 to 7424 s/mm². The diffusion encoding was applied in one direction, and the duration of each diffusion-encoding gradient was δ=15 ms and the time between the starting points of the diffusion-encoding gradients was Δ=130 ms. The imaging parameters were echo time 200 ms, field of view 220 mm, matrix size 128×128 and slice thickness 10 mm. The diffusion coefficient of n-decane (approximately $1.3 \cdot 10^{-3}$ mm²s^{-1} at 20°C) is relatively close to what can be observed in human brain tissue.

Volunteer experiment

In one healthy volunteer, diffusion MRI was performed with δ=15 ms and Δ=113 ms along 6 different directions (xz, -xz, yz, y-z, xy and -xy) using 46 different b-values in the range 0 to 4251 s mm^{-2}. The imaging parameters were as follows: repetition time 6000 ms, echo time 180 ms, field of view 220 mm, matrix size 128×128 and slice thickness 3 mm. Data, before and after noise reduction, were analysed with respect to anisotropy, and maps of the fractional anisotropy (FA) index were calculated.

3 Results

4.1 Simulations

The diagram in Fig. 10 shows the logarithm of signals intensities as a function of increasing b-value in the low diffusivity region ($ADC_1 = 0.1 \times 10^{-3}\ mm^2 s^{-1}$). The mean of the background, in the region in absence of any true signal, is presented as well. Since the ADC value is low, the signal never reaches the noise floor and no filtering effect can be observed. Non-filtered, filtered and theoretical signals curves overlap each other over the entire b-value range from b=0 to b=5800 mm^2s^{-1}.

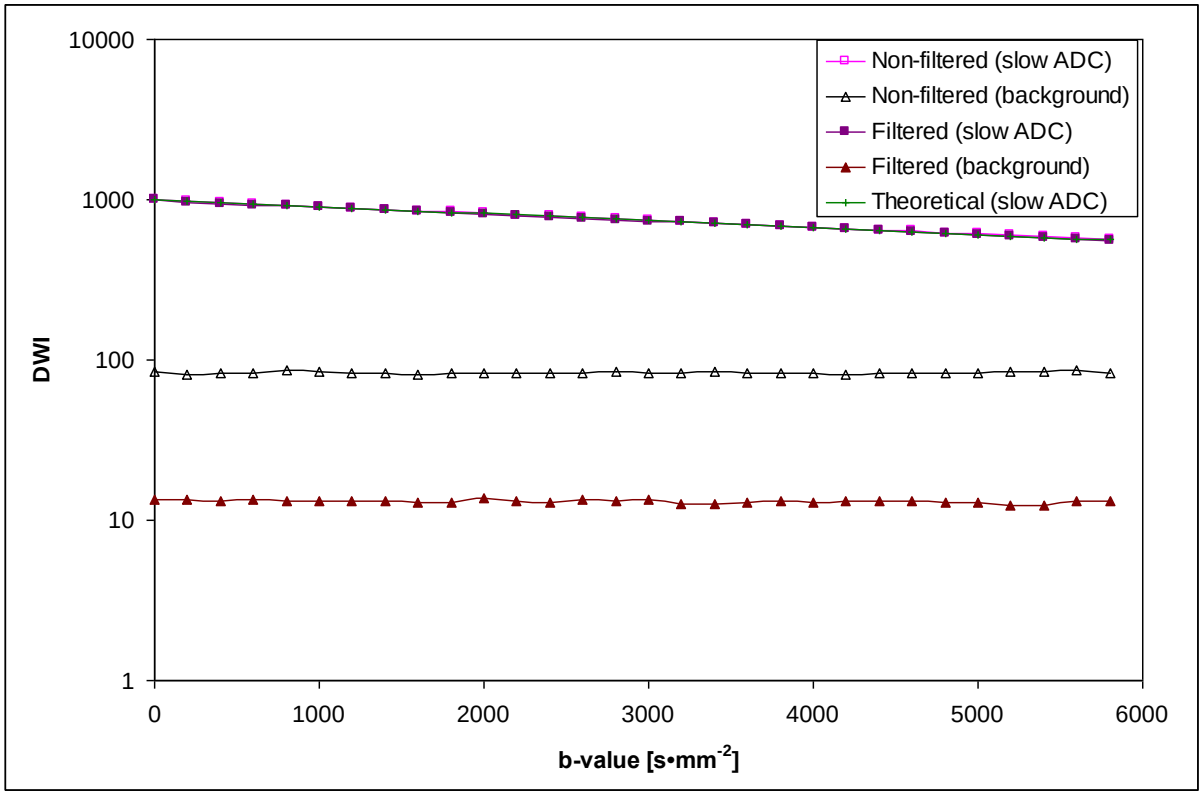


Fig. 10. Mean values of simulated diffusion-weighted image signal intensities in the slow ADC region and mean signal in the background as a function of b-value.

The diagram in Fig 11 shows the standard deviation in the same pixel as in Fig.10 above. After using the proposed scheme, the standard deviation decreased significantly both within the object and in the background region.

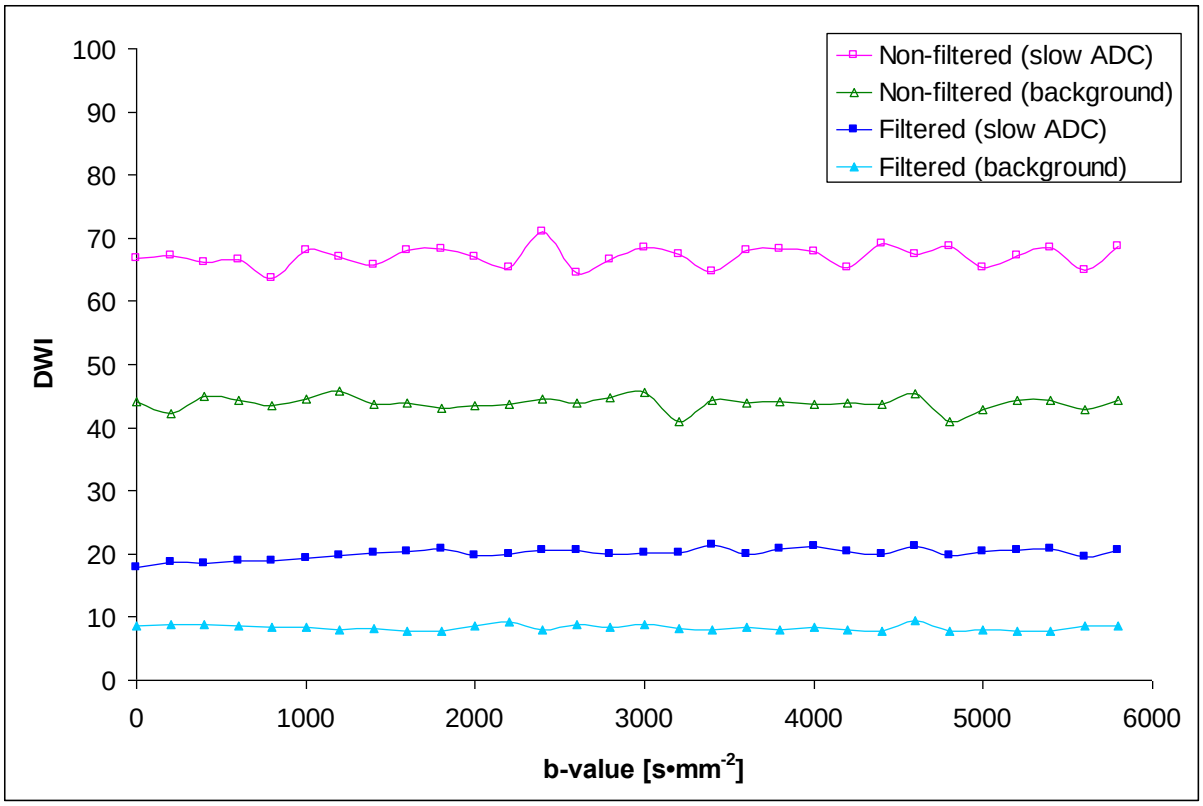


Fig. 11.Signal standard deviations in the simulated slow ADC region and in the background for different b-values.

