Integrated wavelet processing and spatial statistical testing of fMRI data

Dimitri Van De Ville [a,*] Thierry Blu [a] Michael Unser [a]

[a] *Biomedical Imaging Group,*
Swiss Federal Institute of Technology Lausanne (EPFL),
CH-1015 Lausanne, Switzerland

Abstract

We introduce an integrated framework for detecting brain activity from fMRI data, which is based on a 3D spatial discrete wavelet transform. Unlike the standard wavelet-based approach for fMRI analysis, we apply the suitable statistical test procedure in the spatial domain. For a desired significance level, this scheme has one remaining degree of freedom, characterizing the wavelet processing, which is optimized according to the principle of minimal approximation error. This allows us to determine the threshold values in a way that does not depend on data. While developing our framework, we make only conservative assumptions. Consequently the detection of activation is based on strong evidence. We have implemented this framework as a toolbox (WSPM) for the SPM2 software, taking advantage of multiple options and functions of SPM such as the setup of the linear model and the use of the hemodynamic response function. We show by experimental results that our method is able to detect activation patterns; the results are comparable to those obtained by SPM even though statistical assumptions are more conservative.

Key words: discrete wavelet transform, wavelet thresholding, statistical testing, threshold selection, approximation error

Accepted for publication in NeuroImage (YNIMG2708)

* Corresponding author. Telephone +41-21-6935142. Fax +41-21-6933701.
Email addresses: Dimitri.VanDeVille@epfl.ch (Dimitri Van De Ville), Thierry.Blu@epfl.ch (Thierry Blu), Michael.Unser@epfl.ch (Michael Unser).

1 Introduction

Functional magnetic resonance imaging (fMRI) has become a key modality to perform non-invasive studies of brain. Its working principle is based on interaction between neuronal activity and physiology, such as blood oxygenation and blood flow. Through a variation of the magnetic field uniformity, these interactions induce a weak and noisy T2*-contrast signal (Ogawa et al. (1993)).

The most widely deployed and recognized method and associated software package for performing the analysis of fMRI data is Statistical Parametric Mapping (SPM) (Friston et al. (1995); Frackowiak et al. (1997)). SPM is a parametric hypothesis-driven approach: it performs a statistical test on the fitted parameters of a linear model (LM) and detects activation at the spatial locations where the non-activation hypothesis is rejected. One of SPM's main characteristics is the application of a fixed Gaussian prefilter. The rationale of the Gaussian prefilter is twofold. First, it improves the signal-to-noise ratio (SNR) as indicated by the matched-filter theorem (Worsley et al. (1996)). Second, it controls the spatial smoothness of the data which is used to correct for multiple testing by advanced methods based on continuous Gaussian random fields and Euler characteristics (Poline et al. (1997)). However, Gaussian-shaped activation patterns are observed rather rarely in practice. While researchers have also proposed the use of more general filters (Kruggel et al. (1999); Shafie et al. (2003)), the use of fixed Gaussian smoothing is still the most-commonly applied method for fMRI analysis.

As an alternative approach, some researchers have advocated the use of a spatial wavelet transform instead of Gaussian prefiltering (Ruttimann et al. (1998); Turkheimer et al. (2000)). This method exploits the sparsity of the data representation in the wavelet domain to improve the detection sensitivity. In particular, the cluster of voxels that make up an activated region tend to be spatially correlated and can be efficiently encoded using only a few wavelet coefficients. Such a sparse representation increases the SNR, since the noise remains evenly distributed in the wavelet domain. Therefore, a coefficient-wise statistical t-test provides a much higher sensitivity than a voxel-wise approach in the spatial domain, even when multiple testing is compensated by a conservative Bonferroni correction. After the detection phase in the wavelet domain, an inverse wavelet transform is applied to reconstruct an activation map from the coefficients that are designated as significant. While this reconstructed map is very useful for visualization purposes, it does not have a direct statistical interpretation; i.e., the t-values are computed in the wavelet domain and there is no straightforward way to map the statistics to the spatial domain. Some proposed solutions include the application of an ad-hoc threshold to the reconstructed map (e.g., a percentage of the maximal signal level); re-testing

in the spatial domain (without taking into account the effect of the initial test in the wavelet domain) (Fu et al. (1998)); recursively reconstructing the activity map by controlling the false discovery rate (Genovese et al. (2002); Bullmore et al. (2003)). Wavelets have also been applied to fMRI analysis as a pure spatial denoising preprocessing step (Hilton et al. (1996); Wink and Roerdink (2004)). Another important issue is to deal with the temporal correlation of fMRI data (Woolrich et al. (2001)), for which wavelet-based methods have also been proposed (Bullmore et al. (2001); Fadili and Bullmore (2001)). The combination of the proposed method with temporal treatments will be considered in future work.

The previous spatial wavelet-based approach faces a fundamental problem; i.e., approximation and detection are performed at once in the wavelet domain. Therefore, the detected parameter map problematically remains in the wavelet domain, which complicates the spatial interpretation. In this paper, we propose a new framework that is based on two main principles:

(1) Clear separation of the approximation and detection procedures: The approximation procedure is carried out in the wavelet domain, while the statistical testing is done in the spatial domain. The test procedure properly takes into account the effect of the approximation step.
(2) Most faithful reconstruction: After statistical testing in the spatial domain, only a limited number of voxels are detected. Thus, we propose to optimize the wavelet processing so as to minimize the difference between the detected and unprocessed parameter map. Imposing this constraint allows us to uniquely determine the optimal threshold values that characterize the approximation and detection procedures, for a given desired significance level. It turns out that this optimum does not depend on the data.

The key to the proposed framework is that it is completely integrated: eventhough wavelet-domain processing and spatial statistical testing are separate, we take into account their mutual influence. Moreover, the detection procedure of the framework rests on conservative assumptions that comply with a worst-case scenario. Therefore, the framework provides a strong type-I error control; i.e., the rejection of the hypothesis $\mathcal{H}_0$ for no activity is evidence for an activated region since false positives are very well controlled.

The new framework is quite general and can be applied to a wide variety of wavelet-like transforms. In this paper, we restrict ourselves to a proof-of-concept and we demonstrate its usefulness by the way of experiments that involve the well-known B-spline wavelet transform. Despite the conservative assumptions, we find activation maps comparable to those of SPM. The proposed framework has been integrated into SPM2 as a "Wavelet toolbox", allowing the user to have the usual SPM-based analysis and the joint spatio-

wavelet analysis side by side.

The paper is organized as follows: In Sect. 2, we briefly review the standard wavelet-based method. Next, in Sect. 3, we introduce the new framework. Experimental results are shown in Sect. 5.

2 Background

We start this section with a brief review of the wavelet transform. We then present the standard wavelet-based method for fMRI analysis, extended with the linear model (LM) to easily setup experiments and incorporate the effect of the hemodynamic response function (HRF) (Friston et al. (1995); Van De Ville et al. (2003); Mueller et al. (2003)). The statistical inference is performed by a coefficient-wise t-test (Turkheimer et al. (2000); Feilner et al. (2000)).

2.1 The discrete wavelet transform

The discrete wavelet transform (DWT) is a popular tool for multi-resolution analysis (Mallat (1989, 1998)). Its two-channel filterbank implementation is well-known and has found its way into a wide range of applications. To provide the best insight, we first formulate the transform in the continuous domain by decomposing a function into a sum of shifted and scaled continuously defined wavelets.

Let us consider a 1D continuously defined function $v(x)$, which is known by its sample values $v[n]$, $n \in \mathbb{Z}$. We start by the representation of this function in a shift-invariant signal space $\mathcal{V}_0(\varphi) = \mathrm{span}_{n\in\mathbb{Z}}\{\varphi(x-n)\}$. For the decomposition to be numerically stable and unambiuous, the function φ needs to generate a Riesz basis. So we have

$$v(x) = \sum_k c_0[k]\varphi(x-k), \tag{1}$$

where the coefficients $c_0[k]$ can be chosen such that $v(n) = v[n]$. Performing one stage of the wavelet transform consists of nothing else than splitting the signal space $\mathcal{V}_0(\varphi)$ into $\mathcal{V}_1(\varphi) = \mathrm{span}_n\{\varphi(x/2-n)\}$, characterized by dilated basis functions on a coarser grid, and its orthogonal complement in $\mathcal{V}_0$, $\mathcal{W}_1(\psi) = \mathrm{span}_n\{\psi(x/2-n)\}$, characterized by the wavelet ψ. Consequently, the function (1) can be decomposed as

$$v(x) = \sqrt{2}\sum_k c_1[k]\varphi(x/2-k) + \sqrt{2}\sum_k w_1[k]\psi(x/2-k), \tag{2}$$

where the coefficients $c_1[k]$ and $w_1[k]$ fully describe the orthogonal projection of $v(x)$ into the spaces $\mathcal{V}_1$ and $\mathcal{W}_1$, respectively. Specifically, we obtain $c_1[k]$ as $\langle v(\cdot), \tilde{\varphi}(\cdot/2 - k)\rangle$, where $\tilde{\varphi}$ denotes the dual function of φ; i.e., the unique function of $\mathcal{V}_0$ such that $\langle \tilde{\varphi}(\cdot - k), \varphi(\cdot - l)\rangle = \delta_{k-l}$. Similarly, the "detail coefficients" $w_1[k]$ are obtained using the dual wavelet $\tilde{\psi}(\cdot/2 - k)$. Figure 1 (a) shows a flowchart of a one-level decomposition. Using the same principle, the signal (2) can be decomposed further, J_w times, as in

$$v(x) = 2^{J_w/2} \sum_k c_{J_w}[k]\varphi(x/2^{J_w} - k) + \sum_{j=1}^{J_w} 2^{j/2} \sum_k w_j[k]\psi(x/2^j - k). \qquad (3)$$

We now introduce a shorthand notation for a wavelet decomposition such as (3):

$$v(x) = \sum_k w_k \psi_k(x), \qquad (4)$$

where w_k and its index k runs over all scales of the decomposition, while the functions ψ_k corresponds to the associated scaled, shifted, and dilated versions of the scaling function or of the wavelet.

In practice, the DWT is implemented efficiently through an iterated filterbank, which directly operates on the discrete coefficients $c_j[k]$. The filters are derived from the dual scaling function and wavelet. For example, the scaling coefficients $c_1[k]$ after one iteration can be found by filtering and subsampling with the scaling filter $\tilde{h}$ at the analysis side:

$$c_1[k] = (c_0 * \tilde{h})[2k]. \qquad (5)$$

Usually, we characterize the filters in the z-domain; e.g., $\tilde{H}(z) = \sum_k \tilde{h}[k]z^{-k}$. In Fig. 1 (b), we show the flowchart of the filterbank representation for one iteration of the DWT.

The DWT is non-redundant; i.e., the number of coefficients is always equal to the initial number of samples. The wavelet transform of piecewise smooth signals tends to be quite sparse, a property that is very useful for many applications.

The decomposition above is easily extended to the multi-dimensional case by using tensor-product basis functions; i.e., the 3D scaling function $\varphi(\mathbf{x})$ corresponds to $\varphi(x_1)\varphi(x_2)\varphi(x_3)$. Specifically, we have one scaling function and 7 wavelets, while the overall subsampling ratio at each iteration is 8.

