Denoising functional MR images: a comparison of wavelet denoising and Gaussian smoothing

Alle Meije Wink[1,2] and Jos B. T. M. Roerdink[1,2], *Senior Member, IEEE*
[1] Institute for Mathematics and Computing Science
[2]Institute for Behavioral and Cognitive Neurosciences
University of Groningen,The Netherlands
E-mail: {wink,roe}@cs.rug.nl

***Abstract*— We present a general wavelet-based denoising scheme for functional magnetic resonance imaging (fMRI) data and compare it to Gaussian smoothing, the traditional denoising method used in fMRI analysis. One-dimensional WaveLab thresholding routines were adapted to two-dimensional images, and applied to 2D wavelet coefficients. To test the effect of these methods on the signal-to-noise ratio (SNR), we compared the SNR of 2D fMRI images before and after denoising, using both Gaussian smoothing and wavelet-based methods. We simulated a fMRI series with a time signal in an active spot, and tested the methods on noisy copies of it. The denoising methods were evaluated in two ways: by the average temporal SNR inside the original activated spot, and by the shape of the spot detected by thresholding the temporal SNR maps. Denoising methods that introduce much smoothness are better suited for low SNRs, but for images of reasonable quality they are not preferable, because they introduce heavy deformations. Wavelet-based denoising methods that introduce less smoothing preserve the sharpness of the images and retain the original shapes of active regions. We also performed statistical parametric mapping (SPM) on the denoised simulated time series, as well as on a real fMRI data set. False discovery rate control was used to correct for multiple comparisons. The results show that the methods that produce smooth images introduce more false positives. The less smoothing wavelet-based methods, although generating more false negatives, produce a smaller total number of errors than Gaussian smoothing or wavelet-based methods with a large smoothing effect.**

***Index Terms*— Functional neuroimaging, wavelet-based denoising, Gaussian smoothing, statistical parametric mapping, false discovery rate control.**

I. Introduction

Functional neuroimages often need preprocessing before being subjected to statistical analysis. A common preprocessing step is denoising, which is usually done via Gaussian smoothing. Smoothing suppresses noise, but it also changes the intensity variation of the underlying image. This suppresses, or even removes, detailed features of the original image. In this paper, we study wavelet-based denoising as a possible alternative to Gaussian smoothing. Wavelet-based denoising has the advantage over low-pass filtering that relevant detail information is retained, while small details, due to noise, are discarded. The performance of both approaches is compared with respect to (*i*) the improvement of the signal-to-noise ratio (SNR), (*ii*) the preservation of the shapes of active regions during the denoising process, and (*iii*) the improvement in the statistical analysis via statistical parametric mapping (SPM).

The focus of this paper is on functional magnetic resonance imaging (fMRI) time series. In an fMRI experiment, a person lying inside an MRI scanner is asked to perform a certain task while a series of scans of the brain are made. Brain regions involved in this task show increased concentrations of oxygenated blood, inducing local signal changes [1]. These signal changes are referred to as the blood oxygenation level dependent (BOLD) contrast, and detecting and characterising these changes is the main goal of fMRI time series analysis.

Most of the standard statistical tests assume Gaussian distributed noise. However, in the MR literature, noise in MR images is shown to be Rician distributed [2]. We analyse the BOLD contrast as the difference between two MR images (active minus baseline) both containing Rician distributed noise, and show that the distribution of BOLD noise is a close approximation of a Gaussian distribution. Thus, the standard tests requiring normally distributed noise can still be used.

The use of wavelets for the statistical analysis of fMRI and positron emission tomography (PET) studies is not new. Feilner et al. [3] use the wavelet transforms of difference images constructed from epoch-related fMRI experiments. Assuming a normal distribution of values in the difference images, activation is found by applying a t-test to the wavelet coefficients, using Bonferroni correction for multiple testing. The statistical map is found by applying the inverse wavelet transform. Ruttimann et al. [4] follow a similar approach. Their algorithm performs a two-stage test in the wavelet domain. The first test analyses the wavelet coefficients per direction channel: the coefficients are ordered by resolution and by direction (horizontal, vertical, and diagonal). It assumes the cumulative energy in each direction channel to be χ^2-distributed. All coefficients in a direction channel at a certain resolution are discarded if its cumulative energy is lower than the value predicted via this χ^2-distribution. The second test thresholds the wavelet coefficients in the remaining channels individually via a two-sided z-test. Both the channelwise test and the voxelwise test use the Bonferroni correction for multiple testing. The inverse wavelet transform is applied to the output of the second test, yielding an activation map. Raz et al. [5] perform an analysis of variance (ANOVA) in the wavelet domain, by thresholding the wavelet coefficients according to their score in a statistical test. The testing is done blockwise: at the lowest resolution, each coefficient is a block, and at higher resolutions the same number of blocks is used. The false discovery rate (FDR) is used to correct for multiple testing. Hilton et al. [6] use a wavelet-based denoising procedure known from the WaveLab project [7], an open source collection of wavelet routines, and compare this to their own data analytic thresholding procedure. The denoised time series are subjected to statistical testing by means of a voxelwise t-test. Turkheimer et al. [8] model PET images in wavelet space by applying statistical models to the frame-by-frame wavelet transformations of PET time series.

The main novelty of this paper is an extensive comparison of wavelet-based denoising and Gaussian smoothing, which is the standard denoising tool for functional neuroimages. All wavelet-based denoising methods mentioned above except Hilton's [6] perform the ensuing statistical tests in the wavelet domain. We favour the approach used by Hilton et al. for two reasons. First, performing a statistical test in the original domain enables a comparison between the wavelet-based methods and Gaussian smoothing as preprocessing steps. The statistical analysis process is exactly the same for all data

sets and can be kept outside the discussion. Secondly, performing the statistical test in the wavelet domain requires an inverse wavelet transform afterwards, which spreads out the activation in the final statistical map. Whether or not another threshold is needed on this map before display is questionable. Separating the denoising and the statistical analysis has another advantage. The data sets used in this study only require a simple statistical test, but most recent fMRI experiments often require much more complex procedures. It is not likely that all these tests can be done in the wavelet domain. However, if the denoised images are transformed back to the original domain, this problem does not occur.

Another difference between the current study and previous publications on this subject is that we include tests on simulated time series of which the SNRs and noise characteristics are known. Our definition of the BOLD signal allows a very precise characterisation of the noise in all test cases, so that the effect of each method on the SNR can be accurately determined.

Thirdly, we simulate brain activity in the time series by superimposing a time signal on a selected area. From the difference between the shape of the original active spot and the shape of the spot detected by statistical parametric mapping, we can make quantitative analyses of the denoising methods in terms of false positive and false negative error rates.

The remainder of this paper is organised as follows. Section II reviews a number of procedures to correct for multiple hypothesis testing. Section III first describes the noise in MR images, and introduces BOLD noise as the noise in the difference of two MR images. In section IV, we present the wavelet-based denoising methods available in WaveLab [7]. Adjustments have been made to these methods, to (*i*) make hem suitable for processing 2D images and (*ii*) to support noise with unknown autocorrelations. These denoising methods, as well as various degrees of Gaussian smoothing, are tested on 2D images in section V. In sections VI and VII, the wavelet-based and Gaussian methods are tested on an artificial time series, and compared in terms of their effects on the temporal SNR of the denoised time series and on the quality of the statistical parametric map. Finally, we compare the effects of these methods in a statistical analysis of a real fMRI data set in section VIII. Section IX contains some general conclusions.

II. Thresholding statistical maps: multiple hypotheses

Neuroimage analysis often entails hypothesis testing. Consider an experiment in which a subject is asked to perform a task while being recorded by the MRI scanner. The null hypothesis H_0 states that a brain region is not involved in that task. There may be more than one alternative hypothesis, indicating different patterns of activity. In general, rejecting H_0 means that brain activation related to the experiment has been detected. If a large number of hypothesis tests are done simultaneously, the expected number of rejected null hypotheses increases. This introduces the risk of false positives, also called type I errors (see Table I).

TABLE I
CLASSIFICATIONS AND MISCLASSIFICATIONS IN STATISTICAL TESTS.

	inactive voxel	active voxel
keep H_0	correct	type II error
reject H_0	type I error	correct

Statistical parametric mapping (SPM, [9]) is the common method to analyse functional neuroimages. SPM is based on the general linear model, which states that the response in an fMRI experiment can be written as a weighted sum of explanatory signals. Let the matrix $\boldsymbol{Y}_{[T\times N]}$ denote the fMRI data measured in the experiment, where each matrix element $\boldsymbol{y}_{ij}$ denotes the value measured at time $i = 1,\ldots,T$ and voxel location $j = 1,\ldots,N$. According to the general linear model,

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}, \tag{1}$$

where $\boldsymbol{X}_{[T\times M]}$ is a matrix, called the design matrix, whose M column vectors are the signals that represent the modelled effects, called the explanatory variables. The row vectors of the matrix $\boldsymbol{\beta}_{[M\times N]}$ are the weighting factors for those signals, and the values in the matrix $\boldsymbol{e}_{[T\times N]}$ are the residual errors of each voxel in each scan. A least-squares estimate $\boldsymbol{b}$ for $\boldsymbol{\beta}$ is given by $(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$. Given a model of $\boldsymbol{e}$, the significance of the coefficients of $\boldsymbol{b}$, and thus of the modelled effects, can be found in each voxel via hypothesis testing.

A statistical parametric map of N voxels consists of the p-values $p_i,\ 1 \le i \le N$. Given a distribution of outcomes, a p-value is the probability of getting an outcome at least as extreme as the one observed when the null hypothesis H_0 is correct. The SPM method allows for many statistical tests (t-tests, analysis of (co-)variance, regression analysis). In this paper, we will only discuss the one-sample t-test. The temporal noise in fMRI data sets is assumed to be Gaussian distributed, $N(\mu,\sigma^2)$. The null hypothesis H_0 states that $\mu = 0$. We test for increased activation, which means that we perform a one-sided test: H_1 states that $\mu > 0$. We do not know the real σ^2 of the temporal noise distribution, so it must be estimated via the sample variance s^2, which can be computed using the residual time signals in $\boldsymbol{e}$. Using this estimate, we can test for increased activation via the t-test. BOLD contrasts are constructed as linear combinations of rows of $\boldsymbol{b}$ (each of which is an image of N voxels), and their values are t-distributed. The relation between p-values and t-values is as follows. If a t-value in the BOLD contrast is in the upper α % of the distribution, it p-value is below α. In other words: a small p-value provides strong evidence against the null hypothesis. Active voxels are those with p-values below a significance level α. For one test, α is the probability of erroneously rejecting H_0.

Testing multiple independent hypotheses with the same significance level α leads to false positives. For one test, a level of 0.05 is acceptable, but for N simultaneous tests, approximately $0.05\ N$ detected activations will be false positives. Simultaneous tests deal with the 'omnibus' null hypothesis [10], which states that there is no activation in any of the individual tests. Testing the omnibus null hypothesis at level α can be used to decide if there is activity in the image, but not where it is. The omnibus null hypothesis is said to have weak type I error control.

One way to deal with this is Bonferroni correction, where the significance level α is replaced by α/N. This guarantees that the proportion of false positives does not exceed α in any subset of the simultaneous tests. Bonferroni is therefore said to have strong type I error control [11], meaning that rejecting H_0 in a certain region in the brain is evidence for activation in that very region. Bonferroni correction not only affects the number of type I errors: reducing the probability of rejecting the omnibus null hypothesis also affects the number of true positives. This introduces false negatives, or type II errors (see Table I). For fMRI signals, which are spatially correlated, Bonferroni correction may be too conservative.