Figure 12 shows the signal as a function of b-value in the high diffusivity area ($ADC_2 = 1.0 \times 10^{-3}\ mm^2 s^{-1}$) and in the background area. In this case the ADC is high enough to result in a significant signal decay; at b-values larger than approximately 1600 s·mm^{-2} the signal starts to approach the noise floor. Hence, the non-filtered signal starts to diverge from the theoretical value and, as the b-value increases further, the diffusion weighted signal coincides well with the noise floor. However, by filtering the image series, the noise floor was reduced by more than a factor of four. The filtered diffusion weighted signal thus follows the theoretical signal up to a b-value higher than 3000 s·mm^{-2}.

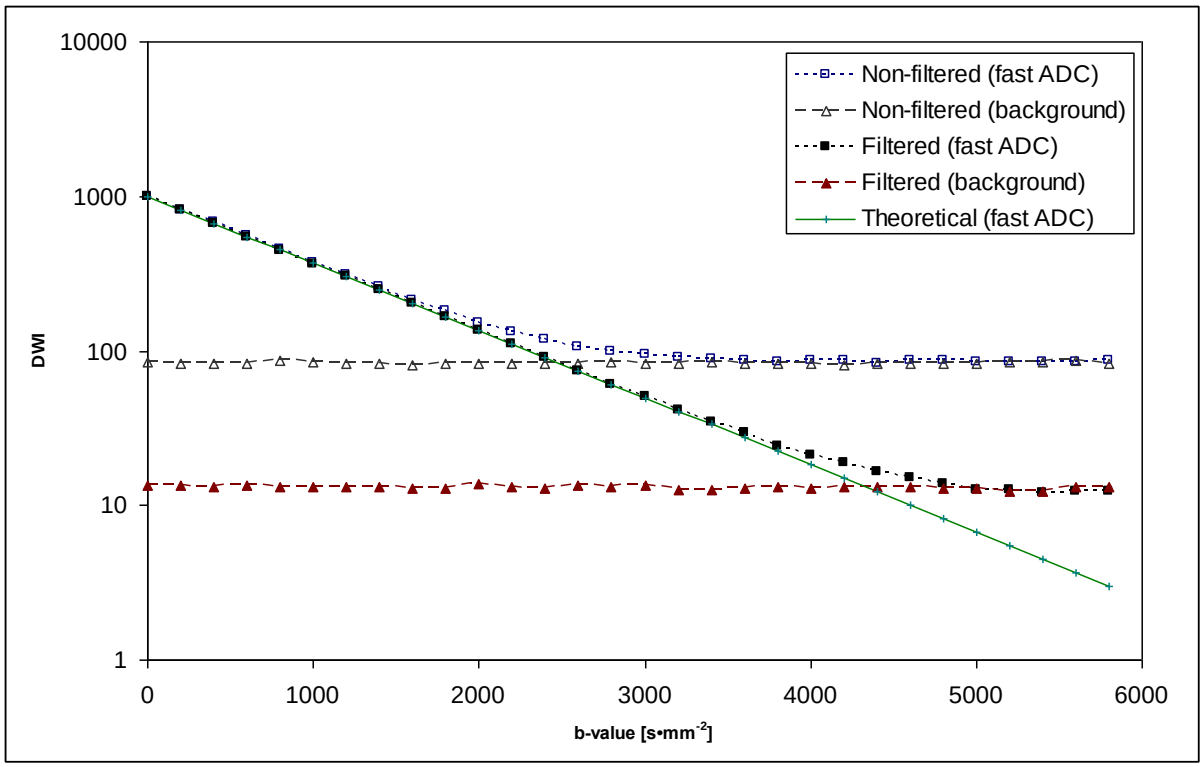


Fig. 12. Mean values of simulated diffusion-weighted signal intensities in the fast ADC region and mean signal in the background as a function of b-value.

Similar to the standard deviation in the low-diffusivity area, the SD in the high-diffusivity area was significantly reduced (Fig 13 below).

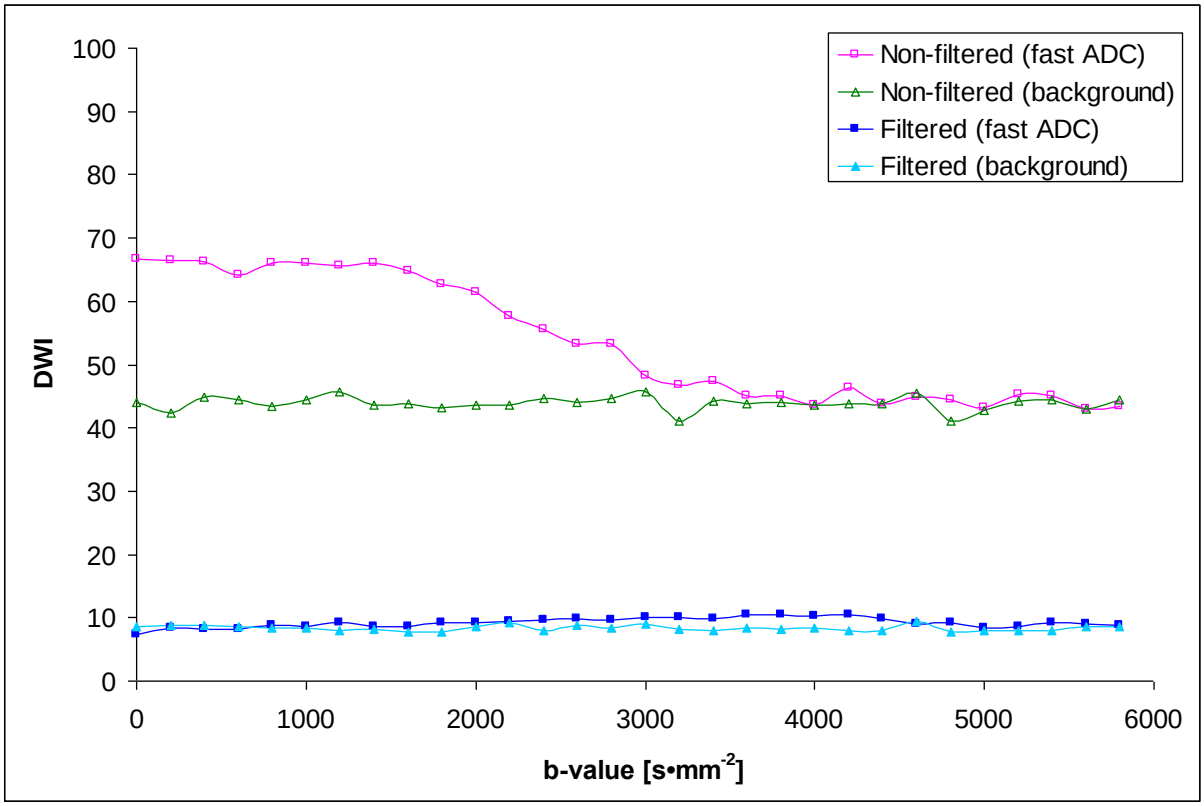


Fig. 13. Signal standard deviations in the simulated fast ADC region and in the background for different b-values.