2.2 Standard wavelet-based statistical analysis of fMRI data

An fMRI dataset is denoted as $v^{(t)}[\mathbf{n}]$, $\mathbf{n} \in \mathbb{Z}^3$, $t \in \mathbb{Z}$, where $\mathbf{n}$ and $t = 1, \ldots, N_t$ are the 3D-spatial and temporal indices, respectively. The non-redundant spa-

tial 3D DWT of a volume, $v^{(t)}[\mathbf{n}]$, yields the coefficients $w_{\mathbf{k}}^{(t)}$, where the index $\mathbf{k}$ addresses all subbands and orientations. As in (4), we compactly denote the wavelet decomposition as

$$v^{(t)}[\mathbf{n}] = \sum_{\mathbf{k}} w_{\mathbf{k}}^{(t)} \psi_{\mathbf{k}}(\mathbf{n}). \tag{6}$$

Now we introduce the time-series vector of length N_t for a wavelet coefficient; i.e., $\mathbf{w_k} = [w_{\mathbf{k}}^{(1)} \dots w_{\mathbf{k}}^{(N_t)}]^{\mathrm{T}}$. The temporal behavior of the wavelet coefficient is described by a LM, which represent the vector $\mathbf{w_k}$ as [1]

$$\mathbf{w_k} = \mathbf{X}\mathbf{y_k} + \mathbf{e_k}, \tag{7}$$

where $\mathbf{X}$ is the $N_t \times L$ design matrix, and $\mathbf{e_k}$ is the residual error. The matrix $\mathbf{X}$ contains the desired signal models. For example, in the case of a simple block-based experiment, $\mathbf{X}$ contains two columns: one for the on-off stimulus and one for the background. Usually, one convolves the stimulus with a model for the HRF. Under the assumption of independently and identically Gaussian-distributed residual error $\mathbf{e_k}$, the optimal unbiased estimate of $\mathbf{y_k}$ is the least-squares estimate given by $\hat{\mathbf{y}}_\mathbf{k} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{w_k}$. The residual error of this estimate is obtained as $\hat{\mathbf{e}}_\mathbf{k} = \mathbf{w_k} - \mathbf{X}\hat{\mathbf{y}}_\mathbf{k}$. Next, the information of interest is extracted from $\hat{\mathbf{y}}_\mathbf{k}$ by a contrast vector $\mathbf{c}$ (e.g., $\mathbf{c} = [1\ 0]^{\mathrm{T}}$ for our simple example). At this stage, we obtain two scalar values for the $\mathbf{k}$-th wavelet coefficient:

$$g_\mathbf{k} = \mathbf{c}^{\mathrm{T}}\hat{\mathbf{y}}_\mathbf{k}, \tag{8}$$
$$s_\mathbf{k}^2 = \hat{\mathbf{e}}_\mathbf{k}^{\mathrm{T}}\hat{\mathbf{e}}_\mathbf{k}\mathbf{c}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{c}, \tag{9}$$

where $g_\mathbf{k}$ and $s_\mathbf{k}^2$ follow a Gaussian and a χ^2 distribution, respectively. The hypothesis to test is whether the coefficient $\mathbf{k}$ is activated and thus has a non-zero mean:

$$\mathcal{H}_0 : \mu_\mathbf{k} = 0, \tag{10}$$
$$\mathcal{H}_1 : \mu_\mathbf{k} \neq 0. \tag{11}$$

The t-value for each wavelet coefficient can be found as

$$t_\mathbf{k} = \frac{g_\mathbf{k}}{\sqrt{s_\mathbf{k}^2/J}}, \quad \text{with } J = N_t - \mathrm{rank}(\mathbf{X}), \tag{12}$$

which can be tested against a threshold τ_w. This value τ_w is determined by the desired significance level α (e.g., 5%) for a two-sided t-test (since both positive

[1] Since the DWT is a spatial linear operator and the LM analysis a temporal linear operator, it is equivalent to apply the LM to $v^{(t)}[\mathbf{n}]$ or to its wavelet coefficients $w_\mathbf{k}^{(t)}$. The link between the respective parameters $y[\mathbf{n}]$ and $y_\mathbf{k}$ is simply the DWT.

and negative wavelet coefficients can contribute to an increased signal in the spatial domain). The significance level α needs to be corrected for multiple hypothesis testing. One typically applies a conservative Bonferroni correction, which reduces the significance level to $\alpha_B = \alpha/N_c$, where N_c is the number of statistical tests. The value N_c can be chosen as the number of intracranial wavelet coefficients, which corresponds closely to the number of intracranial voxels.

The wavelet coefficients $g_{\mathbf{k}}$ that survived the statistical test $|t_{\mathbf{k}}| > \tau_w$ can be reconstructed as

$$r[\mathbf{n}] = \sum_{\mathbf{k}} H(|t_{\mathbf{k}}| - \tau_w)\, g_{\mathbf{k}} \psi_{\mathbf{k}}(\mathbf{n}), \tag{13}$$

where $H(t)$ is the Heaviside step function defined as

$$H(t) = \begin{cases} 0, \text{ when } t < 0, \\ 1, \text{ otherwise.} \end{cases} \tag{14}$$

In other words, the term $H(|t_{\mathbf{k}}| - \tau_w)$ in (13) acts as an indicator function which is equal to 1 for $|t_{\mathbf{k}}| \geq \tau_w$ and 0 otherwise. Depending on the support of the wavelet $\psi_{\mathbf{k}}$, which can be infinite, the reconstructed volume $r[\mathbf{n}]$ will contain many non-zero voxels. Often, an ad-hoc spatial threshold is applied to obtain an "acceptable" activation map. However, such a threshold does not take into account the variability of the underlying voxels as compared to the t-value in the wavelet domain. It is therefore not possible to associate a clear statistical meaning to the reconstructed volume $r[\mathbf{n}]$; this constitutes the main disadvantage of the wavelet-based approach.

Figure 2 (a) shows schematically the standard wavelet procedure. The desired significance level α_B directly determines the threshold value τ_w for the t-test in the wavelet domain. The detection procedure in the wavelet domain is statistically sound and quantitative; unfortunately, the transposition of the results back into the spatial domain is qualitative only.

The non-redundancy of the DWT is an important property for the multiple-testing correction. Indeed, a redundant transform would require the use of a higher threshold value and would therefore reduce the sensitivity of the approach.

3 Method: Integrated wavelet processing and spatial statistical testing

3.1 Main idea

The major advantage of the wavelet-based method is its apparent high sensitivity, even though the conservative Bonferroni-correction for multiple testing is used. The underlying reason is the sparsity property of the wavelet transform. In other words, the activation patterns exhibit a local correlation which is compactly encoded by the wavelet basis functions. Therefore, thresholding the t-values in the wavelet domain is an efficient way to improve the SNR without removing any information. It follows that this approach has the potential to detect activations with a high spatial resolution.

Here, we also propose to rely on thresholding in the wavelet domain. However, this procedure is considered as a pre-processing step (parametrized by the wavelet threshold), and the statistical test is ultimately performed in the spatial domain, taking into account the processing that has been done before. Figure 2 (b) summarizes this basic philosophy. The threshold τ_w is still applied to the t-values in the wavelet domain but is treated as a general parameter of a nonlinear algorithm. The final testing in the spatial domain is implemented via a threshold τ_s. Intuitively, it is clear that there is an infinity of possible combinations (τ_w, τ_s) that will establish the same desired significance level α_B. We introduce the principle of "minimal approximation error" between the unprocessed data (after the LM) and the final result. Our method does not require any additional hypotheses. In particular, the wavelet coefficients do not need to be decorrelated.

The proposed framework consists of two major parts. The first one links the processing of the t-values in the wavelet domain, which is parameterized by a threshold τ_w, to the statistical hypothesis testing in the spatial domain, characterized by a threshold τ_s. Second, we develop the principle of minimizing the approximation error between the data after LM fitting and the result after applying both thresholds. It is this second part of our framework that will provide the right way to determine the optimal values of the wavelet and spatial thresholds. These optimal thresholds can be computed off-line and are data-independent; they are a function of the desired significance level and the number of intracranial voxels.

3.2 Simplified case—true $\sigma_{\mathbf{k}}$ are known

To facilitate the exposition of our method, we first consider the case where N_t is sufficiently large, so that the normalizing term in (12), $s_{\mathbf{k}}/\sqrt{J}$, is essentially equivalent to $\sigma_{\mathbf{k}}$, the standard deviation of $g_{\mathbf{k}}$. In this limiting case, the t-values $t_{\mathbf{k}}$ are equivalent to z-scores and are normally distributed. Therefore, we denote $g_{\mathbf{k}}/\sigma_{\mathbf{k}}$ as $z_{\mathbf{k}}$.

3.2.1 Part I: Effect of wavelet processing

We follow the standard procedure of Sect. 2 up to the calculation of the t-value in the wavelet domain, which is this time equivalent to a z-score under the assumption $\sigma_{\mathbf{k}} = s_{\mathbf{k}}/\sqrt{J}$. Similarly, we apply a threshold to the values $g_{\mathbf{k}}$ to obtain $H(|z_{\mathbf{k}}| - \tau_w)\, g_{\mathbf{k}}$. However, we ignore the interpretation as a statistical test. Our goal is to derive a spatially-varying threshold $q[\mathbf{n}]$ such that, under the null hypothesis, the desired significance level α_B is not exceeded by the probability of the reconstruction of the processed wavelet coefficients that contribute to the value of the voxel $\mathbf{n}$:

$$P\left[\sum_{\mathbf{k}} H(|z_{\mathbf{k}}| - \tau_w)\, g_{\mathbf{k}} \psi_{\mathbf{k}}(\mathbf{n}) \geq q[\mathbf{n}]\right] \leq \alpha_B. \tag{15}$$

For this purpose, we will make use of Theorem 1 in Appendix A, which provides an upper probability bound for a convex sum of random variables. Therefore, we take a closer look at the reconstructed volume and manipulate it to obtain a convex sum of random variables that follow the same probability density function. Specifically, by introducing $\sigma_{\mathbf{k}}$, which is the true standard deviation of $g_{\mathbf{k}}$, we can rewrite the reconstructed map as

$$\begin{aligned}
\sum_{\mathbf{k}} H(|z_{\mathbf{k}}| - \tau_w)\, g_{\mathbf{k}} \psi_{\mathbf{k}}(\mathbf{n}) &= \sum_{\mathbf{k}} H(|z_{\mathbf{k}}| - \tau_w) \overbrace{\frac{g_{\mathbf{k}}}{\sigma_{\mathbf{k}}}}^{z_{\mathbf{k}}} \operatorname{sign}(\psi_{\mathbf{k}}(\mathbf{n}))\ \sigma_{\mathbf{k}} |\psi_{\mathbf{k}}(\mathbf{n})| \\
&= \Lambda[\mathbf{n}] \sum_{\mathbf{k}} \lambda_{\mathbf{k}} H(|z_{\mathbf{k}}| - \tau_w)\, z_{\mathbf{k}} \operatorname{sign}(\psi_{\mathbf{k}}(\mathbf{n})) \\
&= \Lambda[\mathbf{n}] \sum_{\mathbf{k}} \lambda_{\mathbf{k}} \xi_{\mathbf{k}},
\end{aligned} \tag{16}$$

with $\Lambda[\mathbf{n}] = \sum_{\mathbf{k}'} \sigma_{\mathbf{k}'} |\psi_{\mathbf{k}'}(\mathbf{n})|$, $\lambda_{\mathbf{k}} = \sigma_{\mathbf{k}} |\psi_{\mathbf{k}}(\mathbf{n})| / \Lambda[\mathbf{n}]$, and where we have introduced the random variables

$$\xi_{\mathbf{k}} = H(|z_{\mathbf{k}}| - \tau_w)\, z_{\mathbf{k}} \operatorname{sign}(\psi_{\mathbf{k}}(\mathbf{n})). \tag{17}$$

It is important to notice that $\xi_{\mathbf{k}}$ follows a normalized probability density function independent of $\mathbf{k}$ because $z_{\mathbf{k}}$ follows a normalized Gaussian distribution,

which is symmetric. Furthermore, we see that the factors $\lambda_{\mathbf{k}}$ are positive such that $\sum_{\mathbf{k}} \lambda_{\mathbf{k}} = 1$. Therefore, $\Lambda[\mathbf{n}]$ can be considered as a normalization of the reconstructed volume and the non-stationary threshold can be chosen as $q[\mathbf{n}] = \tau_s \Lambda[\mathbf{n}]$.