In most neuroimage analysis programs that use the SPM method for the statistical part of the analysis, images are smoothed with Gaussian filters [12]. The motivation for this is twofold: (*i*) it increases the signal-to-noise ratio (SNR), and (*ii*) it controls the smoothness of the noise in the images when viewed as a lattice representation of a continuous, stationary Gaussian random field

(GRF). In order to use GRF theory, smoothing may be necessary to bring the data more into agreement with the model assumptions. Once the smoothness of an image is known or controlled via filtering, threshold values for statistical maps can be computed using the Euler characteristic of Gaussian random fields to correct for multiple testing [13], [14]. The method has strong type I error control. This approach has two drawbacks. Firstly, even after filtering, the noise in the smoothed images often still differs from a GRF. Noise and signal are smoothed together, so smoothing makes it even more difficult to separate signal from noise. The underlying image is not likely to represent a continuous GRF, so the corrected threshold is likely to be biased. This will influence all corrected p-values. This problem is even more serious when the smoothing kernel has another full width at half maximum (FWHM) than the intrinsic FWHM of the underlying images. Secondly, the smoothing process suppresses and removes details in the images. This hampers the detection of detailed regions during subsequent analysis.

The false discovery rate (FDR) is another alternative for multiple test correction, that is also applied in functional neuroimaging [11]. It does not require spatial smoothness. The FDR is defined as the expected proportion of false positives among the rejected null hypotheses [15]:

$$\text{FDR} = E\left(\frac{\#\text{type I errors} \quad (=\#\text{false positives})}{\#H_0 \text{ rejected} \quad (=\#\text{detections}),}\right) \qquad (2)$$

E denoting expectation, and is identical to 0 when $\#\text{detections} = 0$. The following algorithm results in an FDR approximately equal to q, with $0 \leq q \leq 1$. Given N voxels with p-values $p_1, \ldots, p_N$ and an FDR parameter q, an FDR-controlling threshold selection procedure is given by:

1: define

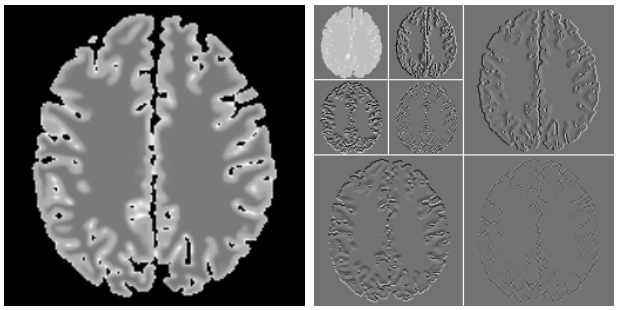
$$\eta(N) = \begin{cases} 1, & \text{when the } p\text{-values are uncorrelated} \\ \sum_{i=1}^{N} 1/i, & \text{otherwise} \end{cases}$$

2: order the p-values so that $p_i \leq p_{i+1}$ for every $0 < i < N$;
3: let r be the largest i for which $p_i \leq qi/N\eta(N)$;
4: reject the null hypotheses of the voxels with $p_i \leq p_r$

This method has weak type I error control. Notice that, when $\eta(N) = 1$, the graph of qi/N vs. i is a straight line from $(1, q/N)$ to (N, q), and that q/N and q correspond to the Bonferroni-corrected threshold and the 'omnibus' threshold, respectively. If $\eta(N) = \sum_{i=1}^{N} 1/i$, the method is much more conservative [11], so the $\eta(N) = 1$ option is preferable if it is allowed.

This correction method has a number of advantages over Bonferroni correction and correction based on GRF theory. It is less conservative than Bonferroni correction and it does not require smoothing, in contrast to GRF theory. The most important advantage is its adaptivity: the threshold is selected on the basis of the distribution of p-values, so after hypothesis testing. Therefore, it can be applied to any set of p-values resulting from a statistical test. It is independent of the type of test and the number of hypotheses, so that comparisons between studies with equal FDRs are possible.

The value $\eta = 1$ is not only valid for uncorrelated p-values, but also for sets of p-values that are positive regression dependent within subsets (PRDS). Genovese et al. [11] explain the PRDS property briefly, and they argue that statistical parametric maps have this property. In section VII we discuss the distribution of the p-values under the null hypothesis and their spatial correlation in greater detail. A uniform distribution of p-values under the null hypothesis proves the validity of the statistical test. The spatial correlation of the residual noise is tested because a time series with Gaussian noise that is positively correlated among voxels, is PRDS [11].

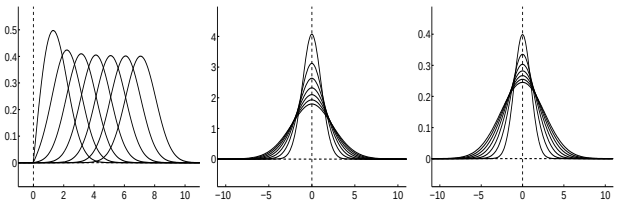


(a) (b)

Fig. 1. (a) A simulated MR image. (b) A 2D nonstandard fast wavelet transform (FWT) of (a).

III. Noise models for fMRI

The computation of p-values in fMRI research is usually done with standard tests, such as the t-test or F-test. The use of these tests is justified by assuming the BOLD noise to be Gaussian under the null hypothesis. In the MR literature, however, the noise in MR images is assumed to be Rician distributed [16], [2], [17]. Rician noise differs from Gaussian noise in that it is multiplicative instead of additive, *i.e.* it depends on the signal intensity, and the probability density function (pdf) of the noise is very asymmetric for low signal intensities.



(a) (b) (c)

Fig. 2. (a) Rician pdfs for different signal intensities. Higher intensities have noise distributions similar to a Gaussian. (b) pdfs of the difference of two Rician distributed sets for a fixed signal intensity, with different standard deviations. (c) Gaussian pdfs, with different standard deviations.

The difference between two images with Rician distributed noise has a symmetric distribution (see Fig. 2). In such difference images, the distribution of noise is very close to Gaussian noise, as can be seen in table II. This table shows, for each listed image size, the mean p-values of the Kolmogorov-Smirnov (KS) test statistic on 32 images of that size, for (*i*) the difference between two images with Rician noise and signal intensity one, (*b*) the difference between two images with Rician noise and signal intensity five, and (*c*) an image containing $N(0, 1)$ noise. The null hypothesis of the KS test is that the noise is normally distributed, and it is rejected if the p-value of the KS test statistic is below 0.05. For very low signal intensities, a deviation from Gaussianity is noticeable only in very large images. We conclude that it is safe to use techniques based on the assumption of Gaussian noise for BOLD images. For a more detailed analysis, we refer to [18].

The BOLD effect involves spatial autocorrelation due to the spatial extent of neuronal events, but this autocorrelation is not exactly known [19], [20]. We tested two types of spatial correlation: white noise, and $1/f$ noise, which has a $1/f$ power spectrum. The motivation for the latter type of noise is that, due to the MR frequency encoding, a unit pulse gets the shape of a peak with exponential slopes [21]. Section V describes how we simulated MR noise.

TABLE II
p-VALUES PRODUCED BY THE KS TEST FOR THE DIFFERENCE BETWEEN IMAGES WITH RICIAN DISTRIBUTED NOISE (R), WITH SIGNAL AMPLITUDES (A) OF 1 AND 5, AND FOR IMAGES OF THE SAME SIZE WITH $N(0,1)$-NOISE (N).

size	p-value R(A=1)	p-value R(A=5)	p-value N
2×2	0.6573	0.5607	0.4569
4×4	0.5761	0.5565	0.4249
8×8	0.5511	0.5493	0.4894
16×16	0.5801	0.5564	0.5854
32×32	0.5833	0.5378	0.5946
64×64	0.5629	0.4869	0.4816
128×128	0.5270	0.5426	0.5147
256×256	0.4390	0.5554	0.5225
512×512	0.3210	0.5219	0.4006
1024×1024	0.0587	0.5236	0.5037

IV. WAVELET-BASED DENOISING

Wavelet bases are bases of nested function spaces, which can be used to analyse signals at multiple scales. Wavelet coefficients carry both time and frequency information, as the basis functions vary in position and scale. The fast wavelet transform (FWT) efficiently converts a signal to its wavelet representation [22]. In a one-level FWT, a signal c_0 is split into an approximation part c_1 and a detail part d_1. In a multilevel FWT, each subsequent c_i is split into an approximation c_{i+1} and detail d_{i+1}. For 2D images, each c_i is split into an approximation c_{i+1} and three detail channels d^1_{i+1}, d^2_{i+1}, and d^3_{i+1}, for horizontally, vertically, and diagonally oriented details, respectively (see Figs. 1b and 4a). The inverse FWT (IFWT) reconstructs each c_i from c_{i+1} and d_{i+1}. If the wavelet basis functions do not have compact support, the FWT is computed most efficiently in the frequency domain. This transform and its inverse are called the Fourier-wavelet decomposition (FWD) and Fourier-wavelet reconstruction (FWR), respectively, see [23] for more details.

A. Wavelet bases

As it is difficult to mathematically characterise functional brain signals, a basis with general properties is preferable. Of the common wavelet bases, like Daubechies wavelets [24], symmlets, coiflets [25], and splines, spline bases have been shown to possess the best approximation properties, such as the smallest L_2 error [26]. Because of their smoothness, splines are well localised in both the frequency and time domains. Earlier studies about the use of wavelets in fMRI analysis [4], [27] favour the use of symmetric wavelets and scaling functions, because they do not introduce phase distortions. Orthogonal bases are recommended, because they transform white noise into white noise [28]. Unser et al. have proposed an FWT that uses fractional spline wavelets [29], [30]. Fractional splines are splines of a real-valued degree, which can be used to produce wavelet bases. They come in many flavours, such as symmetric and causal, orthogonal and biorthogonal.

In view of the above, symmetric, orthonormal cubic spline wavelets (see Fig. 4b) are the best choice for this study. Symmetric, orthogonal, smooth wavelet basis functions cannot have compact support but exponential decay [31]. For this reason, we use a frequency domain implementation via the FWD to compute the FWR.

B. Denoising images by wavelet domain thresholding

The WaveLab package by Donoho et al. [7] contains a number of schemes for wavelet-based denoising, including HybridThresh, InvShrink, MinMaxThresh, MultiMAD, SUREThresh, VisuShrink,

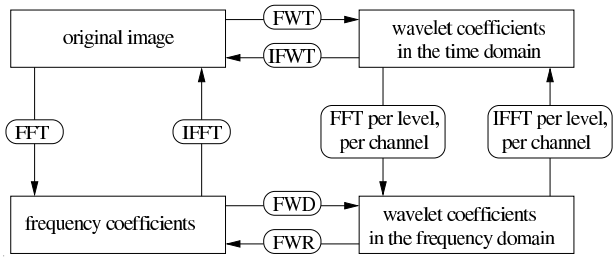


Fig. 3. The FWT and FWD of a signal are interchangeable via the fast Fourier transform (FFT).

and WaveJS [32], [33]. These routines are based on thresholding detail coefficients in the wavelet domain. An important characteristic of these schemes is the amount of smoothness they introduce in the denoised image [6].

InvShrink uses the VisuThresh threshold, which is $\sqrt{2 \log N}$ for a vector d_i of detail coefficients of length N. The signal is scaled before thresholding so that it has unit standard deviation. In multilevel transforms, the height of threshold is doubled for each subsequent level. MultiMAD also uses VisuThresh, and rescales the d_i of each level so that its median absolute coefficient value is 0.6745, which is the median absolute deviation (MAD) of an $N(0,1)$-distribution. MinMaxThresh uses a minimax threshold [33], which minimises the maximum risk. SUREThresh uses Stein's Unbiased Risk Estimate [32]. VisuShrink uses VisuThresh with shrinkage of small coefficients, called soft thresholding, as the default, but hard thresholding, *i.e.* removal of small coefficients, is also used. HybridThresh uses VisuThresh for sparse vectors and SUREThresh for dense vectors. WaveJS uses a threshold based on the James-Stein estimate [32]. InvShrink and MultiMAD change the threshold for each decomposition level, while MinMaxThresh, SUREThresh, VisuShrink, HybridThresh, and WaveJS use one global threshold.