The diagram in Fig. 14 shows the simulated signal as a function of b-value in the two-component ADC region (expected to show bi-exponential signal decay). The fitting of a bi-exponential function to the simulated dataset resulted in A_{noise}=703, B_{noise}=297, $ADC_{A,noise}$=1.22·10^{-9} m^2/s, $ADC_{B,noise}$=0.140·10^{-9} before denoising and $A_{denoise}$=677, $B_{denoise}$=323, $ADC_{A,denoise}$=1.27·10^{-9} m^2/s, $ADC_{B,denoise}$=0.180·10^{-9} after denoising. The relative errors in the four parameters A, B, ADC_A and ADC_B reduced from 3.3%, 7.9%, 2.5% and 29%, respectively, before denoising, to 0.4%, 0.8%, 1.4% and less than 0.01%, respectively, after denoising.

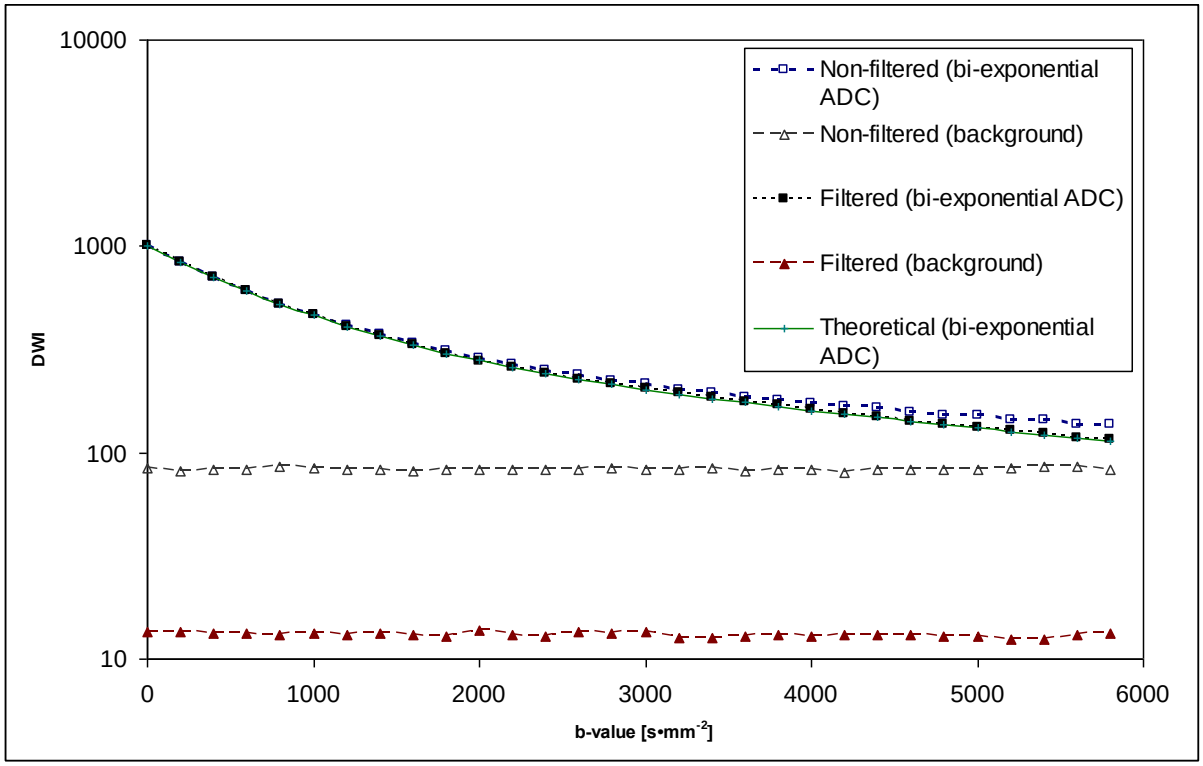


Fig. 14. Mean values of simulated diffusion-weighted intensities in the two-component (bi-exponential) ADC region and mean signal in the background as a function of b-value.

Figure 15 shows the signal SD as a function of b-value for the bi-exponential case.

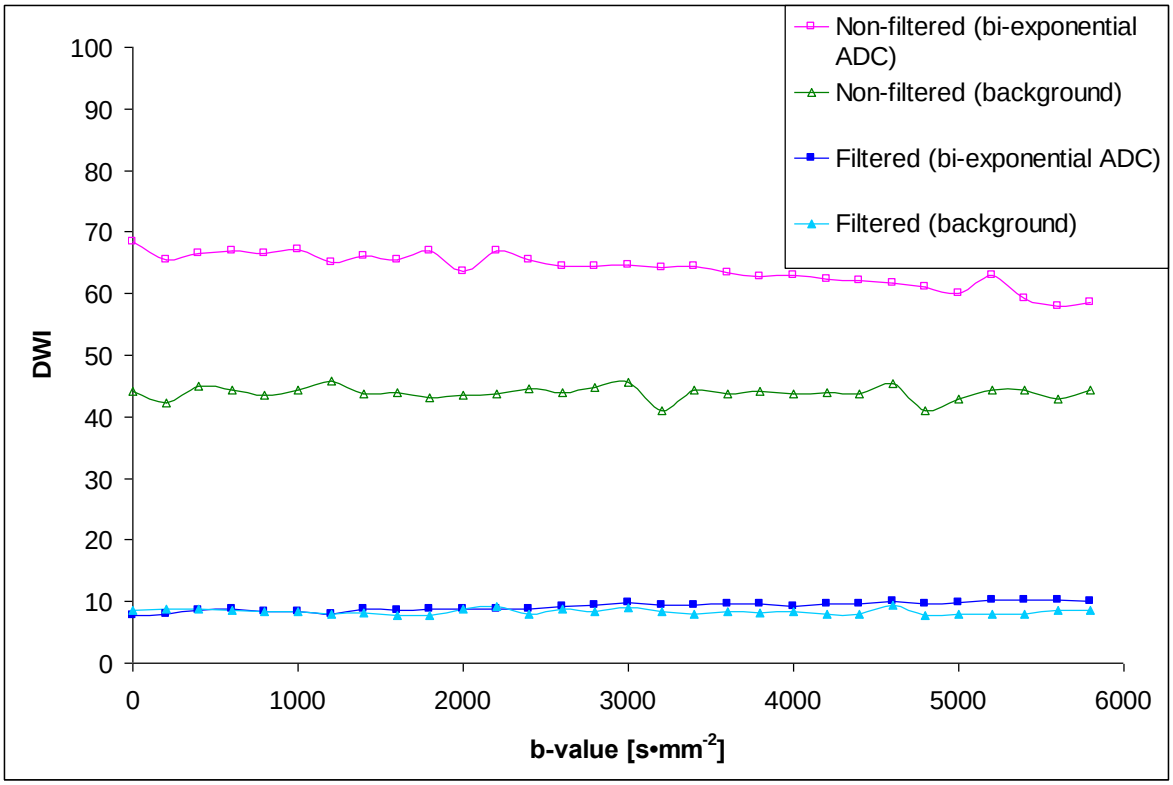


Fig. 15. Signal standard deviations in the simulated bi-exponential ADC case and in the background for different b-values.