Using (16) and (17), we can then simplify Expression (15) as

$$P\left[\sum_{\mathbf{k}} H(|z_{\mathbf{k}}| - \tau_w)\, g_{\mathbf{k}} \psi_{\mathbf{k}}(\mathbf{n}) \geq q[\mathbf{n}]\right] = P\left[\sum_{\mathbf{k}} \lambda_{\mathbf{k}}(\xi_{\mathbf{k}} - \tau_s) \geq 0\right]. \tag{18}$$

Since the sum on the right-hand side of Eq. (18) satisfies the conditions of Theorem 1, we obtain the following probability bound for the convex sum:

$$P\left[\sum_{\mathbf{k}} \lambda_{\mathbf{k}}(\xi_{\mathbf{k}} - \tau_s) \geq 0\right] \leq \min_{a>0} E[(1 + a(\xi - \tau_s))_+], \tag{19}$$

where the notation $(x)_+$ stands for $\max(0, x)$. Going back to Eq. (17), we see that ξ follows a truncated normalized Gaussian distribution with a Dirac peak at the origin, as illustrated in Fig. 3. The effect of the threshold is to map all small coefficients to zero, while the sign function has no influence on a symmetric distribution.

Usually, the value of a that provides the sharpest probability estimate (i.e., which minimizes the right-hand side of (19)) is $a^* = 1/\tau_s$. The probability can then be bounded as

$$P\left[\frac{\sum_{\mathbf{k}} H(|z_{\mathbf{k}}| - \tau_w)\, g_{\mathbf{k}} \psi_{\mathbf{k}}(\mathbf{n})}{\sum_{\mathbf{k}} \sigma_{\mathbf{k}} |\psi_{\mathbf{k}}(\mathbf{n})|} \geq \tau_s\right] \leq \frac{E[(\xi)_+]}{\tau_s},$$

which finally provides the conservative spatial threshold

$$\tau_s = \frac{E[(\xi)_+]}{\alpha_B} \tag{20}$$

given the significance level α_B.

The expectation value of the right-hand side can be computed as a function of the wavelet threshold τ_w, independently from the data. Figure 4 (a) shows the surface described by the triplet $(\tau_s, \tau_w, \alpha_B)$.

The main importance of the bound of (19), is that it allows us to derive sharp threshold values, in particular for short-tailed distributions such as the Gaussian. For an illustrative example, we refer to App. B.

3.2.2 Part II: Minimizing the approximation error

Now we consider the second approximation aspect of our framework. Equation (20) does not present a complete solution to our problem. Indeed, infinityly many combinations of τ_w and τ_s achieve the same desired significance level α_B. These solutions differ, however, by the quality of approximation of the wavelet processing: the better the approximation—i.e., the smaller the wavelet threshold τ_w—the larger the spatial threshold τ_s; on the contrary, if the approximation is poor (i.e., large wavelet threshold), the few spatial detections cannot reasonably be localized and identified on the unprocessed volume.

We resolve this issue by searching for the solution that minimizes the worst-case approximation error between the unprocessed and detected parameter map. In other words, we want the final result to be as close as possible to the original data. We will see how this provides a simple but elegant way to determine optimal threshold values.

The basic problem is outlined in Fig. 5. Top left, we have the wavelet coefficients $g_{\mathbf{k}}$, obtained after fitting the LM. From these data, we can obtain the three following reconstructions in the spatial domain:

(1) The reconstruction of the raw wavelet coefficients without any processing:

$$u[\mathbf{n}] = \sum_{\mathbf{k}} g_{\mathbf{k}} \psi_{\mathbf{k}}(\mathbf{n}). \tag{21}$$

Notice that the same $u[\mathbf{n}]$ would be obtained by fitting the LM directly in the spatial domain.

(2) The reconstruction after thresholding the wavelet coefficients according to their t-values in the wavelet domain:

$$r[\mathbf{n}] = \sum_{\mathbf{k}} H(|z_{\mathbf{k}}| - \tau_w)\, g_{\mathbf{k}} \psi_{\mathbf{k}}(\mathbf{n}). \tag{22}$$

(3) The final result after statistical testing of $r[\mathbf{n}]$ in the spatial domain:

$$r'[\mathbf{n}] = H(r[\mathbf{n}] - \tau_s \Lambda[\mathbf{n}])\, r[\mathbf{n}]. \tag{23}$$

Our guiding principle is to minimize the difference between the non-processed reconstruction $u[\mathbf{n}]$ and the final result $r'[\mathbf{n}]$. To this end, we express the different between $u[\mathbf{n}]$ and $r'[\mathbf{n}]$ as

$$\begin{aligned} |u[\mathbf{n}] - r'[\mathbf{n}]| &= |u[\mathbf{n}] - r[\mathbf{n}] + r[\mathbf{n}] - r'[\mathbf{n}]| \\ &\leq |u[\mathbf{n}] - r[\mathbf{n}]| + |r[\mathbf{n}] - r'[\mathbf{n}]|, \end{aligned}$$

where the first term can be bounded by

$$
\begin{aligned}
|u[\mathbf{n}] - r[\mathbf{n}]| &= \left| \sum_{\mathbf{k}} (1 - H(|z_{\mathbf{k}}| - \tau_w)) g[\mathbf{k}] \psi_{\mathbf{k}}[\mathbf{n}] \right| \\
&\leq \tau_w \sigma_{\mathbf{k}} \, |\psi_{\mathbf{k}}[\mathbf{n}]| \\
&= \tau_w \Lambda[\mathbf{n}],
\end{aligned}
$$

and the second one by

$$
\begin{aligned}
|r[\mathbf{n}] - r'[\mathbf{n}]| &= (1 - H(r[\mathbf{n}] - \tau_s \Lambda[\mathbf{n}])) \, |r[\mathbf{n}]| \\
&\leq \tau_s \Lambda[\mathbf{n}].
\end{aligned}
$$

So we obtain

$$
|u[\mathbf{n}] - r'[\mathbf{n}]| \leq (\tau_w + \tau_s) \Lambda[\mathbf{n}], \tag{24}
$$

which is valid pointwise for each voxel $\mathbf{n}$, and therefore also conservative, in the sense that it holds for a worst-case scenario. It also important to observe that this is a sharp worst-case bound since it can be attained by some configuration of realizations of $g_{\mathbf{k}}$. Consequently, the optimal values for the thresholds τ_w and τ_s are obtained by simply minimizing their sum:

$$
(\tau_w, \tau_s) = \arg\min_{\tau_w, \tau_s} \left\{ \tau_w + \tau_s, \text{ subject to } \tau_s = \frac{E[(\xi)_+]}{\alpha_B} \right\}. \tag{25}
$$

We now apply this optimization to the results obtained from Sect. 3.2.1. First, we derive the relationship between τ_w and τ_s by using Eq. (20), which yields

$$
\tau_s = \frac{1}{\alpha_B} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\tau_w^2}{2} \right). \tag{26}
$$

Furthermore, we can show that the minimization of the sum $\tau_w + \tau_s$ leads us to the optimal values

$$
\tau_s = 1/\tau_w, \quad \tau_w = \sqrt{-W_{-1}\left(-2\pi\alpha_B^2\right)}, \tag{27}
$$

where $W_{-1}(\cdot)$ is the -1-branch of the Lambert W-function (also called Omega function); i.e., the inverse of the function $f(W) = W \exp(W)$. This function can be evaluated numerically (Corless et al. (1996)). For a detailed derivation of (26) and (27), we refer to Appendix C.

On the probability surface of Fig. 4 (a), the circular dots mark the optimal threshold values for the covered range of α_B. In Fig. 4 (b), we show the optimal threshold values directly as a function of α_B. As a typical example, we could consider $\alpha = 5\%$ and $N_c = 10^4$ (number of intracranial voxels), resulting in $\alpha_B = 5 \times 10^{-6}$. Notice that the surface of Fig. 4 (a) can also be used to determine p-values; i.e., for a fixed threshold τ_w, the value of τ_s can be chosen equal to $r[\mathbf{n}]/\Lambda[\mathbf{n}]$, resulting into a p-value $N_c \times \alpha_B(\tau_s, \tau_w)$.

3.3 General case—true $\sigma_{\mathbf{k}}$ are unknown

In the general case, the true $\sigma_{\mathbf{k}}$ are unknown and their estimation by $s_{\mathbf{k}}/\sqrt{J}$ should be taken into consideration for the testing procedure. Consequently, the number of volumes N_t and the related degrees of freedom J will influence the threshold values.

The strategy of our approach is very similar to Sect. 3.2. However, this time the spatially-varying threshold becomes

$$q[\mathbf{n}] = \tau_s \sum_{\mathbf{k}} \frac{s_{\mathbf{k}}}{\sqrt{J}} \left|\psi_{\mathbf{k}}(\mathbf{n})\right|, \tag{28}$$

where $s_{\mathbf{k}}/\sqrt{J}$ is the estimate for $\sigma_{\mathbf{k}}$, meaning that we can no longer reduce t-values to z-scores in the inequality inside the probability of (15). We write

$$\begin{aligned}
&\sum_{\mathbf{k}} H(|t_{\mathbf{k}}| - \tau_w)\, g_{\mathbf{k}} \psi_{\mathbf{k}}(\mathbf{n}) - q[\mathbf{n}] \geq 0 \\
\iff &\sum_{\mathbf{k}} H(|t_{\mathbf{k}}| - \tau_w)\, g_{\mathbf{k}} \psi_{\mathbf{k}}(\mathbf{n}) - \tau_s \frac{s_{\mathbf{k}}}{\sqrt{J}} \left|\psi_{\mathbf{k}}(\mathbf{n})\right| \geq 0 \\
\iff &\sum_{\mathbf{k}} \sigma_{\mathbf{k}} \left|\psi_{\mathbf{k}}(\mathbf{n})\right| \left(H(|t_{\mathbf{k}}| - \tau_w)\, \frac{g_{\mathbf{k}}}{\sigma_{\mathbf{k}}}\, \mathrm{sign}(\psi_{\mathbf{k}}(\mathbf{n})) - \tau_s \frac{s_{\mathbf{k}}/\sqrt{J}}{\sigma_{\mathbf{k}}} \right) \geq 0 \\
\iff &\sum_{\mathbf{k}} \lambda_{\mathbf{k}} \left(H(|t_{\mathbf{k}}| - \tau_w)\, \frac{g_{\mathbf{k}}}{\sigma_{\mathbf{k}}}\, \mathrm{sign}(\psi_{\mathbf{k}}(\mathbf{n})) - \tau_s \frac{s_{\mathbf{k}}/\sqrt{J}}{\sigma_{\mathbf{k}}} \right) \geq 0 \\
\iff &\sum_{\mathbf{k}} \lambda_{\mathbf{k}} (\xi_{\mathbf{k}} - \tau_s \varsigma_{\mathbf{k}}) \geq 0,
\end{aligned} \tag{29}$$

where $\lambda_{\mathbf{k}} = \sigma_{\mathbf{k}} \left|\psi_{\mathbf{k}}(\mathbf{n})\right| / \sum_{\mathbf{l}} \sigma_{\mathbf{l}} \left|\psi_{\mathbf{l}}(\mathbf{n})\right|$, and where we have introduced the random variables

$$\xi_{\mathbf{k}} = H(|t_{\mathbf{k}}| - \tau_w)\, \frac{g_{\mathbf{k}}}{\sigma_{\mathbf{k}}} \mathrm{sign}(\psi_{\mathbf{k}}(\mathbf{n})), \tag{30}$$

$$\varsigma_{\mathbf{k}} = \frac{s_{\mathbf{k}}/\sqrt{J}}{\sigma_{\mathbf{k}}}. \tag{31}$$

Once again, it is important to notice that all $\xi_{\mathbf{k}}$ and $\varsigma_{\mathbf{k}}$ follow the same probability density distributions ξ and ς, respectively. These are defined as some reference random variables

$$\xi = H\left(\left| \frac{g'}{s'/\sqrt{J}} \right| - \tau_w \right) g', \tag{32}$$

$$\varsigma = s'/\sqrt{J}, \tag{33}$$

where g' follows a normalized Gaussian distribution, and where s'^2 follows a normalized χ^2-distribution with J degrees of freedom, and is statistically independent of g'.