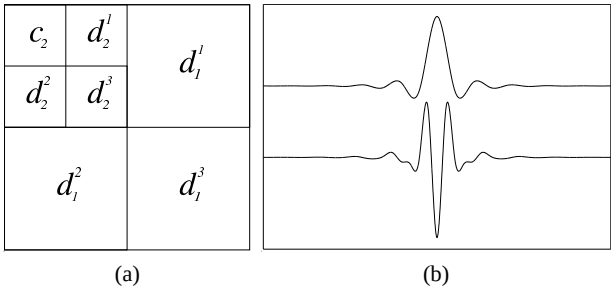


Fig. 4. (a) Ordering of the approximation and detail coefficients of a two-level 2D nonstandard FWT. (b) Symmetric orthonormal cubic spline scaling function (top) and corresponding wavelet (bottom).

If a 1D threshold selection scheme were used in a 2D FWT, for example by applying a 1D thresholding scheme in both spatial dimensions, assumptions used by the WaveLab routines would be violated, because the threshold for detail coefficients would in some cases be determined from both approximation and detail coefficients, and in some cases from detail coefficients only (see Fig. 4a). So it is necessary to respect the ordering of coefficients in a 2D FWT. The WaveLab thresholding schemes are based on the assumption of white Gaussian noise. If the autocorrelation of the noise is unknown, a level-dependent threshold should be used [34]. To meet these requirements, we have made 2D versions of the denoising routines in which all channels with detail coefficients (see Fig. 4a) can be thresholded individually. Each direction channel at each resolution (each square

in Fig. 4a) is thresholded independently using the WaveLab routines. This setting works for each FWT of stationary noise with an unknown autocorrelation: there is a difference in variance between detail channels at different resolutions, but within each channel there is constant variance [34].

V. Denoising 2D images

The BOLD contrast is defined as the difference between an MR image of a brain with increased local activity and an image of the same brain under resting conditions [35]. We used the BrainWeb Simulator [36] to obtain a noise-free T2*-weighted MR image template. The parameters of the simulator were as follows. Modality: T2; voxel size: $1\times1\times1$ mm^3; Noise: 0 %; intensity non-uniformity: 0 %. The Brain Extraction Tool [37] was used to remove non-brain voxels, by setting their intensities to 0. One slice (slice no. 108) of this image was selected, and used as a noise-free MR brain template.

Rician noise was added to this template as follows. Let $m(\boldsymbol{x})$ denote the template slice, and σ_m its standard deviation. Two images $n_1(\boldsymbol{x})$ and $n_2(\boldsymbol{x})$ containing i.i.d. $N(0,\sigma_n^2)$-distributed noise with a known standard deviation σ_n were made, and the noisy MR image $\tilde{m}$ was computed as $\tilde{m}(\boldsymbol{x}) = \sqrt{(m(\boldsymbol{x})+n_1(\boldsymbol{x}))^2 + n_2(\boldsymbol{x})^2}$. The Rician distributed noise in $\tilde{m}$, computed as $r(\boldsymbol{x}) = \tilde{m}(\boldsymbol{x}) - m(\boldsymbol{x})$, has a standard deviation σ_r with approximately $\sigma_r = \sigma_n\sqrt{2-\pi/2}$ [2]. This approximation was used to create noisy MR images with a known signal-to-noise ratio (SNR). The SNR of the noisy images was computed as:

$$\mathrm{SNR}_{\tilde{m}} = 10\ \log_{10}\frac{\sigma_m}{\sigma_r}, \tag{3}$$

Gaussian noise with a $1/f$ power spectrum was produced by transforming both n_1 and n_2 to the frequency domain and multiplying their frequency spectra with a $1/\sqrt{f}$ mask, yielding power spectra with a $1/f$ falloff. Both the real and imaginary parts of the spectra were multiplied, so the phase spectra did not change. Multiplication in the frequency domain is equivalent to convolution in the spatial domain. Because this is a linear operation, the noise distribution in the $1/f$ versions of the n_1 and n_2 images remains Gaussian. These transformed versions of n_1 and n_2 were used to obtain Rician distributed noise with a $1/f$ power spectrum.

The BOLD image was constructed from two of these noisy MR images as follows:

1: create two noisy MR images $\tilde{m}_1$ and $\tilde{m}_2$ using the above procedure;
2: define an 'active region' inside the template brain and create a noise-free BOLD image f_0 as:
$$f_0(\boldsymbol{x}) = \begin{cases} 1, & \text{if } \boldsymbol{x} \text{ is inside the active region} \\ 0, & \text{otherwise} \end{cases}$$
3: add activity to $\tilde{m}_2$ by adding $c\ f_0$, where c = 5% of the maximum intensity of the MR template.
4: compute a BOLD image f_1 as $f_1(\boldsymbol{x}) = \tilde{m}_2(\boldsymbol{x}) - \tilde{m}_1(\boldsymbol{x})$

The top row of Fig. 5 shows the active spot (bright area), and two noisy images $\tilde{m}_1$, with white and $1/f$ noise, respectively. The bottom row shows the noise-free BOLD image, and BOLD images with white noise and $1/f$ noise, respectively with an SNR of 15 dB. Because there is hardly any signal in the noise-free BOLD image image f_0 (see Fig. 5d), the BOLD images have a much lower SNR than the MR images used to construct them. MR images with an SNR of {5, 10, 15, 20, 25, 30, 35} dB yield BOLD images with an SNR of {-15.2, -10.2, -5.2, -0.2, 4.8, 9.8, 14.8} dB, respectively.

For these input images, we compared wavelet-based denoising and various degrees of Gaussian smoothing. Each of the wavelet-based schemes started with a 2D FWT of f_1, computed as shown in Fig. 4b. Symmetric orthonormal cubic spline basis functions and a

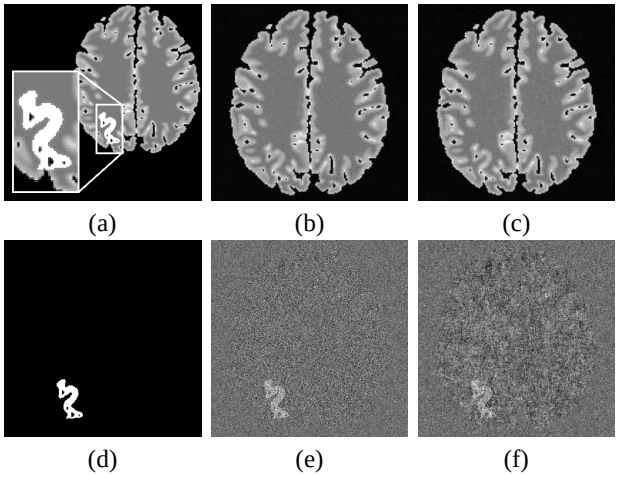


Fig. 5. Image without noise and the active spot magnified and shown in white (a). Noisy MR images with increased intensities inside the active spot, containing white (b) and $1/f$ (c) Rician noise of 15 dB, respectively. The noise-free BOLD image (d) and BOLD images created from noisy MR images with white (e) and $1/f$ (f) noise of 15 dB, respectively.

decomposition level of 4 were used for all tests. After denoising with one of the methods listed in subsection IV-B, a 2D IFWT yielded the denoised image f_2. Denoting the standard deviation of an arbitrary image f by σ_f, the following procedure was carried out for each of the tested methods:

1: the noise $\epsilon_1(\boldsymbol{x})$ present in $f_1(\boldsymbol{x})$ before denoising was computed as $\epsilon_1(\boldsymbol{x}) = f_1(\boldsymbol{x}) - f_0(\boldsymbol{x})$
2: the SNR before denoising, denoted SNR_1, was computed as:
$$\mathrm{SNR}_1 = 10\ \log_{10}\frac{\sigma_{f_0}}{\sigma_{\epsilon_1}}, \tag{4}$$
3: the residual noise $\epsilon_2(\boldsymbol{x})$ after denoising was obtained as $\epsilon_2(\boldsymbol{x}) = f_2(\boldsymbol{x}) - f_0(\boldsymbol{x})$
4: the SNR of the denoised image, denoted SNR_2, was computed as
$$\mathrm{SNR}_2 = 10\ \log_{10}\frac{\sigma_{f_0}}{\sigma_{\epsilon_2}}, \tag{5}$$

Figure 7 shows SNR_2 plotted against the input SNR of the MR images.

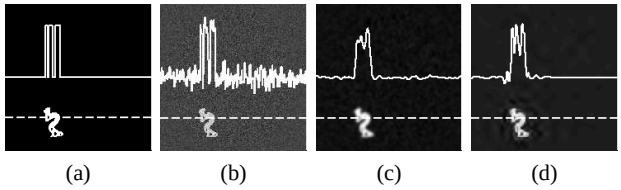


Fig. 6. Images from the 2D denoising test, each with a cross section (solid line) of a line in the image (dotted line) plotted inside: (a) Original (f_0), (b) noisy (f_1), (c) denoised (f_2) with Gaussian smoothing, FWHM = 3 pixels, (d) denoised (f_2) with VisuThresh (s).

Of the Gaussian smoothing methods, the wider kernels perform better for low input SNRs, and smaller kernels perform better for higher input SNRs. The maximum input SNR where Gaussian smoothing still shows SNR improvement decreases as the filter width increases. Figs. 7a+c show that even for Gaussian smoothing with an FWHM of one pixel, the maximum output SNR is about 7 dB.

The wavelet methods perform as well as Gaussian smoothing for low SNRs, and better than Gaussian smoothing for higher SNRs. All wavelet-based methods show maximum output SNRs above 10 dB. For white noise and low SNRs, there is a marked difference in output SNRs. HybridThresh, InvShrink, MultiMAD, and VisuThresh

with both hard (h) and soft (s) thresholding increase the SNR most. WaveJS, MinMaxThresh, and SUREThresh (both h and s) thresholding increase the SNR less. For higher SNRs the differences are smaller, but InvShrink, WaveJS, SUREThresh (h), and VisuThresh (s) now produce visibly lower SNRs. There is another difference between these methods: The images produced by HybridThresh, WaveJS, and VisuThresh (s) smear the active spot out much more than the other methods do. We refer to these schemes as 'smoothing wavelet methods'. MinMaxThresh and SUREThresh (both h and s) produce sharp output images. The other methods produce images of intermediate smoothness. In general, the smoothing wavelet methods perform better for low input SNRs, but the less smoothing wavelet methods are better when the input SNR is higher. In this experiment SUREThresh (h) performs bad with a low input SNR, but with both noise types it performs best for higher input SNR.

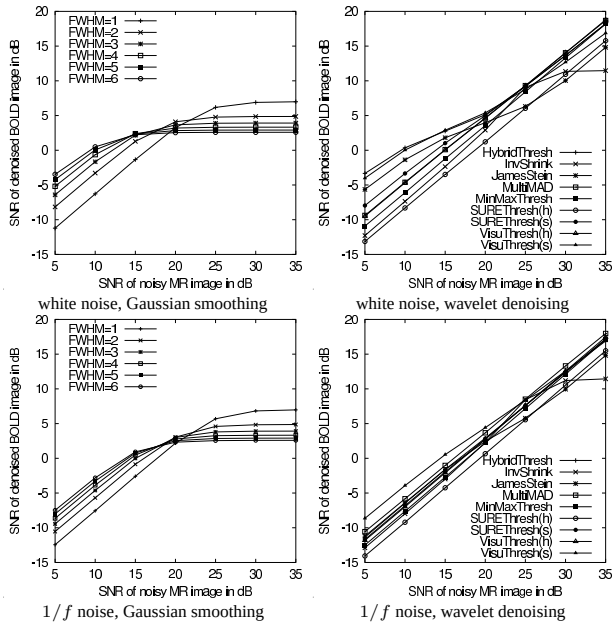


Fig. 7. Performance of various wavelet denoising schemes, and Gaussian smoothing for several values for the FWHM parameter. The SNR of the denoised image is plotted against the SNR of the noisy image.