The diagrams in Fig. 16 show profiles through the centre of a simulated homogeneous image for four different object signal to average background ratios (S/η ratios). The theoretical value of η in the unfiltered images was calculated to be 83.55 (according to Eq. 5).

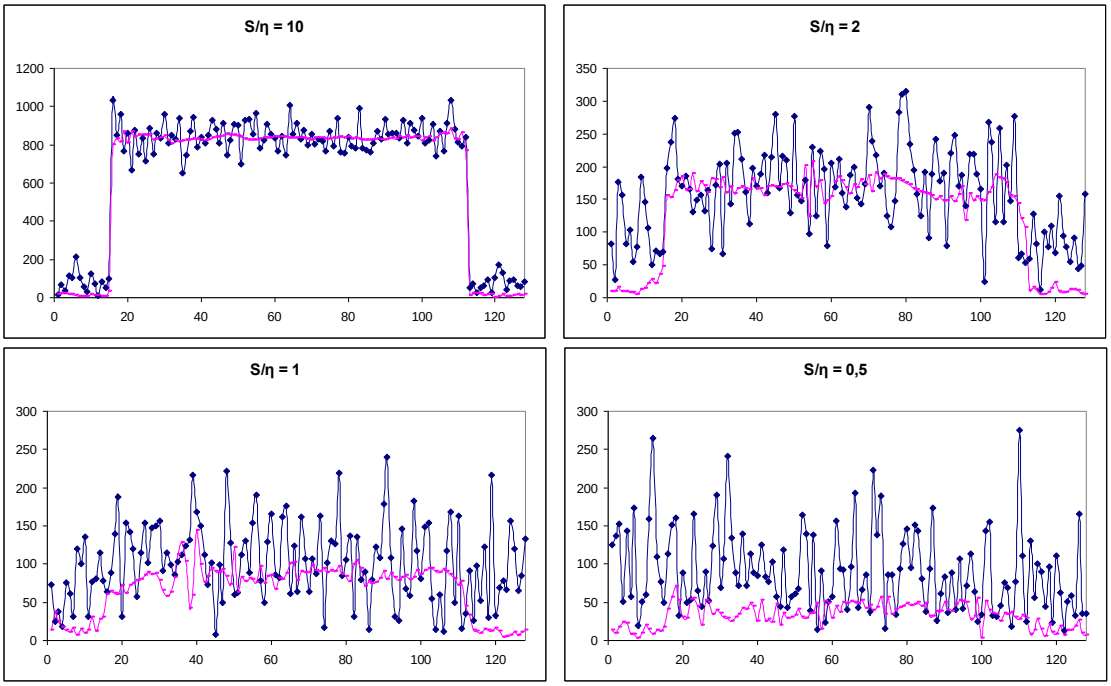


Fig. 16. Profiles through the centre of the simulated images (i.e. signal vs. position) for different S/η ratios, before (in blue) and after (in pink) wavelet-domain filtering.

The images in Fig. 17 show the homogenous simulated object with added Rice distributed noise (at the same S/η ratios as in Fig. 16 above) before and after denoising.

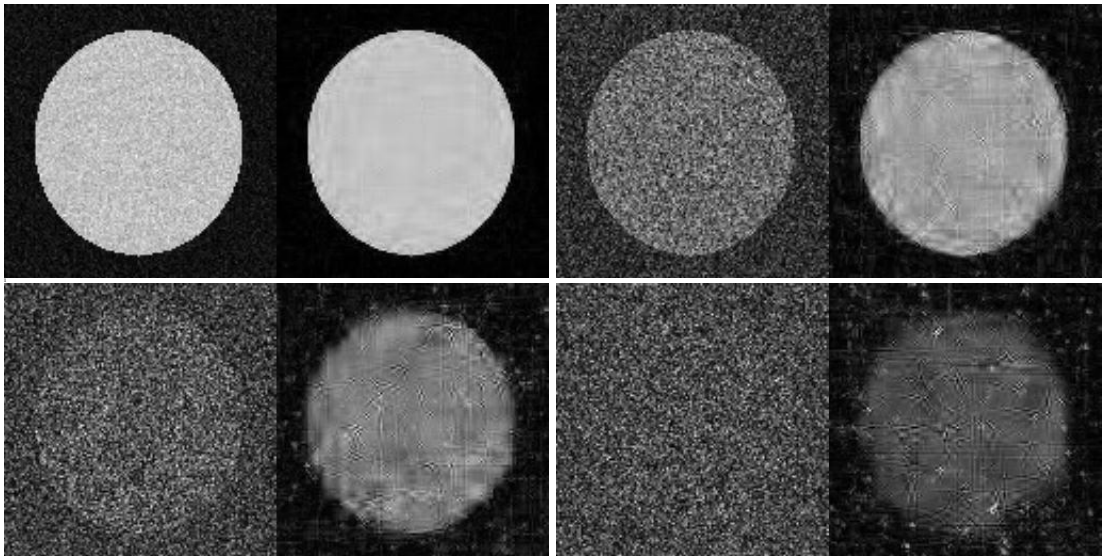


Fig. 17. Noisy and denoised images of a simulated homogenous object at different S/η ratios.

In Table 1, the mean values in the homogeneous object and in the background are displayed before and after filtering. Contrast was calculated according to Eq. 27, and the ratio of contrast after filtering to original contrast was also calculated.

S/η	Object (mean)	Background (mean)	Contrast	(Filtered)/(Unfiltered) Contrast Ratio
10 - unfiltered	838	84.1	0.818	1.17
10 - filtered	833	17.7	0.958	
2 - unfiltered	181	81.3	0.381	2.20
2 - filtered	167	14.9	0.838	
1 - unfiltered	112	82.4	0.152	4.94
1 - filtered	83.8	11.8	0.752	
0.5 - unfiltered	93.0	83.7	0.053	11.13
0.5 - filtered	45.0	11.6	0.591	

Table 1. Contrast and contrast ratio in noisy and denoised images for different S/η ratios.

4.2 Experimental results

Phantom experiment

The proposed wavelet denoising method was applied to the phantom images. In Fig. 18 the diffusion-weighted signal intensities and the mean of the background are plotted as a function of the b-value. All evaluated ROIs contained more than 1000 pixels.