Then, the probability of (29) can be bounded again using Theorem 1:

$$P\left[\frac{\sum_{\mathbf{k}} H(|t_{\mathbf{k}}| - \tau_w)\, g_{\mathbf{k}} \psi_{\mathbf{k}}(\mathbf{n})}{\sum_{\mathbf{k}} s_{\mathbf{k}}/\sqrt{J}\, |\psi_{\mathbf{k}}(\mathbf{n})|} \geq \tau_s\right] = P\left[\sum_{\mathbf{k}} \lambda_{\mathbf{k}} \left(\xi_{\mathbf{k}} - \tau_s \varsigma_{\mathbf{k}}\right) \geq 0\right]$$
$$\leq \min_{a>0} E\left[(1 + a(\xi - \tau_s \varsigma))_+\right], \qquad (34)$$

which finally provides the following (conservative) relation between the spatial and the wavelet threshold

$$\min_{a>0} E\left[(1 + a(\xi - \tau_s \varsigma))_+\right] = \alpha_B \qquad (35)$$

given the significance level, α_B.

In general, the exact computation of the bound (34) is quite involved. A sharp upper value is given in Appendix D; it is also used to determine the optimal value a^* by a numerical optimization procedure. Figure 6 (a) shows the probability surface that we finally obtain. Again, the optimal threshold values τ_w and τ_s can be found by minimizing their sum. The result of the complete optimization is provided in Fig. 6 (b) and (c), where we show the optimal threshold values as function of the desired significance level for $N_t = 50$ and $N_t = 150$, respectively. Notice how, for the second case, the threshold values approach the idealized case of Sect. 3.2. The calculation of the optimal thresholds could be done offline since their values are data-independent. The computation of one pair (τ_s, τ_w), including the optimization of a, requires the degrees of freedom J and the significance level α_B, and takes a few seconds in MATLAB™.

3.4 Summary of the proposed approach

Having discussed the different modules in Figs. 2 (b) and 5, we now briefly summarize the main computational steps of the proposed approach.

(1) The threshold values τ_w and τ_s are determined as a function of the desired significance level α/N_c and the degrees of freedom J.
(2) The spatial DWT is computed for each volume $v^{(t)}[\mathbf{n}]$ of the time-series, resulting in the coefficients $w_{\mathbf{k}}^{(t)}$.
(3) For each time-serie of wavelet coefficient $w_{\mathbf{k}}^{(t)}$, the LM is applied and the parameter of interest is extracted. This way, we obtain the parameter's estimate $g_{\mathbf{k}}$ and its estimated standard deviation $s_{\mathbf{k}}^2$.

(4) After wavelet processing (i.e., applying the threshold τ_w to the $g_{\mathbf{k}}$'s), we use the inverse DWT to reconstruct the volume $r[\mathbf{n}]$.
(5) The values $s_{\mathbf{k}}/\sqrt{J}$ are reconstructed by a modified inverse DWT algorithm, which corresponds to putting the absolute value of the wavelet. We obtain the volume $\Lambda[\mathbf{n}]$.
(6) The detected parameter map is obtained by applying the threshold τ_s to $r[\mathbf{n}]/\Lambda[\mathbf{n}]$.

4 Implementation: A new toolbox for SPM

Our new approach has been implemented as a "Wavelet toolbox", called WSPM, for SPM2. In this way, the user can setup his experiments as usual using SPM's extensive features for preprocessing (e.g., registration) and LM specification, including the HRF modelling. Next to the standard analysis performed by SPM, the toolbox allows to use our joint spatio-wavelet statistical testing. Its results are added as new "contrasts" to the SPM structure related to the experiment, and they can be explored using SPM's extensive visualization features.

5 Experimental results

The aim of this section is to provide a proof-of-concept of the proposed framework, rather than a full coverage and fine-tuning of each possible parameter offered by the method and its wavelet transform. Therefore, unless mentioned otherwise, the results of wavelet-based approach are obtained using the orthogonal B-spline wavelets of degree 1 with $J_w = 1$ iteration in 3D. Essentially, we want to show that we are able to detect activation patterns similar to SPM, while making rather conservative assumptions. Further exploration of the parameter space will be presented in a future paper.

5.1 Type I error control in null data set experiments

Explicit testing of formal type I error control can be studied by null data set experiments, e.g. as in Fadili and Bullmore (2001). For a given significance level—corresponding to the expected false positives fraction (FPF)—the observed FPF is evaluated. Usually, these experiments are carried out for very low significance levels (e.g., a high value $\alpha_B = 0.1$), in order to determine an estimate of the error control rate from only a few null data sets.

However, the integrated framework requires the use of high significance levels. First, the tightness of the detection bound (34), gets better as α_B decreases. Second, the calculation of the thresholds τ_w and τ_s, based on the bound derived in App. D, is also sharper for small α_B. Due to these effects, the results for very high values of α_B are not very instructive since the thresholds are overestimated. Therefore, we conducted a type I error control experiment for synthetic null data sets, which allows us to regenerate easily new data sets and consequently to operate at the (usual) high significance levels.

The synthetic null data sets are generated for a time-series of $N_t = 120$ volumes of $64 \times 64 \times 22$ voxels. The voxels contain Gaussian white noise and the design matrix contained a dummy on-off activation with 5 volumes per epoch. The range of the significance level was chosen $\alpha_B = 10^{-6}, \ldots, 10^{-3}$. The average of the observed FPF curve over 200 experiments is shown in Fig. 7. As expected, the spatial t-test is exactly calibrated by the type I error. Clearly, the new wavelet approach has a more conservative behavior for these null data sets. This can be explained by the non-linear nature of the wavelet thresholding procedure, which suppresses the null-signal extremely well. As we will show in the next sections, this conservative behavior for null data sets does not prevent our method from detecting correctly activation regions.

5.2 Software phantom study

Following Desco et al. (2001), we have performed 3D software phantom studies. The clusters shown in Fig. 9 (a) served as seeds for activation regions. Each of these activation patterns was embedded at different signal levels (4%, 2% and 1% higher than the background) inside an intracranial mask of a $64 \times 64 \times 22$ volume. To obtain more realistic activations, the initial activation map was smoothed by a Gaussian filter (FWHM=2 voxels). The central slice of both the mask and the 12 activation regions, is shown in Fig. 9 (b) and (c), respectively. The activation map was then transformed into a time-series of 80 volumes according to a block-paradigm of 4 cycles with 10 volumes per epoch, taking into account the HRF used by SPM (assuming TR=3s). Each volumes was corrupted by Gaussian white noise of 2% of the background level.

The design matrix used to detect activation incorporated the exact knowledge of the temporal course of activated voxels. The desired significance level was chosen to be $\alpha = 5\%$. The Bonferroni-corrected significance level $\alpha_B = \alpha / N_c$ was corrected by the number of intracranial voxels $N_c = 16087$. The detected activation patterns for the various methods are shown in Fig. 9, and the number of detected voxels per cluster are summarized in Table 1. For detected voxels, the colors (red-yellow-white) are attributed according to the statistical parameter map; i.e., the estimated t-values for SPM and the spatial t-test, the

normalized value $r[\mathbf{n}]/\Lambda[\mathbf{n}]$ for the wavelet approach. None of the methods detected the single voxel activation of cluster 1. A comparison of SPM's results using various levels of smoothing, shows that increasing the width of the Gaussian filter improves the number of detected clusters, but also degrades their localization. The voxel-wise t-test with Bonferroni correction only detects a few voxels. The proposed wavelet approach using the orthogonal B-spline wavelets closely resembles the result obtained by SPM with FWHM=1.5 voxels, but misses the second activation region of cluster 3. As suggested in Van De Ville et al. (2003), the DWT using the dual B-spline wavelet has a "pure" B-spline at the analysis side, which corresponds more closely to a SPM-like Gaussian prefiltering. Interestingly, the results using the dual B-spline wavelet with the same parameters as the orthogonal one (degree 1, 1 iteration), do detect the second activation region of cluster 3, and at the same time renders more concentrated activation patterns. This example shows that the choice of the wavelet basis influences the results. An extensive comparison will be shown in a future paper.

5.3 Block-based experiment

Our example is an fMRI experiment with auditory stimulation following a block-based paradigm (Rees and Friston (1999)). The data were obtained on a 2T Siemens Magneton, 7s repetition time, providing volumes of $64 \times 64 \times 64$ isotropic voxels of 3mm×3mm×3mm. The number of volumes used for the data analysis is $N_t = 84$. The setup of the design matrix has been done using SPM and incorporates a model for the HRF. In Fig. 10, we show the on-off stimulus function and the modelled activation response, which is obtained by convolution with a HRF. The significance level α has been fixed at 5%, which gives, after correction for multiple testing by the number of intracranial voxels, $\alpha_B = 7.1 \times 10^{-7}$. The activation patterns obtained by SPM for the slices located around the auditory cortex are shown in Fig. 11. We used the Gaussian prefilter for typical values recommended by SPM; i.e., the FWHM in the left and right column is equal to 4mm and 6mm, respectively. Clearly, for a higher FWHM, stronger smoothing emphasizes large activation patterns. Specifically, the number of detected voxels is 348 and 617, respectively. The activated voxels are colored according to their t-value above SPM's threshold.

For the desired significance level α_B and N_t, the thresholds as determined by our optimization are $\tau_w = 6.058$ and $\tau_s = 0.234$. To visualize the detected voxels, we color them according to their ratio with the right-hand side of Eq. (34) since this normalization can be considered to be comparable to a t-value.

In the left column of Fig. 12, we show the activation patterns detected by

our method. The total number of detected voxels is 408, which is between the sensitivity reported using SPM with FWHM=4mm and FWHM=6mm. Given the strong type-I error control of the method, this is a promising result.

We also want to indicate the influence of the wavelet coefficients in the highpass subbands on the results. In the right column, we show the result of our approach if we only consider the detected coefficients from the lowpass subband. The activation patterns are more "conventional" (larger and more connected regions) since they originate from lowpass-filtered data. The total number of detected voxels now increases to 460. We observe that including the wavelet coefficients from the highpass subbands contributes to the spatial resolution of the detection map; i.e., activation patterns are refined, which indicates that those highpass wavelet coefficients contain essential information and that the true activation patterns are well localized.

The threshold values τ_w and τ_s have been selected optimally according to the second principle of our framework. As an experiment, we now lower the threshold τ_w at a fixed value 5.6, and compute the remaining threshold τ_s so as to satisfy the desired significance level α_B. This way, we obtain $\tau_s = 1.75$. While this setting surely detects more wavelet coefficients, the spatial statistical testing retains almost nothing: only a few voxels are detected. This clearly shows that our way of choosing the threshold values significantly differs from the standard wavelet approach, which suggests to select τ_w according to a two-sided t-test for the coefficients as in (12) in order to obtain the desired significance level. For this experiment, such a procedure would lower τ_w to 5.20 while further augmenting τ_s and finally detect no activation at all. These observations confirm the conservative nature of our framework and also indicate the optimality of the threshold selection and its criterion based on the approximation quality of the reconstructed data.

6 Conclusions

We have presented an integrated framework for statistical analysis of fMRI data using a joint spatio-wavelet approach. The major disadvantage of the standard wavelet-based method is that it does not yield a statistical interpretation in the spatial domain. Therefore, our new framework addresses two important issues: (1) the link between the wavelet and the spatial domain to properly model the effect of processing wavelet coefficients; (2) the optimal selection of the two threshold values characterizing the processing in the wavelet domain and the statistical testing in the spatial domain. The proposed framework makes conservative assumptions; in particular, it applies a Bonferroni correction for multiple testing, and therefore has a strong type-I error control. As a proof-of-concept, we included experimental results of a block-based fMRI

experiment which are quite comparable to those of SPM. Future research is needed to further explore the full potential of the new framework; e.g., the influence of parameters of the wavelet transform such as the degree/order of wavelet, the number of iterations, and the type of wavelet transform (orthogonal, semi-orthogonal).

The proposed framework incorporates only limited knowledge about the data (i.e., the HRF). Recently published methodologies propose to also take into account aspects of neural dynamics (Friston et al. (2003)). The extension of wavelet-based methods in this direction remains a challenge for future research.