The differences in performance is smaller for $1/f$ noise than for white noise. This holds for the wavelet methods as well as for Gaussian smoothing. For all wavelet methods except InvShrink, WaveJS, and VisuTresh(s), the output SNR is a linear function of the input SNR: unlike Gaussian smoothing, the wavelet methods improve the SNR of input images that already have a high SNR. This suggests that in terms of SNR improvement, wavelets are an attractive alternative to Gaussian smoothing. With white noise and for low SNRs, the less smoothing wavelet methods, such as MinMaxThresh and SUREThresh (h and s), produce relatively lower output SNRs than the other methods. This indicates that introducing smoothness, thereby discarding image features, is necessary to improve images with very low SNRs. Of the methods mentioned above, MultiMAD and VisuThresh(h) give good results for all tested SNRs.

VI. Denoising a simulated time series

In most neuroimaging applications it is not possible to separate signal and noise, so the SNR is not known. Therefore, a simulation study was performed in which the SNR is known a priori. We constructed an artificial time series of 64 copies of the MR template image of the previous experiment, and superimposed noise on each image according to the procedure described in section V. A block signal $b(t)$ (see Fig. 8) was superimposed on the time signals at the voxel locations inside the active region (see Fig. 5a). The sign of the superimposed signal altered after every 8^{th} time point. The size of the original active region was 762 pixels.

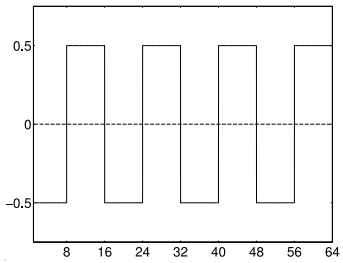


Fig. 8. The time signal in the active spot of the simulated time series.

The time series consisted of 8 blocks of 8 images: 4 blocks were labelled 'rest', and 4 were labelled 'task'. The 'task' blocks were those in which the time signal is positive (see Fig. 8). The amplitude of the time signal was set to 1% of the maximum intensity in the MR template. Starting with the image $f_0(\boldsymbol{x})$ from section V, we use $\mathcal{F}_0$ to denote the original time series with the time signal $b(t)$ superimposed on it, but without the noise:

$$\mathcal{F}_0(\boldsymbol{x},t) = \begin{cases} f_0(\boldsymbol{x}) + b(t) & \text{in the active spot} \\ f_0(\boldsymbol{x}) & \text{in the rest of the image} \end{cases} \tag{6}$$

The noisy time series $\mathcal{F}_1$ was computed as:

$$\mathcal{F}_1(\boldsymbol{x},t) = \mathcal{F}_0(\boldsymbol{x},t) + \epsilon_1(\boldsymbol{x},t), \tag{7}$$

where $\mathcal{F}_1$ and $\epsilon_1(\boldsymbol{x},t)$ denote the noisy time series and the value of the input noise, respectively. A BOLD image was computed from each individual image by subtracting the time series mean. As demonstrated in section III, the noise distribution in such difference images is approximately Gaussian. The BOLD image was denoised using the methods from the previous section, after which the time series mean was added to the denoised image.

Let $\mathcal{F}_2$ denote the denoised time series. After denoising, we tested each voxel location $\boldsymbol{x}$ for the presence of the signal $b(t)$. The residual noise ϵ_2 was computed as:

$$\epsilon_2(\boldsymbol{x},t) = \mathcal{F}_2(\boldsymbol{x},t) - f_0(\boldsymbol{x}) - b(t). \tag{8}$$

We denote the temporal residual noise in a voxel $\boldsymbol{x}$ as a function of t by $\epsilon_2^{\boldsymbol{x}}(t) = \epsilon_2(\boldsymbol{x},t)$. The temporal SNR in a voxel $\boldsymbol{x}$ after denoising was computed as:

$$\text{SNR}_2(\boldsymbol{x}) = 10\ \log_{10} \frac{\sigma_b}{\sigma_{\epsilon_2^{\boldsymbol{x}}}}, \tag{9}$$

where σ_b and $\sigma_{\epsilon_2^{\boldsymbol{x}}}$ are the standard deviations of $b(t)$ and $\epsilon_2^{\boldsymbol{x}}(t)$, respectively.

A. Effect on the temporal SNR

Figure 9 shows $\text{SNR}_2(\boldsymbol{x})$, averaged over all locations $\boldsymbol{x}$ inside the active spot, plotted against the input spatial SNR. The graphs for Gaussian denoising show the same behaviour as in the 2D image experiment, *i.e.* the SNR curve eventually reaches a plateau value. The wavelet-based methods improve the temporal SNR both for low and high input SNR. The same relation observed in the previous experiment between smoothness of the output image and the output SNR is visible here. The smoothing wavelet methods and wide Gaussian smoothing filters produce the highest temporal SNRs for low input SNRs, and the less smoothing wavelet methods and narrow Gaussian filters perform better for high input SNRs.

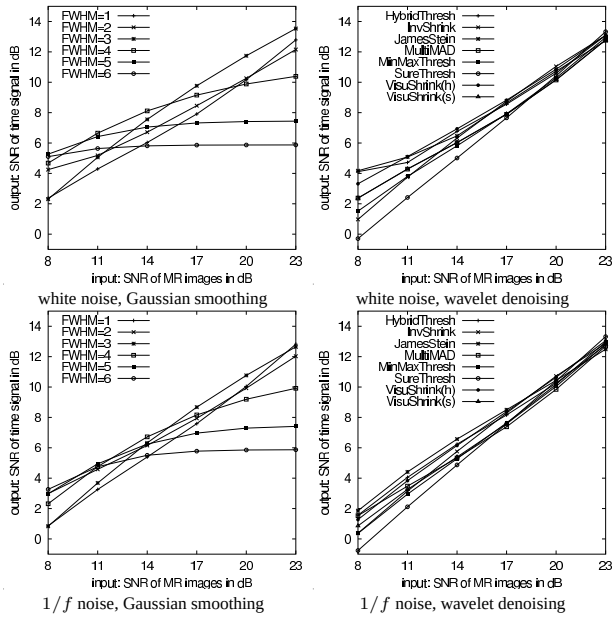


Fig. 9. Performance of the wavelet denoising schemes, as well as Gaussian smoothing for six FWHM values. The average temporal SNR inside the original active spot in the denoised image is plotted against the spatial SNR of the noisy input image.

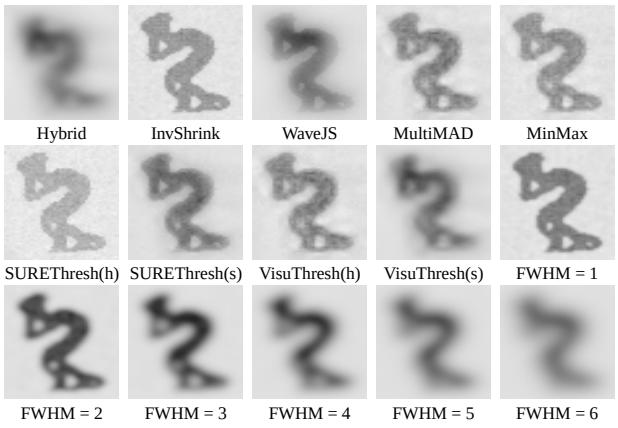


Fig. 10. Temporal SNR maps (inverted) of the area around the active spot. The original images contained white noise with a spatial SNR of 11 dB.

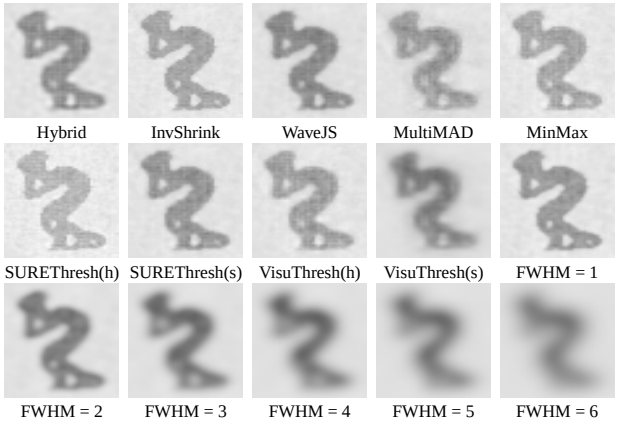


Fig. 11. Temporal SNR maps (inverted) of the area around the active spot. The original images contained $1/f$ noise with a spatial SNR of 11 dB.

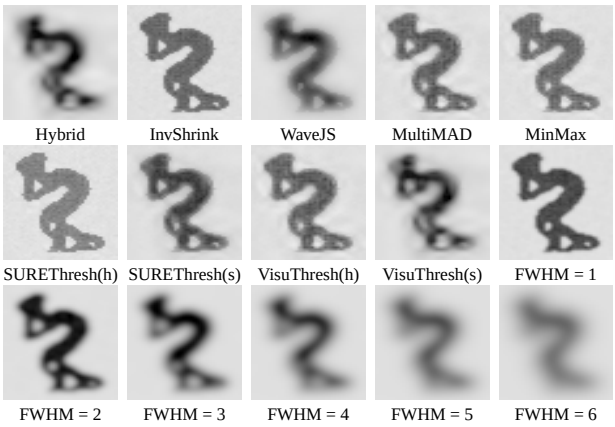


Fig. 12. Temporal SNR maps (inverted) of the area around the active spot. The original images contained white noise with a spatial SNR of 14 dB.

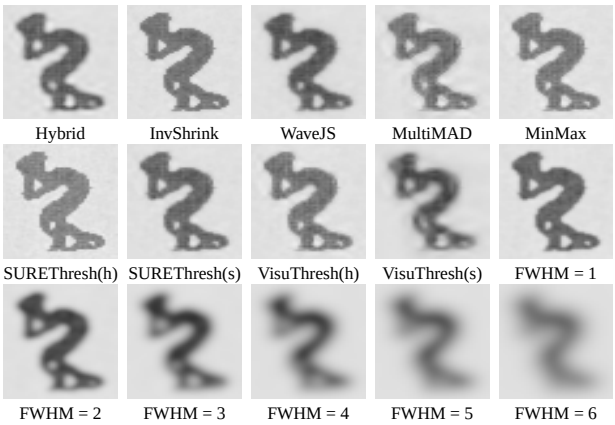


Fig. 13. Temporal SNR maps (inverted) of the area around the active spot. The original images contained $1/f$ noise with a spatial SNR of 14 dB.

B. Effect on the shape of the detected spots

Apart from comparing the average temporal SNRs in the active spot, we also look at spatial maps of temporal SNR values of the denoised time series. Ideally, these maps should have high values inside the active spot and low values outside it. Figures 10 - 13 show the temporal SNRs in the area containing the active spot for white noise and $1/f$ noise of 11 dB and 14 dB, respectively. Note that the images were inverted (reverse video mode) for enhanced display purposes.