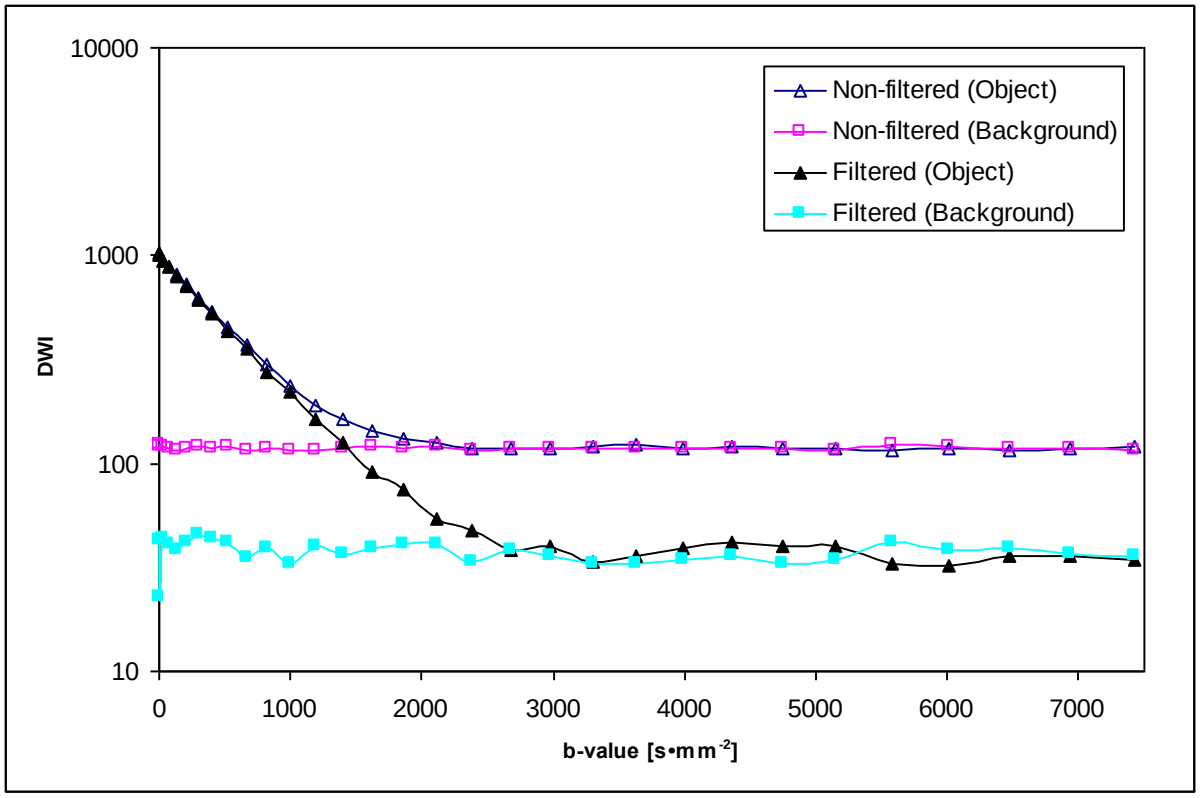


Fig. 18. Mean values of diffusion-weighted signal intensities in the n-decane phantom, and mean signal in the background as a function of b-value.

Figure 19 shows the standard deviations in the same ROIs as those in Fig 18 above. After filtering by use of the proposed wavelet denoising scheme, the standard deviation decreased significantly both in the object and in the background region.

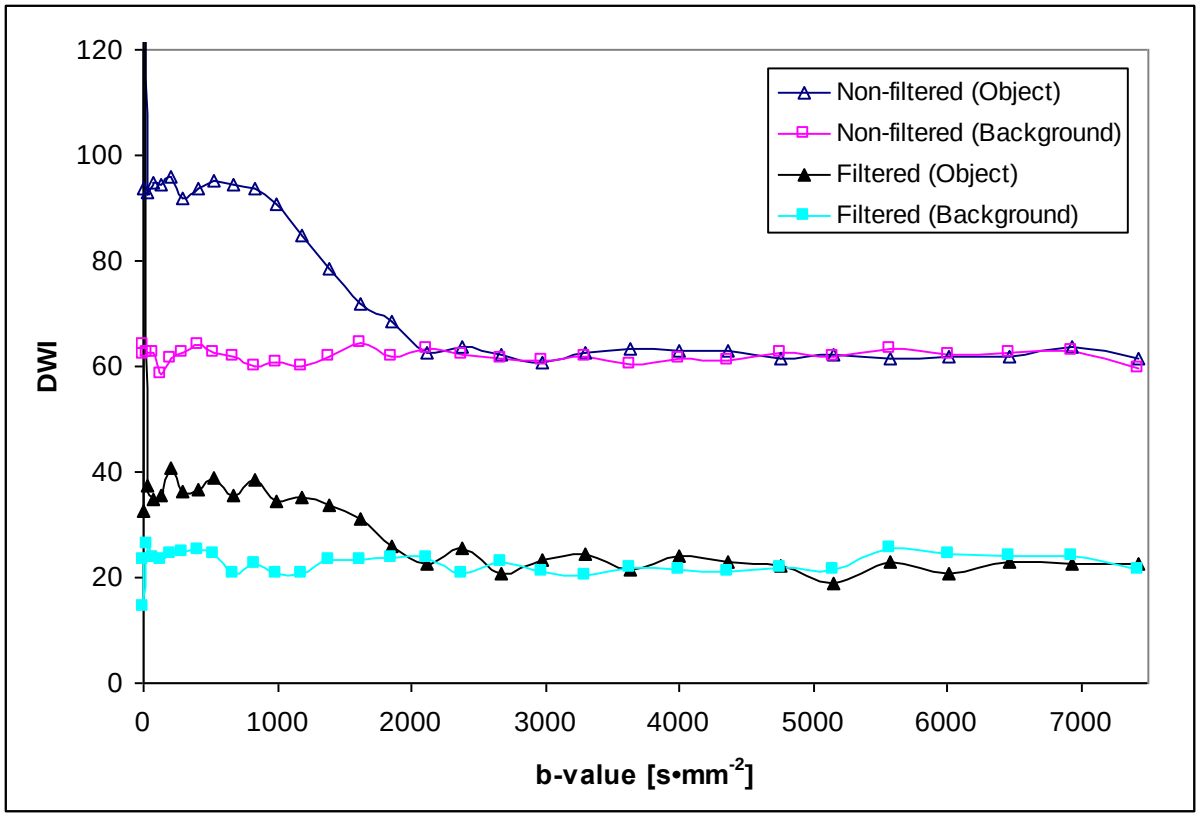
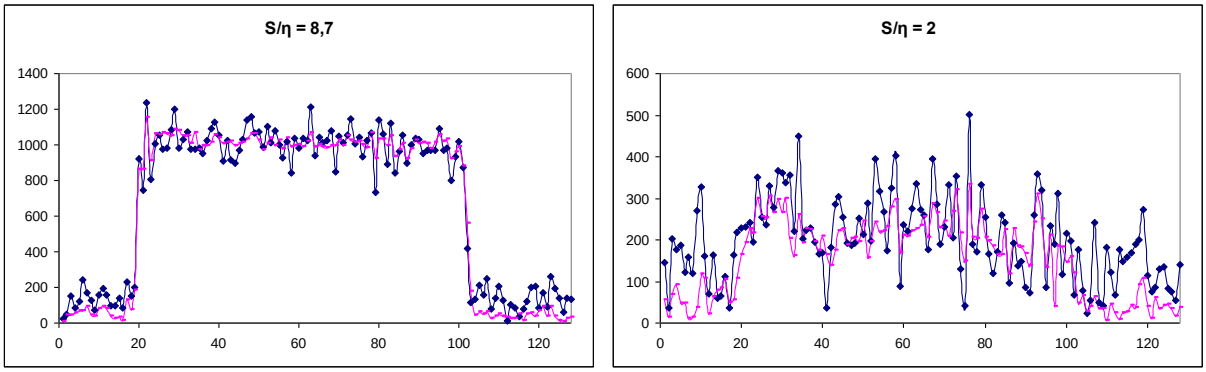


Fig. 19. Signal standard deviations in the n-decane phantom and in the background for different b-values, before and after wavelet denoising.

Figure 20 shows signal profiles through the image centre at four different signal to average background signal ratios (S/η ratios). The η value (the mean value of the noise, measured in the absence of true signal) in the unfiltered images was measured to be approximately 120 units. Only S/η ratios corresponding to the experimentally employed b-values could be investigated. The S/η ratios were, however, approximately the same as those investigated for the simulated images.