A Bound for a convex sum of random variables

Theorem 1 *Consider the random variables x_k, $k = 1, \ldots, N$, following the same probability law as a generic random variable x, and the weighted sum $\sum_{k=1}^{N} \lambda_k x_k$, with $\lambda_k \geq 0$ and $\sum_{k=1}^{N} \lambda_k = 1$. Then the probability that this sum be positive is bounded by*

$$P\left[\sum_{k=1}^{N} \lambda_k x_k \geq 0\right] \leq \min_{a>0} E\left[(1+ax)_+\right]. \tag{A.1}$$

Proof: Obviously, this upper bound should still be valid without the min operand, provided that the parameter a is positive. Let $f(x)$ be the function $(1+ax)_+$, where a is a positive real parameter, and where $(x)_+$ is defined as $\max(0, x)$. We then have

$$\begin{aligned}
E\left[f\left(\sum_{k=1}^{N} \lambda_k x_k\right)\right] &= E\left[f\left(\sum_{k=1}^{N} \lambda_k x_k\right) \middle| \sum_{k=1}^{N} \lambda_k x_k \geq 0\right] P\left[\sum_{k=1}^{N} \lambda_k x_k \geq 0\right] \\
&\quad + E\left[f\left(\sum_{k=1}^{N} \lambda_k x_k\right) \middle| \sum_{k=1}^{N} \lambda_k x_k < 0\right] P\left[\sum_{k=1}^{N} \lambda_k x_k < 0\right] \\
&\quad \text{(Bayes' rule)} \\
&\geq E\left[f\left(\sum_{k=1}^{N} \lambda_k x_k\right) \middle| \sum_{k=1}^{N} \lambda_k x_k \geq 0\right] P\left[\sum_{k=1}^{N} \lambda_k x_k \geq 0\right] \\
&\quad \text{(since } f(x) \geq 0\text{)} \\
&\geq P\left[\sum_{k=1}^{N} \lambda_k x_k \geq 0\right]. \\
&\quad \text{(since } \inf_{x \geq 0} f(x) = 1\text{)}
\end{aligned} \tag{A.2}$$

Moreover, the function $f(x)$ is convex, which implies Jensen's inequality (Jensen (1906):

$$f\left(\sum_{k=1}^{N}\lambda_k x_k\right) \leq \sum_{k=1}^{N}\lambda_k f(x_k). \tag{A.3}$$

Combining (A.2) and (A.3) results into

$$\begin{aligned} P\left[\sum_{k=1}^{N}\lambda_k x_k \geq 0\right] \leq \sum_{k=1}^{N}\lambda_k E\left[f(x_k)\right] &= \left(\sum_{k=1}^{N}\lambda_k\right) E\left[f(x)\right] \\ &= E\left[f(x)\right] \\ &= E\left[(1+ax)_+\right], \end{aligned}$$

which is the bound of (A.1). Finally, the parameter $a > 0$ can be optimized to make the right-hand side as small as possible.

The optimality of the choice $f(x) = (1+ax)_+$ over all possible convex functions will be shown in another paper.

B On the sharpness of the probability bound

By a simple example of a normalized Gaussian random variable x, we can show that the bound (A.1) can be advantageously used to obtain "sharp" thresholds. According to Theorem 1, the probability that a realization of x exceeds the threshold τ is bounded by

$$P[x \geq \tau] \leq \min_{a>0} E[(1+a(x-\tau))_+]. \tag{B.1}$$

The optimal value a^* is found by setting the derivative of the right-hand side to 0,

$$\frac{\mathrm{d}}{\mathrm{d}a}E[(1+a(x-\tau))_+] \equiv E[(x-\tau)H(1+a(x-\tau))] = 0, \tag{B.2}$$

which yields the optimal value $a^* \approx \tau$, for large τ.
It is instructive to rewrite the bound (B.1) as

$$\begin{aligned} P[x \geq \tau] &\leq E[(1+a^*(x-\tau))_+] \\ &= E[(1+a^*(x-\tau))(1+a^*(x-\tau))_+^0] \\ &= E[(1+a^*(x-\tau))_+^0], \quad \text{thanks to (B.2)} \\ &= P\left[x \geq \tau - \frac{1}{a^*}\right]. \end{aligned}$$

We clearly see that the estimated threshold τ becomes sharper as $1/a^*$ gets smaller, which is the case for large threshold values since $a^* \approx \tau$. This is particularly true for short-tailed probability distributions such as the Gaussian distribution.

Notice the difference with the optimal value $a^* = 1/\tau_s$ in Sect. 3.2.1, which corresponds to the case of a *truncated* Gaussian random variables.

C Optimal values of τ_s and τ_w in Sect. 3.2.2

The relationship between τ_s and τ_w, given in (26), can be derived by computing the expectation in (20):

$$\begin{aligned}\tau_s = \frac{E[(\xi)_+]}{\alpha_B} &= \frac{1}{\alpha_B}\int_{\tau_w}^{+\infty} \frac{x}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \mathrm{d}x \\ &= \frac{1}{\alpha_B}\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\tau_w^2}{2}\right).\end{aligned}$$

To minimize $\tau_s + \tau_w$, we first rewrite this sum using (26) as

$$\frac{1}{\alpha_B}\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\tau_w^2}{2}\right) + \tau_w,$$

and set the derivative with respect to τ_w equal to zero

$$-\frac{\tau_w}{\alpha_B}\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\tau_w^2}{2}\right) + 1 = 0.$$

This finally yields

$$-\tau_w^2 \exp\left(-\tau_w^2\right) = -2\pi\alpha_B^2.$$

The term $-\tau_w^2$ can be identified as the inverse of the function $f(W) = W\exp(W)$ for $f(W) < 0$, which is known as the -1-branch of the Lambert W-function (also called Omega function) and can be evaluated numerically (Corless et al. (1996)). In this way, we obtain the optimal threshold values

$$\tau_s = 1/\tau_w, \quad \tau_w = \sqrt{-W_{-1}\left(-2\pi\alpha_B^2\right)}.$$

D Calculation of the bound in Sect. 3.3

In this Appendix, we approximate the bound of Eq. (34) by a larger upper-bound. We recall the definition of the random variables

$$\begin{aligned}\xi &= H\left(\left|\frac{g'}{s'/\sqrt{J}}\right| - \tau_w\right) g',\\ \varsigma &= s'/\sqrt{J},\end{aligned}$$

where g' is distributed according to a normalized Gaussian, and s' according to a χ-distribution with J degrees of freedom. We also introduce the variable $t' = g'/(s'/\sqrt{J})$, which follows a t-distribution with J degrees of freedom. Now, we can split the bound into three parts. We write

$$\begin{aligned}&E[(1+a(\xi-\tau_s\varsigma))_+] =\\ &E[(1-a\tau_s\varsigma)_+ H(-|t'|-\tau_w)] + &\text{(D.1)}\\ &E[(1+a(g'-\tau_s\varsigma))_+ H(t'-\tau_w)] + &\text{(D.2)}\\ &E[(1+a(g'-\tau_s\varsigma))_+ H(-t'-\tau_w)], &\text{(D.3)}\end{aligned}$$

where each part can be more easily calculated or bounded as follows:

$$\begin{aligned}&\text{(D.1)} \le E[(1-a\tau_s\varsigma)_+] &\text{(D.4)}\\ &\text{(D.2)} = E[(1+a(g'-\tau_s\varsigma))H(t'-\tau_w)] &\text{(D.5)}\\ &\text{(D.3)} \le E[H(-t'-\tau_w)]. &\text{(D.6)}\end{aligned}$$

Equality (D.5) holds as long as $\tau_s < \tau_w$, which is a very reasonable constraint for the threshold we expect. The first term (D.4) and the third one (D.6) can be calculated as

$$\begin{aligned}\text{(D.4)} &= \frac{\gamma\left(\frac{J}{2}, \frac{J}{2(a\tau_s)^2}\right) - a\tau_s\sqrt{\frac{2}{J}}\,\gamma\left(\frac{J+1}{2}, \frac{J}{2(a\tau_s)^2}\right)}{\Gamma\left(\frac{J}{2}\right)},\\ \text{(D.6)} &= \alpha_{\tau_w},\end{aligned}$$

where $\Gamma(a) = \int_0^{+\infty} x^{a-1}\exp(-x)\mathrm{d}x$ is the Gamma function, and

$$\gamma(a,b) = \int_0^b x^{a-1}\exp(-x)\mathrm{d}x \qquad \text{(D.7)}$$

is the related lower incomplete Gamma function[2], and where α_{τ_w} is the complementary cumulative t-distribution $\int_{\tau_w}^{+\infty} p_t(t')\mathrm{d}t'$. Furthermore, by changing variables from $(g',\, \varsigma)$ to $(t',\, x = \varsigma\sqrt{J + t'^2})$, we can also compute the second term (D.5) as

$$(\mathrm{D.5}) = \alpha_{\tau_w} + \frac{a}{\sqrt{2\pi}\,(1+\tau_w^2/J)^{J/2}} - \frac{a\tau_s}{2}\sqrt{\frac{J}{2\pi}}\,B_{\frac{1}{1+\tau_w^2/J}}\left(\frac{J+1}{2}, \frac{1}{2}\right),$$

where $B_b(a_0, a_1) = \int_0^b x^{a_0-1}(1-x)^{a_1-1}\mathrm{d}x$ is the incomplete Beta function[2].

References

Bullmore, E., Fadili, J., Breakspear, M., Salvador, R., Suckling, J., Brammer, M., 2003. Wavelets and statistical analysis of functional magnetic resonance images of the human brain. Statistical methods in medical research 12 (5), 375–399.

Bullmore, E., Long, C., Suckling, J., Fadili, J., Calvert, G., Zelaya, F., Carpenter, T., Brammer, M., 2001. Colored noise and computational inference in neurophysiological time series analysis: Resampling methods in time and wavelet domains. Human Brain Mapping 12, 61–78.

Corless, R. M., Gonnet, G. H., Hare, E. G., Knuth, D. E., 1996. On the Lambert W function. Advances in Computational Mathematics 5 (4), 329–359.

Desco, M., Hernandez, J., Santos, A., Brammer, M., 2001. Multiresolution analysis in fMRI: Sensitivity and specificity in the detection of brain activation. Human Brain Mapping 14, 16–27.

Fadili, M. J., Bullmore, E., 2001. Wavelet-generalised least squares: a new BLU estimator of linear regression models with 1/f errors. NeuroImage 15, 217–232.

Feilner, M., Blu, T., Unser, M., 2000. Analysis of fMRI data using spline wavelets. In: Proceedings of the Tenth European Signal Processing Conference (EUSIPCO'00). Vol. IV. Tampere, Finland, pp. 2013–2016.

Frackowiak, R., Friston, K., Frith, C., Dolan, R., Mazziotta, J., 1997. Human Brain Function. Academic Press.

Friston, K. J., Harrison, L., Penny, W. D., 2003. Dynamic causal modeling. NeuroImage 19, 1273–1302.

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J. P., Frith, C. D., Frackowiak, R. S. J., 1995. Statistical parametric maps in functional imaging: A general linear approach. Human Brain Mapping 2, 189–210.

[2] Note that MATLAB™ computes a normalized version of the incomplete Gamma function, namely $\gamma(a,b)/\Gamma(a)$, and of the incomplete Beta function, namely $B_b(a_0,a_1)/B(a_0,a_1)$.

Fu, Z., Hui, Y., Liang, Z.-P., 1998. Joint spatiotemporal statistical analysis of functional MRI data. In: Proceedings ICIP. pp. 709–713.

Genovese, C. R., Lazar, N. A., Nicols, T. E., 2002. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. NeuroImage 15, 772–786.

Hilton, M., Ogden, T., Hattery, D., Eden, G., Jawerth, B., 1996. Wavelets in Biology and Medicine. CRC Press, Ch. Wavelet Denoising of Functional MRI Data, pp. 93–114.

Jensen, J. L. W. V., 1906. Sur les fonctions convexes et les inégalités entre les valeurs moyennes. Acta Math. 30, 175–193.

Kruggel, F., von Cramon, D., Descombes, X., 1999. Comparison of filtering methods for fMRI datasets. NeuroImage 10 (5), 530–543.