Gaussian smoothing with small smoothing kernels and the smoothing wavelet methods show bright spots, even for a low input spatial SNR like 11 dB. Wider kernels, with FWHM > 3 pixels, produce maps with a very smooth spot, which is less bright. The smoothing wavelet methods show bright spots, while those produced by the less smoothing wavelet methods are darker, cf. Fig. 10. The smoother the output image, the more the shape and the SNR value distribution of the visible spot differs from the original active spot. Temporal SNR maps of methods that produce smooth images (both Gaussian and wavelet-based) show spots with a somewhat elliptic shape and a peaked (non-uniform) intensity distribution. The less smoothing wavelet methods retain the shape of the original spot and its uniform intensity distribution. For noise of 14 dB, InvShrink, MinMaxThresh, SUREThresh and Gaussian smoothing with FWHM = 1 return almost exactly the original spot, with a very uniform distribution of temporal SNR values. Other less smoothing wavelet methods, such as MultiMAD and VisuThresh (h), and Gaussian smoothing with

FWHM = 2, retain the shape of the spot quite well, with most of the changes in the temporal SNR values near the contour of the spot.

C. Segmentation via SNR thresholding

Segmentation of MR images based on thresholding is a commonly used technique, and it has also been used on statistical parametric maps, see [38], [39] for examples and references. We assumed the temporal SNR maps to have bimodal histograms: one peak of low values for the background and another peak of high values for the active spot. This assumption was used to segment the maps into a 'non-active' area and an 'active' area. We used the following steps to determine a threshold.

1: Smooth the histogram with a moving average filter
2: Take the logarithm of each entry in the smoothed histogram
3: Model the log-histogram as the sum of two Gaussian peaks
4: Use the 99.9% level of the cumulative histogram of the left peak as a threshold

Filtering the histogram was implemented by applying a three-tap averaging filter ten times. Taking the logarithm was applied to amplify the second peak: the number of background pixels is generally much larger than the size of the active spot, and large values decrease more by taking their logarithm than small values. The histogram was approximated by the sum of two Gaussian peaks with the Levenberg-Marquardt curve fitting algorithm. The threshold was based on the distribution of the noise.

This is a very simple method, based on a simple assumption of the bimodality of the histogram. We demonstrate its performance in a number of cases, with SNRs ranging from low to high. If the temporal SNR is very low, the histogram of the temporal SNR map is equal to the histogram of an image just containing noise. As it is not possible to distinguish two peaks in this case, the threshold is determined incorrectly (see Fig. 14c).

In the worst case we tested (Fig. 14a), the temporal SNR map itself also has a very low SNR of -10 dB, and the histogram of the SNR map has the shape of the noise distribution, so that separation of signal and noise is not possible. The histogram in Fig. 14b yields a sensible threshold, though the noise prevents a better detection (see Fig. 14e). Figure 14c-d show that temporal SNR maps with an SNR of at least 0 dB can be segmented well with this technique.

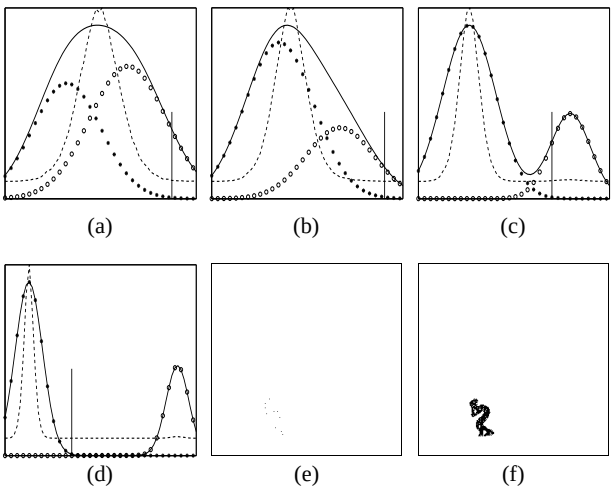


Fig. 14. Six situations in which the temporal SNR threshold was determined. Dashed line: histogram entries; solid line: log-histogram; vertical line: threshold; $*$: background intensities; $\circ$: activation intensities. BOLD images: SNR of -10 dB (a), -5 dB (b), 0 dB (c), 5 dB (d). (e) Activation (inverted) detected from (b), (f) activation (inverted) detected from (c).

In the experiment, we looked at two measures: the number of false positive classifications, *i.e.* points outside the original active spot labelled 'active', and false negative classifications *i.e.* points inside the original active spot labelled 'non-active'.

TABLE III
NUMBER OF FALSE POSITIVE CLASSIFICATIONS FOR WHITE (LEFT) AND $1/f$ (RIGHT) NOISE. THE SNR MAPS WERE ASSUMED TO HAVE BIMODAL HISTOGRAMS.

	spatial SNR					
method	8	11	14	17	20	23
Hybrid	767	651	356	168	74	0
InvShrink	0	1	3	0	0	0
JAMES	599	589	78	2	154	75
MultiMAD	54	123	96	9	0	0
MinMax	2	42	42	3	0	0
SURE(h)	47	0	0	0	0	0
SURE(s)	236	292	165	1	0	0
Visu(h)	52	120	95	10	0	0
Visu(s)	680	552	313	165	35	0
FWHM = 1	2	41	10	0	0	0
FWHM = 2	198	62	15	35	24	23
FWHM = 3	348	35	65	98	85	86
FWHM = 4	116	68	148	155	182	166
FWHM = 5	84	104	188	221	200	229
FWHM = 6	99	132	216	245	249	265

	spatial SNR					
method	8	11	14	17	20	23
Hybrid	2	114	164	150	59	0
InvShrink	17965	1	1	0	0	0
JAMES	2	107	169	21	15	85
MultiMAD	9	129	102	7	0	0
MinMax	23649	10	27	4	0	0
SURE(h)	306	0	1	0	0	0
SURE(s)	0	51	102	43	0	0
Visu(h)	0	42	76	13	0	0
Visu(s)	225	386	293	151	29	0
FWHM = 1	0	8	86	0	0	0
FWHM = 2	28	209	18	17	35	26
FWHM = 3	130	337	39	58	87	82
FWHM = 4	262	411	59	105	172	150
FWHM = 5	359	339	85	141	225	203
FWHM = 6	460	240	104	165	249	243

TABLE IV
NUMBER OF FALSE NEGATIVE CLASSIFICATIONS FOR WHITE (LEFT) AND $1/f$ (RIGHT) NOISE. THE SNR MAPS WERE ASSUMED TO HAVE BIMODAL HISTOGRAMS.

	spatial SNR					
method	8	11	14	17	20	23
Hybrid	0	0	0	25	0	0
InvShrink	754	0	0	0	0	0
JAMES	0	0	158	202	0	0
MultiMAD	135	1	0	0	0	0
MinMax	568	1	0	0	0	0
SURE(h)	684	15	0	0	0	0
SURE(s)	2	0	0	0	0	0
Visu(h)	136	1	0	0	0	0
Visu(s)	0	0	0	0	0	0
FWHM = 1	45	0	1	374	3	158
FWHM = 2	0	37	78	18	55	54
FWHM = 3	0	144	43	15	33	27
FWHM = 4	103	124	22	28	14	22
FWHM = 5	167	116	31	21	31	22
FWHM = 6	203	123	39	30	31	28

	spatial SNR					
method	8	11	14	17	20	23
Hybrid	560	1	0	0	0	0
InvShrink	52	0	0	0	0	0
JAMES	313	0	0	38	3	0
MultiMAD	510	1	0	0	0	0
MinMax	63	5	0	0	0	0
SURE(h)	730	88	0	0	0	0
SURE(s)	723	1	0	0	0	0
Visu(h)	757	3	0	0	0	0
Visu(s)	10	0	0	0	0	0
FWHM = 1	759	1	0	14	101	158
FWHM = 2	80	0	78	66	18	45
FWHM = 3	14	0	125	61	34	46
FWHM = 4	6	4	160	53	25	36
FWHM = 5	4	43	162	61	19	29
FWHM = 6	3	89	185	76	27	33

Tables III and IV show the false positive and false negative classifications, respectively. Images with spatial SNRs of 8 dB do not yield SNR maps that can be analysed in this way, because the SNRs of the BOLD images, as well as the SNRs of the temporal SNR maps, are too low (see the list of BOLD SNRs in section V and Fig. 14a-b). They either yield many false positives or many false negatives. In general, the denoising methods that introduce much smoothness yield more false positive classifications for higher SNRs, while the less smoothing methods yield many false negatives for the lowest SNR. Of the wavelet-based methods, InvShrink and SUREThresh (h) perform well for both noise types, and MultiMAD, MinMaxThresh, and VisuThresh (h) yield good results for moderate and high SNRs. The relatively high numbers of type II errors for Gaussian smoothing with large FWHM relate to the blurring effect visible in Figs. 10-13. The intensity distribution in the spot changes from uniform to peaked, which influences detections close to the boundary of the spot. InvShrink, SUREThresh (h) and Gaussian smoothing with FWHM = 1 yield good results. MultiMAD and MinMaxThresh also perform well, the other methods yield more errors.

VII. STATISTICAL TESTS ON THE SIMULATED TIME SERIES

We also performed a standard statistical analysis on the denoised time series with the SPM method [9]. The design matrix, $\boldsymbol{X}$ in

(1), had two columns: a block signal like in Fig. 8 and a column containing a constant signal to capture the time series mean. In terms of (1), the matrix $\boldsymbol{\beta}$ contained two columns, each representing an image. The first image, $\boldsymbol{\beta}_1(\boldsymbol{x})$, contained the covariance of the block signal with the time signal at each location $\boldsymbol{x}$. Image $\boldsymbol{\beta}_2(\boldsymbol{x})$ contained the time series mean of each voxel.

Although the noise in the MR images is Rician distributed, the noise in the BOLD image has, to very good approximation, a Gaussian distribution, as explained in section III. If the temporal noise is Gaussian distributed, the values in $\boldsymbol{\beta}(\boldsymbol{x})$ are t-distributed. Using the sample variance s^2 (see section II) we can compute a statistical parametric map of p-values by means of the t-test. We used FDR correction as described in section II, with an FDR parameter q of 0.05, to threshold $\boldsymbol{\beta}_1$. This yielded the statistically significant activations for all denoising methods. To obtain more robust results, this experiment was repeated 20 times and the outcomes of the individual experiments were averaged.

Two important issues are critical to the validity of this method. First, a t-test is only appropriate if, after denoising, the temporal noise is still Gaussian. Second, to use the $\eta = 1$ setting described in section II, the data is required to be PRDS. The validity of these two assumptions is discussed in the next subsections.

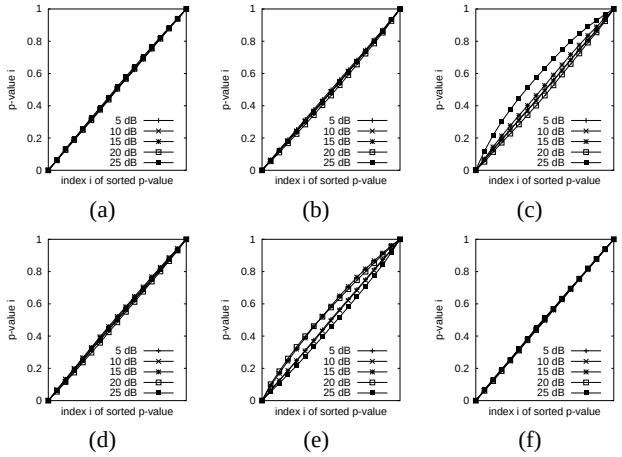


Fig. 15. Sorted p-values in the statistical map of denoised time series with white noise, without activation. (a) No denoising, (b) FWHM = 1, (c) FWHM = 3, (d) InvShrink, (e) MultiMAD, (f) MinMaxThresh. The symbols {o, x, +, *, □} represent the time series of images with an input SNR of {5, 10, 15, 20, 25} dB.