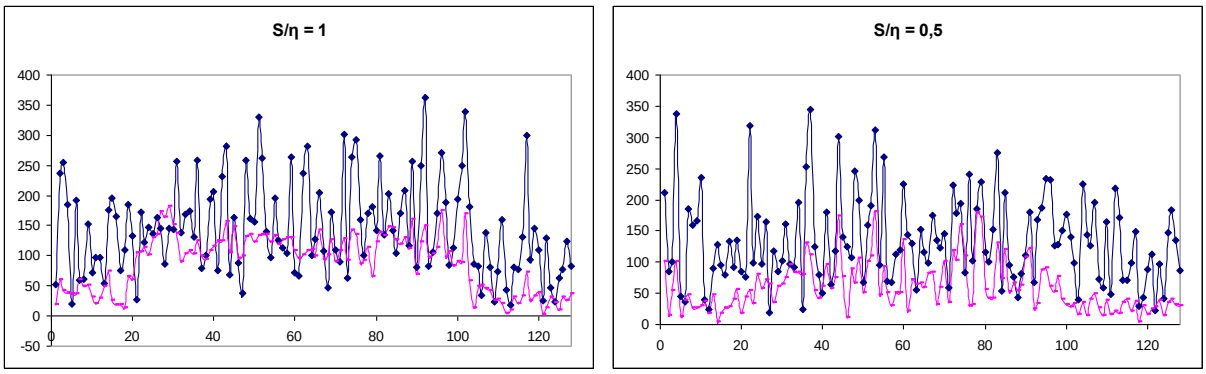


Fig. 20. Profiles through the centre of the images (i.e. signal vs. position) from the n-decane phantom for different S/η ratios, before (in blue) and after (in pink) wavelet-domain filtering.

The images in Fig. 21 below show n-decane phantom images influenced by Rice distributed noise before and after de-noising for the different S/η ratios presented in the diagrams in Fig. 20 above.

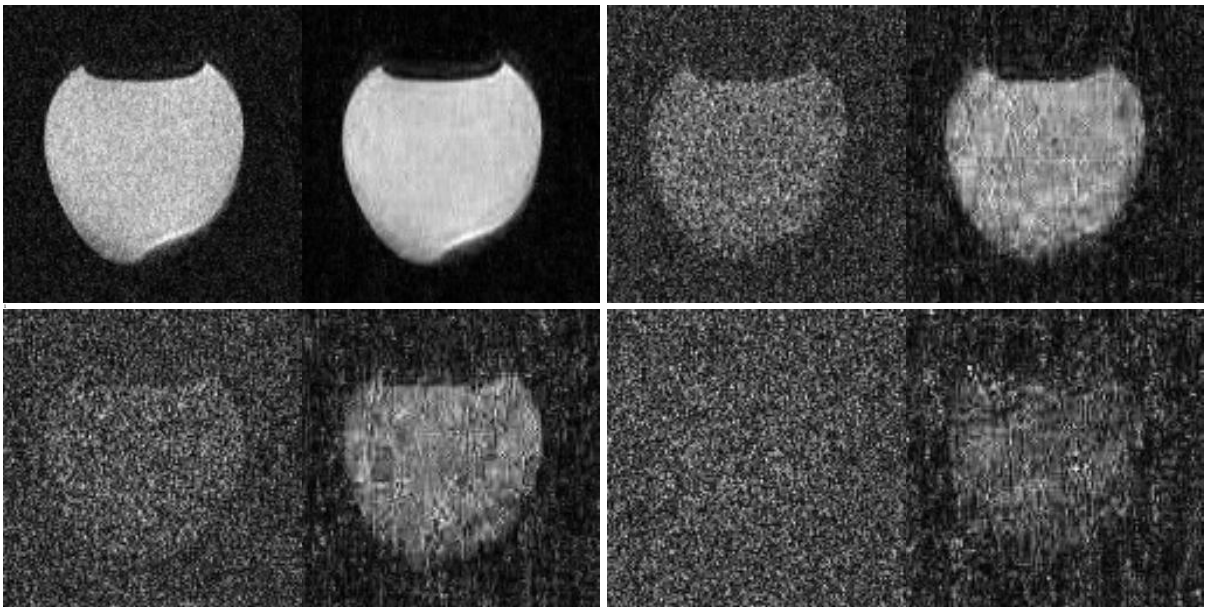


Fig. 21. Noisy and denoised images of the n-decane phantom at different S/η ratios.

In Table 2, the mean values within the object and in the background are given before and after filtering. Contrast was calculated to Eq xxx, and the ratio of contrasts after denoising to original contrast is also presented.

S/η	Object (mean)	Background (mean)	Contrast	(Filtered)/(Unfiltered) Contrast Ratio
8.7 - unfiltered	1020	121	0.789	1.17
8.7 - filtered	1020	42.8	0.919	
2 - unfiltered	242	119	0.339	2.01

2 - filtered	222	42.1	0.681	
1 - unfiltered	164	114	0.178	3.23
1 - filtered	128	34.4	0.576	
0.5 - unfiltered	136	115	0.082	4.48
0.5 - filtered	73.9	34.2	0.367	

Table. 2 Contrast and contrast ratio in noisy and de-noised phantom images for different S/η ratios.

Volunteer results

Figs. 22-23 show original and denoised images from the volunteer. The mean value in the background decreased more than four times, while the standard deviation in the same background ROI decreased about 3.5 times.

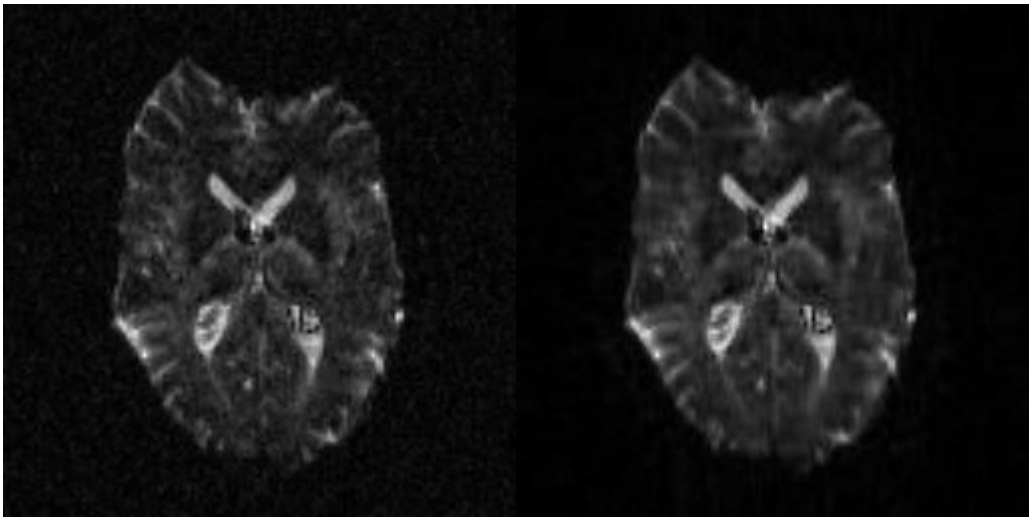


Fig. 22 Original (left) and denoised (right) S_0 images (at b=0) from a normal volunteer.