Mallat, S., 1989. A theory for multiresolution signal decomposition: The wavelet decomposition. IEEE Trans. Pattern Anal. Mach. Intell. 11, 674–693.

Mallat, S., 1998. A Wavelet Tour of Signal Processing. Academic Press, San Diego (CA).

Mueller, K., Lohmann, G., Zysset, S., von Carmon, Y., 2003. Wavelet statistics of functional MRI data and the general linear model. Journal of Magnetic Resonance Imaging 17, 20–30.

Ogawa, S., Menon, R., Tank, D., Kim, S., Merkle, H., Ellerman, J., Ugurbil, K., 1993. Functional brain mapping by blood oxygenation level-dependent constrast magnetic resonance imaging. Biophysical Journal 64, 803–812.

Poline, J., Worsley, K., Evans, A., Friston, K., 1997. Combining spatial extent and peak intensity to test for activations in functional imaging. NeuroImage 5 (2), 83—96.

Rees, G., Friston, K., 1999. Single subject epoch (block) auditory fMRI activation data. http://www.fil.ion.ucl.ac.uk/spm/data/.

Ruttimann, U., Unser, M., Rawlings, R., Rio, D., Ramsey, N., Mattay, V., Hommer, D., Frank, J., Weinberger, D., 1998. Statistical analysis of functional MRI data in the wavelet domain. IEEE Transactions on Medical Imaging 17 (2), 142–154.

Shafie, K., Sigal, B., Siegmund, D., Worsley, K., 2003. Rotation space random fields with an application to fMRI data. Annals of Statistics 31, 1732–1771.

Turkheimer, F. E., Brett, M., Aston, J. A. D., Leff, A. P., Sargent, P. A., Wise, R. J., Grasby, P. M., Cunningham, V. J., 2000. Statistical modelling of positron emission tomography images in wavelet space. Journal of Cebebral Blood Flow and Metabolism 20, 1610–1618.

Van De Ville, D., Blu, T., Unser, M., 2003. Wavelets versus resels in the context of fMRI: establishing the link with SPM. In: SPIE's Symposium on Optical Science and Technology: Wavelets X. Vol. 5207. SPIE, San Diego CA, USA, pp. 417–425.

Wink, A. M., Roerdink, J. B. T. M., 2004. Denoising functional MR images: a comparison of wavelet denoising and Gaussian smoothing. IEEE Transactions on Medical Imaging.

Woolrich, M. W., Ripley, B. D., Brady, M., Smith, S. M., 2001. Temporal autocorrelation in univariate linear modeling of fMRI data. NeuroImage 14 (6), 1370–1386.

Worsley, K., Marrett, S., Neelin, P., Evans, A., 1996. Searching scale space for activation in PET images. Human Brain Mapping 4 (1), 74–90.

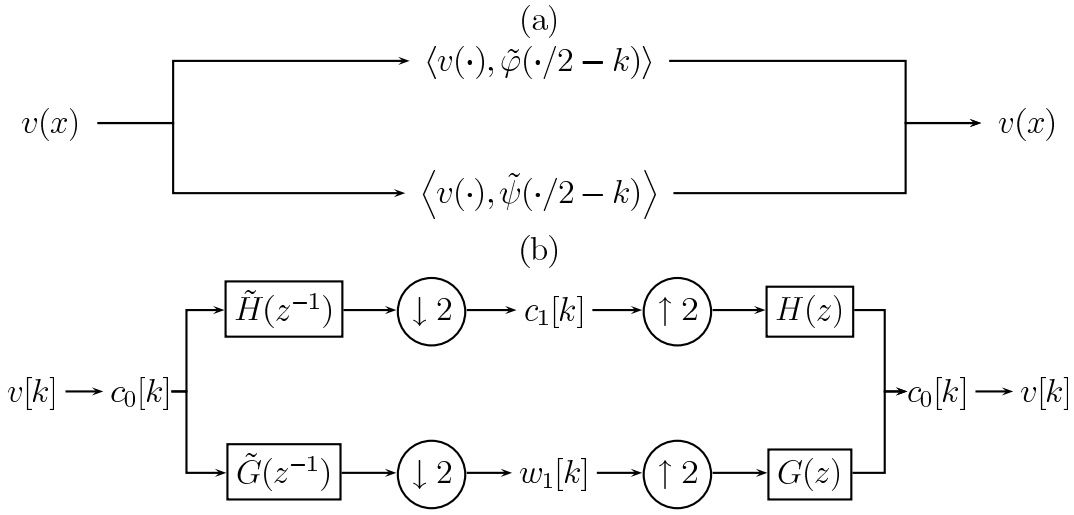


Fig. 1. The discrete wavelet transform for one iteration. (a) The continuous-domain representation. (b) The filterbank representation.

Method	cluster 2			cluster 3			cluster 4		
	4%	2%	1%	4%	2%	1%	4%	2%	1%
SPM (FWHM=1 voxel)	2	none detected	none detected	4	0	none detected	26	4	0
SPM (FWHM=1.5 voxels)	6			9	2		47	18	1
SPM (FWHM=2 voxels)	15			16	4		75	37	7
Spatial t-test	0			1	0		13	0	0
Wavelet (orthogonal)	7			17	0		41	9	1
Wavelet (dual)	1			6	2		30	6	2

Table 1
Overview of the number of voxels detected for each activation cluster, corresponding to the results of Fig. 9.

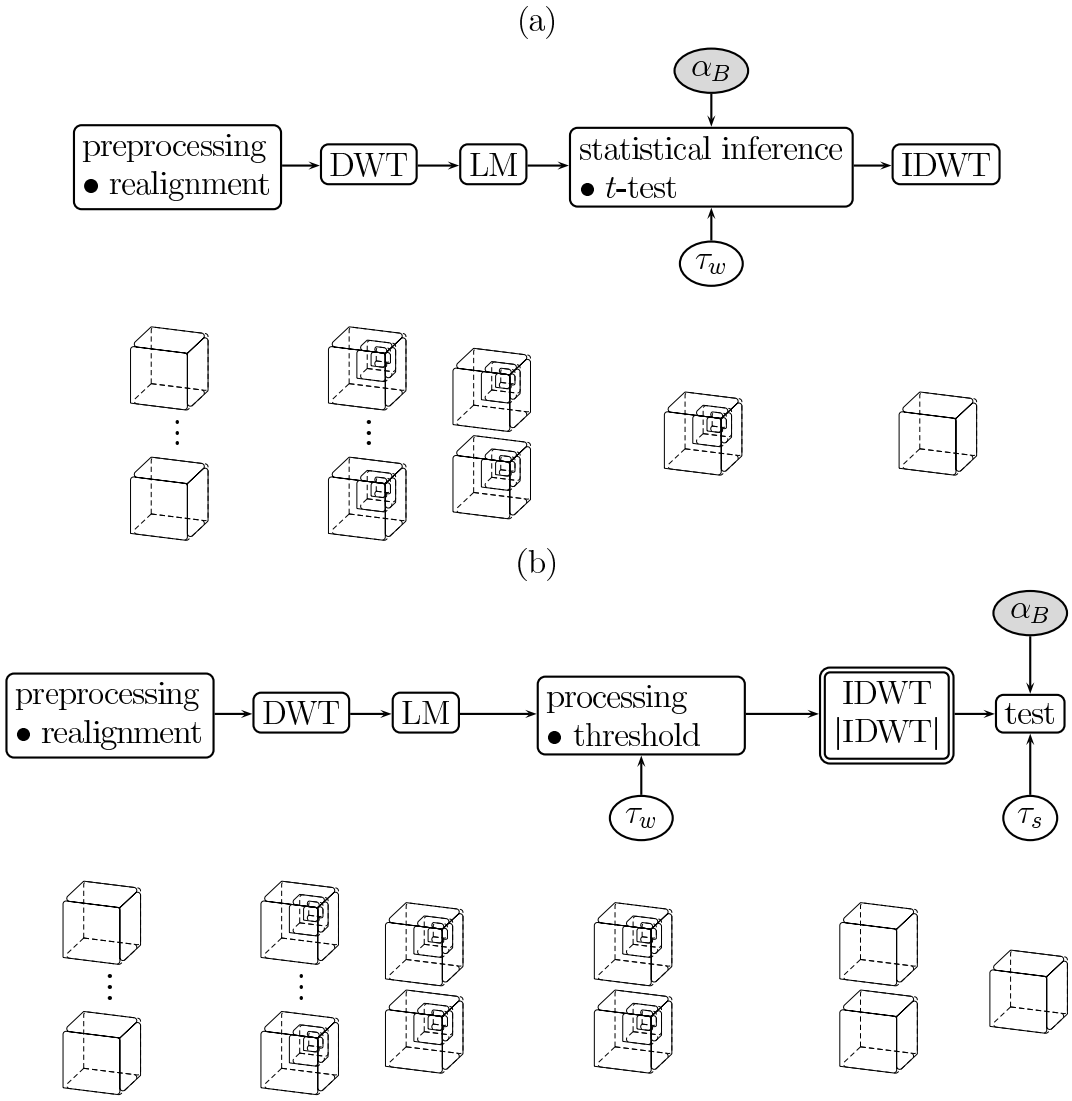


Fig. 2. (a) The standard wavelet-based approach. The desired significance level α_B is fed to the statistical inference stage in the wavelet domain. (b) The approach of the present paper is to reinterprete the wavelet threshold as a general algorithm parameter. After reconstruction, a statistical test is performed in the spatial domain, taking into account the wavelet processing.

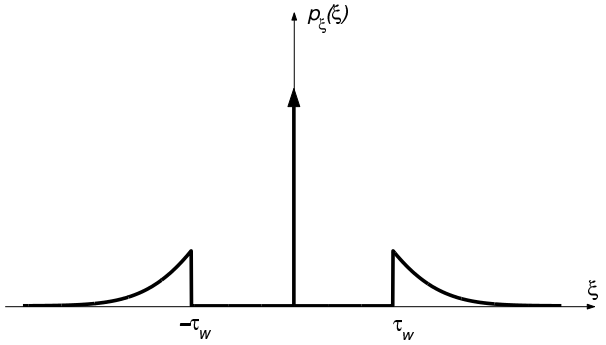
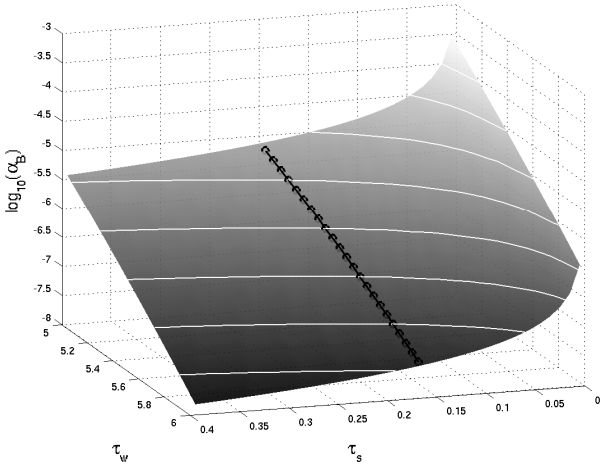


Fig. 3. The truncated Gaussian law $p_\xi(\xi)$.

(a)



(b)

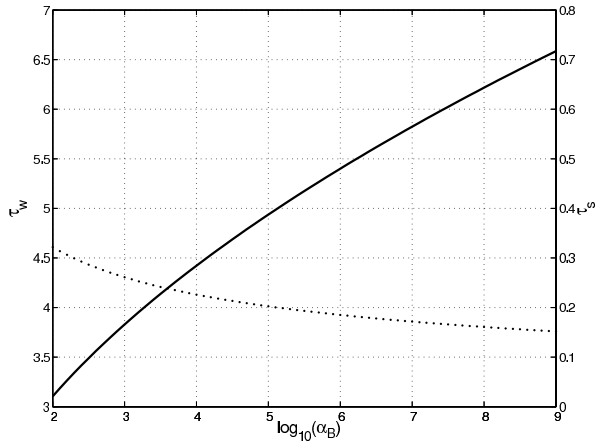


Fig. 4. Simplified case where $\sigma_{\mathbf{k}}$ is assumed to be known. (a) Probability surface. Equiprobable contour lines are in white. The circular dots mark the optimal threshold values. (b) The optimal threshold values τ_w (full line) and τ_s (dotted line).