A. Impact of spatial filtering on the distribution of temporal noise

The p-values resulting from a set of statistical tests are uniformly distributed on [0,1] if the 'omnibus' null hypothesis is true [40]. The sequence of ordered p-values from that set of tests should lie on a straight line. We tested this by constructing time series similar to those previously described, but without activation: the null hypothesis was true for all voxels. We applied all denoising methods to these time series and sorted the p-values acquired in the statistical analysis. Figure 15 shows representative results of both wavelet methods and Gaussian smoothing. Some methods produce Gaussian temporal noise, others introduce a deviation from Gaussianity. The top row shows that Gaussian smoothing with FWHM = 1 yields uniformly distributed p-values under the null hypothesis, while for FWHM = 3 or higher, non-uniformly distributed p-values are obtained. The plots in the bottom row show results for three wavelet methods. For InvShrink and MinMaxThresh the distribution of p-values is uniform, but for MultiMAD it is non-uniform.

The fact that even for Gaussian smoothing the distribution of the noise may become non-Gaussian may seem puzzling, but can be explained by the fact that, for Rician noise, a higher intensity in the image leads to a larger noise amplitude. This gives a kind of spatial structure to the noise, which is observable in the (BOLD) difference images. Denoising methods that produce smoother images change this structure, thus introducing errors. Although the deviation from normality varies between methods, we chose to keep all methods in the statistical analysis, since the t-test is quite robust to deviations from normality.

B. Positive regression dependence of the p-values

Benjamini et al. [41] show that the setting $\eta = 1$ can be used in the FDR-controlling procedure if the data are PRDS, and that multivariate positively correlated normally distributed data are PRDS. Genovese et al. [11] argue that most fMRI data sets satisfy this condition.

To test the spatial correlation of the noise after applying a denoising method, we observed the time series (without activation) of the residual noise in the GLM, *i.e.* the $\boldsymbol{e}$ images in (1). We used the *SPMd* toolbox [42] to compute a normalised residual time series $\mathcal{E}$. The noise in this time series was $N(0,1)$-distributed. We tested for a positive correlation as follows. Let $\mathcal{E}_2^{\boldsymbol{x}}(t)$ denote the normalised residual time signal (the matrix $\boldsymbol{e}$ in (1)) at location $\boldsymbol{x}$ as a function of t. We assumed the autocorrelation function to be localised, and for each location $\boldsymbol{x}$ in the image, the amount of spatial correlation $\mathrm{a}(\boldsymbol{x})$ was estimated by averaging the covariances of the voxel's time signal with those of a number of neighbouring voxels:

$$\mathrm{a}(\boldsymbol{x}) = \frac{1}{N_v} \sum_{\boldsymbol{x}_i} \mathrm{cov}\left(\mathcal{E}_2^{\boldsymbol{x}}, \mathcal{E}_2^{\boldsymbol{x}_i}\right), \tag{10}$$

with $\boldsymbol{x}_i$ from a small neighbourhood of $\boldsymbol{x}$ of size N_v (in our case, N_v was 11×11 voxels). Figure 16 shows this function for a number of settings. The top rows show the amount of correlation found without applying denoising. Wavelet methods introduce positive spatial correlations for lower input SNRs, and hardly any negative correlations for higher input SNRs. Gaussian smoothing introduces strong positive correlations for all SNRs.

Another way to characterise the autocorrelation function is to look at statistics of the distribution of $\mathrm{a}(\boldsymbol{x})$. Figure 16 and Table V show that every denoising method changes the spatial correlations in the residual time series. All methods, except MultiMAD, introduce significantly more positive correlations than negative ones. Wavelet methods change the spatial correlation much less than Gaussian smoothing. We assume that without denoising, the residuals do not have significant negative correlations. In our test data (spatially white or $1/f$ Gaussian noise) we know that this is the case. Because the only significant correlations introduced by denoising are positive, the residuals are either uncorrelated, or positively correlated in space. These results indicate that the $\eta = 1$ setting is allowed in the statistical tests.

C. Results

In this experiment, we investigate the effect of denoising on the outcomes of the usual statistical analysis. In particular, we looked at two measures: the number of false positives and the number of false negatives. It is important to realize that denoising has two effects: first, the desired effect of noise reduction, and second, an unwanted but unavoidable change of the shape of the active spot. In order to take the latter effect into account, false positives/negatives were defined as points outside/inside the *original* active spot (see Fig. 5d) after denoising, which are marked 'active'/'inactive' in the t-test with FDR control ($q = 0.05$). These numbers are shown in Tables VI and VII.

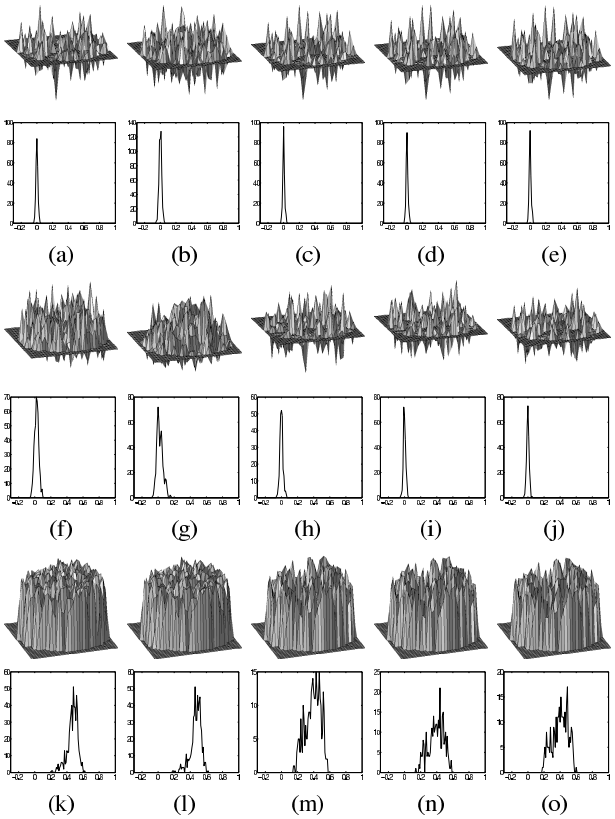


Fig. 16. Surface plots of the spatial autocorrelation function (top) and histograms (bottom) of individual correlations computed in (10) of residual time series. The original time series contained white noise. (a-e) SNR input images = {5, 10, 15, 20, 25} dB, without denoising. (f-j) Idem, denoised with MinMaxThresh. (k-o) Idem, after Gaussian smoothing with FWHM = 3 voxels.

TABLE V

MINIMAL, MAXIMAL, MEAN, AND MEDIAN VALUES AND THE STANDARD DEVIATIONS OF THE TEMPORAL STATISTICAL CORRELATION OF VOXELS WITH THEIR NEIGHBOURS. THE INPUT TIME SERIES CONTAINED NO ACTIVATION, AND THE SNR WAS 15 DB. LEFT: WHITE NOISE, RIGHT: $1/f$ NOISE.

	statistic				
method	min	mean	max	σ	med
No Denoising	-0.03	0.00	0.04	0.01	-0.00
Hybrid	-0.05	0.00	0.07	0.02	0.00
InvShrink	-0.07	0.06	0.28	0.06	0.04
WaveJS	-0.04	-0.00	0.04	0.01	-0.00
MultiMAD	-0.18	0.08	0.83	0.14	0.03
MinMax	-0.06	-0.00	0.06	0.02	-0.00
SURE(h)	-0.04	-0.00	0.04	0.01	-0.00
SURE(s)	-0.05	0.00	0.07	0.02	0.00
Visu(h)	-0.08	-0.00	0.10	0.02	0.00
Visu(s)	-0.08	0.03	0.14	0.04	0.02
FWHM = 1	-0.03	0.03	0.12	0.03	0.03
FWHM = 2	0.01	0.20	0.35	0.06	0.20
FWHM = 3	0.15	0.39	0.57	0.09	0.40
FWHM = 4	0.25	0.55	0.74	0.12	0.58
FWHM = 5	0.32	0.66	0.85	0.13	0.70
FWHM = 6	0.34	0.74	0.91	0.13	0.80

	statistic				
method	min	mean	max	σ	med
No Denoising	-0.04	0.04	0.12	0.03	0.04
Hybrid	-0.02	0.07	0.19	0.04	0.06
InvShrink	-0.01	0.18	0.51	0.10	0.16
WaveJS	-0.04	0.05	0.13	0.03	0.04
MultiMAD	-0.16	0.13	0.88	0.16	0.09
MinMax	-0.03	0.05	0.16	0.04	0.04
SURE(h)	-0.04	0.04	0.13	0.03	0.04
SURE(s)	-0.02	0.07	0.19	0.04	0.06
Visu(h)	-0.03	0.05	0.19	0.04	0.05
Visu(s)	-0.03	0.12	0.34	0.07	0.11
FWHM = 1	0.02	0.16	0.32	0.05	0.15
FWHM = 2	0.13	0.37	0.57	0.09	0.37
FWHM = 3	0.26	0.53	0.73	0.12	0.55
FWHM = 4	0.29	0.65	0.84	0.14	0.67
FWHM = 5	0.35	0.73	0.93	0.15	0.78
FWHM = 6	0.36	0.79	0.96	0.14	0.86

TABLE VI

NUMBER OF TYPE I ERRORS IN THE SPM ANALYSIS WITH FDR CONTROL ($q = 0.05$) FOR WHITE (LEFT) AND $1/f$ (RIGHT) NOISE.

	spatial SNR					
method	8	11	14	17	20	23
Hybrid	105	161	12	14	10	7
InvShrink	35	37	3	1	4	24
JAMES	114	199	13	14	14	10
MultiMAD	206	234	27	19	20	15
MinMax	53	71	2	1	1	0
SURE(h)	28	29	3	0	0	0
SURE(s)	77	119	6	5	5	9
Visu(h)	80	109	4	4	5	3
Visu(s)	477	519	55	68	62	46
FWHM = 1	49	99	10	7	7	8
FWHM = 2	162	250	13	13	14	16
FWHM = 3	299	400	18	23	29	36
FWHM = 4	431	536	30	47	63	74
FWHM = 5	563	671	60	100	127	145
FWHM = 6	689	798	122	178	210	230

	spatial SNR					
method	8	11	14	17	20	23
Hybrid	1400	1047	164	100	52	25
InvShrink	35	47	3	1	6	28
JAMES	1037	853	69	29	21	13
MultiMAD	278	238	27	23	19	15
MinMax	102	116	5	5	8	3
SURE(h)	30	17	2	0	0	0
SURE(s)	553	494	48	41	28	25
Visu(h)	268	226	22	20	17	12
Visu(s)	1178	900	177	175	135	79
FWHM = 1	75	138	8	7	8	8
FWHM = 2	277	327	13	14	16	19
FWHM = 3	476	510	24	30	38	41
FWHM = 4	668	673	53	68	77	85
FWHM = 5	864	828	116	139	152	165
FWHM = 6	1060	975	202	225	239	248

False discovery rates can be obtained from this table by computing the number of false positives, divided by the number of detections; the latter number equals the size of active spot(=762 pixels), plus the number of false positives, minus the number of false negatives. A consequence of taking the *original* active spot as a reference is that the observed false discovery rates after denoising may exceed the 5% threshold imposed by the FDR controlling procedure.

These tables show that the smoothing methods produce more false positives, whereas InvShrink, MinMaxThresh, SUREThresh(h), VisuThresh(h) and Gaussian smoothing with FWHM = 1 produce very few false positives. The other wavelet methods and Gaussian smoothing with an FWHM of two voxels also perform well. For larger Gaussian filters, the number of type I errors increases with the filter width. The number of type I errors is larger for $1/f$ noise than for white noise. The less smoothing wavelet-based methods and Gaussian smoothing with an FWHM of one voxel produce more type II errors than the other methods. With $1/f$ noise, this effect is worse than with white noise. In general, the wavelet-based methods and Gaussian smoothing with an FWHM of one voxel introduce more type II errors, the other Gaussian filters introduces more type I errors. Figures 17 - 20 show statistical parametric maps built from the denoised time series with white noise and $1/f$ noise of 11dB and 14dB, respectively. Generally, the less smoothing methods produce spots that are closest to the original. The spots detected after InvShrink, MinMaxThresh, and SUREThresh (h) denoising and Gaussian smoothing with an FWHM of one voxel are closest to the original spot (see Figs. 17-20). Because the boundary voxels are not detected, the resulting active spot is smaller than the original (type II errors). HybridThresh, WaveJS, and VisuThresh(s) and all the Gaussian smoothing methods produce larger spots (type I errors).