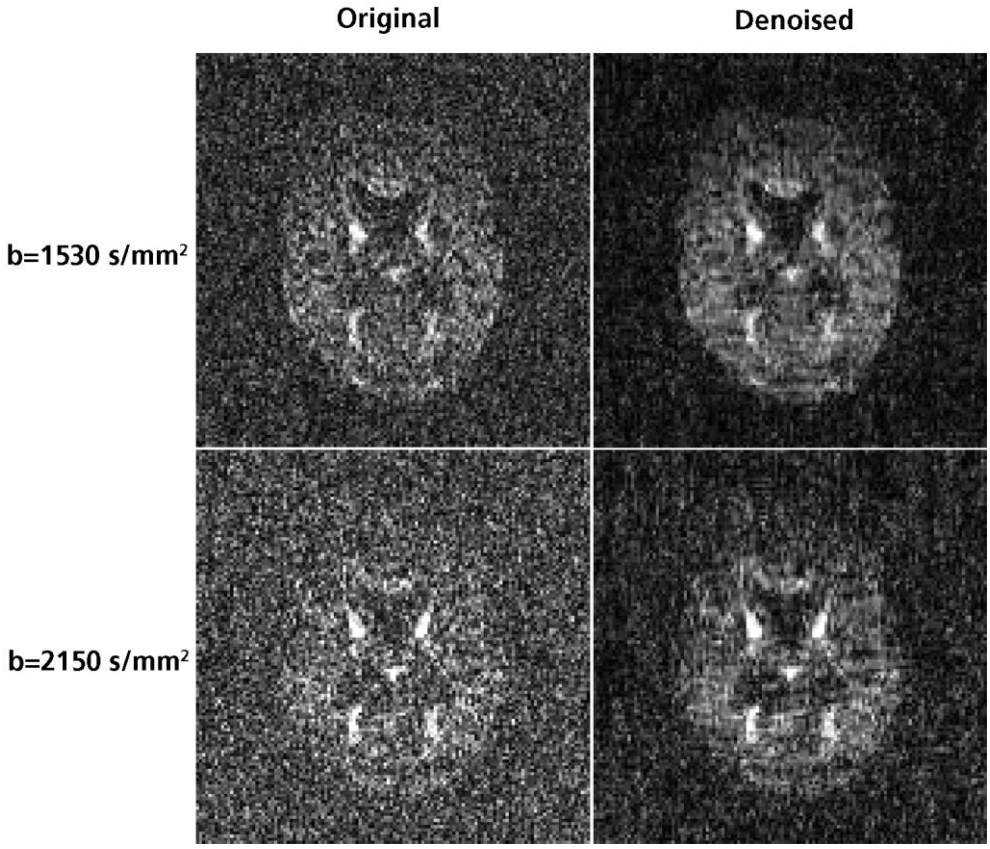


Fig. 23. Diffusion-weighted images from a normal volunteer, obtained at b-values of 1530 and 2150 s/mm2, before and after wavelet-based denoising.

In Fig. 24 the corresponding FA images are shown. FA based on b-values in the range from 0 to 930 s/mm² decreased when denoising was applied: FA recorded from ROIs in low-anisotropy regions (peripheral grey matter/cerebrospinal fluid) was reduced from 0.27 to 0.18 while white-matter FA values were reduced from 0.75 to 0.71, i.e. approximately 31% and 6%, respectively.

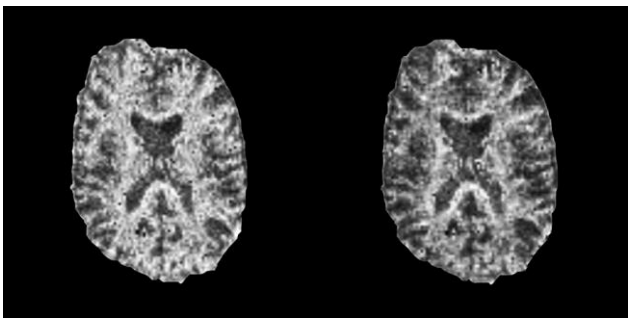


Fig. 24. FA maps from the volunteer, calculated using b-values in the range 0-930 s/mm^2. Left: Original data. Right: Denoised data. Background noise was removed by masking.

5. Future methodological improvements

In order to obtain an equivalent denoising through the whole series of images it is important to find the correct value of the standard deviation σ. The method employed in this study (and presented in previous publications) can provide somewhat higher values than the true standard deviation. The reason is that the finest scale wavelet coefficients do not contain only noise, but information about the fine details of the object signal as well. Hence, the images with higher SNR and finer details (i.e. the very first images in the DWI series) can be affected more than the rest. The phenomenon gives rise to a higher thresholding of the very first image than in the rest of the images, and a high empiric threshold factor ρ (see Eq. 14) can result in blurring or loss of fine details in the image. To obtain a more correct value of the standard deviation, the following method is proposed: The noise in the finest scale wavelet coefficients is Gaussian and distributed around zero. The wavelet coefficients from the true signal are few in relation to the wavelet coefficients from the noise. However, some of the coefficients from true signal are very high and they are the ones that influence the estimation of SD the most, while the small ones, in principle, drown in the noise (see Fig. 25).

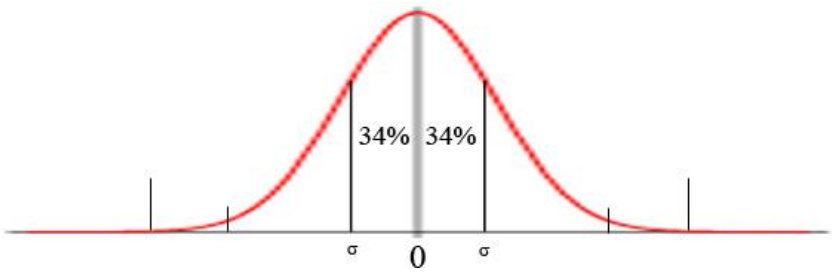


Fig. 25. Distribution of finest scale wavelet coefficients.

One solution is therefore to make a frequency distribution (histogram) of the coefficients and then display the absolute values (Fig. 26). Assuming a Gaussian distribution, the standard deviation can be found at the 68% level.

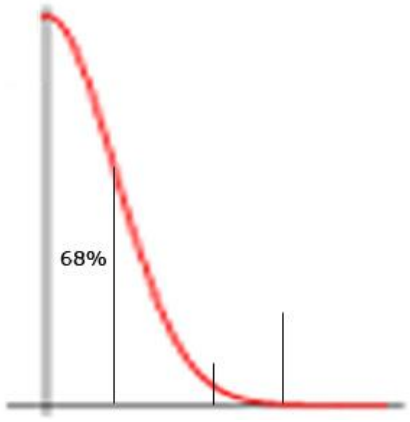


Fig. 26 Histogram of the absolute values of the finest scale wavelet coefficients.

Coefficients originating from true signal will, obviously, still be present in the distribution, but mainly at higher values than σ (illustrated by two additional high-value contibutions in Fig. 26 above). Hence, the σ estimation will in this case be less influenced by coefficients from the true signal.