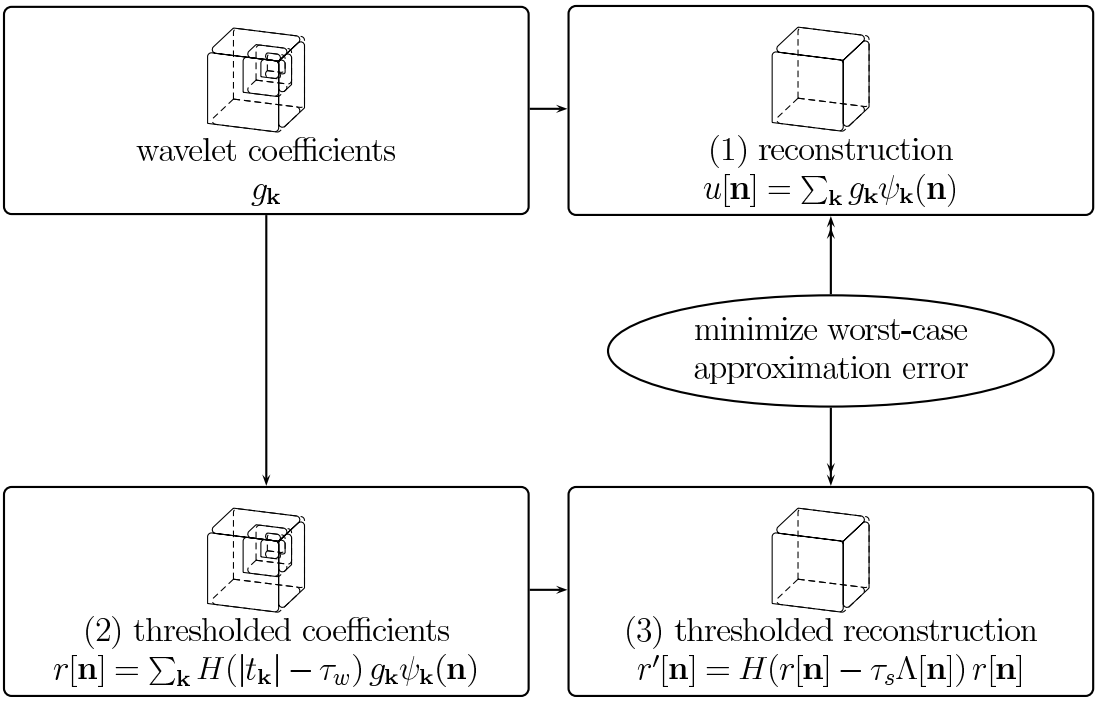
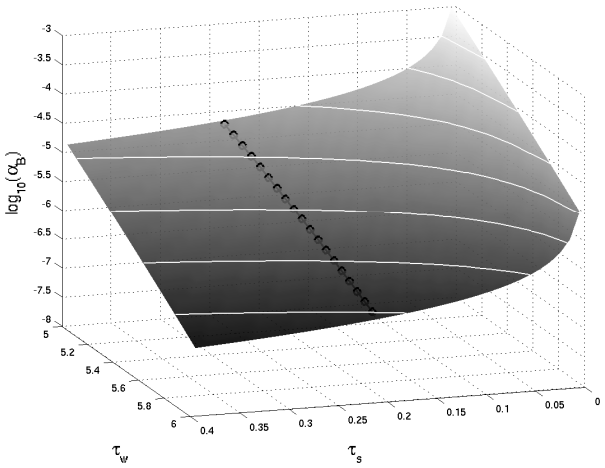
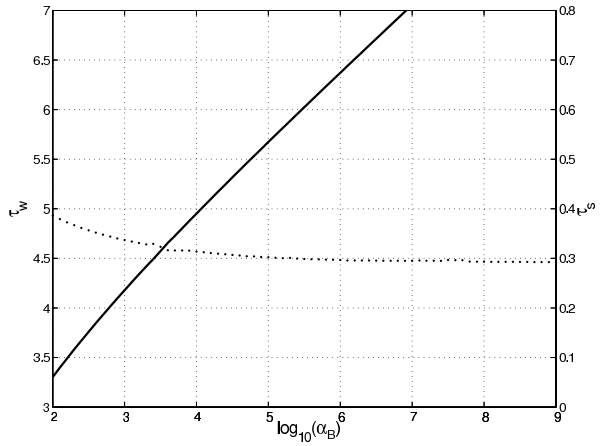


Fig. 5. Principle of minimizing the worst-case approximation error between the detected and the unprocessed parameter map, represented by $r'[\mathbf{n}]$ and $u[\mathbf{n}]$, respectively.

(a)



(b)



(c)

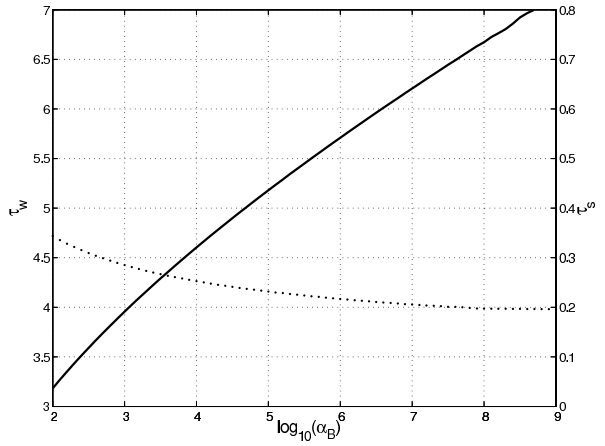


Fig. 6. General case taking into account the estimate of $\sigma_{\mathbf{k}}$. (a) Probability map. The white contour lines are equiprobable. The circular dots mark the optimal threshold values. (b) Optimal threshold values for $N_t = 50$. (c) Optimal threshold values for $N_t = 150$.

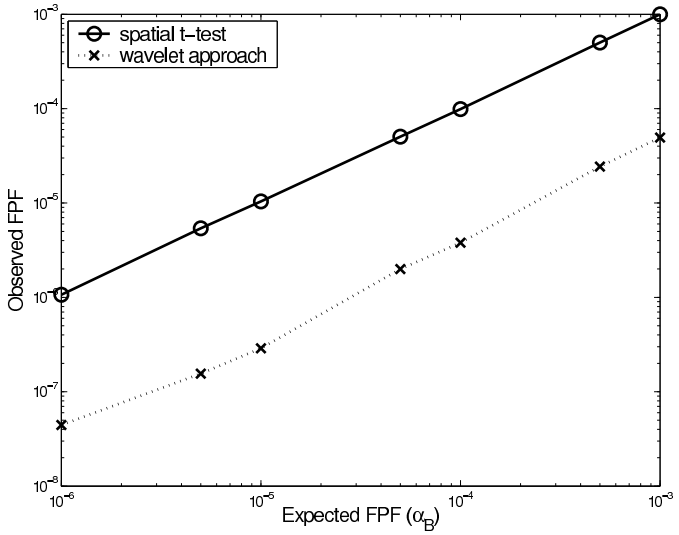


Fig. 7. Detection results for synthetic null data sets. The curves show the observed versus expected FPF as an average over 200 experiments. Each experiment corresponds to a time-series of $N_t = 120$ volumes ($64 \times 64 \times 22$ voxels). The design matrix contained a dummy on-off activation with 5 volumes per epoch.

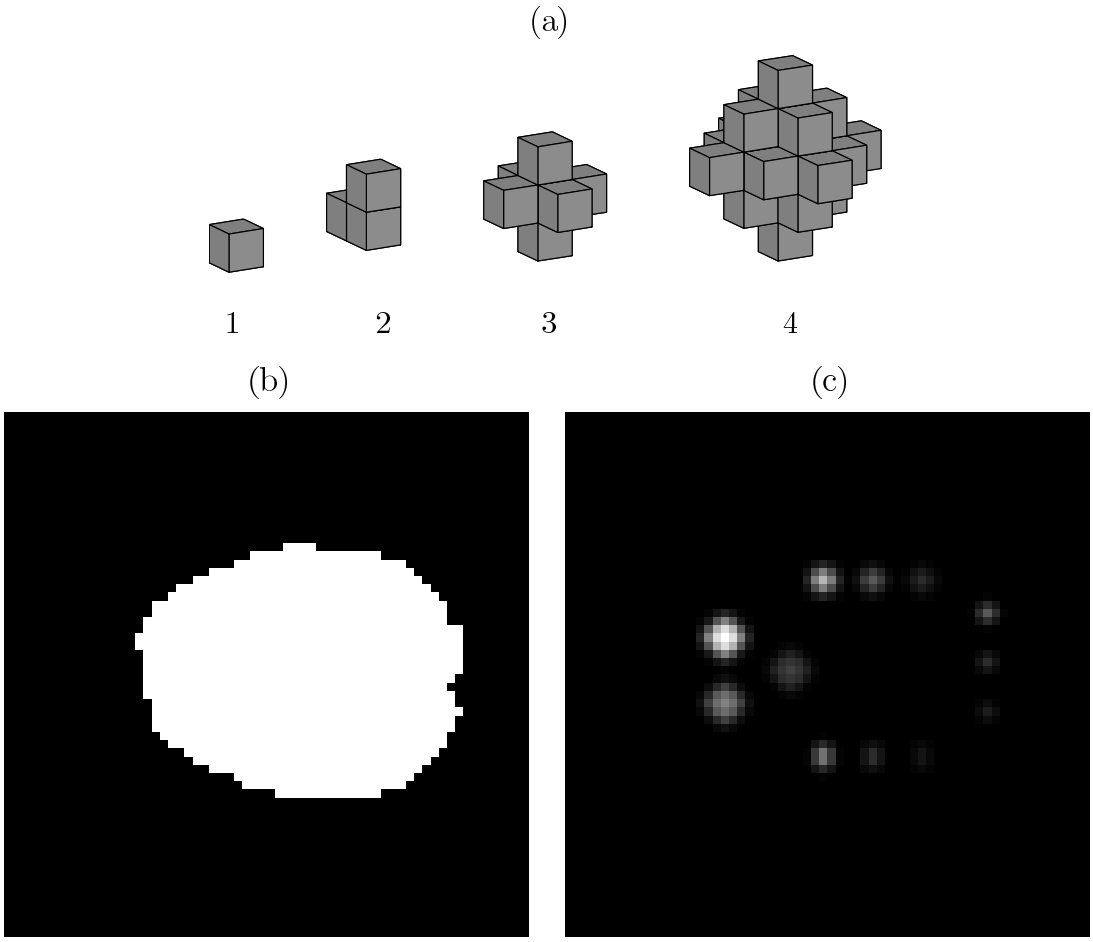


Fig. 8. The software phantom was constructed using the initial 3D activation regions of (a), which where embedded with three different signal levels into the intracranial mask of (b). The initial activation regions contain 1, 3, 7, and 25 voxels, respectively. Next, these activation regions where smoothed by a Gaussian filter with FWHM=2 voxels, resulting into the activation map shown in (c); right: cluster 1, bottom: cluster 2, top: cluster 3, left: cluster 4. Subfigures (b) and (c) show the central slice of the volume.