VIII. STATISTICAL TESTS ON A REAL FMRI DATA SET

To test the denoising methods on real data, we used an example fMRI data set provided by the Dartmouth Brain Imaging Center [43]. This is a recording of an experiment in which a subject was scanned for 4 minutes with a TR of 2000 ms. The subject's condition switched every 30 s (15 scans) between 'rest' and 'task', starting with 'rest'. During the 'task' periods, the subject had to perform an object manipulation task. The data set consists of 120 volumes with a resolution of $64\times64\times27$ voxels. Each voxel has a volume of $3.75\times3.75\times5.50$ mm^3.

The 3D volumes, each consisting of 27 axial planes of 64×64 voxels, were transformed plane-by-plane to the wavelet domain. The decomposition level was set to 4. Denoising was done by both the wavelet-based methods and Gaussian smoothing. For the latter we used smoothing kernels of $5\times5\times5.5$ mm^3, $10\times10\times5.5$, and mm^3, $15\times15\times5.5$ mm^3. We compared the activation images, using the activation map of the original data without preprocessing (see Fig. 21) as a reference: the shape of the active region detected after denoising should not differ too much from that detected from the original time series. The data underwent the same statistical analysis as the simulated time series. Figure 22 shows the voxels in a selected plane whose t-statistic was above the FDR threshold, for all denoising methods, overlayed on the first image of the original time series.

TABLE VII
NUMBER OF TYPE II ERRORS IN THE SPM ANALYSIS WITH FDR CONTROL ($q = 0.05$) FOR WHITE (LEFT) AND $1/f$ (RIGHT) NOISE.

	spatial SNR					
method	8	11	14	17	20	23
Hybrid	0	0	67	70	85	112
InvShrink	3	0	168	139	122	113
JAMES	0	0	72	71	77	84
MultiMAD	5	0	163	151	136	123
MinMax	9	0	179	157	137	123
SURE(h)	72	0	198	164	140	125
SURE(s)	0	0	121	114	113	112
Visu(h)	5	0	163	151	136	123
Visu(s)	0	0	78	82	92	97
FWHM = 1	0	0	123	93	79	77
FWHM = 2	0	0	73	64	56	53
FWHM = 3	0	0	62	55	50	45
FWHM = 4	0	0	57	51	47	44
FWHM = 5	0	0	52	47	44	43
FWHM = 6	0	0	47	43	40	40

	spatial SNR					
method	8	11	14	17	20	23
Hybrid	9	0	131	101	107	109
InvShrink	17	0	175	144	124	114
JAMES	2	0	119	91	83	91
MultiMAD	15	1	169	153	137	123
MinMax	33	0	191	161	139	123
SURE(h)	76	0	191	166	140	124
SURE(s)	8	0	155	129	119	115
Visu(h)	24	1	183	157	138	122
Visu(s)	1	0	102	95	97	101
FWHM = 1	10	0	147	104	84	78
FWHM = 2	2	0	96	74	63	56
FWHM = 3	1	0	80	66	57	51
FWHM = 4	1	0	72	61	54	48
FWHM = 5	1	0	66	57	51	45
FWHM = 6	0	0	62	53	46	42

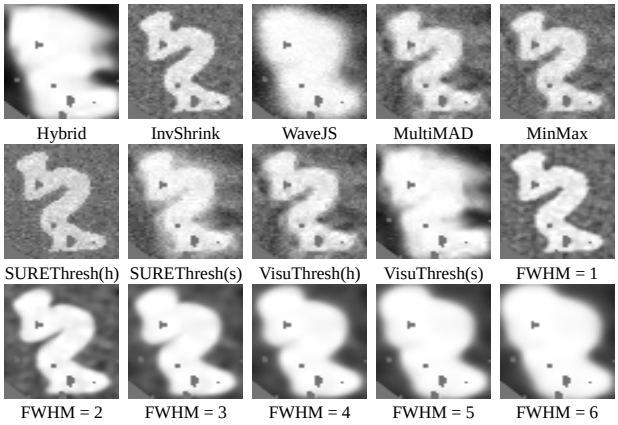


Fig. 17. Statistical parametric maps of the area surrounding the active spot. The original images contained white noise with a spatial SNR of 11dB.

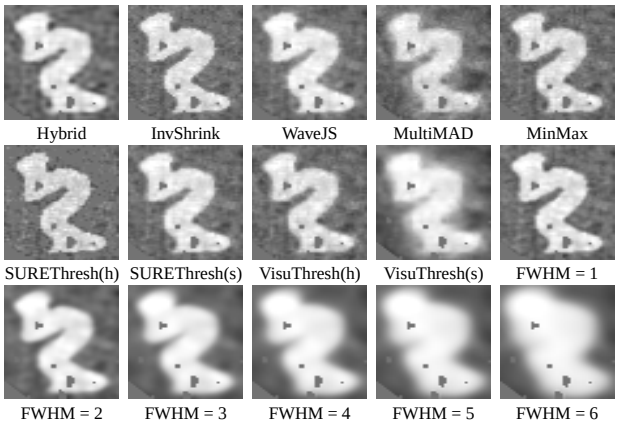


Fig. 18. Statistical parametric maps of the area surrounding the active spot. The original images contained $1/f$ noise with a spatial SNR of 11dB.

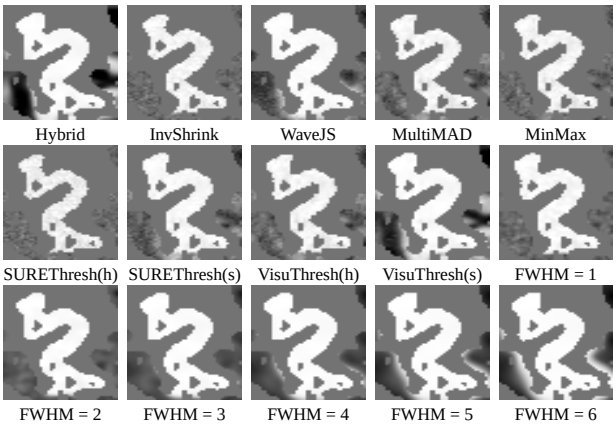


Fig. 19. Statistical parametric maps of the area surrounding the active spot. The original images contained white noise with a spatial SNR of 14dB.

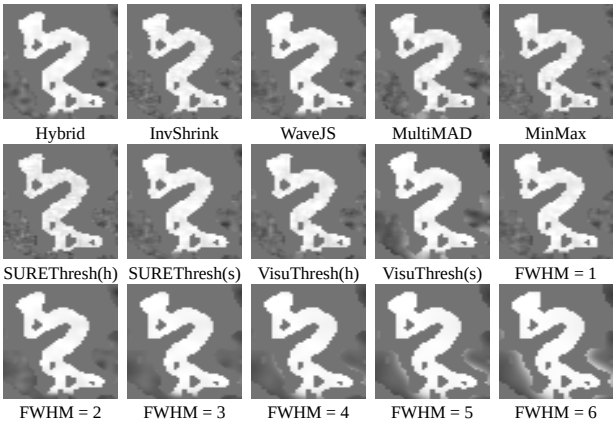


Fig. 20. Statistical parametric maps of the area surrounding the active spot. The original images contained $1/f$ noise with a spatial SNR of 14dB.

As in the case of the simulated time series, the active spot takes an elliptic shape for Gaussian smoothing with large FWHMs. The spots detected from the data sets denoised with MinMaxThresh and SUREThresh look very similar, and those found with WaveJS, InvShrink, and VisuThresh(h) are also similar. HybridThresh, MultiMAD and VisuThresh(s) yield rather different maps. After smoothing with a Gaussian kernel with an FWHM of $5\times5\times5.5$ mm^3, the detected spot resembles the ones found after InvShrink and VisuThresh with hard thresholding. The other smoothing kernels yield heavily deformed maps and show active spots very different from the one in the reference image.

IX. CONCLUSIONS

We have compared wavelet denoising and Gaussian smoothing in the context of functional MRI in three settings: (*i*) 2D images and (*ii*) time series of 2D images, both contaminated by white or $1/f$ noise with a known SNR, and (*iii*) a real fMRI data set with an unknown noise type and SNR. The noise in BOLD images was described as the difference of two MR images containing Rician noise, and shown to have a Gaussian-like distribution. The denoising methods were compared with respect to SNR improvement, effect on the shape of activated regions, and the effect on the quality of statistical parametric maps. In contrast to most previous wavelet-based denoising schemes, we have chosen to do the subsequent statistical analysis in the spatial domain. This allowed us to directly compare the results of Gaussian and wavelet-based methods.

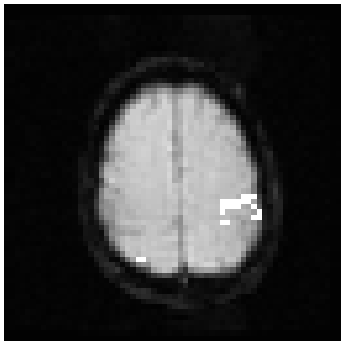


Fig. 21. Activation detected by the SPM method in the original fMRI time series, after FDR thresholding with the FDR parameter set to $q = 0.05$.

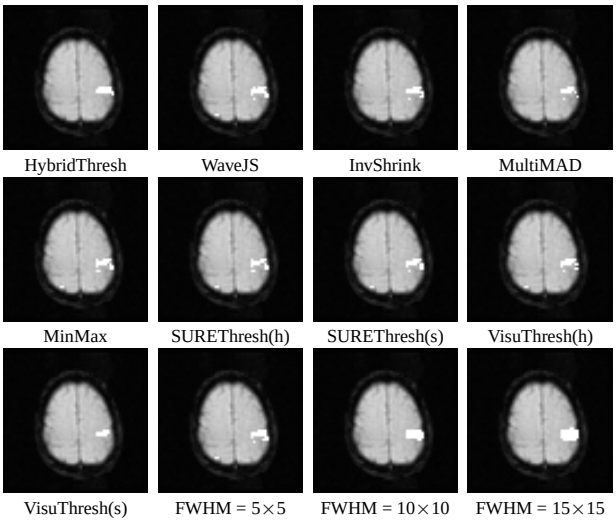


Fig. 22. Activation detected with the SPM method in the denoised fMRI time series, after FDR thresholding with the FDR parameter set to $q = 0.05$.

A discriminating characteristic of all tested denoising methods is the amount of smoothing they introduce. This characteristic plays a significant role in the applicability of the methods. When the input SNR is very low, denoising schemes that produce smoother images are preferred, and the gain in SNR is highest. However, when the images have moderate to high SNRs, these denoising schemes change the shapes of objects in the images. The more smoothing is introduced, the heavier the deformation, and in this case less smoothing wavelet-based denoising methods are preferred. Gaussian smoothing may be the best choice for SNRs which are too low even for smoothing wavelet-based methods, but the resulting SNR may still not be high enough for a reliable analysis.