6. Discussion and conclusions

Noise is an intrinsic problem in MRI, and it is of particular importance in diffusion MRI since diffusion encoding inevitably leads to reduced signal and since quantification is of relevance. This problem has been recognised, for example, in the assessment of fractional anisotropy [16, 17, 18], and additional complications occur in high b-value diffusion MRI due to the rectified noise floor in magnitude MR images [1]. With respect to wavelet denoising in MRI, most studies have focused on the qualitative improvement of image quality and/or contrast, but the effort spent on applications in which the quantitative assessment of signal levels is crucial has been limited. In diffusion MRI, the gain is two-fold; the accuracy and precision of diffusion parameters are influenced by the rectified noise floor as well as the signal SD, and both are reduced by the suggested wavelet-domain filtering procedure.

Denoising by wiener-like filtering in the wavelet domain is an established method, used, for example, for filtering of old movies, satellite and photographic images. The method is successful when noise is Gaussian distributed, and therefore it is very important to employ wavelet denoising on MRI data with Gaussian noise. Denoising of k–space data is, in principle, feasible but tends to cause image artefacts and blurred images. The main reasons are that independent filtering of the real and imaginary channels may cause phase errors, and since the Fourier space lacks the spatial resolution the fine details (high frequencies) are mixed with the noise and parts of the fine details are likely to be removed by the denoising. Filtering of MRI data in the complex image domain is therefore a better alternative, since it provides spatially encoded images that are suitable for application of the wavelet denoising tool as the noise is still Gaussian distributed in the real and imaginary images.

The choice of wavelet family and wavelet order is also important. Different wavelet families and orders do not separate the signal from the noise in equal manners. Hence, a mixture of different families and orders provides the best result. The optimal wavelet combination is image- and signal dependent and it is therefore impossible to specify the ultimate combination of wavelet families and orders. The selections made in the present work were based on empirical tests.

The observed reduction of the rectified noise floor, after denoising, was more than a factor 6 and the SD reduction was up to 85-90% in the simulated images. Of particular importance is the preserved ln(S) versus b-value linearity at higher b-values after denoising. The analysis of the dataset with bi-exponential signal decay implies that the behaviour of the noisy signal, in the interval in which the noise transfers from Gaussian to Rician, can indeed influence the obtained fast and slow ADC components as well as the estimated fraction of each component, even if the signal does not descend as low as to the noise floor. The improvement was, as expected, larger for parameters corresponding to the slow diffusion component. Furthermore, the denoising improved the contrast of the simulated images, especially in low-SNR regions (Table 1). After filtering, the contrast increased more than ten times in the low-SNR regions.

The general patterns of noise-floor and SD reduction were very similar for the

experimental data, as indicated by the n-decane results, although the effects of denoising were somewhat diminished compared with the simulated case. The SD of an ROI in a phantom can, obviously, be influenced by other factors than noise, but still the SD was reduced by 74% in the S_0 image, and the rectified noise floor was decreased by almost a factor 6. The quality of images obtained at high b-values was considerably improved by the filtering procedure, without introduction of any significant artefacts or smoothing effects. At the cost of somewhat more evident artefacts, even further lowering of the rectified noise floor, with retained ln(S) versus b-value linearity, can be obtained if ρ is increased.

The *in vivo* results showed that the noise-related image quality at high b-values was clearly improved by the wavelet denoising procedure. Furthermore, it is well established that noise leads to an overestimation of FA (when FA is calculated from images with intermediate diffusion weighting), and the degree of overestimation, i.e., the SNR dependence, is more pronounced at low anisotropy [16, 18]. In the present study, FA values (calculated from the 0-930 s/mm^2 b-value range) decreased after noise reduction, and the reduction in a low-anisotropy region was considerably larger.

Every filtering method has its own area of application, with advantages and disadvantages, and the proposed method is no exception. Hard threshold filtering in wavelet domain removes not only noise but even a part of the signal especially in high-frequency area and the wiener filter affects almost all coefficients, therefore a strong filtering can cause some artifacts in the output images like smoothing effects and visible line structures.

To conclude, the proposed model-free denoising algorithm works well for decreasing the standard deviation and background noise floor in diffusion weighted MR images. Furthermore, image contrast in low-SNR images can be considerably improved. By filtering DW images possible misinterpretations at medium or high b-values can be avoided. In the presence of noise, a true mono-exponential signal decay of the DW signal might, for example, be misinterpreted as a multi-exponential due to the signal behaviour when the noise distribution transfers from Gaussian to Rician. Finally, the method does not require any powerful computer systems and can be used in all MRI applications where the complex raw data (k-space data) are available.

7.References

1. Jones D, Basser P. “Squashing Peanuts and Smashing Pumpkins”: How Noise Distorts Diffusion-Weighted MR Data. Magn Reson Med 2004; 52:979-993

2. Wirestam R, Ståhlberg F. Wavelet-based noise reduction for improved deconvolution of time-series data in dynamic susceptibility contrast MRI. MAGMA. 2005;18:113-118.

3. Ghael S, Sayeed A, Baraniuk R. Improved wavelet denoising via empirical wiener filtering,;Proceedings of SPIE, San Diego, July 1997

4. J.Bergh, F. Ekstedt, M. Lindberg, “Wavelets”, Studentliteratur, Lund 1999

5. H. Choi, R. Baraniuk, Analysis of Wavelet-Domain Wiener Filters,; Department of Electrical and Computer Engineering, Rice University, Houston TX 77251-1892, USA

6. V. Bruni, D. Vitulano, “Old Movies Noise Reduction via Wavelets and Wiener Filter”, Instituto per le Applicazioni del Calcolo, Rome, Italy

7. Gudbjartsson H, Patz S., The Rician Distribution of Noisy MRI Data. Magn Reson Med 1996;34:910-4

8. Lund University, MR Physics, Internet hompage: http://www.jubileum.lu.se/MR_physics/

9. Benson H. University Physics. ISBN:0471006890, John Wiley & Sons LTD 1996, p.409

10. Charles K. Chui. An Introduction to Wavelets. ISBN:0121745848, Academic Press INC. (London) LTD, 1992.

11. Amara Graps, Astronomer at The American University of Rome, homepage, **http://www.amara.com**

12. Nowak RD, Baraniuk RG. Wavelet-domain Filtering for Photon Imaging System. IEEE Transactions on Image Processing 1999; Vol. 8, Nr. 5, p. 666-678

13. Nowak RD. Wavelet-Based Rician Noise Removal for Magnetic Resonance Imaging. IEEE Transactions on Image Processing 1999; Vol. 8, Nr. 10, p. 1408-1419

14. Alexander ME, Baumgartner R, Summers AR, Windischberger C,

Klarhoefer M, Moser E, Somorjai RL. A Wavelet-based method for improving signal to noise ratio and contrast in MR images. Magn Reson Imaging 2000; 18:169-82

15. Gudbjartsson H, Patz S. The Rician Distribution on Noisy MRI Data. Magn Reson Med 1995; 34:910-4

16. Pierpaoli C, Basser PJ. Toward a quantitative assessment of diffusion anisotropy. Magn Reson Med 1996;36:893-906

17. Skare S, Hedehus M, Moseley ME, Li TQ. Condition number as a measure of noise performance of diffusion tensor data acquisition schemes with MRI. J Magn Reson 2000;147:340-352

18. Armitage PA, Bastin ME. Selecting an appropriate anisotropy index for displaying diffusion tensor imaging data with improved contrast and sensitivity. Magn Resn Med 2000;44:117-121

19. Mulkern RV, Vajapeyam S, Haker SJ, Maier SE. Magnetization transfer studies of the fast and slow tissue water diffusion components in the human brain. NMR Biomed 2005;18:186-194

20. Papoulis A. The Fourier Integral and its Applications. New York: Mc-Graw Hill, 1962