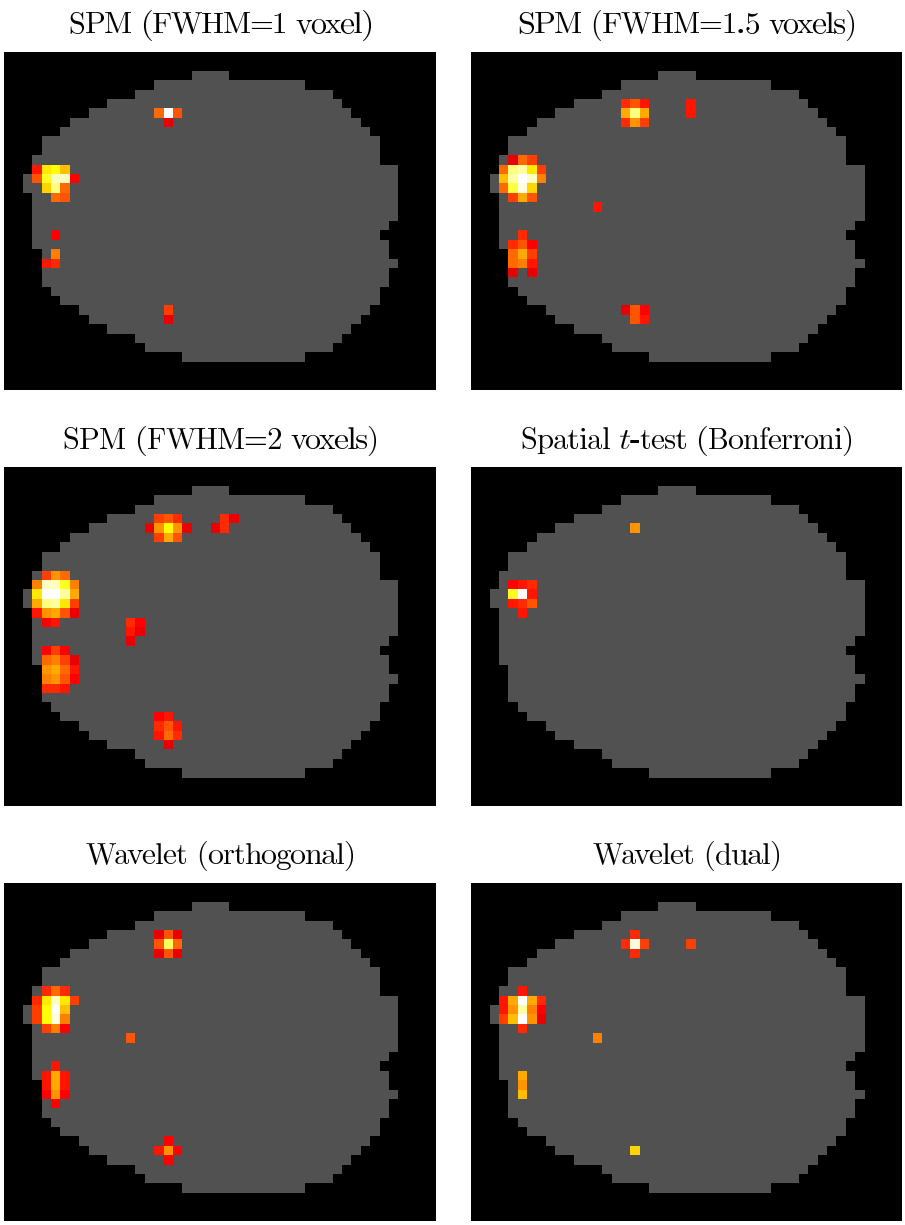


Fig. 9. Activation patterns obtained by the various methods for the central slice of the software phantom.

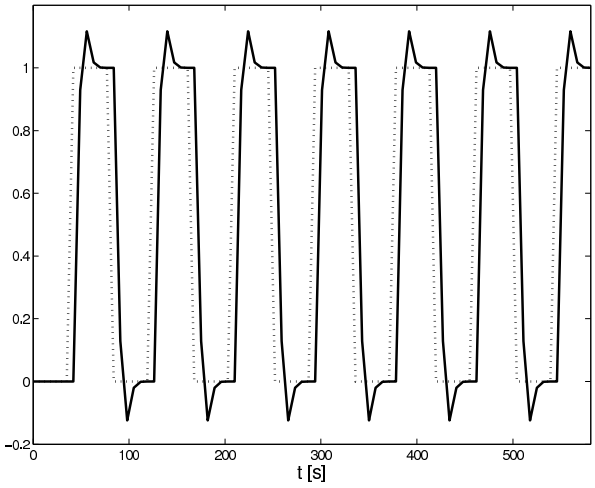
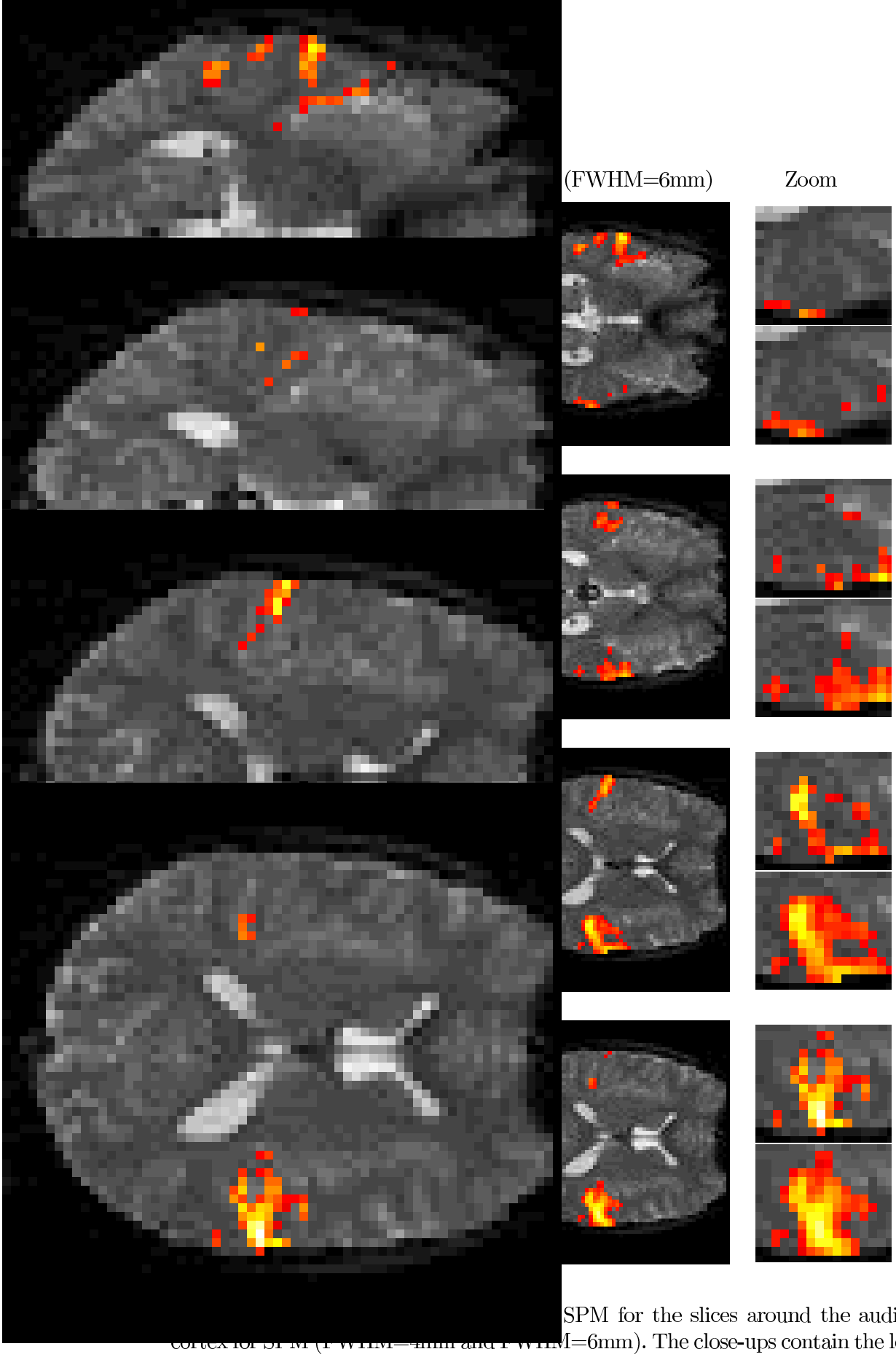
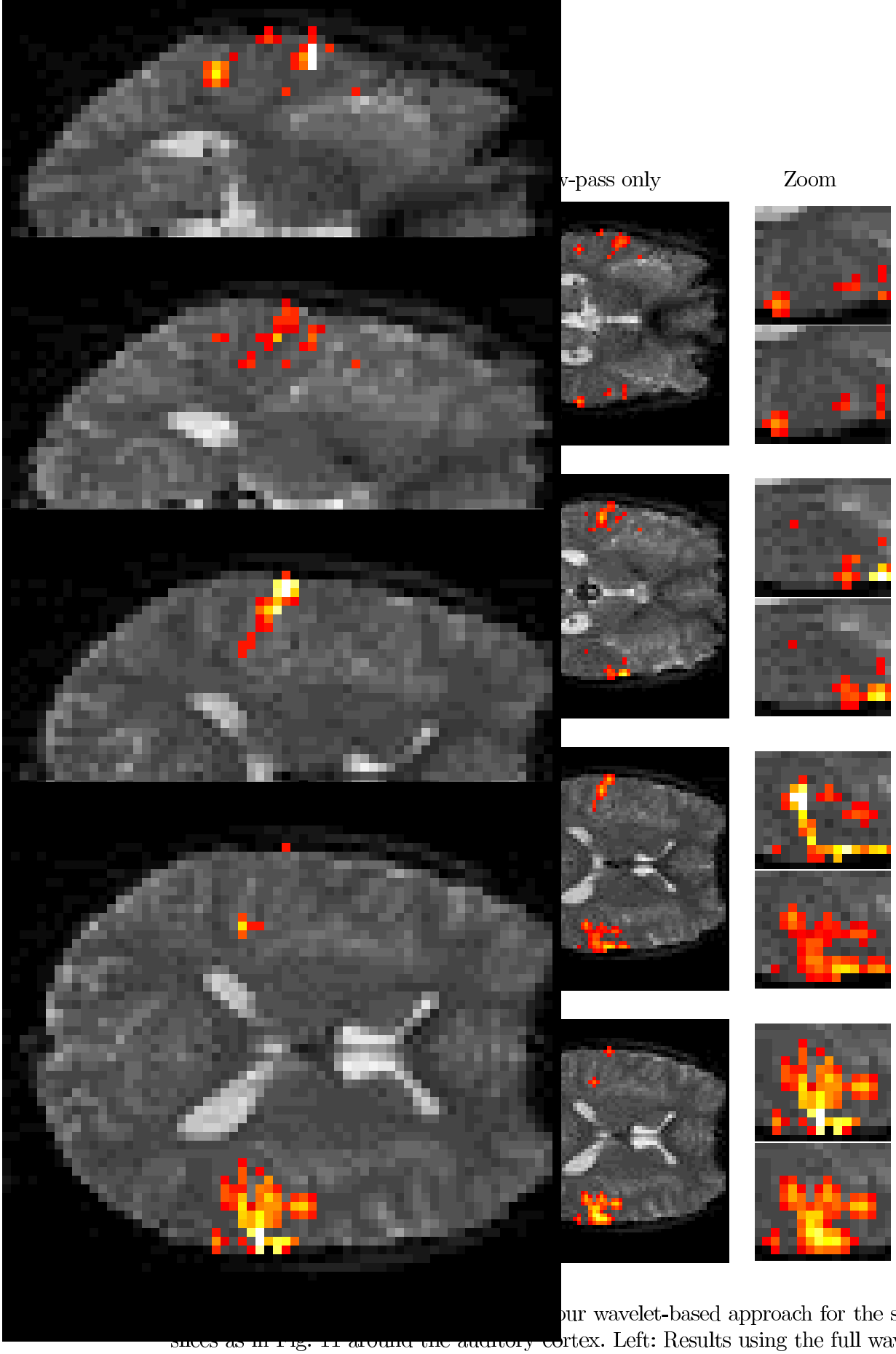


Fig. 10. Stimulus function (dotted line) and model for activation response (full line).



SPM for the slices around the auditory cortex for SPM (FWHM=4mm and FWHM=6mm). The close-ups contain the lower part of the activation.



our wavelet-based approach for the same slices as in Fig. 11 around the auditory cortex. Left: Results using the full wavelet decomposition; Right: Results using only the low-pass subband.