The experiment with artificial time series showed that all denoising schemes have an effect on the shape of the activation spot. Gaussian smoothing and the more smoothing wavelet-based methods introduce severe deformations and blur the edges of the active spot. We used spatial maps of the temporal SNR as a diagnostic to compare the denoising methods. Segmentations based on the temporal SNR maps showed that heavy smoothing obscures the border regions of the active spot, introducing false negatives, while for low SNRs the less smoothing methods lead to false positives. In the intermediate SNR range, wavelet methods generally show smaller numbers of errors than Gaussian smoothing. The same was observed in the statistical analysis. Via plots of the distribution of p-values under the null hypothesis, we have shown that after the less smoothing wavelet-based denoising methods and after modest Gaussian smoothing, fMRI data do not violate the assumption of normally distributed temporal noise. All tested denoising method preserved the PRDS property of fMRI data, which allowed us to use the favourable $\eta = 1$ setting for the FDR controlling procedure.

For the real fMRI data set, only the smallest Gaussian smoothing kernel yielded reliable results. The wide smoothing kernels yield much larger detected areas (meaning more type I errors), in contrast to those obtained via less smoothing wavelet denoising methods.

Summarising all of these results, wavelet denoising methods that introduce relatively little smoothness are generally preferable over Gaussian smoothing for denoising fMRI time series. In particular, InvShrink, MinMaxThresh or SUREThresh (h) are safe choices. For low SNRs, the methods MultiMAD and SUREThresh (s) are best applied.

We expect to find similar results for PET data, although there are differences with fMRI regarding noise models and the SNR. We did not use *temporal* denoising of the time series in this study, but wavelet denoising may prove a good alternative to smoothing in time as well. This will be the subject of future work.

X. Acknowledgements

This research is part of the project 'Wavelets and their applications', funded by the Dutch National Science Foundation (NWO), project no. 613.006.570. We thank the Dartmouth Brain Imaging Center for providing the example fMRI data set. The authors would like to thank Thomas E. Nichols for his helpful comments on FDR and PRDS.

References

[1] R. Turner, A. Howseman, G. E. Rees, O. Josephs, and K. J. Friston, "Functional magnetic resonance imaging of the human brain: data acquisition and analysis", *Experimental Brain Research*, vol. 123, pp. 5–12, 1998.

[2] H. Gudbjartsson and S. Patz, "The Rician distribution of noisy MRI data", *Magnetic Resonance in Medicine*, vol. 34, pp. 910–914, 1995.

[3] M. Feilner, T. Blu, and M. Unser, "Statistical analysis of fMRI data using orthogonal filterbanks", in *Proc. SPIE: Wavelet Applications in Signal and Image Processing*, vol. 3813, pp. 551–560, 1999.

[4] U. E. Ruttimann, M. Unser, R. R. Rawlings, D. Rio, N. F. Ramsey, V. S. Mattay, D. W. Hommer, J. A. Frank, and D. R. Weinberger, "Statistical analysis of functional MRI data in the wavelet domain", *IEEE Transactions on Medical Imaging*, vol. 17, pp. 142–154, 1998.

[5] J. Raz and B. Turetsky, "Wavelet ANOVA and fMRI", in *Proc. SPIE: Wavelet Applications in Signal and Image Processing*, vol. 3813, pp. 561–570, 1999.

[6] M. Hilton, T. Ogden, D. Hattery, G. Eden, and B. Jawerth, "Wavelet processing of functional MRI data", in A. Aldroubi and M. Unser, editors, *Wavelets in Biology and Medicine*. CRC Press, 1996.

[7] J. B. Buckheit and D. L. Donoho, "Wavelab and reproducible research", Technical Report 474, Dept. of statistics, Stanford University, 1995, http://www-stat.stanford.edu/~wavelab.

[8] F. E. Turkheimer, R. B. Banati, D. Visvikis, J. A. D. Aston, R. N. Gunn, and V. J. Cunningham, "Modeling dynamic PET-SPECT studies in the wavelet domain", *Journal of Cerebral Blood Flow and Metabolism*, vol. 20, pp. 879–893, 2000.

[9] K. J. Friston, A. P. Holmes, K. J. Worsley, J. P. Poline, C. D. Frith, and R. S. J. Frackowiak, "Statistical parametric maps in functional imaging: A general linear approach", *Human Brain Mapping*, vol. 2, pp. 189–210, 1995, http://www.fil.ion.ucl.ac.uk/spm.

[10] A. P. Holmes, *Statistical Issues in functional Brain Mapping*, PhD thesis, University of Glasgow, 1995.

[11] C. R. Genovese, N. A. Lazar, and T. E. Nichols, "Thresholding of statistical maps in functional neuroimaging using the false discovery rate", *NeuroImage*, vol. 15, pp. 772–786, 2002, http://www.sph.umich.edu/~nichols/FDR.

[12] S. Gold, B. Christian, S. Arndt, G. Zeien, T. Cizadlo, D. L. Johnson, M. Flaum, and N. C. Andreasen, "Functional MRI statistical software packages: A comparative analysis", *Human Brain Mapping*, vol. 6, pp. 73–84, 1998.

[13] K. J. Friston, K. J. Worsley, R. S. J. Frackowiak, J. C. Mazziotta, and A. C. Evans, "Assessing the significance of focal activations using their spatial extent", *Human Brain Mapping*, vol. 1, pp. 214–220, 1994.
[14] K. J. Worsley, S. Marrett, P. Neelin, A. C. Vandal, K. J. Friston, and A. C. Evans, "A unified statistical approach for determining significant signals in images of cerebral activation", *Human Brain Mapping*, vol. 4, pp. 58–73, 1996.
[15] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing", *Journal of the Royal Statistical Society*, vol. 57, pp. 289–300, 1995.
[16] W. A. Edelstein, P. A. Bottomley, and P. M. Pfeifer, "A signal-to-noise calibration procedure for NMR imaging systems", *Medical Physics*, vol. 11, pp. 180–185, 1983.
[17] J. Sijbers, A. J. den Dekker, J. Van Audekerke, M. Verhoye, and D. Van Dyck, "Estimation of the noise in magnitude in MR images", *Magnetic Resonance Imaging*, vol. 16, pp. 87–90, 1998.
[18] A. M. Wink and J. B. T. M. Roerdink, "BOLD noise assumptions", 2003, in preparation.
[19] D. Malonek and A. Grinvald, "Interactions between electrical activity and cortical microcirculation revealed by imaging spectroscopy; implications for functional brain imaging", *Science*, vol. 272, pp. 551–554, 1996.
[20] K. J. Worsley, M. Wolforth M, and A. C. Evans, "Scale space searches for a periodic signal in fMRI data with spatially varying hemodynamic response", in *Proceedings of BrainMap'96*, 1997.
[21] L. P. Panych, "Theoretical comparison of fourier and wavelet encoding in magnetic resonance imaging", *IEEE Transactions on Medical Imaging*, vol. 15, pp. 141–153, 1996.
[22] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, pp. 674–693, 1989.
[23] M. A. Westenberg and J. B. T. M. Roerdink, "Frequency domain volume rendering by the wavelet X-ray transform", *IEEE Transactions on Image Processing*, vol. 9, pp. 1249–1261, July 2000.
[24] I. Daubechies, "Orthonormal bases of compactly supported wavelets", *Communications on Pure and Applied Mathematics*, vol. 41, pp. 909–996, 1988.
[25] I. Daubechies, "Orthonormal bases of compactly supported wavelets: II. variations on a theme", *SIAM Journal on Mathematical Analysis*, vol. 24, pp. 499–519, 1993.
[26] M. Unser, "Splines: A perfect fit for signal and image processing", *IEEE Signal Processing Magazine*, vol. 16, pp. 22–38, 1999.
[27] F. E. Turkheimer, M. Brett, D. Visvikis, and V. J. Cunningham, "Multiresolution analysis of emission tomography images in the wavelet domain", *Journal of Cerebral Blood Flow and Metabolism*, vol. 19, pp. 1189–1208, 1999.
[28] M. Jansen and A. Bultheel, "Empirical bayes approach to improve wavelet thresholding for image noise reduction", *Journal of the American Statistical Association*, vol. 96, pp. 629–639, 2001.
[29] M. Unser and T. Blu, "Fractional splines and wavelets", *SIAM Review*, vol. 42, pp. 43–67, 2000.
[30] T. Blu and M. Unser, "The fractional spline wavelet transform: Definition and implementation", in *Proc. IEEE: International Conference on Acoustics, Speech, and Signal Processing.*, pp. 512–515, 2000, http://bigwww.epfl.ch/blu/fractsplinewavelets.
[31] A. Cohen, I. Daubechies, and J. C. Feauveau, "Biorthogonal bases of compactly supported wavelets", *Communications on Pure and Applied Mathematics*, vol. 45, pp. 485–560, 1992.
[32] D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness by wavelet shrinkage", *Journal of the American Statistical Association*, vol. 90, pp. 1200–1224, 1995.
[33] D. L. Donoho and I. M. Johnstone, "Ideal denoising in an orthonormal basis chosen from a library of bases", *Comptes Rendus de l'Académie des Sciences, Series A*, vol. 319, pp. 1317–1322, 1994.
[34] I. M. Johnstone and B. W. Silverman, "Wavelet threshold estimators for data with correlated noise", *Journal of the Royal Statistical Society*, vol. 59, pp. 319–351, 1997.
[35] S. Ogawa, T. M. Lee, A. R. Kay, and D. W. Tank, "Brain magnetic resonance imaging with contrast dependent on blood oxygenation", *Proceedings of the National Academy of Sciences*, vol. 87, pp. 9868–9872, 1990.
[36] R. K-S. Kwan, A. C. Evans, and G. B. Pike, "An extensible MRI simulator for post-processing evaluation", in *Proc. Visualization in Biomedical Computing*, vol. 1131 of *Lecture Notes in Computer Science*, pp. 135–140, 1996, http://www.bic.mni.mcgill.ca/brainweb.
[37] S. M. Smith, "Fast automated brain extraction", *Human Brain Mapping*, vol. 17, pp. 143–155, 2002.
[38] J. C. Rajapakse, J. Giedd, and J. L. Rapoport, "Statistical approach to segmentation of single-channel MR images", *IEEE Transactions on Medical Imaging*, vol. 16, pp. 176–186, 1997.
[39] J. C. Rajapakse and J. Piyaratna, "Bayesian approach to segmentation of statistical parametric maps", *IEEE Transactions on Biomedical Engineering*, vol. 48, pp. 1186–1194, 2001.
[40] R. M. J. Donahue, "A note on information seldom reported via the P value", *The American Statistician*, vol. 53, pp. 303–306, 1999.
[41] Y. Benjamini and D. Yekutieli, "The control of the false discovery rate in multiple testing under dependency", *Annals of Statistics*, vol. 29, pp. 1165–1188, 2001.
[42] W-L. Luo and T. E. Nichols, "Diagnosis and exploration of massively univariate fMRI models", Technical report, University of Michigan, 2002, http://www.sph.umich.edu/~nichols/Docs/fMRIvis.pdf.
[43] Dartmouth Brain Imaging Center, "Example fMRI data set", http://dbic.dartmouth.edu/researcher/data.php.

Alle Meije Wink received his M. Sc. ('99) in computing science from the University of Twente, the Netherlands. After a short period of work in medical image processing at the University Hospital of Groningen, the Netherlands, he is currently pursuing a Ph. D. in computing science at the University of Groningen. His research interests include biomedical image processing, functional imaging and wavelets.

Jos B. T. M. Roerdink received his M.Sc. (1979) in theoretical physics from the University of Nijmegen, the Netherlands. Following his Ph.D. (1983) from the University of Utrecht and a two-year position (1983-1985) as a Postdoctoral Fellow at the University of California, San Diego, both in the area of stochastic processes, he joined the Centre for Mathematics and Computer Science in Amsterdam. There he worked from 1986-1992 on image processing and tomographic reconstruction. He was appointed associate professor (1992) and full professor (2003), respectively, at the Institute for Mathematics and Computing Science of the University of Groningen, where he currently holds a chair in Scientific Visualization and Computer Graphics. His current research interests include biomedical visualization, neuroimaging and bioinformatics.